

**TRƯỜNG ĐẠI HỌC KINH TẾ QUỐC DÂN
KHOA TOÁN KINH TẾ**



**BÀI TẬP LỚN
PHÂN TÍCH VÀ DỰ BÁO CHUỖI THỜI GIAN
TRONG TÀI CHÍNH**

Lĩnh vực: Tiêu dùng thiết yếu và Bất động sản

Giảng viên	:	Ths. Bùi Dương Hải
Nhóm thực hiện	:	09
Lớp	:	Toán Kinh tế 64

Hà Nội, 2024

MỤC LỤC

DANH MỤC BẢNG BIỂU	1
DANH MỤC HÌNH ẢNH	2
I. GIỚI THIỆU CHUNG	3
II. PHẦN CÁ NHÂN	4
1. Phân tích chuỗi SBT – Ngô Thị Minh Hạnh.....	4
1.1.Khái quát về công ty cổ phần Thành Thành Công – Biên Hoà	4
1.2.Phân tích chuỗi tài chính	4
1.3.Phân tích và dự báo giá cổ phiếu theo mô hình ARIMA.....	5
1.3.1.Phân tích và định giá giá cổ phiếu	5
1.3.2.Mô hình dự báo giá cổ phiếu cho 10 phiên đầu năm 2025	5
2. Phân tích chuỗi DIG – Trần Đức Hiền.....	10
2.1.Giới thiệu về Tổng Công ty Cổ phần Đầu tư Phát triển Xây dựng (DIC Corp) và mã cổ phiếu DIG.....	10
2.2.Phân tích và dự báo chuỗi doanh thu thuần.....	10
2.2.1.Mô hình hồi quy giản đơn theo xu thế.	11
2.2.2.Mô hình hồi quy có yếu tố mùa vụ	11
2.3. Phân tích và dự báo chuỗi giá cổ phiếu	12
3. Phân tích chuỗi VIC – Lê Tuấn Minh	16
3.1.Giới thiệu về dữ liệu	16
3.2.Các mô hình với chuỗi doanh thu.....	16
3.3.Các mô hình với giá cổ phiếu	18
3.3.1.Các mô hình với chuỗi Log return	18
3.3.2.Các mô hình với chuỗi giá	19
3.3.3.Kết quả.....	20
4. Phân tích chuỗi LSS – Phạm Phương Thảo	21
4.1. Giới thiệu mã cổ phiếu LSS	21
4.2. Phân tích chuỗi tài chính	21
4.3. Phân tích chuỗi giá đóng cửa và chuỗi log-return của LSS.....	23
4.3.1. Chuỗi giá	23
4.3.2. Chuỗi log-return	24
III. BÀI TẬP NHÓM	27
1. Phân tích tương quan giữa các cổ phiếu.....	27

2. Xây dựng danh mục tối ưu.....	27
2.1. Bài toán tối đa hóa lợi suất kỳ vọng của danh mục trong điều kiện phương sai danh mục bằng trung bình nhân phương sai của lợi suất các cổ phiếu	27
2.2. Bài toán thiểu hóa rủi ro danh mục trong điều kiện lợi suất kỳ vọng của danh mục bằng lợi suất của thị trường	28
3. Dự báo lợi suất danh mục	28
4. Phân tích đồng tích hợp.....	31
5. Phân tích VAR	32
5.1. Ước lượng mô hình VAR	32
5.2. Hàm phản ứng IRF	34
5.3. Phân rã phương sai	37
IV. KẾT LUẬN	40
PHỤ LỤC	41
1. Chuỗi SBT – Ngô Thị Minh Hạnh.....	41
2. Chuỗi DIC – Trần Đức Hiền	44
3. Chuỗi VIC – Lê Tuấn Minh	46
4. Chuỗi LSS – Phạm Phương Thảo.....	78
PHỤ LỤC NHÓM	93

DANH SÁCH THÀNH VIÊN

STT	Họ và tên	Mã sinh viên	Đóng góp	Công việc đã làm
1	Ngô Thị Minh Hạnh	11222150	25%	<ul style="list-style-type: none">- Xây dựng danh mục tối ưu: bài toán tối thiểu hoá rủi ro- Hoàn thiện bản word, viết introduction, viết kết luận
2	Trần Đức Hiền	11222280	25%	<ul style="list-style-type: none">- Xây dựng danh mục tối ưu: bài toán tối thiểu hóa rủi ro.- Trình bày kết quả về phân tích đồng tích hợp, mô hình VAR
3	Lê Tuấn Minh	11224210	25%	<ul style="list-style-type: none">- Code các phần: Dự báo chuỗi lợi suất danh mục theo cách 2; Đồng tích hợp; VAR.- Viết phụ lục nhóm.
4	Phạm Phương Thảo	11225952	25%	<ul style="list-style-type: none">- Xây dựng danh mục tối ưu: bài toán tối đa hóa lợi suất- Trình bày kết quả phân tích tương quan- Trình bày kết quả xây dựng danh mục và dự báo lợi suất danh mục- Double-check data, kết quả

DANH MỤC BẢNG BIỂU

Bảng 1.3a: Dự báo giá cổ phiếu SBT cho 10 phiên đầu năm 2025.....	8
Bảng 1.3b: Dự báo log return và giá cổ phiếu SBT cho 10 phiên đầu năm 2025	9
Bảng 2.1. Kết quả sai số dự báo thu được từ 4 mô hình	11
Bảng 2.2. Kết quả dự báo từ mô hình lin-log.....	11
Bảng 2.3. Kết quả sai số dự báo thu được từ 5 mô hình	12
Bảng 2.4. Kết quả mô hình Holtwinter dạng cộng và nhân	12
Bảng 2.5. Kết quả dự báo cho 4 quý 2025.....	12
Bảng 2.6 Kết quả kiểm định ADF	13
Bảng 3.1. Kết quả từ các mô hình đối với chuỗi doanh thu	17
Bảng 3.2. Kết quả dự báo cho năm 2025.....	18
Bảng 3.3. Sai số dự báo RMSE từ các mô hình	20
Bảng 3.4. Log return dự báo từ mô hình ARIMA(3,1,1) với chuỗi giá	20
Bảng 4.1: Bảng kết quả ước lượng mô hình	21
Bảng 4.2: Kết quả dự báo doanh thu thuần theo mô hình Holt-Winter dạng nhân	22
Bảng 4.3: Kết quả dự báo giá và thực tế	26
Bảng 5.1: Ma trận Covariance giữa lợi suất các mã cổ phiếu.....	27
Bảng 5.2.: Ma trận Correlation giữa lợi suất các mã cổ phiếu	27
Bảng 5.3: Kết quả tỷ trọng đầu tư mỗi mã cổ phiếu của bài toán tối đa hóa lợi suất.....	28
Bảng 5.4: Kết quả tỷ trọng đầu tư mỗi mã cổ phiếu của bài toán tối thiểu hóa rủi ro.....	28
Bảng 5.5 : Chuỗi lợi suất thực tế của danh mục, dự báo lợi suất danh mục từ dự báo mỗi lợi suất cổ phiếu, dự báo lợi suất danh mục từ lợi suất thực tế của danh mục trong quá khứ	
31	
Bảng 5.6 Kết quả kiểm định Johansen theo thống kê maximum eigenvalue.....	31
Bảng 5.7. Sai số dự báo RMSE từ hai mô hình ARIMA và VAR dành cho chuỗi log_return	
34	

DANH MỤC HÌNH ẢNH

Hình 1.2. Biểu đồ doanh thu thuần trong giai đoạn từ Q1 năm 2010 đến Q3 năm 2024 của chuỗi SBT.....	4
Hình 1.3a. Đồ thị chuỗi giá và chuỗi log return SBT.....	5
Hình 1.3b. Đồ thị chuỗi sai phân của giá SBT	6
Hình 1.3c. Lược đồ tương quan của chuỗi sai phân của chuỗi giá	7
Hình 1.3d. Lược đồ tương quan của chuỗi log return SBT	8
Hình 2.1. Doanh thu thuần của DIC Corp giai đoạn 2010Q1 – 2024Q4 (đơn vị: triệu Việt Nam Đồng).....	10
Hình 2.2. Chuỗi giá đóng cửa và log_return của mã cổ phiếu DIG.....	13
Hình 2.3. Đồ thị kiểm định tính dừng qua nghiệm nghịch đảo mô hình ARIMA(2,1,2).....	15
Hình 2.4. Đồ thị kiểm định tính dừng qua nghiệm nghịch đảo mô hình ARIMA(1,0,1).....	15
Hình 3.1. Đồ thị của chuỗi doanh thu	16
Hình 3.2. Đồ thị của mô hình Holt-Winter Additive với log doanh thu	18
Hình 3.3. Đồ thị của chuỗi log return.....	18
Hình 3.4. Lược đồ tương quan của chuỗi log return	19
Hình 3.5. Đồ thị chuỗi giá và chuỗi sai phân của chuỗi giá.....	19
Hình 3.6. Lược đồ tương quan của chuỗi sai phân của chuỗi giá	20
Hình 4.1: Biểu đồ doanh thu thuần LSS theo quý giai đoạn 2010-2024.....	21
Hình 4.2: Đồ thị chuỗi giá cổ phiếu LSS giai đoạn 2023-2024	23
Hình 4.3: Biểu đồ chuỗi log-return của cổ phiếu LSS giai đoạn 2023-2024	24
Hình 5.1: Kiểm định Box-test cho chuỗi lợi suất danh mục	30
Hình 5.2. Kết quả dự báo chuỗi log_return dựa trên mô hình VAR.....	33
Hình 5.3. Phản ứng của các biến khi có sốc xảy ra với rSBT	34
Hình 5.4. Phản ứng của các biến khi có sốc xảy ra với rDIG	35
Hình 5.5. Phản ứng của các biến khi có sốc xảy ra với rVIC	35
Hình 5.6. Phản ứng của các biến khi có sốc xảy ra với rLSS	36
Hình 5.7. Đồ thị phân rã phương sai của rSBT	37
Hình 5.8. Đồ thị phân rã phương sai của rDIG	37
Hình 5.9. Đồ thị phân rã phương sai của rVIC	38
Hình 5.10. Đồ thị phân rã phương sai của rLSS	38

I. GIỚI THIỆU CHUNG

Ngành tiêu dùng thiết yếu và bất động sản đều đóng vai trò quan trọng trong nền kinh tế quốc dân, ảnh hưởng trực tiếp đến chất lượng sống của người dân và tốc độ phát triển của đất nước. Trong đó, lĩnh vực tiêu dùng thiết yếu đảm bảo nguồn cung ổn định các sản phẩm như thực phẩm, đồ uống, và hàng tiêu dùng hàng ngày – những mặt hàng luôn có nhu cầu cao, kể cả trong điều kiện kinh tế khó khăn. Đây là ngành có tính ổn định cao, ít bị ảnh hưởng bởi chu kỳ kinh tế và thường thu hút sự quan tâm của các nhà đầu tư dài hạn.

Ngược lại, ngành bất động sản lại thể hiện sự nhạy cảm rõ rệt với tình hình kinh tế vĩ mô, chính sách lãi suất và dòng tiền trong nền kinh tế. Trong những năm gần đây, thị trường bất động sản Việt Nam trải qua nhiều biến động: giai đoạn tăng trưởng nóng bị kìm hãm bởi chính sách siết tín dụng, lãi suất tăng, và sự sụt giảm niềm tin của nhà đầu tư. Tuy nhiên, với tầm quan trọng chiến lược trong phát triển hạ tầng, đô thị và nhu cầu nhà ở vẫn đang gia tăng, bất động sản vẫn là một ngành được kỳ vọng phục hồi và phát triển trong trung – dài hạn.

Nhằm hiểu rõ hơn về tình hình hoạt động và triển vọng phát triển của hai lĩnh vực trên, nhóm em đã thực hiện phân tích các doanh nghiệp đại diện: Công ty CP Thành Thành Công - Biên Hòa (SBT) và Công ty CP Mía đường Lam Sơn (LSS) trong lĩnh vực tiêu dùng thiết yếu; Tổng công ty CP Đầu tư Phát triển Xây dựng (DIG) và Tập đoàn Vingroup (VIC) trong lĩnh vực bất động sản. Đây đều là các doanh nghiệp niêm yết có quy mô lớn, dữ liệu minh bạch và có vai trò nhất định trong ngành, qua đó giúp nhóm đưa ra các đánh giá và dự báo sát với thực tiễn thị trường hiện nay.

Cụ thể, mục tiêu của nghiên cứu này là:

- Phân tích và sử dụng các mô hình để dự báo doanh thu thuần theo từng quý của 4 công ty cổ phần.
- Áp dụng các phương pháp dự báo để dự báo giá hoặc lợi suất cổ phiếu trong 10 phiên giao dịch đầu tiên từ năm 2025, dựa trên dữ liệu lịch sử từ năm 2023 và năm 2024.
- Xây dựng danh mục tối ưu từ 4 mã cổ phiếu và dự báo lợi suất của danh mục.
- Đánh giá độ chính xác của các mô hình dự báo thông qua việc phân tích sai số và đưa ra những nhận định tương ứng.

II. PHÂN CÁ NHÂN

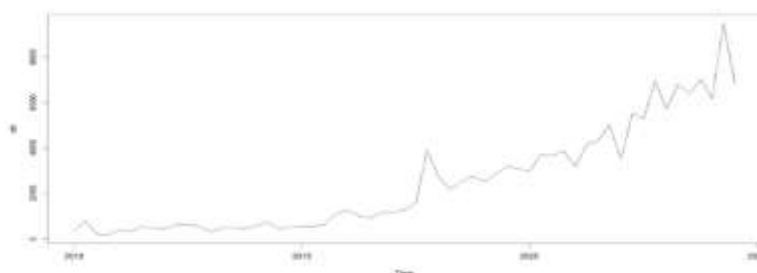
1. Phân tích chuỗi SBT – Ngô Thị Minh Hạnh

1.1. Khái quát về công ty cổ phần Thành Thành Công – Biên Hoà

Công ty Cổ phần Thành Thành Công - Biên Hòa là một trong những doanh nghiệp hàng đầu trong lĩnh vực sản xuất và kinh doanh đường tại Việt Nam. Được thành lập vào năm 1979, công ty có trụ sở chính tại Thành phố Hồ Chí Minh và hoạt động chủ yếu trong các lĩnh vực: sản xuất và kinh doanh đường mía, sản xuất điện sinh khối từ bã mía, sản xuất phân bón và các sản phẩm phụ từ quá trình sản xuất đường.. Trên thị trường chứng khoán, cổ phiếu SBT được niêm yết và giao dịch trên Sở Giao dịch Chứng khoán Thành phố Hồ Chí Minh (HOSE), thu hút sự quan tâm của nhiều nhà đầu tư trong và ngoài nước.

1.2. Phân tích chuỗi tài chính

Dưới đây là biểu đồ doanh thu thuần của CTCP Thành Thành Công - Biên Hòa (SBT) trong giai đoạn từ Q1/2010 đến Q3/2024. Có thể thấy, doanh thu thuần của công ty có xu hướng tăng trưởng mạnh mẽ trong giai đoạn này. Trong bối cảnh ngành mía đường chịu nhiều áp lực từ biến động giá nguyên liệu và chính sách bảo hộ của các quốc gia, SBT vẫn duy trì được mức tăng trưởng ổn định nhờ chiến lược kinh doanh hiệu quả và tối ưu hóa chuỗi cung ứng.



Hình 1.2. Biểu đồ doanh thu thuần trong giai đoạn từ Q1 năm 2010 đến Q3 năm 2024 của chuỗi SBT

Đồ thị thể hiện doanh thu thuần của SBT, nhìn vào đồ thị có thể thấy rằng doanh thu thuần có xu hướng tăng dần qua thời gian, đặc biệt rõ rệt từ khoảng 2018 trở đi. Giai đoạn từ 2010 – 2017 doanh thu tăng chậm, gần như đi ngang hoặc tăng nhẹ. Giai đoạn từ 2018 – 2024 doanh thu tăng mạnh theo xu hướng dốc đứng, cho thấy sự mở rộng, tăng trưởng quy mô. Mức độ biến động tăng dần cho thấy chuỗi đang trong giai đoạn tăng trưởng nhanh nhưng chưa hoàn toàn ổn định, vẫn chịu ảnh hưởng bởi các yếu tố bên ngoài. Nếu xu hướng này tiếp tục, SBT có tiềm năng tăng trưởng dài hạn, nhưng cũng cần kiểm soát rủi ro từ biến động ngắn hạn.

Thực hiện phân tích dựa trên dữ liệu doanh thu thuần của SBT (đtt: triệu đồng) được lấy theo quý từ Q1 năm 2010 đến Q4 năm 2023 (56 quan sát). Sử dụng dữ liệu từ Q1 năm 2010 đến hết Q4 năm 2023 (56 quan sát), với 10 mô hình phổ biến, dự báo cho 2024. Sau đó so sánh các kết quả đó với số liệu thực tế của 2024 và chọn ra mô hình tốt nhất, đồng thời dự báo cho 4 quý của năm 2025. Để lựa chọn mô hình tốt nhất thì chọn mô hình có MAPE thấp nhất. Dựa vào bảng kết quả ở phần phụ lục (phụ lục 1), theo mô hình Holt- Winters mùa vụ dạng nhân có: $MAPE = 7.59\%$ là mô hình có giá trị MAPE nhỏ nhất.

⇒ Mô hình Holt- Winters mùa vụ dạng nhân là mô hình tốt nhất được lựa chọn để dự báo cho doanh thu thuần 4 quý của năm 2025 có dạng:

$$\widehat{dtt}_t = (6361.226 + 167.821 \cdot k) \cdot s_{\text{tương ứng}}$$

Với $s_1 = 0.89860$, $s_2 = 1.07874$, $s_3 = 1.01022$, $s_4 = 1.12264$

Kết quả dự báo cho 4 quý năm 2025:

$$dtt_{2025Q1(t=61)}^F = (6361.226 + 167.821 \cdot k) \cdot s_{\text{tương ứng}} = 6469.506$$

$$dtt_{2025Q2(t=62)}^F = (6361.226 + 167.821 \cdot k) \cdot s_{\text{tương ứng}} = 7948.348$$

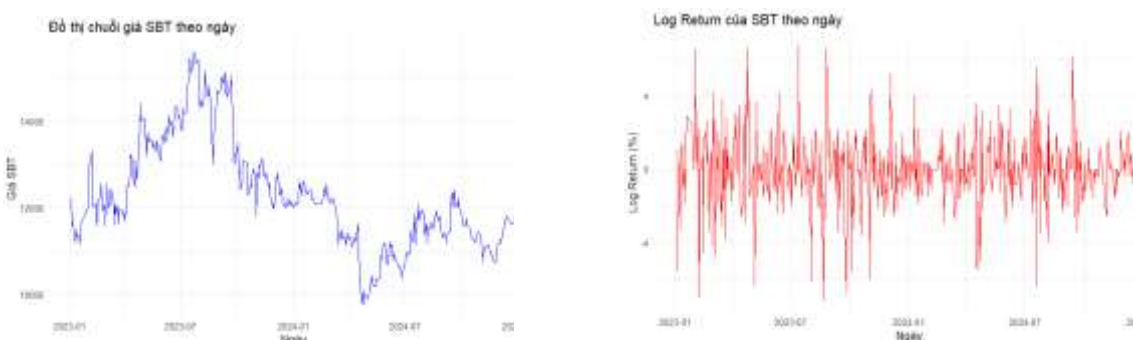
$$dtt_{2025Q3(t=63)}^F = (6361.226 + 167.821 \cdot k) \cdot s_{\text{tương ứng}} = 7612.976$$

$$dtt_{2025Q4(t=64)}^F = (6361.226 + 167.821 \cdot k) \cdot s_{\text{tương ứng}} = 8648.560$$

1.3. Phân tích và dự báo giá cổ phiếu theo mô hình ARIMA

1.3.1. Phân tích và định giá cổ phiếu

Số liệu: Chuỗi Log-return của mã chứng khoán SBT từ năm 2023 – 2024 (tính bằng chuỗi giá đóng cửa theo ngày, bao gồm 498 quan sát giá đóng cửa, đơn vị: VNĐ), bao gồm 497 quan sát, đơn vị %



Hình 1.3a. Đồ thị chuỗi giá và chuỗi log return SBT

Nhận xét: Nhìn vào hai đồ thị ta thấy, có thể cho rằng giá đóng cửa SBT không có tính dừng trong giai đoạn quan sát, khi thể hiện rõ xu hướng biến động theo thời gian. Thời gian những tháng đầu năm 2023, giá cổ phiếu SBT dao động quanh mức khoảng 11000 – 12000 đồng/cổ phiếu, sau đó có xu hướng tăng mạnh và đạt đỉnh gần 15500 đồng/cổ phiếu. Tuy nhiên, ở giai đoạn tiếp theo từ 8/2023 – 4/2024 giá giảm đáng kể và trải qua nhiều biến động phức tạp, có thời điểm chạm mức thấp dưới 10000 đồng/cổ phiếu, phản ánh những khó khăn trong hoạt động kinh doanh hoặc tác động từ điều kiện thị trường chung từ tình hình kinh tế sau đại dịch Covid-19. Trong khoảng thời gian tiếp theo, giá cổ phiếu bắt đầu phục hồi nhẹ và giao động trong khoảng 11000 – 13000 đồng/cổ phiếu, cho thấy SBT đang dần ổn định trở lại sau giai đoạn giảm sâu. Đây là dấu hiệu tích cực cho thấy niềm tin của nhà đầu tư đang dần được củng cố, cũng như sự kì vọng vào khả năng phục hồi của doanh nghiệp trong tương lai. Điều này cũng phản ánh phần nào sự thích nghi của SBT với biến động của thị trường và các yếu tố bên ngoài.

Trong khi đó, chuỗi log return của SBT có thể được xem là chuỗi dừng, khi dao động ổn định quanh mức trung bình 0, với biên độ chủ yếu trong khoảng (-6%, 6%). Điều này phản ánh rằng mặc dù giá cổ phiếu có biến động đáng kể theo thời gian, nhưng lợi suất hàng ngày mà nhà đầu tư nhận được không quá bất thường, và không có xu thế tăng hoặc giảm rõ rệt trong dài hạn.

1.3.2. Mô hình dự báo giá cổ phiếu cho 10 phiên đầu năm 2025

Bước 1: Kiểm định tính dừng của chuỗi giá SBT ban đầu \Rightarrow Chuỗi không dừng

- Kiểm định ADF có xu thế : Với cặp giả thuyết

$$\begin{cases} H_0: \text{Chuỗi có nghiệm đơn vị (Chuỗi không dừng)} \\ H_1: \text{Chuỗi không có nghiệm đơn vị (Chuỗi dừng vây quanh xu thế)} \end{cases}$$

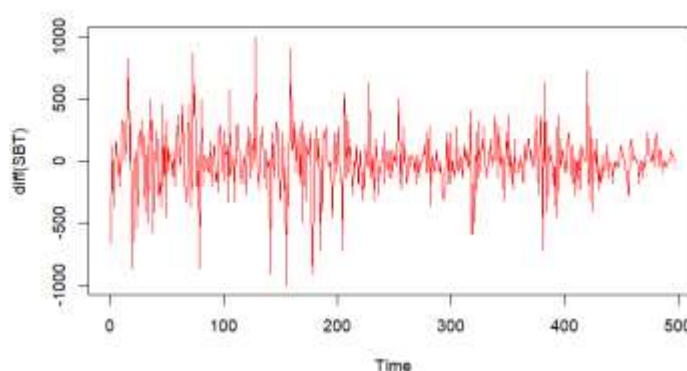
Kết quả thu được:

	Estimate	Std. Error	t value	Pr (> t)
Intercept	427.0	145.4	2.937	0.00347
z.lag.1	- 0.03094	0.01061	-2.916	0.00371
tt	0.1849	0.09252	-1.998	0.04626
z.diff.lag	-0.0007637	0.04465	-0.017	0.98636
Value of test-statistic is: -2.9159 2.8850 4.3265				
Critical values for test statistics: 1pct 5pct 10pct				
Tau1 -3.98 -3.42 -3.13				

Với mức ý nghĩa 5%, $p - value = 0.04626 < 0.05$, cho thấy hệ số xu thế có ý nghĩa thống kê.

Vì $|\tau_{qs}| = 2.9159 < |\tau_{0.1}| = 3.13$, chưa bác bỏ H_0 , do đó chuỗi giá SBT ban đầu không dừng.

Bước 2: Kiểm định tính dừng cho chuỗi sai phân ΔSBT_t



Hình 1.3b. Đồ thị chuỗi sai phân của giá SBT

- Kiểm định ADF có xu thế, thu được kết quả như sau:

	Estimate	Std. Error	t value	Pr (> t)
Intercept	6.87431	21.97820	0.313	0.755
z.lag.1	-1.00215	0.06407	-15.642	<2e-16
tt	-0.02654	0.07656	-0.347	0.729
z.diff.lag	-0.01089	0.04478	-0.243	0.808

Do $p - value = 0.729 > 0.05$, do đó xu thế không có ý nghĩa thống kê.

- Kiểm định ADF chỉ có hệ số chặn, thu được kết quả như sau:

	Estimate	Std. Error	t value	Pr (> t)
Intercept	0.26682	10.92807	0.024	0.981
z.lag.1	-1.00175	0.06400	-15.653	<2e-16
z.diff.lag	-0.01103	0.04474	-0.246	0.805

Do $p - value = 0.981 > 0.05$, do đó xu thế không có ý nghĩa thống kê.

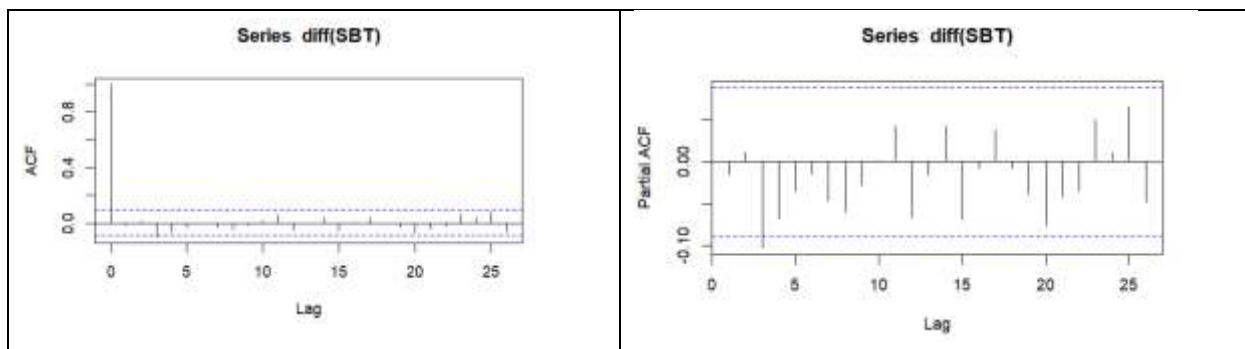
- Kiểm định None, thu được kết quả như sau:

Value of test-statistic is: -15.6684
 Critical values for test statistics: 1pct 5pct 10pct
 Tau1 -2.58 -1.95 -1.62

Như vậy, với mức ý nghĩa 1%, chuỗi ΔSBT_t dừng $\Rightarrow SBT_t \sim I(1)$

Bước 3: Chọn mô hình ARIMA phù hợp để ước lượng, đánh giá và dự báo

Để xác định bậc $ARMA(p,q)$ của chuỗi ΔSBT_t , ta cần vẽ đồ thị ACF và PACF của chuỗi.



Hình 1.3c. Ước đồ tương quan của chuỗi sai phân của chuỗi giá

Từ đồ thị của ACF và PACF, ta đề xuất 2 mô hình tiêu biểu để ước lượng chuỗi giá đóng cửa SBT_t là $ARIMA(5,1,1)$, $ARIMA(3,1,3)$. Thực hiện ước lượng 2 mô hình ARIMA, kết quả ước lượng mô hình được ghi trong phụ lục. Ta sẽ đi xét ý nghĩa thống kê của các hệ số AR theo công thức sau:

$$|t_{qs}| = \left| \frac{coef}{se} \right|$$

Và so sánh với giá trị tới hạn $z_{0.05} = 1.96$. Nếu giá trị quan sát lớn hơn 1.96 thì hệ số đó có ý nghĩa thống kê và ngược lại, ta thu được kết quả sau:

Mô hình	ARIMA(5,1,1)	ARIMA(3,1,3)
Hệ số có ý nghĩa thống kê	AR(1), MA(1)	AR(1), AR(2), AR(3), MA(1), MA(2), MA(3)
AIC	6878.59	6877.191
MAPE	1.356553	1.354277

Nhận xét: Dựa vào các kết quả từ mô hình thì mô hình $ARIMA(3,1,3)$ có AIC và MAPE nhỏ nhất, tất cả các hệ số của các biến đều có ý nghĩa thống kê. Do vậy, ta có thể nhận định rằng mô hình $ARIMA(3,1,3)$ là mô hình dự báo giá cổ phiếu SBT tốt nhất

Bước 4: Kiểm định về mô hình

Kết quả phân tích được trình bày dưới phần phụ lục 1.2 cho thấy rằng cả 2 mô hình ARIMA sử dụng đều có nghiệm nằm trong đường tròn đơn vị và không có tự tương quan giữa các bậc, phần dư hay nhiễu trắng.

Bước 5: Dự báo giá cổ phiếu SBT cho 10 phiên đầu năm 2025

Bảng 1.3a: Dự báo giá cổ phiếu SBT cho 10 phiên đầu năm 2025

Ngày	ARIMA(5,1,1)	ARIMA(3,1,3)	Giá thực tế
2/1/2025	11800	11800	12850
3/1/2025	11783	11789	12550
6/1/2025	11770	11776	12150
7/1/2025	11761	11765	12150
8/1/2025	11755	11757	12300
9/1/2025	11750	11752	12600
10/1/2025	11746	11747	11700
13/1/2025	11742	11742	11800
14/1/2025	11739	11738	11800
15/1/2025	11737	11736	11850
RMSE	550.2731	548.278	
MAPE	3.414413 %	3.400734 %	

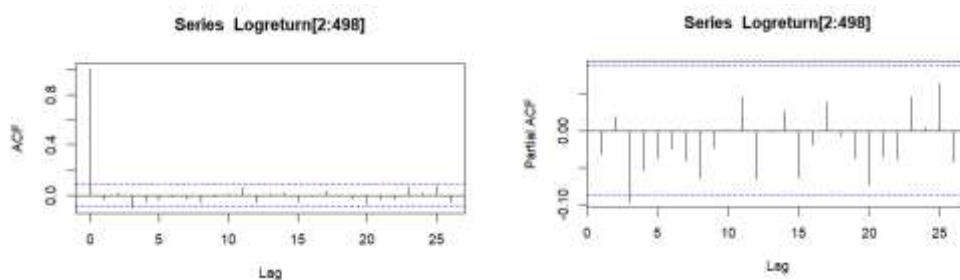
Nhận xét: Từ bảng trên, ta thấy mô hình ARIMA(3,1,3) là mô hình dự báo tốt nhất với sai số RMSE và MAPE đều nhỏ nhất trong 2 mô hình. Điều này cho thấy nhận xét ở bước 3 là hoàn toàn chính xác.

• **Đối với chuỗi log return**

Bước 1: Kiểm định tính dừng của chuỗi log return \Rightarrow Chuỗi dừng

Thực hiện các kiểm định tương tự như chuỗi giá đóng cửa, ta thu được kết quả ở phụ lục 2.2. Kết quả cho thấy chuỗi log return dừng, hay log return $\sim I(0)$.

Bước 2: Chọn mô hình ARIMA phù hợp để ước lượng, kiểm tra, đánh giá và dự báo.



Hình 1.3d. Lược đồ tương quan của chuỗi log return SBT

Từ đồ thị của ACF và PACF, ta đề xuất 2 mô hình tiêu biểu để ước lượng chuỗi giá đóng cửa SBT_t là ARIMA(1,0,0), ARIMA(3,0,0), ước lượng mô hình thu được kết quả

Mô hình	ARIMA(1,0,0)	ARIMA(1,0,1)
AIC	2075.62	2076.52
MAPE	1.940936%	1.93878%

Nhận xét: Dựa vào các kết quả cho thấy mô hình mặc dù mô hình ARIMA(1,0,1) có AIC cao hơn nhưng tất cả các biến đều có ý nghĩa thống kê, do vậy có thể mô hình ARIMA(1,0,1) cho dự báo tốt hơn.

Bước 3: Kiểm định mô hình

Kết quả phân tích được trình bày dưới phần phụ lục 1.3 cho thấy rằng cả 2 mô hình ARIMA sử dụng đều có nghiệm nằm trong đường tròn đơn vị và không có tự tương quan giữa các bậc, phần dư hay nhiễu trắng.

Bước 4: Dự báo log return cho 10 phiên đầu năm 2025

Bảng 1.3b: Dự báo log return và giá cổ phiếu SBT cho 10 phiên đầu năm 2025

Phiên	ARIMA (1,0,0)	Giá dự báo	ARIMA (1,0,1)	Giá dự báo	Giá thực tế
449	-0.006779	11738	0.004338	11869	12850
500	-0.006560	11661	-0.014169	11702	12550
501	-0.006567	11585	-0.001375	11686	12150
502	-0.006567	11509	-0.010219	11567	12150
503	-0.006567	11434	-0.004105	11520	12300
504	-0.006567	11359	-0.008332	11424	12600
505	-0.006567	11285	-0.005410	11363	11700
506	-0.006567	11211	-0.007430	11279	11800
507	-0.006567	11137	-0.006033	11211	11800
508	-0.006567	11064	-0.006999	11133	11850
RMSE		813.85		739.868	

Từ bảng trên ta thấy mô hình ARIMA(1,0,1) là mô hình dự báo tốt nhất cho log return với sai số dự báo RMSE nhỏ hơn so với mô hình ARIMA(1,0,0). Điều này cho thấy nhận xét ở bước 2 là chính xác.

Nhận xét: Từ 2 cách dự báo 10 phiên giá đóng cửa đầu năm 2025 của SBT ta thấy:

- Với cách dự báo trực tiếp, ta thấy mô hình ARIMA(3,1,3) là mô hình dự báo tốt nhất với sai số dự báo $RMSE = 548.278$, $MAPE = 3.400734\%$
- Với cách dự báo thông qua dự báo log return, mô hình ARIMA(1,0,1) là mô hình dự báo tốt nhất với sai số dự báo $RMSE = 739.868$

Như vậy, cách dự báo trực tiếp chuỗi giá đóng cửa cho SBT là tối ưu nhất với sai số dự báo của cách này nhỏ hơn so với dự báo thông qua dự báo log return.

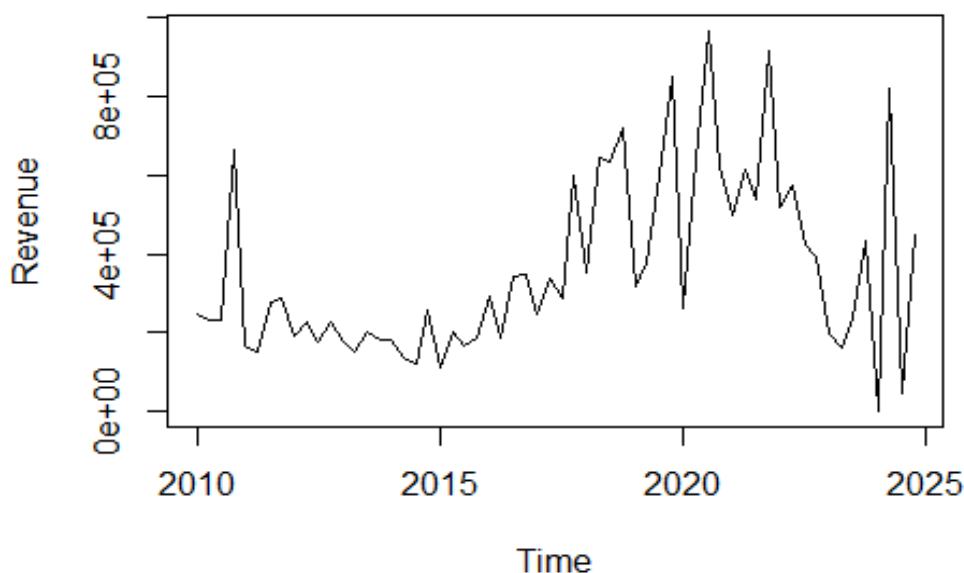
2. Phân tích chuỗi DIG – Trần Đức Hiền

2.1. Giới thiệu về Tổng Công ty Cổ phần Đầu tư Phát triển Xây dựng (DIC Corp) và mã cổ phiếu DIG

Tổng Công ty Cổ phần Đầu tư Phát triển Xây dựng (DIC Corp), một doanh nghiệp lâu đời và có tiếng trong lĩnh vực bất động sản và xây dựng tại Việt Nam. Công ty hoạt động đa ngành, trong đó phát triển đô thị, đầu tư hạ tầng và kinh doanh bất động sản là trọng tâm. Các dự án lớn như Khu đô thị Long Tân, Gateway Vũng Tàu hay Nam Vĩnh Yên đã giúp DIG khẳng định vị thế trong ngành. Được niêm yết trên sàn Sở Giao dịch Chứng khoán TP. Hồ Chí Minh (HOSE) từ năm 2006, cổ phiếu DIG thường thu hút sự quan tâm của nhà đầu tư nhờ tiềm năng tăng trưởng, tính thanh khoản tốt và khả năng hưởng lợi từ chu kỳ phục hồi bất động sản. Tuy nhiên, DIG cũng đối mặt với một số rủi ro như biến động theo thị trường, áp lực tài chính lớn do đầu tư dàn trải và phụ thuộc vào chính sách tín dụng. Với nền tảng phát triển vững chắc và chiến lược mở rộng hợp lý, DIG được xem là cổ phiếu tiềm năng trong nhóm bất động sản trung hạn và dài hạn.

2.2. Phân tích và dự báo chuỗi doanh thu thuần

Đối với chuỗi doanh thu thuần, thực hiện chia data set thành hai phần: train set bao gồm dữ liệu từ 2010Q1 – 2023Q4 và validation set bao gồm dữ liệu 4 quý 2024. Sử dụng dữ liệu từ tập train để xây dựng mô hình dự báo cho 4 quý 2024, từ đó so sánh đối chiếu với dữ liệu thật đã có từ tập validation để lựa chọn mô hình dự báo tốt nhất dựa trên tiêu chí sai số dự báo RMSE hoặc MAPE là nhỏ nhất. Sau đó, sử dụng dữ liệu của toàn bộ giai đoạn để dự báo cho năm 2025.



Hình 2.1. Doanh thu thuần của DIC Corp giai đoạn 2010Q1 – 2024Q4 (đơn vị: triệu Việt Nam Đồng)

Dựa vào hình 2.1, có thể thấy doanh thu của DIC Corp có xu hướng tăng trưởng đều qua các năm, đặc biệt bùng nổ trong giai đoạn 2018 – 2021. Tuy có xu hướng suy thoái vào những

năm 2022 và 2023 do ảnh hưởng của Covid-19 và suy thoái kinh tế toàn cầu, nhưng hoạt động kinh doanh của công ty này cũng đã có xu hướng phục hồi. Đặc biệt, có thể thấy rõ tính mùa vụ trong chuỗi doanh thu thuần của công ty này.

2.2.1. Mô hình hồi quy giản đơn theo xu thế.

Trong phần này, thực hiện hồi quy chuỗi doanh thu thuần theo biến xu thế thời gian theo 4 loại mô hình: lin-lin, log-log, lin-log và log-lin. Từ đó, lựa chọn mô hình với RMSE nhỏ nhất.

Sau khi thực hiện hồi quy theo cả bốn phương pháp thu được kết quả RMSE và MAPE như sau:

Bảng 2.1. Kết quả sai số dự báo thu được từ 4 mô hình

	Lin-lin	Log-log	Lin-log	Log-lin
RMSE	361609.3	467742	377001.4	467741.9
MAPE	89.74833	0.9938915	27.87442	0.9937319

Có thể thấy, có sự khác biệt trong hai chỉ số RMSE và MAPE. Cụ thể dù cho hai mô hình lin-lin và lin-log có giá trị sai số RMSE nhỏ hơn nhưng MAPE lại lớn hơn rất nhiều so với hai mô hình log-log và log-lin. Điều này có thể do sự khác biệt trong cách đo lường của hai chỉ số này. Cụ thể: RMSE là phép đo sai số tuyệt đối, còn MAPE là phép đo sai số tương đối vì vậy RMSE dễ nhạy cảm với các giá trị ngoại lai (Outliers). Đặc biệt, trong mẫu dữ liệu thu thập được có phát hiện giá trị ngoại lai, giá trị quan sát doanh thu thuần vào quý 1 và quý 3 2024 chỉ đạt 489 và 47281 (triệu đồng), trong khi tất cả các quan sát còn lại đều không dưới 100000, có thể đây là lý do dẫn đến sự sai lệch trong sai số đo từ RMSE.

Tuy vậy, vẫn thực hiện lựa chọn mô hình dựa trên hai phép đo này, cụ thể mô hình lin-log sẽ được lựa chọn do cả RMSE và MAPE thu được từ mô hình này đều không lớn nhất so với các mô hình còn lại. Sau khi đã lựa chọn được mô hình, thực hiện hồi quy doanh thu thuần theo biến xu thế thời gian đối với toàn bộ dữ liệu hết năm 2024. Kết quả thu được như sau:

$$\widehat{Revenue}_t = 76160 + 90126 * \ln(t)$$

Việc giải thích ý nghĩa của hệ số biến xu thế thời gian là không có ý nghĩa, vì vậy sẽ chỉ sử dụng kết quả để dự báo cho 4 quý năm 2025.

Bảng 2.2. Kết quả dự báo từ mô hình lin-log

	2025Q1	2025Q2	2025Q3	2025Q4
Giá trị dự báo	446656.618	448122.11	449564.16	450983.5

Tuy nhiên kết quả trên có vẻ không quá đúng với thực tế vì, quan sát mẫu dữ liệu cho thấy dữ liệu về doanh thu thuần của công ty này có yếu tố mùa vụ.

2.2.2. Mô hình hồi quy có yếu tố mùa vụ

Do ở phần trên mô hình lin-log được lựa chọn, vì vậy ở phần này vẫn lựa chọn mô hình n hồi quy theo xu thế thời gian và có yếu tố mùa vụ dạng lin-log, tổng cộng thực hiện hồi quy 5 mô hình: mô hình dạng cộng, mô hình dạng nhân, mô hình kết hợp dạng cộng và nhân, mô hình Holtwinter dạng cộng và mô hình Holtwinter dạng nhân.

Sau khi thực hiện hồi quy cả năm mô hình thu được kết quả RMSE và MAPE như sau:

Bảng 2.3. Kết quả sai số dự báo thu được từ 5 mô hình

	Dạng cộng	Dạng nhân	Kết hợp	Holtwinter dạng cộng	Holtwinter dạng nhân
RMSE	367745.9	371760.2	391206.6	453000.5	334062.8
MAPE	4.653	29.76	66.34	7.203	99.45

Kết quả thu được từ các mô hình vẫn cho ra kết quả giống như phần trên, khi RMSE và MAPE có sự khác biệt. Trong đó: mô hình Holtwinter dạng cộng cho MAPE thấp nhất nhưng RMSE cao nhất và dạng nhân cho RMSE nhỏ nhất và MAPE lớn nhất. Vẫn có thể giải thích nguyên nhân đến từ các giá trị quan sát ngoại lai (outliers).

Do khó khăn trong việc lựa chọn nên sẽ thực hiện dự báo trên cả hai mô hình Holtwinter dạng cộng và dạng nhân dành cho chuỗi doanh thu thuần vào 4 quý năm 2025. Kết quả thu được như sau:

Bảng 2.4. Kết quả mô hình Holtwinter dạng cộng và nhân

Hệ số	Holtwinter dạng cộng	Holtwinter dạng nhân
Intercept	252420.7	2.682574e+05
Trend	-17855.2	-5.305746e+03
S1	-191406.0	5.813094e-01
S2	168379.6	1.416674e+00
S3	-105517.0	8.408176e-01
S4	148978.8	1.561215e+00

Bảng 2.5. Kết quả dự báo cho 4 quý 2025

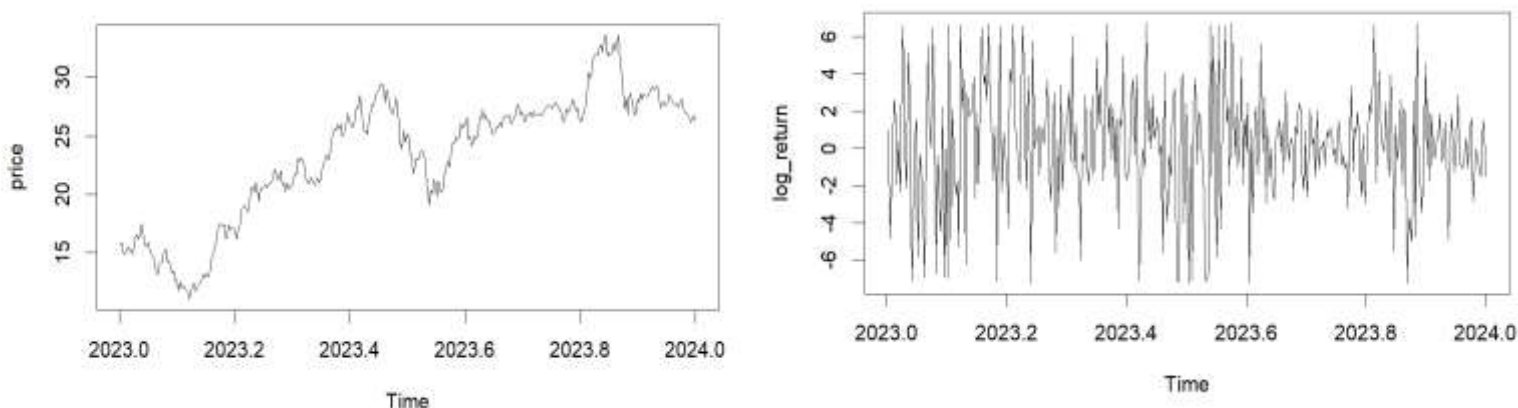
Dự báo	Holtwinter dạng cộng	Holtwinter dạng nhân
2025Q1	43159.54	152856.3
2025Q2	385089.91	365000.1
2025Q3	93338.08	212172.1
2025Q4	329978.67	385673.8

Do chưa có dữ liệu thực tế về doanh thu thuần năm 2025 nên chưa thể thực hiện tính toán sai số dự báo cho hai mô hình.

2.3. Phân tích và dự báo chuỗi giá cổ phiếu

Sử dụng dữ liệu giá đóng của mã cổ phiếu DIG trong hai năm 2023, 2024 để xây dựng mô hình ARIMA, từ đó tiến hành dự báo cho giá của 10 phiên đầu tiên năm 2025. Quy trình như

sau: Thực hiện kiểm định ADF để kiểm tra tính dừng của hai chuỗi giá đóng cửa và log_return, trên cơ sở đó xác định bậc tích hợp d trong mô hình ARIMA. Về bậc p và q, do kết quả thu được từ lược đồ tương quan ACF và PACF đưa đến kết luận rõ ràng nên sẽ thực hiện ước lượng các mô hình ARIMA với bậc p,q bất kì (nhỏ hơn 3) và lựa chọn mô hình có tiêu chí AIC nhỏ nhất.



Hình 2.2. Chuỗi giá đóng cửa và log_return của mã cổ phiếu DIG

Dựa vào hai biểu đồ trên, có thể phần nào nhìn thấy, chuỗi giá đóng cửa là một chuỗi không dừng, trong khi chuỗi log_return có thể là một chuỗi dừng. Thực hiện kiểm định ADF dành cho hai chuỗi trên, kết quả như sau:

Bảng 2.6 Kết quả kiểm định ADF

Chuỗi	Type = trend	Type = drift	Type = none	Kết luận
Giá đóng cửa	Not significant	Significant $ \tau_{qs} = 1.7 < \tau_{0.1} = 2.7$	(-)	Chuỗi không dừng
Log_return	Not significant	Not significant	$ \tau_{qs} = 15.02 > \tau_{0.01} = 2.58$	Chuỗi dừng
Sai phân bậc 1 của giá đóng cửa	Not significant	Not significant	$ \tau_{qs} = 15.06 > \tau_{0.01} = 2.58$	Chuỗi dừng

Dựa vào kết quả của bảng trên, có thể kết luận chuỗi giá đóng cửa là I(1), chuỗi log_return là I(0). Từ đó có thể xác định bậc d cho mô hình ARIMA của hai chuỗi lần lượt là 1 và 0.

Bài tập lớn Phân tích chuỗi thời gian và dự báo

Sau khi xác định được bậc d, sẽ ước lượng mô hình ARIMA với bậc p,q sao cho tiêu chuẩn AIC là nhỏ nhất. Kết quả thu được đối với chuỗi giá đóng cửa là ARIMA (2,1,2) và chuỗi log_return là ARIMA(1,0,1):

$$\Delta \widehat{Price}_t = 0.5966 * \Delta Price_{t-1} + (-0.8785) * \Delta Price_{t-2} + (-0.6341) * u_{t-1} + 0.9577 * u_{t-2} \quad (1)$$

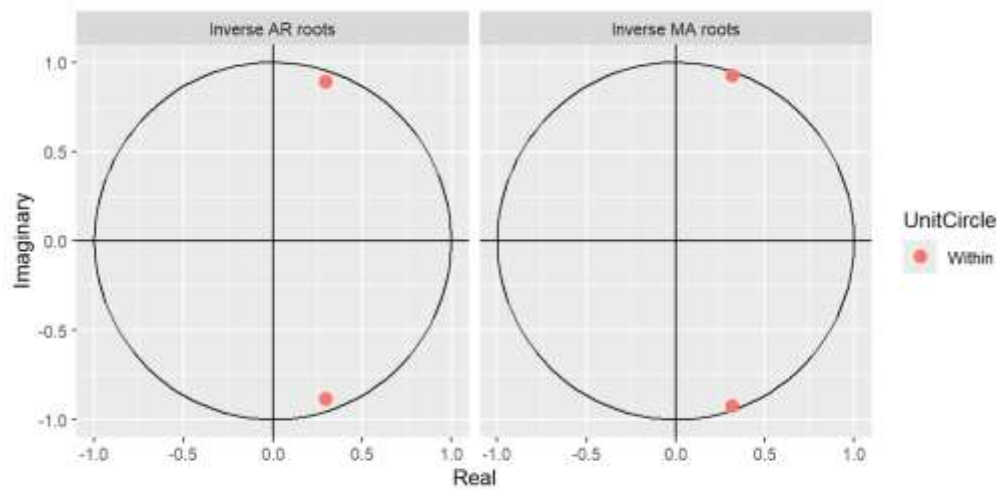
$$\widehat{\logreturn}_t = -0.569 * \logreturn_{t-1} + 0.518 * u_{t-1} \quad (2)$$

Sau đó, thực hiện dự báo 10 phiên đầu tiên trên hai mô hình trên, kết quả như sau:

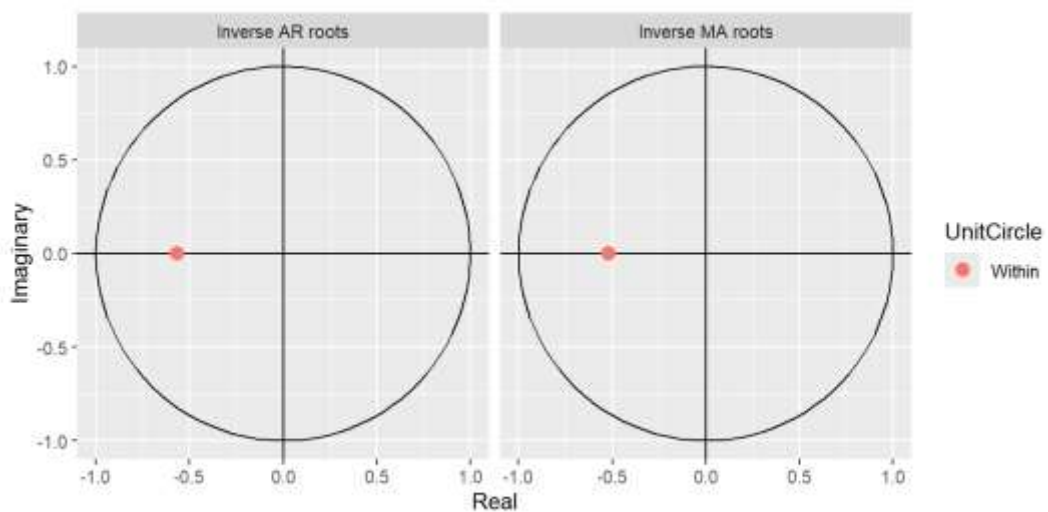
Phiên	(1)	(2)_logreturn	(2)_giá	Giá thực tế
1	19.203	-0.068	19.237	18.9
2	19.14	0.103	19.252	19.3
3	19.188	0.005	19.285	19.15
4	19.27	0.061	19.321	18.75
5	19.28	0.0296	19.36	18.1
6	19.21	0.0477	19.4	18.4
7	19.163	0.0374	19.435	18.15
8	19.195	0.0433	19.473	17.9
9	19.255	0.034	19.511	18
10	19.264	0.042	19.55	17.9
RMSE	0.926		1.1035	

Có thể thấy cả hai mô hình đều dự báo giá tăng đều trong 10 phiên (ngoại trừ phiên đầu tiên) trong khi giá thực tế lại giảm xuống dưới 19 (nghìn VND) kể từ phiên thứ 3. Để so sánh, mô hình dự báo trên chuỗi giá bằng ARIMA (2,1,2) cho kết quả gần với giá thực tế hơn khi có sai số dự báo RMSE nhỏ hơn.

Ngoài ra thực hiện kiểm tra tính dừng của hai mô hình ARIMA trên thông qua nghiệm nghiệm đảo. Kết quả thu được như sau:



Hình 2.3. Đồ thị kiểm định tính dừng qua nghiệm nghịch đảo mô hình $ARIMA(2,1,2)$



Hình 2.4. Đồ thị kiểm định tính dừng qua nghiệm nghịch đảo mô hình $ARIMA(1,0,1)$

Kết quả kiểm định cho thấy nghiệm nghịch đảo của cả hai mô hình dự báo $ARIMA(2,1,2)$ dành cho chuỗi giá đóng cửa và $ARIMA(1,0,1)$ dành cho chuỗi lợi suất đều nằm trong vòng tròn đơn vị, có thể kết luận cả hai chuỗi ước lượng từ hai mô hình đều hội tụ về dừng.

3. Phân tích chuỗi VIC – Lê Tuấn Minh

3.1. Giới thiệu về dữ liệu

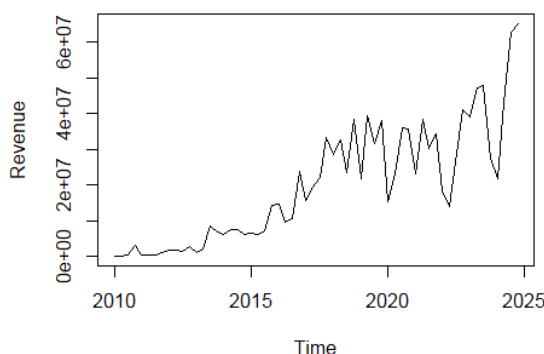
Tập đoàn VINGROUP được thành lập năm 2002, tiền thân là Công ty Cổ phần Thương mại Tổng hợp Việt Nam. Các lĩnh vực kinh doanh chính của công ty bao gồm Kinh doanh bất động sản; Dịch vụ cho thuê văn phòng, nhà ở, máy móc, thiết bị công trình; Kinh doanh khách sạn; Dịch vụ giải trí, vui chơi; Dịch vụ làm đẹp thẩm mỹ; Dịch vụ ăn uống, giải khát; Xây dựng Dân dụng và công nghiệp; Dịch vụ y tế. Mã chứng khoán của công ty là VIC, nằm trong nhóm ngành bất động sản và được niêm yết trên sàn HOSE từ năm 2007.

Trong bài viết này, dữ liệu được lựa chọn là chuỗi giá cổ phiếu của công ty từ phiên giao dịch ngày 3/1/2023 đến 31/12/2024 được lấy trên investing.com tính bằng đơn vị VNĐ và chỉ tiêu tài chính doanh thu thuần (sau đây gọi tắt là doanh thu) từ quý 1/2010 đến quý 4/2024 được lấy trên vietstock tính bằng đơn vị triệu VNĐ. Chuỗi log return là chuỗi được tính từ chuỗi giá cổ phiếu theo công thức:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \times 100 \text{ trong đó } r_t \text{ là log return ngày thứ } t, P_t \text{ là giá cổ phiếu ngày thứ } t$$

3.2. Các mô hình với chuỗi doanh thu

Dưới đây là đồ thị doanh thu của VINGROUP từ 2010 đến 2024. Có thể thấy doanh thu tăng trưởng tương đối ổn định trong giai đoạn 2010 – 2018 sau đó biến động mạnh theo từng quý.



Hình 3.1. Đồ thị của chuỗi doanh thu

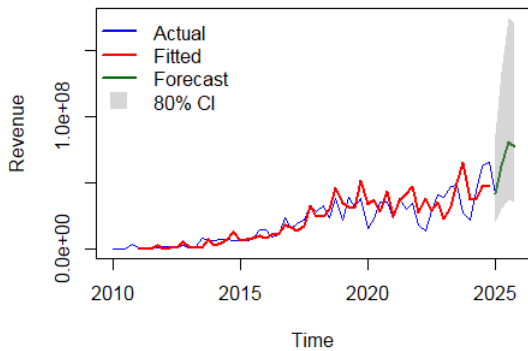
Để dự báo chuỗi doanh thu ta thực hiện chia dữ liệu thành tập train (Q1/2010 – Q4/2023) và tập validation (Q1/2024 – Q4/2024), xây dựng các mô hình khác nhau trên tập train và tính RMSE trên tập validation. Ta chọn mô hình có RMSE nhỏ nhất và cập nhật mô hình đó với dữ liệu từ tập validation rồi dự báo cho năm 2025.

Dựa trên kết quả từ 12 mô hình, kết luận mô hình Holt-Winter Additive đối với log của doanh thu cho kết quả RMSE và MAPE trên tập validation là nhỏ nhất, lần lượt là 12424017 và 0.2372553. Do đó, mô hình này được lựa chọn để dự báo cho năm 2025. Dưới đây là mô hình sau khi chạy lại với dữ liệu từ cả tập train và tập validation:

$$\begin{aligned} \ln(Rev_{t+h}^F) &= 17.58006382 + 0.07079279h + S_{tương ứng} \text{ với } S_1 = -0.09784664, S_2 \\ &= 0.24102779, S_3 = 0.41791477, S_4 = 0.30589349 \end{aligned}$$

Bảng 3.1. Kết quả từ các mô hình đối với chuỗi doanh thu

STT	Mô hình	R-sq	RMSE	MAPE
1	$Rev_t^F = -4622688 + 787311t$	0.7678	18040682	0.3963216
2	$Rev_t^F = -21068646 + \ln(t)$	0.5974	24926613	0.4320232
3	$\ln(Rev_t^F) = 13.528715 + 0.085860t$	0.7587	66703385	1.7528607
4	$\ln(Rev_t^F) = 10.78470 + 1.68665\ln(t)$	0.8847	16252063	0.3999299
5	$Rev_t^F = -7266372 + 778836t + 2758632s_2$ $+ 2937484s_3 + 5844797s_4$	0.788	16521261	0.3530398
6	$Rev_t^F = -3140029 + 626008t - 2391340s_2$ $- 3517209s_3 + 639010s_4$ $+ 189386s_2t + 233116s_3t$ $+ 188809s_4t$	0.798	14394458	0.2802641
7	$\ln(Rev_t^F) = 13.348854 + 0.085042t$ $+ 0.103707 + 0.156518s_3$ $+ 0.552510s_4$	0.7759	73293367	1.6030492
8	$\ln(Rev_t^F) = 13.1247969 + 0.0933405t$ $+ 0.1055050s_2 + 0.1876935$ $+ 1.4631882s_4$ $- 0.0003606s_2t$ $- 0.0016473s_3t$ $- 0.0311858s_4t$	0.7939	68265297	1.7688293
9	Holt-Winter Additive $Rev_{t+h}^F = 36598906.876 + 530299.405h$ $+ S_{\text{trend}}$		18707727	0.4390395
10	Holt-Winter Additive $\ln(Rev_{t+h}^F) = 17.24077979 + 0.07079279h$ $+ S_{\text{trend}}$		12424017	0.2372553
11	Holt-Winter Multiplicative $Rev_{t+h}^F = (-1487344$ $- 5.048134h)S_{\text{trend}}$		16593785	0.3447701
12	Holt-Winter Multiplicative $\ln(Rev_{t+h}^F) = (17.20702970$ $+ 0.07079279h)S_{\text{trend}}$		12585038	0.238489



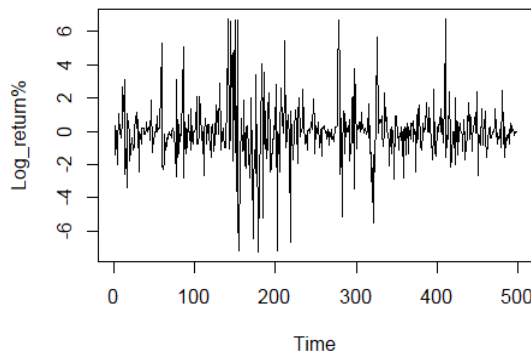
Thời kỳ	Dự báo điểm	Low 80	High 80
Q1	41992852	20339003	86700396
Q2	63254523	29957629	133559793
Q3	81032184	37550761	174862364
Q4	77759282	35279232	171389952

Hình 3.2. Đồ thị của mô hình Holt-Winter Additive với log doanh thu

Bảng 3.2. Kết quả dự báo cho năm 2025 (Đơn vị: Triệu VNĐ)

3.3. Các mô hình với giá cổ phiếu

3.3.1. Các mô hình với chuỗi Log return



Hình 3.3. Đồ thị của chuỗi log return

Nhìn qua đồ thị, ta có thể dự đoán chuỗi log return là dừng quanh 0. Để kiểm định tính dừng của chuỗi, ta sử dụng kiểm định ADF với các bước như sau:

- Bước 1: Kiểm định nghiệm đơn vị theo mô hình sau, với số lag j được lựa chọn qua tiêu chí AIC

$$\Delta r_t = \mu + \delta r_{t-1} + \beta t + \sum \gamma_j \Delta r_{t-j} + v_t$$

$$\begin{cases} H_0: \delta = 0, \text{ có nghiệm đơn vị, chuỗi không dừng} \\ H_1: \delta < 0, \text{ không có nghiệm đơn vị, chuỗi dừng quanh xu thế (chuỗi không dừng)} \end{cases}$$

Kết quả kiểm định cho thấy hệ số của xu thế β không có ý nghĩa thống kê ở mức 5% (p-value = 0.89) nên ta loại bỏ xu thế.

- Bước 2: Kiểm định $\Delta r_t = \mu + \delta r_{t-1} + \sum \gamma_j \Delta r_{t-j} + v_t$

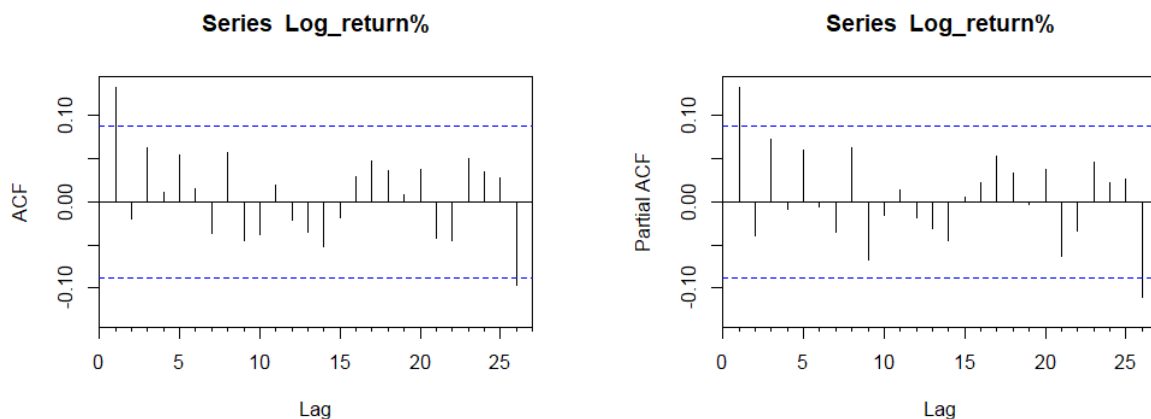
$$\begin{cases} H_0: \delta = 0, \text{ có nghiệm đơn vị, chuỗi không dừng} \\ H_1: \delta < 0, \text{ không có nghiệm đơn vị, chuỗi dừng quanh hệ số chặn} \end{cases}$$

Kết quả kiểm định cho thấy hệ số chặn μ không có ý nghĩa thống kê ở mức 5% (p-value = 0.449) nên ta loại bỏ hệ số chặn.

- Bước 3: Kiểm định $\Delta r_t = \delta r_{t-1} + \sum \gamma_j \Delta r_{t-j} + v_t$

$$\begin{cases} H_0: \delta = 0, \text{có nghiệm đơn vị, chuỗi không dừng} \\ H_1: \delta < 0, \text{không có nghiệm đơn vị, chuỗi dừng} \end{cases}$$

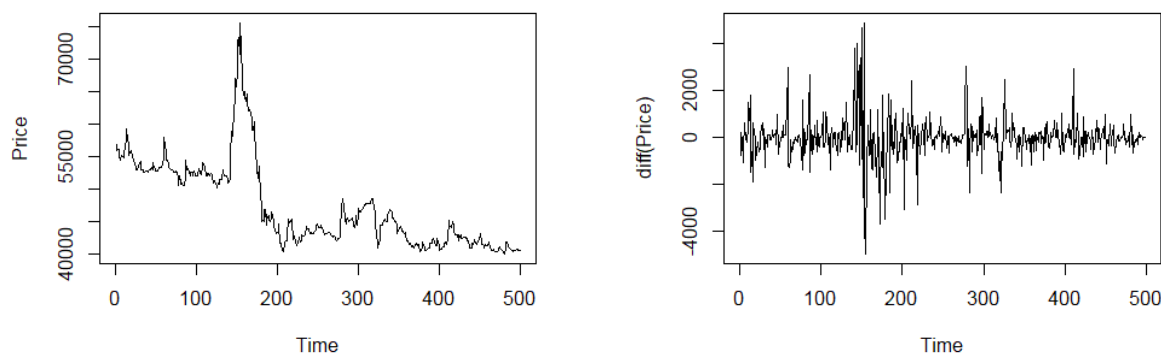
Vì $|\tau_{stat}| = |-15.1925| > |\tau_{0.05}| = |-2.87|$ nên bác bỏ H_0 , kết luận chuỗi dừng ở mức ý nghĩa 5%. Để xác định bậc của arima với log return, ta sử dụng lược đồ tương quan ACF và PACF



Hình 3.4. Lược đồ tương quan của chuỗi log return

Nhìn đồ thị, có thể thấy tương quan cao giữa biến log return với trễ của nó sau 1 thời kỳ, do đó các mô hình AR(1), MA(1) và ARIMA(1,0,1) được cân nhắc.

3.3.2. Các mô hình với chuỗi giá



Hình 3.5. Đồ thị chuỗi giá và chuỗi sai phân của chuỗi giá

Nhìn qua đồ thị, ta có thể nhận thấy chuỗi giá nhiều khả năng là chuỗi không dừng. Kiểm định ADF được thực hiện với chuỗi giá tương tự như chuỗi log return, bắt đầu bằng mô hình có xu thế.

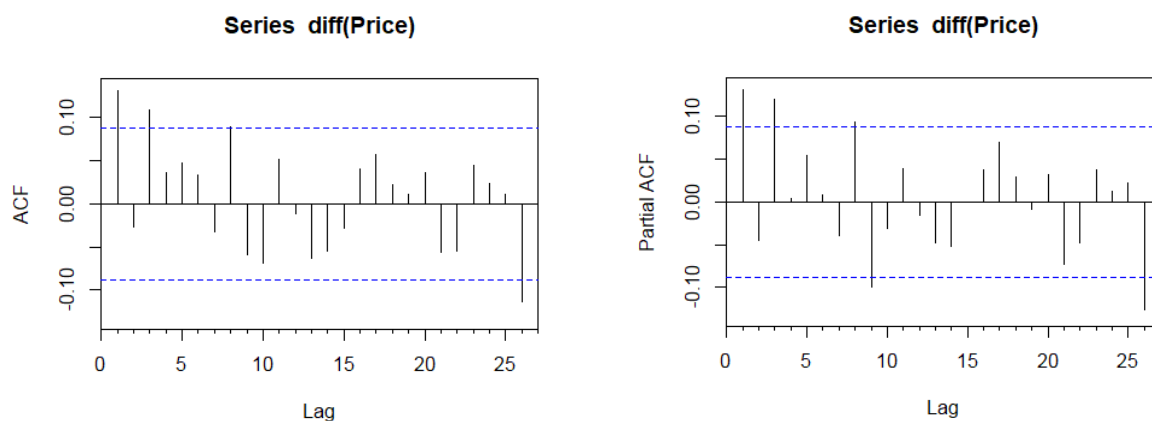
- Bước 1: $\Delta P_t = \mu + \delta P_{t-1} + \beta t + \sum \gamma_j \Delta P_{t-j} + v_t$

$$\begin{cases} H_0: \delta = 0, \text{có nghiệm đơn vị, chuỗi không dừng} \\ H_1: \delta < 0, \text{không có nghiệm đơn vị, chuỗi dừng quanh xu thế (chuỗi không dừng)} \end{cases}$$

Kết quả cho thấy hệ số β của xu thế có ý nghĩa thống kê ở mức 10% (p-value = 0.05137). Do đó ta dừng lại ở ngay bước 1, xét $|\tau_{stat}| = |-2.702| < |\tau_{0.1}| = |-3.13|$ và chưa bác bỏ H_0 , kết luận chuỗi không dừng.

Ta tạo ra chuỗi sai phân của chuỗi giá theo công thức $\Delta P_t = P_t - P_{t-1}$

Quá trình kiểm định ADF với chuỗi sai phân của chuỗi giá được thực hiện tương tự như đã làm với chuỗi log return nên từng bước sẽ không được trình bày chi tiết. Kết quả kiểm định cho thấy ở bước 1, hệ số của xu thế không có ý nghĩa (p – value = 0.873) và ở bước 2, hệ số chặn không có ý nghĩa ở mức 5% (p – value = 0.487). Bước 3, vì $|\tau_{stat}| = |-15.2901| > |\tau_{0.05}| = |-1.95|$ nên bác bỏ H_0 , kết luận chuỗi sai phân dừng ở mức ý nghĩa 5%. Điều này đồng nghĩa với việc chuỗi giá $P_t \sim I(1)$. Sau đây là lược đồ tương quan của chuỗi sai phân:



Hình 3.6. Lược đồ tương quan của chuỗi sai phân của chuỗi giá

Dựa trên đồ thị ta có thể nhận thấy tương quan của ΔP_t với trễ của nó ở các bậc 1,3,8,9,26. Tuy nhiên, tương quan ở các bậc cao là chưa rõ ràng nên chỉ các mô hình đối với P_t đến bậc 3 được cân nhắc, bao gồm ARIMA(1,1,3), ARIMA(3,1,1), ARIMA(3,1,3).

3.3.3. Kết quả

Dữ liệu năm 2023 và 2024 được sử dụng để chạy mô hình còn dữ liệu về log return cổ phiếu 10 phiên đầu năm 2025 được sử dụng làm tập validation để so sánh với dự báo từ các mô hình thông qua chỉ tiêu RMSE. Giá dự báo từ chuỗi giá được chuyển thành log return.

Bảng 3.3. Sai số dự báo RMSE từ các mô hình

Dữ liệu	Chuỗi Log return			Chuỗi giá		
ARIMA	(1,0,0)	(0,0,1)	(1,0,1)	(1,1,3)	(3,1,1)	(3,1,3)
RMSE	0.3916437	0.3916683	0.3921733	0.3919672	0.3919538	0.3929238

Kiểm định tự tương quan đến bậc 10 của phần dư bằng kiểm định Ljung – Box đối với tất cả các mô hình đều cho kết luận không có tự tương quan ở mức ý nghĩa 5%. Kiểm tra nghiệm nghịch đảo của các mô hình đều cho kết quả các nghiệm nghịch đảo nằm trong vòng tròn đơn vị. Kết quả chi tiết của các kiểm định này được trình bày ở phần phụ lục. Trong các mô hình trên, ARIMA(1,0,0) và ARIMA(0,0,1) với chuỗi log return cho RMSE nhỏ nhất, tuy nhiên, các giá trị dự báo hầu như bằng 0. Vì vậy, các giá trị dự báo từ mô hình ARIMA(3,1,1) (có RMSE chỉ lớn hơn khoảng 0.003) đối với chuỗi giá được lựa chọn để trình bày.

Bảng 3.4. Log return dự báo từ mô hình ARIMA(3,1,1) với chuỗi giá

Ngày	1	2	3	4	5	6	7	8	9	10
Log-return	-0.027	-0.01	-0.015	-0.003	-0.002	-0.001	0	0	0	0

4. Phân tích chuỗi LSS – Phạm Phương Thảo

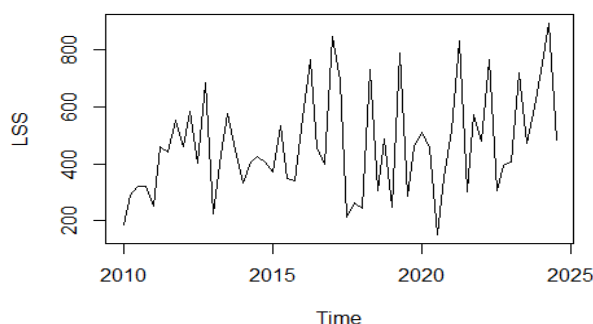
4.1. Giới thiệu mã cổ phiếu LSS

Công ty cổ phần mía đường Lam Sơn tiền thân là nhà máy đường Lam Sơn (31/03/1980.), được thành lập ngày 06 tháng 12 năm 1999, hoạt động trong lĩnh vực sản xuất mía đường và nông sản trong nước và xuất khẩu ra thế giới. Ngày 1/1/2000, Công ty đi vào hoạt động dưới hình thức Công ty cổ phần với vốn điều lệ 150 tỷ đồng. Công ty bắt đầu niêm yết với mã chứng khoán LSS trên Sở giao dịch T.P Hồ Chí Minh ngày 21/12/2007.

4.2. Phân tích chuỗi tài chính

Dữ liệu được sử dụng cho phần này là chuỗi tài chính bao gồm 60 quan sát về doanh thu thuần của LSS theo quý từ năm 2010 đến hết năm 2024 (Đơn vị: tỷ VND) được lấy trên website Fiiinpro.com.

Dưới đây là biểu đồ doanh thu thuần theo quý giai đoạn 2010-2024. Có thể thấy doanh thu thuần của LSS có độ biến động tương đối ổn định.



Hình 4.1: Biểu đồ doanh thu thuần LSS theo quý giai đoạn 2010-2024

Để dự báo doanh thu thuần cho 4 quý của năm 2025, ta sử dụng các dữ liệu từ Quý 1/2010 đến Quý 4/2023 để chạy các mô hình sau:

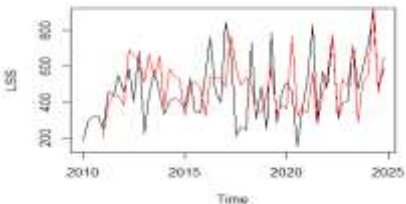
Bảng 4.1: Bảng kết quả ước lượng mô hình

STT	Mô hình	Ước lượng	RMSE	R-sq
1	Lin-lin	$\widehat{Revenue}_t = 386.872 + 2.333t$	215.74	0.031
2	Lin-log	$\widehat{Revenue}_t = 286.66 + 54.16\ln t$	226.64	0.062
3	Log-lin	$\ln \widehat{Revenue}_t = 5.907365 + 0.004774t$	242.85	0.021
4	Log-log	$\ln \widehat{Revenue}_t = 5.66514 + 0.12291\ln t$	250.27	0.061

5	Linear trend + Additive	$\widehat{Revenue}_t = 338.496 + 2.397t + 198.317 * S_2 + (-50.294) * S_3 + 38.166 * S_4$	171.53	0.293
6	Linear trend + Multiplicative	$\widehat{Revenue}_t = 294.725 + 4.018t + 3.581 * S_2 * t + (-6.487) * S_3 * t + (-3.579) * S_4 * t$	157.94	0.395
7	Holt-Winter + Additive	$Revenue_{T+k}^F = (642.7729 + 7.9003 * k) + S_k$ $S_1 = -111.829106$ $S_2 = 147.893398$ $S_3 = -228.556568$ $S_4 = -94.366600$	105.69	
8	Holt-Winter + Multiplicative	$Revenue_{T+k}^F = (716.7293 + 14.2255 * k) * S_k$ $S_1 = 0.7564842$ $S_2 = 1.2162047$ $S_3 = 0.5635833$ $S_4 = 0.7821507$	91.29	

Để đánh giá khả năng dự báo của mô hình, ta sử dụng chỉ tiêu RMSE để chọn ra mô hình dự báo doanh thu thuần 4 quý năm 2024 có sai số so với số liệu thực tế là nhỏ nhất. Theo bảng kết quả trên, mô hình Holt-Winter dạng nhân có RMSE nhỏ nhất nên được chọn làm mô hình để dự báo cho doanh thu thuần 4 quý năm 2025.

Áp dụng mô hình Holt-Winter dạng nhân với số liệu từ Quý 1/2010 đến Quý 4/2024, ta có kết quả ước lượng mô hình như sau:

$Revenue_{T+k}^F = (817.402 + 17.289 * k) * S_k$ $S_1 = 0.8222935$ $S_2 = 1.1836681$ $S_3 = 0.5730897$ $S_4 = 0.7571990$	 <p>Hình 2: Đồ thị dự báo theo mô hình Holt-Winter dạng nhân</p>
--	--

Dự báo doanh thu thuần 4 quý năm 2025 của LSS như sau:

Bảng 4.2: Kết quả dự báo doanh thu thuần theo mô hình Holt-Winter dạng nhân

Quý	Dự báo doanh thu thuần (đơn vị: tỷ VND)
2025Q1	686.362
2025Q2	1008.4639
2025Q3	498.1706
2025Q4	671.3035

4.3. Phân tích chuỗi giá đóng cửa và chuỗi log-return của LSS

4.3.1. Chuỗi giá

Dữ liệu được sử dụng trong phần này là chuỗi giá cổ phiếu LSS giai đoạn 2023-2024 được lấy trên website Fiinpro.com, bao gồm 499 quan sát từ ngày 03/01/2023 đến ngày 31/12/2024.



Hình 4.2: Đồ thị chuỗi giá cổ phiếu LSS giai đoạn 2023-2024

Trong trường hợp này, rất khó để kết luận chuỗi giá cổ phiếu LSS giai đoạn 2023-2024 là chuỗi dừng hay không dừng thông qua việc quan sát đồ thị. Vì vậy ta tiến hành kiểm định Dickey-Fuller để xác định chính xác.

Kiểm định Dickey-Fuller

Bước 1: Kiểm định nghiệm đơn vị có hệ số chặn và xu thế

Mô hình: $\Delta Price = \mu + \beta_t + \delta Price_{t-1} + v_t$

Cặp giả thuyết: $\begin{cases} H_0: \beta_t = 0 \text{ (Xu thế không có ý nghĩa thống kê)} \\ H_1: \beta_t \neq 0 \text{ (Xu thế có ý nghĩa thống kê)} \end{cases}$

Do $P - value = 0.45058 > 0.05 \rightarrow$ Chưa bác bỏ $H_0 \rightarrow$ Xu thế không có ý nghĩa thống kê

Bước 2: Kiểm định nghiệm đơn vị có hệ số chặn

Mô hình: $\Delta Price = \mu + \delta Price_{t-1} + v_t$

Cặp giả thuyết: $\begin{cases} H_0: \mu = 0 \text{ (Hệ số chặn không có ý nghĩa thống kê)} \\ H_1: \mu \neq 0 \text{ (Hệ số chặn có ý nghĩa thống kê)} \end{cases}$

Do $P - value = 0.00659 < 0.05 \rightarrow$ Bác bỏ $H_0 \rightarrow$ Hệ số chặn có ý nghĩa thống kê

Cặp giả thuyết 2: $\begin{cases} H_0: \delta = 0 \text{ (có nghiệm đơn vị hay chuỗi không dừng)} \\ H_1: \delta < 0 \text{ (không có nghiệm đơn vị hay chuỗi dừng)} \end{cases}$

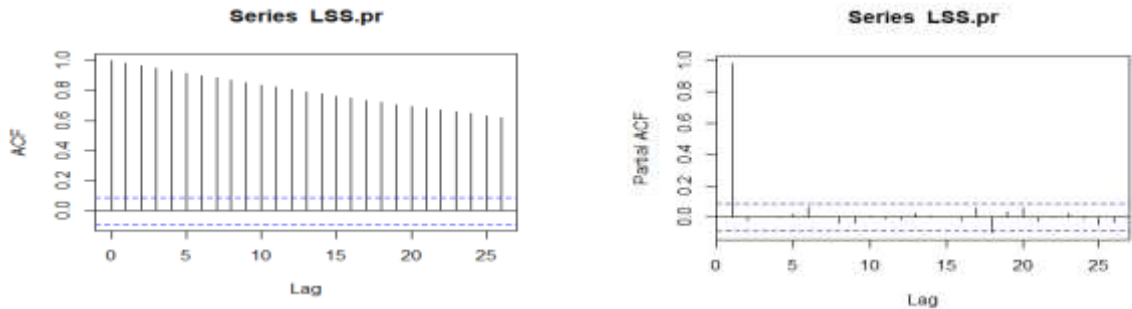
Do $\tau_{qs} = -2.6025$ mà $\tau_{0.1} = -2.57 \rightarrow |\tau_{qs}| > |\tau_{0.1}|$

\rightarrow Kết luận: Với mức ý nghĩa 10%, chuỗi giá dừng $\rightarrow LSS_{price} \sim I(0)$

Như vậy chuỗi giá có thể áp dụng mô hình ARMA(p,q)

Xác định bậc của mô hình ARMA(p,q)

Ta xác định bậc của AR(p) và MA(q) dựa vào đồ thị ACF và PACF



Dựa vào hình dạng đồ thị có thể xác định được đây là mô hình AR(1).

Đánh giá tính dừng của mô hình:

Do nghiệm nghịch đảo nằm bên trong nên có thể kết luận mô hình là dừng.

Đánh giá tính nhiễu trắng của phần dư:

Cặp giả thuyết: $\begin{cases} H_0: \text{Phần dư } u_t \text{ không có tự tương quan} \\ H_1: \text{Phần dư } u_t \text{ có tự tương quan} \end{cases}$

Sử dụng kiểm định Ljung-Box test ta có kết quả P-value = 0.1818 > 0.05 → Chưa bác bỏ H_0 → Kết luận: Phần dư không có tự tương quan

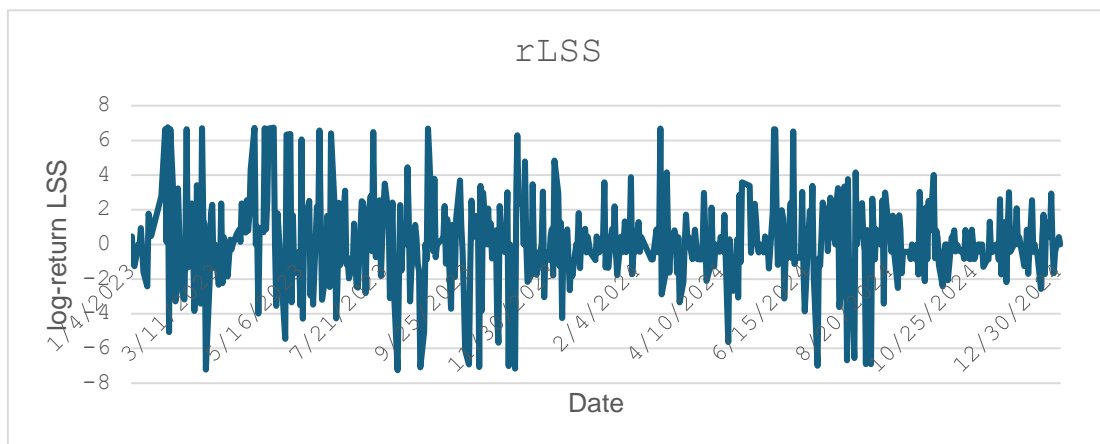
Như vậy có thể kết luận rằng mô hình AR(1) là thích hợp để dự báo cho giá cổ phiếu LSS trong tương lai gần.

Kết quả chạy mô hình AR(1) và những kiểm định trên được trình bày ở phụ lục.

4.3.2. Chuỗi log-return

Chuỗi log-return cổ phiếu LSS giai đoạn 2023-2024 bao gồm 498 quan sát được tính từ chuỗi giá theo công thức sau:

$$\log return = \ln \left(\frac{Price_t}{Price_{t-1}} \right) * 100$$



Hình 4.3: Biểu đồ chuỗi log-return của cổ phiếu LSS giai đoạn 2023-2024

Với các bước tương tự như chuỗi giá, ta tiến hành kiểm định tính dừng của chuỗi log-return

Kiểm định Dickey-Fuller

Bước 1: Kiểm định nghiệm đơn vị có hệ số chặn và xu thế

$P - value = 0.1756 > 0.05 \rightarrow$ Chưa bác bỏ $H_0 \rightarrow$ Xu thế không có ý nghĩa thống kê

Bước 2: Kiểm định nghiệm đơn vị có hệ số chặn

$P - value = 0.249 > 0.05 \rightarrow$ Chưa bác bỏ $H_0 \rightarrow$ Hệ số chặn không có ý nghĩa thống kê

Bước 3: Kiểm định nghiệm đơn vị không có hệ số chặn

Cặp giả thuyết: $\begin{cases} H_0: \delta = 0 \text{ (có nghiệm đơn vị hay chuỗi không dừng)} \\ H_1: \delta < 0 \text{ (không có nghiệm đơn vị hay chuỗi dừng)} \end{cases}$

Do $\tau_{qs} = -21.7211$ mà $\tau_{0.1} = -1.62 \rightarrow |\tau_{qs}| > |\tau_{0.1}| \rightarrow$ Bác bỏ H_0

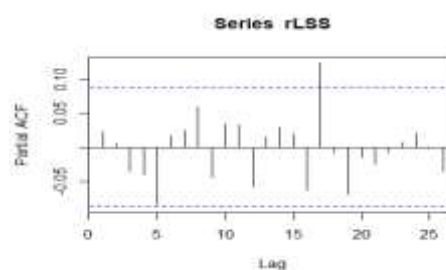
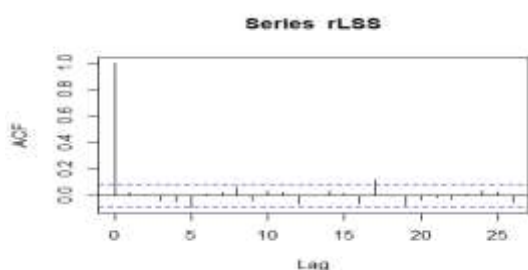
\rightarrow Kết luận: Với mức ý nghĩa 10%, chuỗi log – return dừng $\rightarrow LSS_{\ln(price)} \sim I(0)$

Như vậy chuỗi log-return có thể áp dụng mô hình ARMA(p,q).

Kết quả các kiểm định trên được trình bày ở phụ lục.

Xác định bậc của mô hình ARMA(p,q)

Ta xác định bậc của AR(p) và MA(q) dựa vào đồ thị ACF và PACF



Do hầu hết các giá trị trong cả 2 đồ thị đều nằm trong khoảng tin cậy nên có thể chuỗi lợi suất portfolio là nhiễu trắng. Thực hiện kiểm định Box-test để xác định chính xác hơn.

Cặp giả thuyết: $\begin{cases} H_0: \text{Chuỗi không có tự tương quan đến bậc 10} \\ H_1: \text{Chuỗi có ít nhất 1 bậc tự tương quan đến bậc 10} \end{cases}$

Kết quả cho thấy $P - value = 0.1873 > 0.05 \rightarrow$ Chưa đủ cơ sở bác bỏ H_0

Kết luận: Chuỗi lợi suất LSS không có tự tương quan đến bậc 10, hay chuỗi là nhiễu trắng. Không thể áp dụng mô hình ARIMA cho chuỗi này.

Do yêu cầu thêm từ người hướng dẫn, mô hình ARMA(1,1) sẽ được áp dụng cho dữ liệu chuỗi lợi suất thực tế của LSS trong năm 2023 và 2024 để dự báo cho lợi suất danh mục trong 10 phiên giao dịch tiếp theo năm 2025.

Đánh giá tính dừng của mô hình:

Theo vòng tròn đơn vị, các nghiệm nghịch đảo của các mô hình đều nằm bên trong nên có thể kết luận mô hình là dừng.

Đánh giá tính nhiễu trắng của phần dư:

Cặp giả thuyết: $\begin{cases} H_0: \text{Phần dư } u_t \text{ không có tự tương quan} \\ H_1: \text{Phần dư } u_t \text{ có tự tương quan} \end{cases}$

Bài tập lớn Phân tích chuỗi thời gian và dự báo

Sử dụng kiểm định Ljung-Box test ta có kết quả P-value = 0.41 > 0.05 → Chưa bác bỏ H_0

Kết luận: Phần dư không có tự tương quan

Như vậy có thể tạm thời kết luận rằng mô hình ARMA(1,1) là thích hợp để dự báo cho giá cổ phiếu LSS trong tương lai gần.

Kết quả ước lượng mô hình và các kiểm định được trình bày ở phụ lục.

Kết quả dự báo:

Log-return của 10 phiên giao dịch tiếp theo trong năm 2025 dự báo bằng các mô hình ARMA đề xuất sẽ được quy đổi trở lại giá cổ phiếu dự báo theo công thức:

$$Price_t = Price_{t-1} * \exp\left(\frac{\log - return_t}{100}\right)$$

Kết quả dự báo cho giá cổ phiếu LSS trong 10 phiên giao dịch tiếp theo trong năm 2025 và sai số so với giá thực tế theo chỉ tiêu RMSE khi áp dụng các mô hình dự báo bằng giá và dự báo bằng log-return như sau:

Bảng 4.3: Kết quả dự báo giá và thực tế

Phiên	Giá thực tế	AR(1)	ARMA(1,1)
1	12100.00	10635.12	10651.0924
2	12000.00	10624.57	10656.56649
3	11900.00	10614.10	10662.04603
4	11700.00	10603.71	10667.52845
5	11800.00	10593.40	10673.01369
6	11800.00	10583.16	10678.50176
7	11700.00	10572.99	10683.99264
8	11700.00	10562.90	10689.48635
9	10600.00	10552.88	10694.98288
10	10700.00	10542.94	10700.48224
RMSE		1114.43667	1053.231293

Nếu đánh giá khả năng dự báo của các mô hình theo chỉ tiêu RMSE thì mô hình ARMA(1,1) có khả năng dự báo tốt hơn do có RMSE nhỏ hơn. Như vậy cách dự báo giá cổ phiếu thông qua log-return sẽ có hiệu quả hơn trong ngắn hạn.

III. BÀI TẬP NHÓM

1. Phân tích tương quan giữa các cổ phiếu

Từ dữ liệu giá cổ phiếu của mỗi mã cổ phiếu, tính lợi suất mỗi cổ phiếu theo công thức

$$\log return = \ln \left(\frac{Price_t}{Price_{t-1}} \right) * 100$$

Sử dụng hàm covariance.s trong Excel để tính ma trận Covariance giữa lợi suất các mã cổ phiếu:

Bảng 5.1: Ma trận Covariance giữa lợi suất các mã cổ phiếu

COV	rSBT	rDIG	rVIC	rLSS
rSBT	3.772342345			
rDIG	2.211869122	8.073289746		
rVIC	0.765587148	1.574257355	3.074148058	
rLSS	3.416656527	2.706198758	1.031250359	6.597557572

Sử dụng công cụ Data Analysis trong Excel để tính toán ma trận Correlation giữa lợi suất các mã cổ phiếu:

Bảng 5.2.: Ma trận Correlation giữa lợi suất các mã cổ phiếu

COR	rSBT	rDIG	rVIC	rLSS
rSBT	1			
rDIG	0.400800825	1		
rVIC	0.22481573	0.316000719	1	
rLSS	0.684863748	0.370802648	0.228986631	1

Nhận xét: Nhìn chung, lợi suất của các mã cổ phiếu không tương quan mạnh với nhau, cao nhất là giữa 2 mã cùng lĩnh vực tiêu dùng thiết yếu, SBT và LSS, với hệ số tương quan không đến 0.7.

2. Xây dựng danh mục tối ưu

2.1. Bài toán tối đa hóa lợi suất kỳ vọng của danh mục trong điều kiện phương sai danh mục bằng trung bình nhân phương sai của lợi suất các cổ phiếu

Ta biểu diễn bài toán dưới dạng: Giả sử danh mục $P = \{w_1, w_2, w_3, w_4\}$ là danh mục với tỷ trọng đầu tư vào 4 mã cổ phiếu SBT, DIG, VIC, LSS lần lượt là w_1, w_2, w_3, w_4 .

Ta có lợi suất kỳ vọng và phương sai của danh mục được tính theo công thức như sau:

$$E(r_p) = \sum_{i=1}^n w_i E(r_i)$$

$$var(r_p) = \sum_{i=1}^n \sum_{j=1}^n w_i w_j Cov(r_i, r_j)$$

Dựa vào ma trận covariance tính được trung bình nhân phương sai lợi suất là 4.985312671

Sử dụng công cụ Solver trong Excel để giải bài toán tối ưu $E(r_p) \rightarrow \max$ với điều kiện

$$\begin{cases} var(r_p) = 4.985312671 \\ 0 \leq w_1, w_2, w_3, w_4 \leq 1 \\ w_1 + w_2 + w_3 + w_4 = 1 \end{cases}$$

Ta xây dựng được danh mục với tỷ trọng đầu tư vào các mã như sau:

Bảng 5.1: Kết quả tỷ trọng đầu tư mỗi mã cổ phiếu của bài toán tối đa hóa lợi suất

wSBT	wDIG	wVIC	wLSS
0	0.175297497	0.05829163792	0.76641086475

với lợi suất danh mục:

$$E(r_p) = 0.108646525$$

Nhận xét: Khi kiểm soát mức độ rủi ro sao cho rủi ro danh mục bằng với trung bình mức độ rủi ro của từng cổ phiếu, nếu nhà đầu tư phân bổ vốn vào 4 mã cổ phiếu với tỷ trọng như trên thì sẽ đạt được lợi suất cao nhất.

2.2. Bài toán thiểu hóa rủi ro danh mục trong điều kiện lợi suất kỳ vọng của danh mục bằng lợi suất của thị trường

Lợi suất của thị trường tính theo trung bình lợi suất của mã VNIndex trong năm 2023 và 2024 là 0.038858343.

Sử dụng công cụ Solver trong Excel để giải bài toán tối ưu $var(r_p) \rightarrow \min$ với điều kiện

$$\begin{cases} E(r_p) = 0.038858343 \\ 0 \leq w_1, w_2, w_3, w_4 \leq 1 \\ w_1 + w_2 + w_3 + w_4 = 1 \end{cases}$$

Ta xây dựng được danh mục với tỷ trọng đầu tư vào các mã như sau:

Bảng 5.2: Kết quả tỷ trọng đầu tư mỗi mã cổ phiếu của bài toán tối thiểu hóa rủi ro

wSBT	wDIG	wVIC	wLSS
0.116143266	0.117334781	0.34492430986	0.42159761853

với phương sai danh mục:

$$var(r_p) = 2.851771305$$

Nhận xét: Khi muốn xây dựng một danh mục có lợi suất bằng với lợi suất của thị trường thì nhà đầu tư nên phân bổ vốn vào 4 mã cổ phiếu có tỷ trọng như trên để gặp ít rủi ro nhất.

3. Dự báo lợi suất danh mục

Nhóm sử dụng 02 cách để dự báo lợi suất danh mục cho 10 phiên giao dịch tiếp theo trong năm 2025:

Cách 1: Dự báo lợi suất danh mục dựa trên dự báo từng chuỗi lợi suất cổ phiếu

Nhóm sử dụng kết quả dự báo lợi suất 10 phiên giao dịch tiếp theo trong năm 2025 của 4 mã cổ phiếu bằng mô hình ARIMA từ phần bài cá nhân, kết hợp với kết quả tính tỷ trọng đầu tư từ bài toán thiểu hóa rủi ro danh mục, từ đó tính được dự báo lợi suất danh mục cho 10 phiên giao dịch tiếp theo trong năm 2025.

Cách 2: Dự báo lợi suất danh mục dựa trên chuỗi lợi suất danh mục thực tế

Nhóm sử dụng chuỗi lợi suất thực tế của 4 mã cổ phiếu trong năm 2023 và 2024 kết hợp với tỷ trọng đầu tư từ bài toán thiểu hóa rủi ro danh mục để tính ra chuỗi lợi suất thực tế của danh mục trong năm 2023 và 2024.

Kiểm định Dickey-Fuller: Kiểm định tính dừng của chuỗi lợi suất danh mục

Bước 1: Kiểm định nghiệm đơn vị có hệ số chặn và xu thế

Mô hình: $\Delta Portfolio = \mu + \beta_t + \delta Portfolio_{t-1} + v_t$

Cặp giả thuyết: $\begin{cases} H_0: \beta_t = 0 \text{ (Xu thế không có ý nghĩa thống kê)} \\ H_1: \beta_t \neq 0 \text{ (Xu thế có ý nghĩa thống kê)} \end{cases}$

Do $P - value = 0.265 > 0.05 \rightarrow$ Chưa bác bỏ H_0

\rightarrow Kết luận: Xu thế không có ý nghĩa thống kê

Bước 2: Kiểm định nghiệm đơn vị có hệ số chặn

Mô hình: $\Delta Portfolio = \mu + \delta Portfolio_{t-1} + v_t$

Cặp giả thuyết: $\begin{cases} H_0: \mu = 0 \text{ (Hệ số chặn không có ý nghĩa thống kê)} \\ H_1: \mu \neq 0 \text{ (Hệ số chặn có ý nghĩa thống kê)} \end{cases}$

Do $P - value = 0.616 > 0.05 \rightarrow$ Chưa bác bỏ H_0

\rightarrow Kết luận: Hệ số chặn không có ý nghĩa thống kê

Bước 3: Kiểm định nghiệm đơn vị không có hệ số chặn

Cặp giả thuyết: $\begin{cases} H_0: \delta = 0 \text{ (có nghiệm đơn vị hay chuỗi không dừng)} \\ H_1: \delta < 0 \text{ (không có nghiệm đơn vị hay chuỗi dừng)} \end{cases}$

Do $\tau_{qs} = -14.9955$ mà $\tau_{0.1} = -1.62 \rightarrow |\tau_{qs}| > |\tau_{0.1}|$

\rightarrow Kết luận: Với mức ý nghĩa 10%, chuỗi lợi suất portfolio dừng

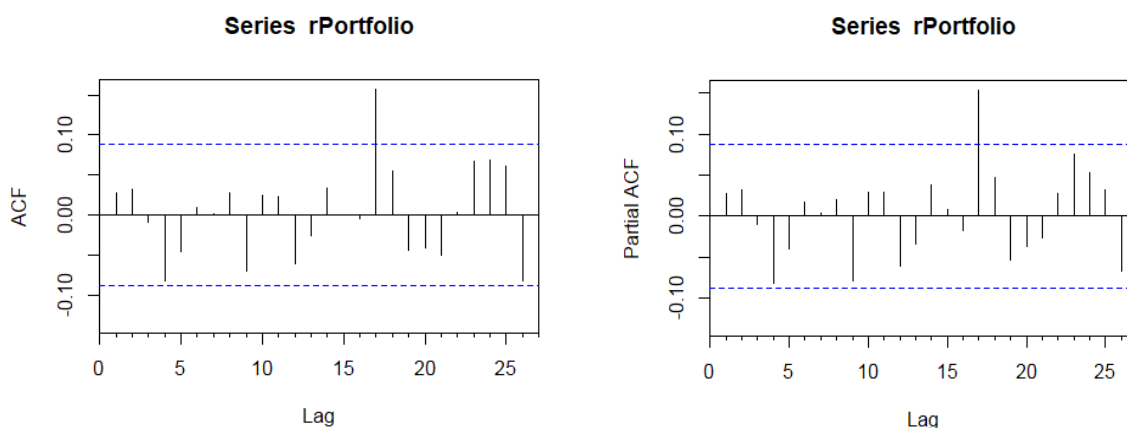
$\rightarrow Portfolio_{log-return} \sim I(0)$

Như vậy chuỗi lợi suất portfolio có thể áp dụng mô hình ARMA(p,q)

Kết quả các kiểm định trên được trình bày ở phụ lục.

Xác định bậc của mô hình ARMA(p,q)

Ta xác định bậc của AR(p) và MA(q) dựa vào đồ thị ACF và PACF



Do hầu hết các giá trị trong cả 2 đồ thị đều nằm trong khoảng tin cậy nên có thể chuỗi lợi suất portfolio là nhiễu trắng. Thực hiện kiểm định Ljung – Box về sự tự tương quan của chuỗi đến bậc 20 để xác định chính xác hơn.

Cặp giả thuyết: $\begin{cases} H_0: \text{Chuỗi không có tự tương quan đến bậc 20} \\ H_1: \text{Chuỗi có ít nhất 1 bậc tương quan đến bậc 20} \end{cases}$

```
Box.test(rPortfolio, lag = 20, type = "Ljung-Box")
##
## Box-Ljung test
##
## data: rPortfolio
## X-squared = 27.769, df = 20, p-value = 0.115
```

Hình 5.3: Kiểm định Box-test cho chuỗi lợi suất danh mục

Kết quả cho thấy $P - value = 0.115 > 0.05 \rightarrow$ Chưa đủ cơ sở bác bỏ H_0

Kết luận: Với mức ý nghĩa 5%, chuỗi lợi suất portfolio không có tự tương quan đến bậc 20. Như vậy, chuỗi dừng quanh 0 và không có tự tương quan nên nhóm nhận định chuỗi là nhiễu trắng.

Theo yêu cầu từ người hướng dẫn, nhóm áp dụng mô hình ARMA(1,1) cho dữ liệu chuỗi lợi suất thực tế của danh mục trong năm 2023 và 2024 để dự báo cho lợi suất danh mục trong 10 phiên giao dịch tiếp theo năm 2025.

Đánh giá tính dừng của mô hình:

Do nghiệm nghịch đảo nằm bên trong nên có thể kết luận mô hình là dừng.

Đánh giá tính nhiễu trắng của phần dư:

Cặp giả thuyết: $\begin{cases} H_0: \text{Phần dư } u_t \text{ không có tự tương quan} \\ H_1: \text{Phần dư } u_t \text{ có tự tương quan} \end{cases}$

Sử dụng kiểm định Ljung-Box test ta có kết quả $P\text{-value} = 0.4207 > 0.05 \rightarrow$ Chưa bác bỏ $H_0 \rightarrow$ Kết luận: Phần dư không có tự tương quan

Như vậy có thể tạm thời kết luận rằng mô hình ARMA(1,1) là thích hợp để dự báo cho giá cổ phiếu LSS trong tương lai gần.

Kết quả chạy mô hình và các kiểm định được trình bày ở phụ lục.

Để so sánh khả năng dự báo chính xác của 2 cách trên, nhóm sử dụng chỉ tiêu RMSE để áp dụng cho từng chuỗi so với chuỗi lợi suất thực tế của danh mục trong 10 phiên giao dịch đầu tiên trong năm 2025. Trong đó chuỗi lợi suất thực tế của danh mục trong 10 phiên giao dịch đầu tiên trong năm 2025 được tính dựa trên chuỗi lợi suất thực tế của 4 mã cổ phiếu trong 10 phiên giao dịch đầu năm 2025 và tỷ trọng đầu tư từ bài toán tối ưu hóa rủi ro danh mục

Kết quả đó chuỗi lợi suất thực tế của danh mục trong 10 phiên giao dịch đầu tiên trong năm 2025, dự báo lợi suất theo 2 cách và RMSE của mỗi cách được tổng hợp trong bảng dưới đây:

Bảng 5.5 3: Chuỗi lợi suất thực tế của danh mục, dự báo lợi suất danh mục từ dự báo mỗi lợi suất cổ phiếu, dự báo lợi suất danh mục từ lợi suất thực tế của danh mục trong quá khứ

Actual log return of portfolio	Forecast 1: from individual forecasted series	Forecast 2: from past values of portfolio
6.61791208	-0.025894365	0.002814024
-0.79977708	0.025864194	0.000819255
-0.975173237	0.041101884	0.000238512
-1.128567686	0.041896507	6.94388E-05
0.646503444	0.043380031	2.02159E-05
0.373919576	0.043551312	5.88552E-06
-1.636822466	0.044031907	1.71347E-06
-0.006541276	0.043827938	4.98847E-07
-4.313700543	0.044060357	1.45231E-07
0.8067879	0.043987403	4.22815E-08
RMSE	2.647589532	2.629099372

Kết luận: Kết quả cho thấy dự báo theo cách thứ 2 sẽ có sai số nhỏ hơn cách thứ nhất.

4. Phân tích đồng tích hợp

Sử dụng kiểm định đồng tích hợp của Johansen đối với 4 chuỗi giá cổ phiếu. Tuy nhiên với kết quả về kiểm định tính dừng của 4 chuỗi giá thì chỉ có 3 trên 4 chuỗi (bao gồm:) là I(1), phù hợp với kiểm định này. Cặp giả thuyết của kiểm định trên như sau như sau.

$$\begin{cases} H_0: \text{số mối quan hệ đồng tích hợp} = 0 \\ H_1: \text{số mối quan hệ đồng tích hợp} > 0 \end{cases}$$

...

$$\begin{cases} H_0: \text{số mối quan hệ đồng tích hợp} \leq 2 \\ H_1: \text{số mối quan hệ đồng tích hợp} > 2 \end{cases}$$

Giả thuyết H_0 bị bác bỏ nếu λ_{Trace} hoặc $\lambda_{max\ eigen} >$ giá trị tới hạn.

Kết quả của kiểm định như sau:

Bảng 5.6. Kết quả kiểm định Johansen theo thống kê maximum eigenvalue

	Giá trị thống kê quan sát	Giá trị thống kê		
		10%	5%	1%
$r = 0$	12.27	18.90	21.07	25.75
$r \leq 1$	6	12.91	14.90	19.19
$r \leq 2$	3.45	6.50	8.18	11.65

Dựa vào kết quả của bảng trên có thể thấy giá trị thống kê ở cặp giả thuyết đầu tiên nhỏ hơn giá trị tới hạn mức ý nghĩa 10% nên chưa bác bỏ H_0 , không có quan hệ đồng tích hợp giữa ba chuỗi giá cổ phiếu.

Ngoài ra, nhóm cũng thực hiện kiểm tra đồng tích hợp đối với các chuỗi cùng ngành, kết quả kiểm định cũng tương tự như trên, không có quan hệ đồng tích hợp giữa các chuỗi cổ phiếu cùng ngành. Đây là kết quả hợp lý, cho thấy các công ty trên có hoạt động độc lập với nhau.

Do các chuỗi không có quan hệ đồng tích hợp nên nhóm sẽ bỏ qua mô hình hiệu chỉnh sai số ECM.

5. Phân tích VAR

5.1. Ước lượng mô hình VAR

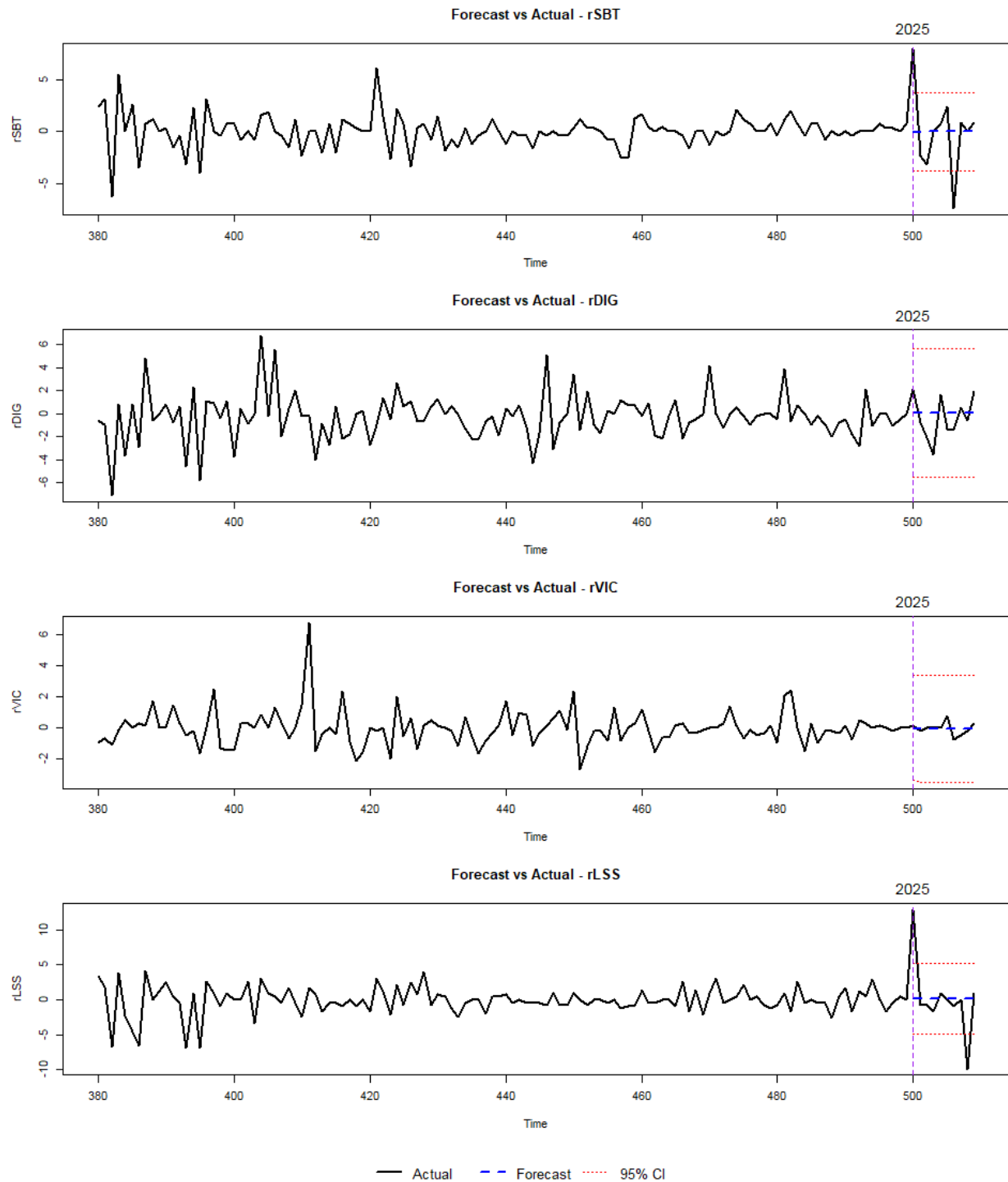
Ở phần này, nhóm thực hiện dự báo chuỗi log_return của 4 mã cổ phiếu dựa vào mô hình VAR (Vector Autoregressive Model). Và so sánh kết quả với kết quả dự báo từ mô hình ARIMA ở trên.

Đầu tiên, thực hiện tìm độ trễ tối ưu cho mô hình VAR dựa trên tiêu chí AIC, kết quả thu được độ trễ 1 là phù hợp đối với mô hình VAR dành cho 4 chuỗi log_return. Sau khi đã có trễ tối ưu, thực hiện ước lượng mô hình VAR, kết quả ước lượng như sau:

$$\begin{aligned}\widehat{rSBT}_t &= 0.0127 + (-0.03) * rSBT_{t-1} + 0.01 * rDIG_{t-1} + 0.06 * rVIC_{t-1} + (-0.02) \\ &\quad * rLSS_{t-1} \\ \widehat{rDIG}_t &= 0.046 + 0.028 * rSBT_{t-1} + (-0.017) * rDIG_{t-1} + (-0.024) * rVIC_{t-1} + (-0.07) \\ &\quad * rLSS_{t-1} \\ \widehat{rVIC}_t &= -0.048 + 0.033 * rSBT_{t-1} + 0.045 * rDIG_{t-1} + 0.125 * rVIC_{t-1} + (-0.07) \\ &\quad * rLSS_{t-1} \\ \widehat{rLSS}_t &= 0.138 + 0.091 * rSBT_{t-1} + 0.04 * rDIG_{t-1} + (-0.03) * rVIC_{t-1} + (-0.036) \\ &\quad * rLSS_{t-1}\end{aligned}$$

Qua kiểm định tự tương quan của phần dư bằng kiểm định Portmanteau, nhóm đưa ra kết luận phần dư không tự tương quan ở mức ý nghĩa 5%

Kết quả dự báo dành cho 4 chuỗi log_return dựa trên mô hình VAR sẽ được trình bày tại Hình 5.2 dưới đây (do dữ liệu dài nên trên đồ thị chỉ trình bày 120 phiên cuối và 10 phiên dự báo)



Hình 5.2. Kết quả dự báo chuỗi log_return dựa trên mô hình VAR

Để so sánh với kết quả dự báo bằng mô hình ARIMA, nhóm sẽ thực hiện tính toán sai số dự báo RMSE cho kết quả dự báo đến từ hai mô hình với giá trị thực tế của 2025. Kết quả thu được như sau:

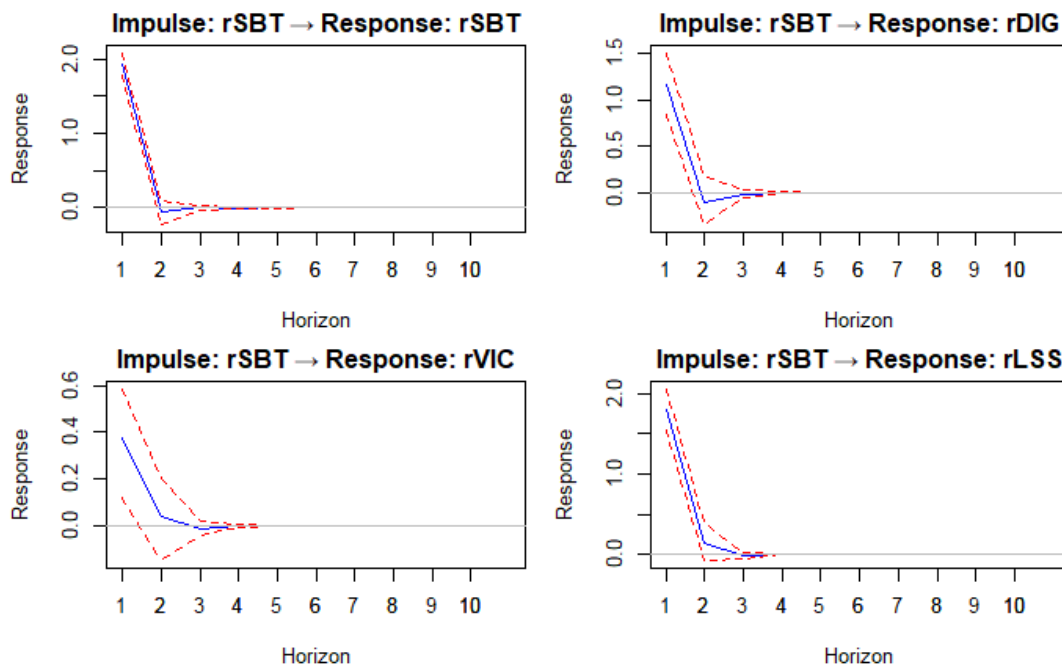
Bảng 5.7. Sai số dự báo RMSE từ hai mô hình ARIMA và VAR dành cho chuỗi log_return

	rSBT	rDIG	rVIC	rLSS
ARIMA	3.775	1.908	0.392	5.178
VAR	3.78	1.814	0.385	5.162

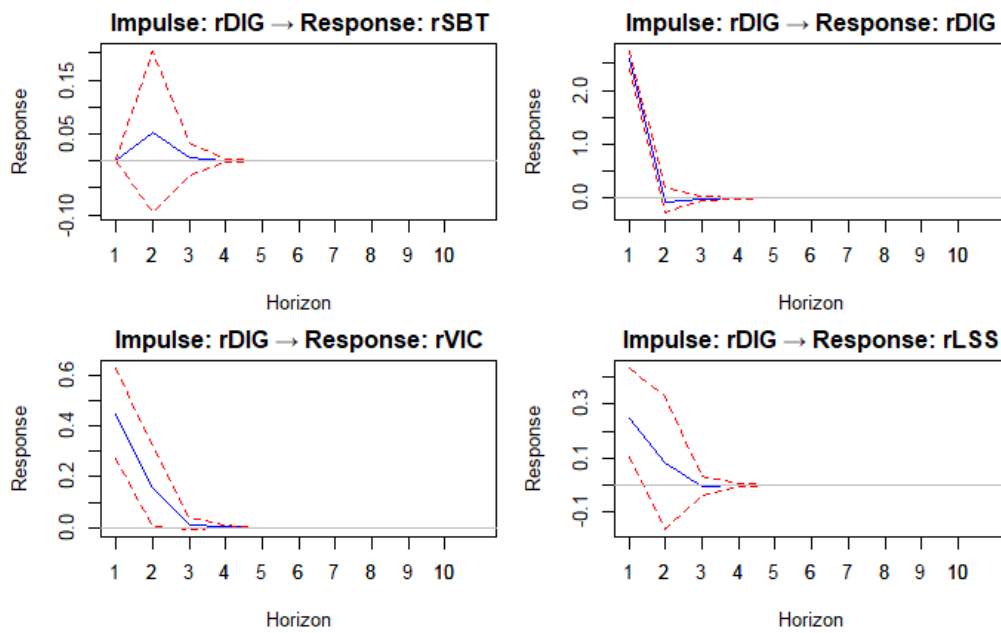
Dựa trên kết quả của bảng 5.2, có thể thấy mô hình VAR có sai số dự báo RMSE nhỏ hơn đối với rDIG, rVIC, rLSS và lớn hơn đối với rSBT, tuy vậy mức chênh lệch là không quá lớn.

5.2. Hàm phản ứng IRF

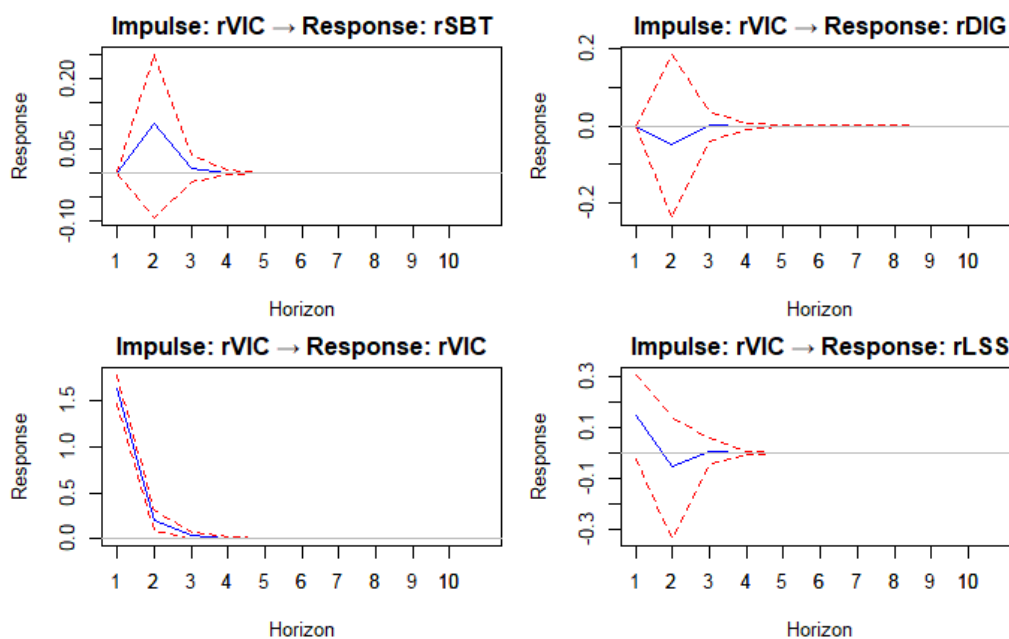
Để phân tích tác động của sốc trong mô hình VAR, ta sử dụng hàm phản ứng (Impulse Response Function). Sau đây là đồ thị của các hàm phản ứng:



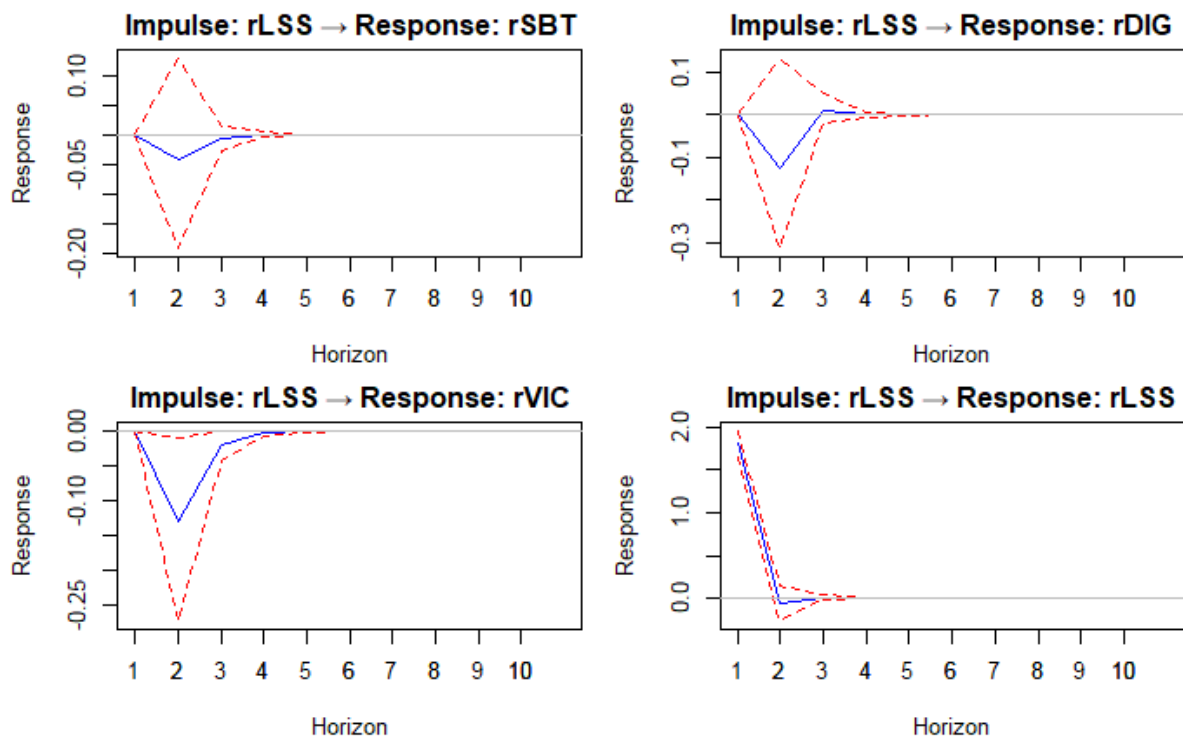
Hình 5.3. Phản ứng của các biến khi có sốc xảy ra với rSBT



Hình 5.4. Phản ứng của các biến khi có sốc xảy ra với $rDIG$



Hình 5.5. Phản ứng của các biến khi có sốc xảy ra với $rVIC$



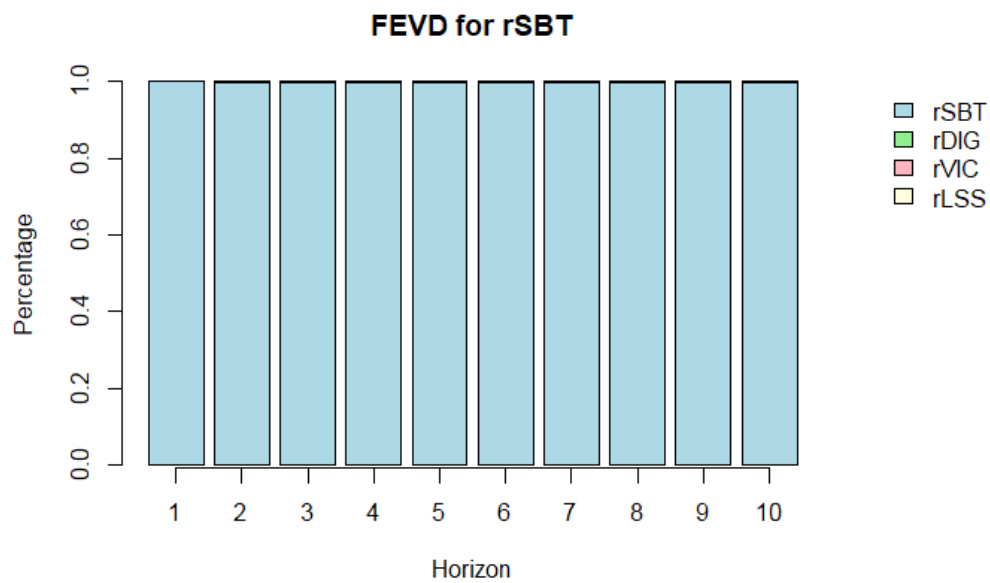
Hình 5.6. Phản ứng của các biến khi có sốc xảy ra với $rLSS$

Nhìn chung, các phản ứng của cả 4 mã cổ phiếu đối với các cú sốc có xu hướng hội tụ về 0 trong khoảng từ 3 đến 4 kỳ sau khi sốc xảy ra. Điều này phản ánh rằng tác động của các cú sốc là có tính ngắn hạn, dần trở về trạng thái cân bằng sau một thời gian tương đối ngắn. Tuy nhiên, có một số ngoại lệ đáng chú ý, cho thấy tính chất động học phức tạp hơn giữa các biến trong hệ thống.

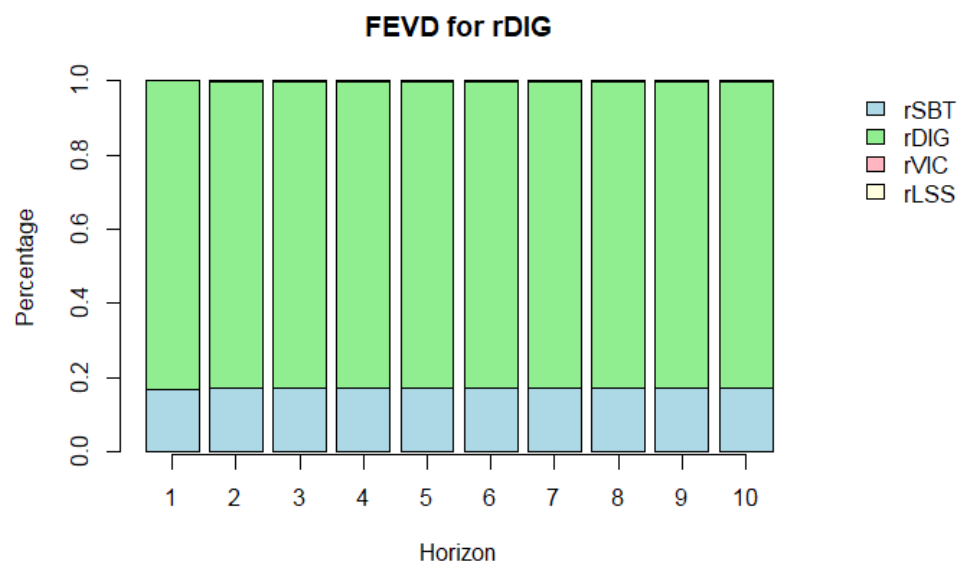
Cụ thể, trong một số trường hợp, cú sốc của một biến không gây ra tác động đáng kể sau kỳ trễ đầu tiên, nhưng bắt đầu xuất hiện ảnh hưởng từ kỳ trễ thứ hai trở đi. Ví dụ điển hình là cú sốc từ $rDIG$ lên $rSBT$, từ $rVIC$ lên cả $rSBT$ và $rDIG$, cũng như cú sốc của $rLSS$ lên cả ba biến còn lại ($rSBT$, $rDIG$, và $rVIC$). Những phản ứng trễ này cho thấy sự lan truyền tác động trong hệ thống là không tức thời, mà có độ trễ nhất định, có thể do thời gian các thông tin được phản ánh vào giá hoặc do sự khác biệt trong chu kỳ phản ứng của các doanh nghiệp.

Đặc biệt, có một trường hợp nổi bật là phản ứng của $rDIG$ trước cú sốc từ $rVIC$, khi mà thời gian để hội tụ về 0 kéo dài đến khoảng 8 kỳ nếu xét theo khoảng tin cậy 95%. Điều này cho thấy tác động của $rVIC$ lên $rDIG$ là không chỉ tồn tại mà còn có tính chất bền vững trong trung hạn. Một giả thuyết hợp lý để lý giải hiện tượng này là do VIC là doanh nghiệp có quy mô vốn hóa lớn nhất trong nhóm bất động sản, do đó các biến động từ VIC có thể tạo ra ảnh hưởng lan tỏa đáng kể đến các doanh nghiệp khác trong ngành, đặc biệt là những doanh nghiệp có mối liên hệ hoặc mức độ tương quan cao như DIG . Tuy nhiên, cần lưu ý rằng đây mới chỉ là một phỏng đoán mang tính trực giác dựa trên quy mô thị trường, và chưa đủ căn cứ định lượng để xác nhận giả thuyết này một cách chắc chắn.

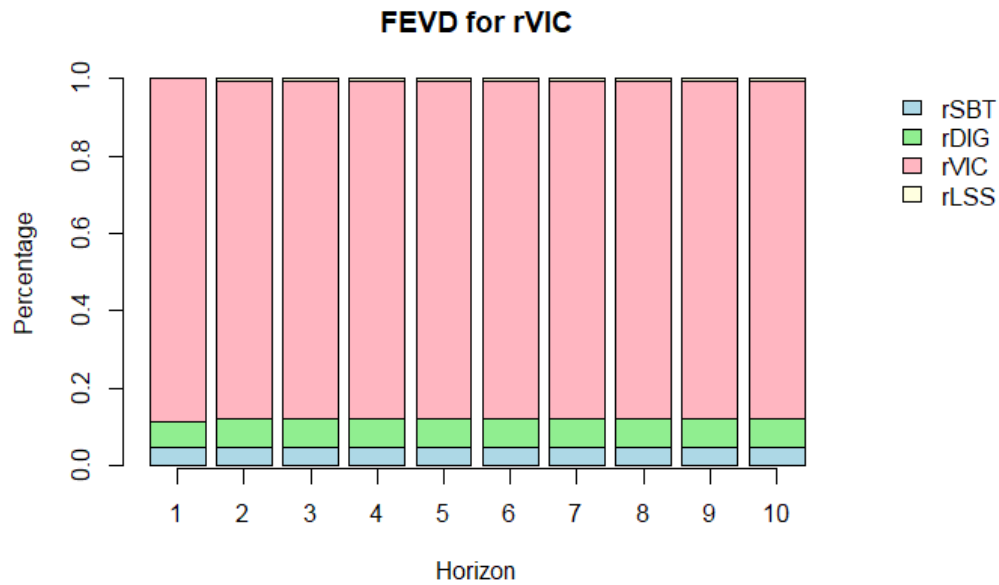
5.3. Phân rã phương sai



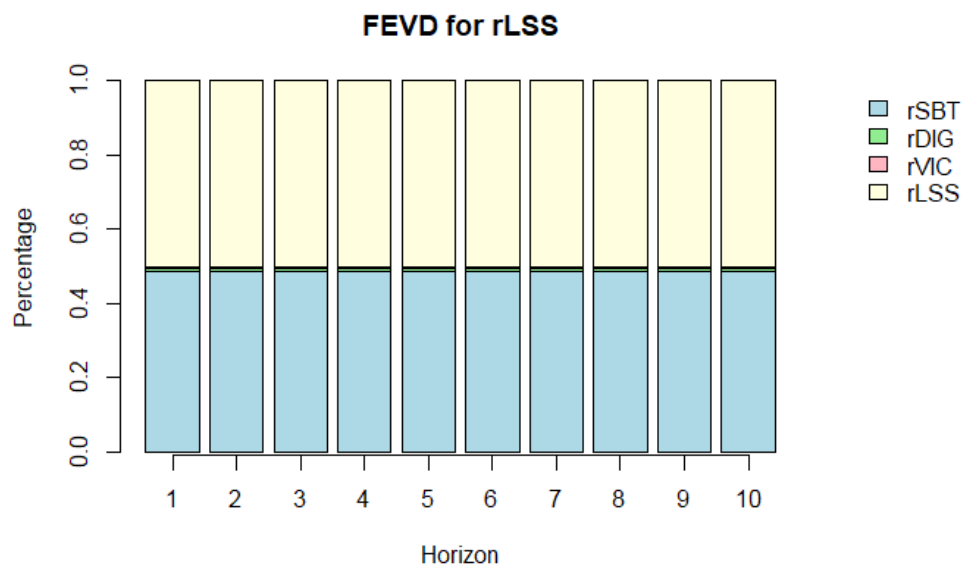
Hình 5.7. Đồ thị phân rã phương sai của rSBT



Hình 5.8. Đồ thị phân rã phương sai của rDIG



Hình 5.9. Đồ thị phân rã phương sai của rVIC



Hình 5.9. Đồ thị phân rã phương sai của rLSS

Quan sát đồ thị, ta nhận thấy rằng gần như toàn bộ sai số bình phương trung bình (MSE) trong việc dự báo biến rSBT trong vòng 10 kỳ đều được giải thích chủ yếu bởi chính cú sốc từ biến này. Điều này cho thấy rSBT có tính tự dự báo rất cao, phản ánh sự chi phối mạnh mẽ của sự tự vận động của nó đối với diễn biến trong tương lai.

Tương tự, đối với biến rDIG, ta cũng thấy rằng các cú sốc từ chính rDIG giải thích hơn 80% MSE trong giai đoạn dự báo. Điều này cho thấy rDIG ít chịu tác động từ các biến khác trong mô hình. Phần còn lại trong MSE của rDIG chỉ được giải thích bởi một tỷ lệ rất nhỏ đến từ cú sốc của rSBT, trong khi ảnh hưởng từ các biến như rVIC và rLSS là không đáng kể, gần như bằng 0.

Trường hợp của rVIC cũng diễn ra theo xu hướng tương tự. Cú sốc từ chính rVIC chiếm tỷ trọng lớn trong việc giải thích MSE của biến này trong suốt 10 kỳ, vượt quá 80%. Điều này cho thấy rVIC có mức độ tự giải thích khá cao, và ít bị chi phối bởi các yếu tố bên ngoài trong mô hình.

Tuy nhiên, điểm khác biệt đáng chú ý nằm ở biến rLSS. Khác với các biến còn lại, rLSS chỉ có khoảng 50% MSE được giải thích bởi cú sốc từ chính nó. Phần còn lại chủ yếu được lý giải bởi tác động từ cú sốc của rSBT, cho thấy rLSS có sự phụ thuộc tương đối lớn vào động thái của rSBT là một mã cổ phiếu cùng ngành. Trong khi đó, ảnh hưởng từ các biến rDIG và rVIC đối với rLSS hầu như không đáng kể.

IV. KẾT LUẬN

Bài viết trên đã tiến hành phân tích về doanh thu thuần của các công ty VIC, DIG, SBT và LSS – đại diện cho hai lĩnh vực bất động sản và tiêu dùng thiết yếu. Thông qua việc áp dụng các mô hình khác nhau để dự báo doanh thu năm 2024 đối chiếu với số liệu thực tế, nghiên cứu đã xác định được mô hình phù hợp nhất cho từng công ty. Các mô hình tối ưu này sau đó được sử dụng để đưa ra các dự báo cho năm 2025, giúp nâng cao tính chính xác và ứng dụng thực tiễn của kết quả

Bên cạnh đó, bài viết cũng đã giải quyết hai bài toán tối ưu danh mục: tối đa hoá lợi suất kì vọng trong điều kiện phương sai danh mục bằng trung bình nhân phương sai của các cổ phiếu, và tối thiểu hoá rủi ro danh mục trong điều kiện lợi suất danh mục bằng lợi suất thị trường. Các kết quả từ hai bài toán đã cho thấy sự đánh đổi giữa lợi suất và rủi ro, đồng thời gợi ý cơ cấu danh mục phù hợp trong từng tình huống đầu tư.

Trong phần dự báo lợi suất, hai phương pháp được so sánh: (1) dự báo từng chuỗi lợi suất cổ phiếu rồi tổng hợp lại thành lợi suất danh mục và (2) dự báo trực tiếp chuỗi lợi suất danh mục thực tế. Kết quả cho thấy phương pháp thứ hai có sai số dự báo nhỏ hơn, phản ánh rằng việc dự báo trực tiếp từ chuỗi lợi suất thực tế của danh mục sẽ mang lại độ chính xác cao hơn việc dự báo từ từng chuỗi thành phần, nguyên nhân có thể do vấn đề sai số tăng lên khi ước lượng và dự báo bằng nhiều mô hình khác nhau.

Bài viết cũng sử dụng hai mô hình ARIMA và VAR để dự báo giá cho từng mã cổ phiếu trong 10 phiên giao dịch đầu năm 2025. Việc so sánh giữa hai mô hình cho thấy rằng không có mô hình nào vượt trội hoàn toàn khi chênh lệch RMSE dự báo từ hai mô hình rất nhỏ.

Ngoài ra, phân tích đồng tích hợp đã được sử dụng để kiểm tra mối quan hệ dài hạn giữa các cổ phiếu trong danh mục. Mô hình VAR được ước lượng tiếp theo, đi kèm với phân tích hàm phản ứng đẩy (IRF) và phân rã phương sai (FEVD). IRF cho thấy cách lợi suất phản ứng với các cú sốc trong ngắn hạn, trong khi FEVD làm rõ nguồn gốc biến động từ nội tại hay từ ảnh hưởng của các cổ phiếu khác. Những kết quả này giúp nhận diện rõ hơn vai trò của từng doanh nghiệp trong hệ thống, đồng thời hỗ trợ việc ra quyết định đầu tư một cách khoa học và hiệu quả hơn.

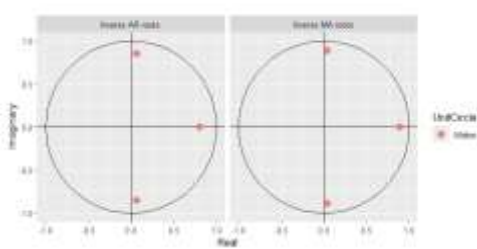
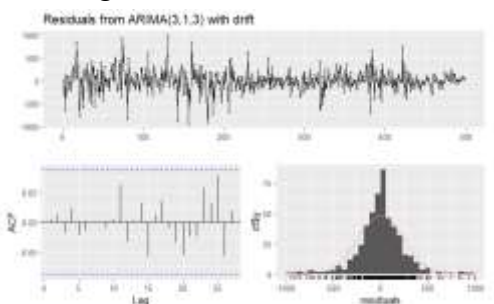
PHỤ LỤC

1. Chuỗi SBT – Ngô Thị Minh Hạnh

1.1. Các mô hình dự báo doanh thu thuần về bán hàng và cung cấp dịch vụ

Tên mô hình	Mô hình ước lượng	MAPE	RMSE
Linear - Linear	$\widehat{dtt}_t = -1012.239 + 113.478*t$	13.26%	1017.531
Linear- Log	$\widehat{dtt}_t = -2721.5 + 1606.2*\ln(t)$	40.75%	2743.55
Log-Log	$\ln(\widehat{dtt}_t) = 4.2482 + 0.9656*\ln(t)$	43.65%	2966.713
MH xu thế tgian TT+ mùa vụ, dạng cộng	$\begin{aligned}\widehat{dtt}_t = & -1217.35 + 112.86*t \\ & + 271.79*s_2 \\ & + 144.64*s_3 \\ & + 474.79*s_4\end{aligned}$	13.53%	1036.097
MH xu thế tgian TT + mùa vụ, dạng nhân	$\begin{aligned}\widehat{dtt}_t = & -593.11 + 89.74*t \\ & + 23.26*t*s_2 \\ & + 24.21*t*s_3 \\ & + 45.02*t*s_4\end{aligned}$	15.64%	1208.754
MH xu thế tgian không TT+ mùa vụ, dạng cộng	$\begin{aligned}\ln(\widehat{dtt}_t) = & 5.415 + 0.0616*t \\ & + 0.1209*s_2 + 0.0061*s_3 \\ & + 0.0678*s_4\end{aligned}$	33.03%	2961.582
MH xu thế tgian không TT + mùa vụ, dạng nhân	$\begin{aligned}\ln(\widehat{dtt}_t) = & 5.5109 + 0.0581*t \\ & - 0.0019*t*s_2 \\ & + 0.046*t*s_3 \\ & + 0.0115*t*s_4\end{aligned}$	49.50%	4373.707
Holt- Winters mùa vụ dạng cộng	$\begin{aligned}\widehat{dtt}_t = & (6388.535 + 207.353*k) \\ & + s_{\text{trong ứng}} \\ \text{Với} \\ s_1 = & -257.493 \\ s_2 = & 541.403 \\ s_3 = & 194.916 \\ s_4 = & 742.575\end{aligned}$	11.40%	755.298
Holt- Winters mùa vụ dạng nhân	$\begin{aligned}\widehat{dtt}_t = & (6361.226 \\ & + 167.821*k)*s_{\text{trong ứng}} \\ s_1 = & 0.89860 \\ s_2 = & 1.07874\end{aligned}$	7.59%	568.663

1.2. Kiểm định đánh giá mô hình chuỗi giá

Đánh giá tính dừng của ARIMA(3,1,3) qua nghiệm nghịch đảo	Kiểm định tính nhiễu trắng phần dư
 <p>Hình . Đồ thị kiểm định tính dừng qua nghiệm nghịch đảo ARIMA(3,1,3) của chuỗi SBT</p> <p>Từ đồ thị trên, ta thấy nghiệm nghịch đảo nằm trong đường tròn đơn vị => Điều này có ý nghĩa rằng quá trình đó không có xu hướng và không dao động quá mức, tức là trung bình và phương sai của quá trình là không đổi theo thời gian. => Ước lượng mô hình chuỗi dừng ARIMA (3,1,3) trên là hợp lý.</p>	<p>Với cặp giả thuyết:</p> <p>H0: Không tồn tại hiện tượng tự tương quan (nhiều trắng)</p> <p>H1: Tồn tại hiện tượng tự tương quan (không nhiều trắng)</p> <p>Với mức ý nghĩa 5%, $p - value = 0.92359 > 0.05$, \Rightarrow Chưa BB H0, không tồn tại hiện tượng tự tương quan, phần dư nhiễu trắng.</p>  <p>Hình . Kiểm định đánh giá phần dư mô hình ARIMA(3,1,3) của chuỗi SBT</p> <p>ACF của phần dư không có sự tương quan ở các bậc được thể hiện qua hình vẽ trên nên việc lựa chọn mô hình ARIMA(3,1,3) là hợp lý và tối ưu.</p>

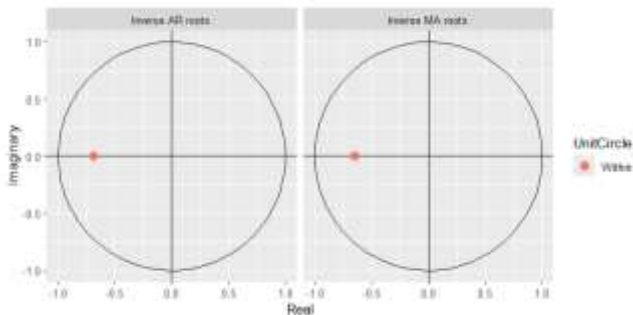
1.3. Kiểm định tính dừng cho chuỗi Log return

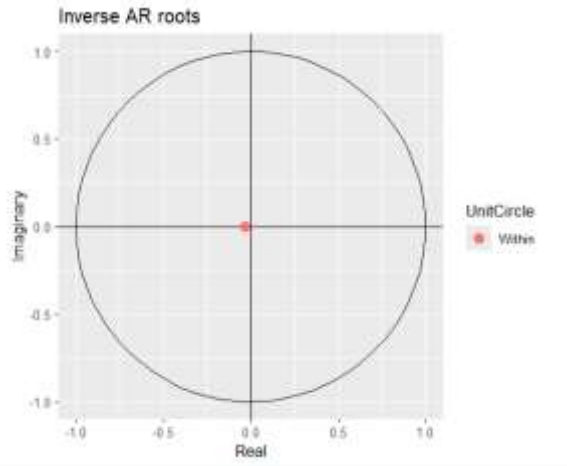
Kiểm định	Log return
Kiểm định ADF có xu thế	<pre> Coefficients: Estimate Std. Error t value Pr(> t) (Intercept) 0.0518905 0.1748137 0.297 0.767 z.lag.1 -1.0099118 0.0645517 -15.645 <2e-16 *** tt -0.0001988 0.0006089 -0.326 0.744 z.diff.lag -0.0189761 0.0447365 -0.424 0.672 --- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 Residual standard error: 1.936 on 491 degrees of freedom Multiple R-squared: 0.5151, Adjusted R-squared: 0.5121 F-statistic: 173.8 on 3 and 491 DF, p-value: < 2.2e-16 Value of test-statistic is: -15.645 81.5908 122.3851 Critical values for test statistics: 1pct 5pct 10pct tau3 -3.98 -3.42 -3.13 phi2 6.15 4.71 4.05 phi3 8.34 6.30 5.36 </pre>

Bài tập lớn Phân tích chuỗi thời gian và dự báo

Kiểm định ADF chỉ có hệ số chặn	<p>Coefficients:</p> <table><thead><tr><th></th><th>Estimate</th><th>Std. Error</th><th>t value</th><th>Pr(> t)</th></tr></thead><tbody><tr><td>(Intercept)</td><td>0.002398</td><td>0.086922</td><td>0.028</td><td>0.978</td></tr><tr><td>z.lag.1</td><td>-1.009560</td><td>0.064484</td><td>-15.656</td><td><2e-16 ***</td></tr><tr><td>z.diff.lag</td><td>-0.019084</td><td>0.044695</td><td>-0.427</td><td>0.670</td></tr></tbody></table> <p>--- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</p> <p>Residual standard error: 1.934 on 492 degrees of freedom Multiple R-squared: 0.515, Adjusted R-squared: 0.513 F-statistic: 261.2 on 2 and 492 DF, p-value: < 2.2e-16</p> <p>Value of test-statistic is: -15.656 122.5555</p> <p>Critical values for test statistics:</p> <table><thead><tr><th></th><th>1pct</th><th>5pct</th><th>10pct</th></tr></thead><tbody><tr><td>tau2</td><td>-3.44</td><td>-2.87</td><td>-2.57</td></tr><tr><td>phi1</td><td>6.47</td><td>4.61</td><td>3.79</td></tr></tbody></table>		Estimate	Std. Error	t value	Pr(> t)	(Intercept)	0.002398	0.086922	0.028	0.978	z.lag.1	-1.009560	0.064484	-15.656	<2e-16 ***	z.diff.lag	-0.019084	0.044695	-0.427	0.670		1pct	5pct	10pct	tau2	-3.44	-2.87	-2.57	phi1	6.47	4.61	3.79
	Estimate	Std. Error	t value	Pr(> t)																													
(Intercept)	0.002398	0.086922	0.028	0.978																													
z.lag.1	-1.009560	0.064484	-15.656	<2e-16 ***																													
z.diff.lag	-0.019084	0.044695	-0.427	0.670																													
	1pct	5pct	10pct																														
tau2	-3.44	-2.87	-2.57																														
phi1	6.47	4.61	3.79																														
Kiểm định none	<p>Coefficients:</p> <table><thead><tr><th></th><th>Estimate</th><th>Std. Error</th><th>t value</th><th>Pr(> t)</th></tr></thead><tbody><tr><td>z.lag.1</td><td>-1.00956</td><td>0.06442</td><td>-15.672</td><td><2e-16 ***</td></tr><tr><td>z.diff.lag</td><td>-0.01908</td><td>0.04465</td><td>-0.427</td><td>0.669</td></tr></tbody></table> <p>--- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</p> <p>Residual standard error: 1.932 on 493 degrees of freedom Multiple R-squared: 0.515, Adjusted R-squared: 0.513 F-statistic: 261.7 on 2 and 493 DF, p-value: < 2.2e-16</p> <p>Value of test-statistic is: -15.6719</p> <p>Critical values for test statistics:</p> <table><thead><tr><th></th><th>1pct</th><th>5pct</th><th>10pct</th></tr></thead><tbody><tr><td>tau1</td><td>-2.58</td><td>-1.95</td><td>-1.62</td></tr></tbody></table>		Estimate	Std. Error	t value	Pr(> t)	z.lag.1	-1.00956	0.06442	-15.672	<2e-16 ***	z.diff.lag	-0.01908	0.04465	-0.427	0.669		1pct	5pct	10pct	tau1	-2.58	-1.95	-1.62									
	Estimate	Std. Error	t value	Pr(> t)																													
z.lag.1	-1.00956	0.06442	-15.672	<2e-16 ***																													
z.diff.lag	-0.01908	0.04465	-0.427	0.669																													
	1pct	5pct	10pct																														
tau1	-2.58	-1.95	-1.62																														

1.4. Kiểm định, đánh giá mô hình chuỗi Log return

Mô hình	Kiểm định tính dừng qua nghiệm nghịch đảo	Kiểm định tính nhiễu trắng của phần dư
ARIMA(1,0,1)		<p>Ljung-Box test</p> <p>data: Residuals from ARIMA(1,0,0) with non-zero mean</p> <p>$Q^* = 9.5026$, $df = 9$, p-value = 0.3922</p> <p>Model df: 1. Total lags used: 10</p>

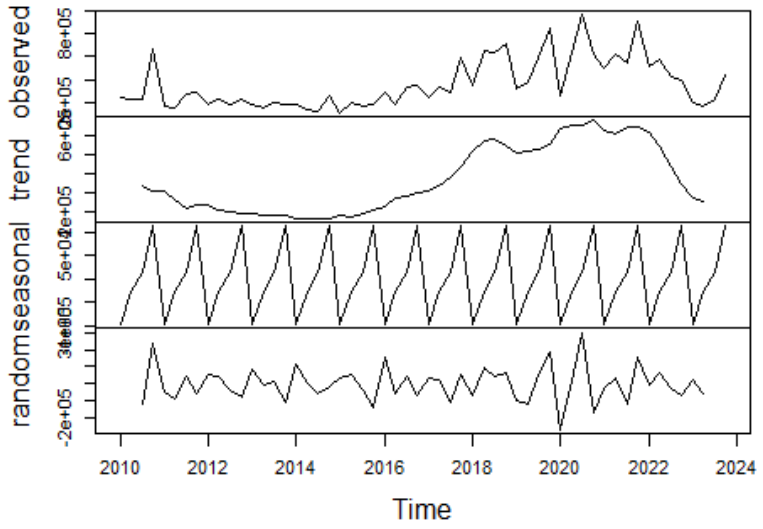
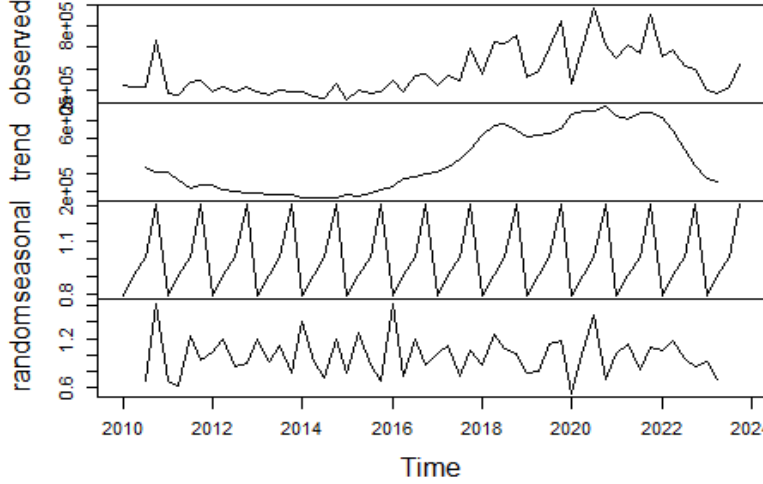
ARIMA(1,0,0)		<p>Ljung-Box test data: Residuals from ARIMA(1,0,1) with non-zero mean $Q^* = 8.3828$, $df = 8$, p-value = 0.397 Model df: 2. Total lags used: 10</p>
--------------	--	--

2. Chuỗi DIC – Trần Đức Hiền

2.1. Kết quả các mô hình dự báo chuỗi doanh thu thuần.

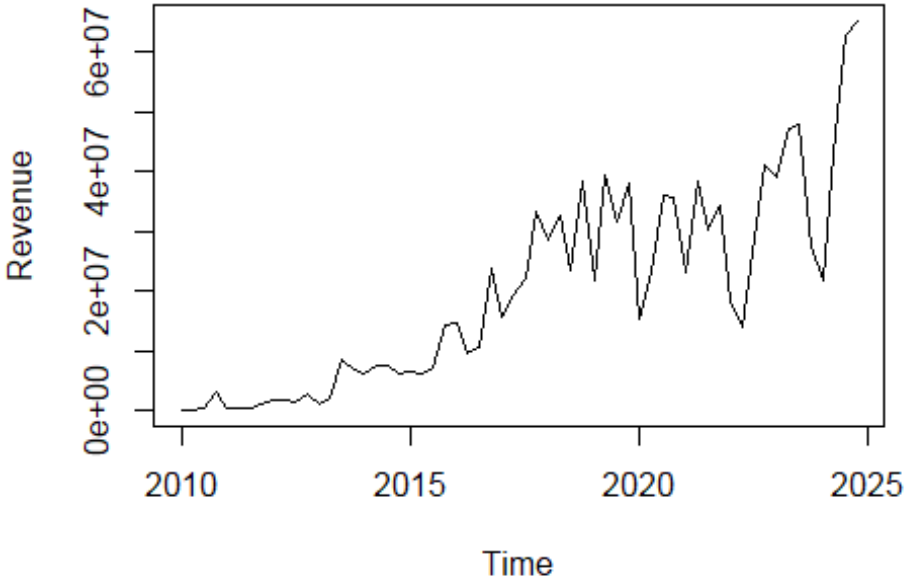
Mô hình	Kết quả ước lượng (với dữ liệu đến hết 2023Q4)
Lin-lin	$\widehat{Revenue}_t = 167101 + 6828*t$
Lin-log	$\widehat{Revenue}_t = 53118 + 100265*\ln(t)$
Log-lin	$\ln(\widehat{Revenue}_t) = 12.095 + 0.019*t$
Log-log	$\ln(\widehat{Revenue}_t) = 11.802 + 0.27*\ln(t)$
MH xu thế thời gian + mùa vụ lin-log, dạng cộng	$\widehat{Revenue}_t = -6643 + 93066*\ln(t) + 53313*s_2 + 86904*s_3 + 187452*s_4$
MH xu thế thời gian + mùa vụ lin-log, dạng nhân	$\widehat{Revenue}_t = 56793 + 68815*\ln(t) + 22736*\ln(t)*s_2 + 34180*\ln(t)*s_3 + 61640*\ln(t)*s_4$
MH xu thế thời gian + mùa vụ lin-log, dạng kết hợp	$\widehat{Revenue}_t = 129198 + 47046*\ln(t) + (-141141*s_2 + -192118*s_3 + 21980*s_4) + (65238*\ln(t)*s_2 + 91859*\ln(t)*s_3 + 55321*\ln(t)*s_4)$
Holt-winter dạng cộng	$\widehat{Revenue}_t = (260706.27 - 76527.62*k) + s_i$ $s_1 = -196229.57; s_2 = -101221.34;$ $s_3 = -45283.78; s_4 = 99791.31$
Holt-winter dạng nhân	$\widehat{Revenue}_t = (2.545*10^5 - 9.1*10^3 k)*s_i$ $s_1 = 7.84*10^{-1}; s_2 = 9.57*10^{-1};$ $s_3 = 1.098; s_4 = 1.488$

2.2. Kết quả phân rã các thành phần của chuỗi thời gian

	Kết quả phân rã
Dạng cộng	<p>Decomposition of additive time series</p>  <p>The plot displays four stacked time series from 2010 to 2024. The top panel shows the 'observed' data with a y-axis ranging from $-2e+05$ to $8e+05$. The second panel shows the 'trend' component, which is a smooth curve starting around $4e+05$ and peaking near $7e+05$ around 2020. The third panel shows the 'seasonal' component, characterized by regular, periodic spikes. The bottom panel shows the 'random' component, which is a noisy line fluctuating around zero with a y-axis ranging from $-2e+05$ to $3e+05$.</p>
Dạng nhân	<p>Decomposition of multiplicative time series</p>  <p>The plot displays four stacked time series from 2010 to 2024. The top panel shows the 'observed' data with a y-axis ranging from 0.6 to $8e+05$. The second panel shows the 'trend' component, a smooth curve starting around 1.1 and peaking near 1.2 around 2020. The third panel shows the 'seasonal' component, with regular periodic spikes. The bottom panel shows the 'random' component, a noisy line fluctuating around one with a y-axis ranging from 0.6 to 1.2.</p>

3. Chuỗi VIC – Lê Tuấn Minh

	TẢI THU VIỆN
	<pre> setwd("D:/Time Series") library(readxl) library(Metrics) ## Warning: package 'Metrics' was built under R version 4.4.2 library(forecast) ## Registered S3 method overwritten by 'quantmod': ## method from ## as.zoo.data.frame zoo ## ## Attaching package: 'forecast' ## The following object is masked from 'package:Metrics': ## ## accuracy library(urca) library(tseries) </pre>
3.1	GIỚI THIỆU VỀ DỮ LIỆU
3.2	CÁC MÔ HÌNH VỚI CHUỖI DOANH THU
	Tải dữ liệu, chia tập và vẽ đồ thị
	<pre> dat = read_excel("Vingroup JSC Stock Price History.xlsx", sheet = "Sheet1", range = "C1:C61") attach(dat) dat = ts(dat, start = c(2010,1), frequency = 4) time = seq_along(dat) t = head(time, -4) t2 = tail(time, 4) Revenue = ts(Revenue, start = c(2010, 1), frequency = 4) plot.ts(Revenue, type = 'l') </pre>

	 <pre> train_dat = dat[1:(length(dat)-4)] val_dat = dat[(length(dat)-3):length(dat)] rev1 = ts(train_dat, start = c(2010,1), frequency = 4) rev = ts(dat, start = c(2010, 1), frequency = 4) s1 <- c(rep(c(1,0,0,0), 14)) s2 <- c(rep(c(0,1,0,0), 14)) s3 <- c(rep(c(0,0,1,0), 14)) s4 <- c(rep(c(0,0,0,1), 14)) </pre>
	<p>Chạy các mô hình với dữ liệu tập train và tính rmse, mape trên tập validation</p>
	<pre> reg_lin_lin = lm(rev1 ~ t) summary(reg_lin_lin) ## ## Call: ## lm(formula = rev1 ~ t) ## ## Residuals: ## Min 1Q Median 3Q Max ## -20889029 -4346495 -405981 4774457 14706637 ## ## Coefficients: ## Estimate Std. Error t value Pr(> t) ## (Intercept) -4622688 1930367 -2.395 0.0201 * ## t 787311 58917 13.363 <2e-16 *** ## --- ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 </pre>

```

1
##
## Residual standard error: 7126000 on 54 degrees of freedom
## Multiple R-squared:  0.7678, Adjusted R-squared:  0.7635
## F-statistic: 178.6 on 1 and 54 DF,  p-value: < 2.2e-16

reg_lin_log = lm(rev1~log(t))
summary(reg_lin_log)

##
## Call:
## lm(formula = rev1 ~ log(t))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14502404  -7803714  -2546705   7893797  21173005
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -21068646    4520944  -4.660 2.11e-05 ***
## log(t)       12634095     1411293   8.952 2.96e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

1
##
## Residual standard error: 9383000 on 54 degrees of freedom
## Multiple R-squared:  0.5974, Adjusted R-squared:  0.59
## F-statistic: 80.14 on 1 and 54 DF,  p-value: 2.958e-12

reg_log_lin = lm(log(rev1)~t)
summary(reg_log_lin)

##
## Call:
## lm(formula = log(rev1) ~ t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
##  -2.0590  -0.4976   0.1844   0.6853   1.1354
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.528715    0.215873  62.67  <2e-16 ***
## t           0.085860    0.006589  13.03  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

1
##
## Residual standard error: 0.7969 on 54 degrees of freedom
## Multiple R-squared:  0.7587, Adjusted R-squared:  0.7543
## F-statistic: 169.8 on 1 and 54 DF,  p-value: < 2.2e-16

```

```
reg_log_log = lm(log(rev1)~log(t))
summary(reg_log_log)

##
## Call:
## lm(formula = log(rev1) ~ log(t))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.57313 -0.24402  0.05693  0.27410  1.87916
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  10.78470    0.26544   40.63  <2e-16 ***
## log(t)        1.68665    0.08286   20.36  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5509 on 54 degrees of freedom
## Multiple R-squared:  0.8847, Adjusted R-squared:  0.8826
## F-statistic: 414.3 on 1 and 54 DF,  p-value: < 2.2e-16

reg_add = lm(rev1~t+s2+s3+s4)
summary(reg_add)

##
## Call:
## lm(formula = rev1 ~ t + s2 + s3 + s4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -20580215  -4252312    246478   3586951  14369363
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7266372    2442354  -2.975  0.00447 **
## t             778836     58069   13.412  < 2e-16 ***
## s2            2758632    2648992   1.041  0.30261
## s3            2937484    2650901   1.108  0.27301
## s4            5844797    2654079   2.202  0.03220 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7007000 on 51 degrees of freedom
## Multiple R-squared:  0.788, Adjusted R-squared:  0.7714
## F-statistic: 47.39 on 4 and 51 DF,  p-value: < 2.2e-16

reg_mul = lm(rev1 ~ t + t*s2 + t*s3+ t*s4)
summary(reg_mul)
```

```
##
## Call:
## lm(formula = rev1 ~ t + t * s2 + t * s3 + t * s4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21384494  -3742516   264086   3700349  14003782
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3140029    3674935  -0.854    0.397
## t             626008     116856    5.357 2.37e-06 ***
## s2          -2391340     5268903  -0.454    0.652
## s3          -3517209     5342256  -0.658    0.513
## s4             639010     5417136    0.118    0.907
## t:s2           189386     165260    1.146    0.257
## t:s3           233116     165260    1.411    0.165
## t:s4           188809     165260    1.142    0.259
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7050000 on 48 degrees of freedom
## Multiple R-squared:  0.798, Adjusted R-squared:  0.7685
## F-statistic: 27.09 on 7 and 48 DF,  p-value: 1.258e-14

reg_add1 = lm(log(rev1)~t+s2+s3+s4)
summary(reg_add1)

##
## Call:
## lm(formula = log(rev1) ~ t + s2 + s3 + s4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.8783 -0.5215  0.1553  0.5773  1.1710
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.348854    0.275454  48.461  <2e-16 ***
## t             0.085042    0.006549  12.985  <2e-16 ***
## s2            0.103707    0.298759   0.347   0.7299
## s3            0.156518    0.298974   0.524   0.6029
## s4            0.552510    0.299333   1.846   0.0707 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7903 on 51 degrees of freedom
```

```
## Multiple R-squared:  0.7759, Adjusted R-squared:  0.7584
## F-statistic: 44.15 on 4 and 51 DF,  p-value: 5.65e-16

reg_mul1 = lm(log(rev1) ~ t + t*s2 + t*s3+ t*s4)
summary(reg_mul1)

##
## Call:
## lm(formula = log(rev1) ~ t + t * s2 + t * s3 + t * s4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6625 -0.5268  0.1439  0.5991  1.2641
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 13.1247969  0.4072611  32.227  < 2e-16 ***
## t            0.0933405  0.0129501   7.208 3.54e-09 ***
## s2           0.1055050  0.5839067   0.181  0.8574
## s3           0.1876935  0.5920358   0.317  0.7526
## s4           1.4631882  0.6003341   2.437  0.0186 *
## t:s2        -0.0003606  0.0183143  -0.020  0.9844
## t:s3        -0.0016473  0.0183143  -0.090  0.9287
## t:s4        -0.0311858  0.0183143  -1.703  0.0951 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7813 on 48 degrees of freedom
## Multiple R-squared:  0.7939, Adjusted R-squared:  0.7638
## F-statistic: 26.41 on 7 and 48 DF,  p-value: 2.02e-14

# 1. Create prediction dataframe
df <- data.frame(
  t = 57:60,
  s2 = c(0, 1, 0, 0),
  s3 = c(0, 0, 1, 0),
  s4 = c(0, 0, 0, 1)
)

# 2. Real validation data
actuals <- val_dat

# 3. List of models
models <- list(
  reg_lin_lin = reg_lin_lin,
  reg_lin_log = reg_lin_log,
  reg_log_lin = reg_log_lin,
  reg_log_log = reg_log_log,
  reg_add = reg_add,
  reg_mul = reg_mul,
  reg_add1 = reg_add1,
  reg_mul1 = reg_mul1
)
```

```
)

# 4. Predict and evaluate
results <- data.frame(Model = character(), RMSE = numeric(), MAPE = numeric())

for (model_name in names(models)) {
  model <- models[[model_name]]

  # Predict
  preds <- predict(model, newdata = df)

  # Back-transform if model is on log scale
  if (model_name %in% c("reg_log_lin", "reg_log_log", "reg_add1", "reg_mul1")) {
    preds <- exp(preds)
  }

  # Evaluate
  model_rmse <- rmse(actuals, preds)
  model_mape <- mape(actuals, preds)

  # Append to results
  results <- rbind(results, data.frame(
    Model = model_name,
    RMSE = model_rmse,
    MAPE = model_mape
  ))
}

# 5. Show results
print(results)

##           Model      RMSE      MAPE
## 1 reg_lin_lin 18040682 0.3963216
## 2 reg_lin_log 24926613 0.4320232
## 3 reg_log_lin 66703385 1.7528607
## 4 reg_log_log 16252063 0.3999299
## 5      reg_add 16521261 0.3530398
## 6      reg_mul 14394458 0.2802641
## 7    reg_add1 73293367 1.6030492
## 8    reg_mul1 68265297 1.7688293

hw.rev.a = HoltWinters(rev1, seasonal = 'a')
hw.rev.a

## Holt-Winters exponential smoothing with trend and additive seasonal component.
##
## Call:
## HoltWinters(x = rev1, seasonal = "a")
##
## Smoothing parameters:
```



```
## alpha: 0.3628351
## beta : 0.02937448
## gamma: 0.04398116
##
## Coefficients:
##          [,1]
## a 36598906.876
## b  530299.405
## s1 -1377991.135
## s2  157858.747
## s3   -1230.095
## s4 2748405.192

pred = forecast(hw.rev.a, h = 4)
rmse(pred$mean, val_dat)

## [1] 18707727

mape(pred$mean, val_dat)

## [1] 0.4390395

hw.rev.a1 = HoltWinters(log(rev1), seasonal = 'a')
hw.rev.a1

## Holt-Winters exponential smoothing with trend and additive s
easonal component.
##
## Call:
## HoltWinters(x = log(rev1), seasonal = "a")
##
## Smoothing parameters:
## alpha: 0.2453319
## beta : 0
## gamma: 0.5397249
##
## Coefficients:
##          [,1]
## a 17.24077979
## b  0.07079279
## s1 0.13026647
## s2 0.19955937
## s3 0.32062663
## s4 0.18787439

pred = forecast(hw.rev.a1, h = 4)
rmse(exp(pred$mean), val_dat)

## [1] 12424017

mape(exp(pred$mean), val_dat)

## [1] 0.2372553
```

```

hw.rev.m = HoltWinters(rev1, seasonal = 'm')
hw.rev.m

## Holt-Winters exponential smoothing with trend and multiplica
tive seasonal component.
##
## Call:
## HoltWinters(x = rev1, seasonal = "m")
##
## Smoothing parameters:
##   alpha: 0
##   beta : 0
##   gamma: 0.6594779
##
## Coefficients:
##           [,1]
## a  -1.487344e+06
## b   -5.048134e+04
## s1  -2.599428e+01
## s2  -2.967443e+01
## s3  -3.145220e+01
## s4  -2.380703e+01

pred = forecast(hw.rev.m, h = 4)
rmse(pred$mean, val_dat)

## [1] 16593785

mape(pred$mean, val_dat)

## [1] 0.3447701

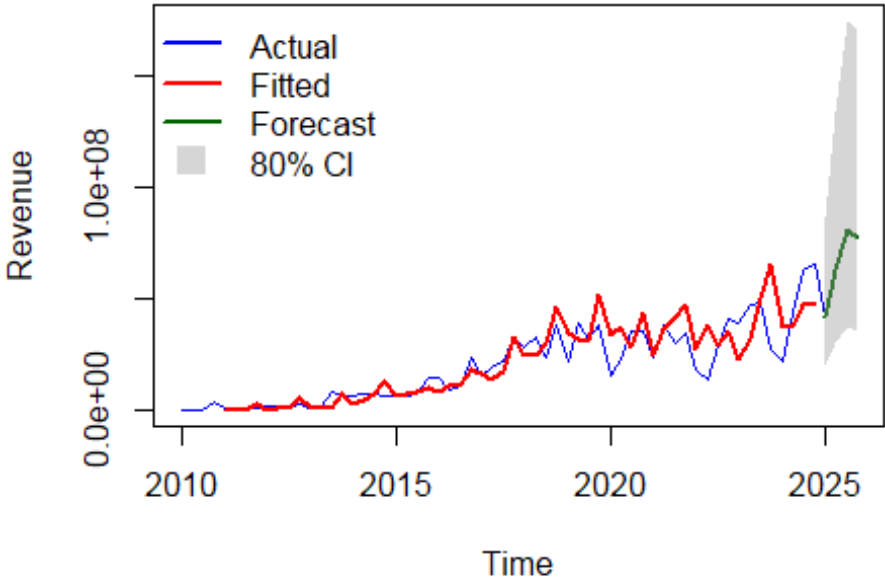
hw.rev.m1 = HoltWinters(log(rev1), seasonal = 'm')
hw.rev.m1

## Holt-Winters exponential smoothing with trend and multiplica
tive seasonal component.
##
## Call:
## HoltWinters(x = log(rev1), seasonal = "m")
##
## Smoothing parameters:
##   alpha: 0.22653
##   beta : 0
##   gamma: 0.5672312
##
## Coefficients:
##           [,1]
## a   17.20702970
## b    0.07079279
## s1   1.01113799
## s2   1.01436861

```

	<pre>## s3 1.02162897 ## s4 1.01253752 pred = forecast(hw.rev.m1, h = 4) rmse(exp(pred\$mean), val_dat) ## [1] 12585038 mape(exp(pred\$mean), val_dat) ## [1] 0.238489</pre>
	<p>Chọn mô hình tốt nhất, cập nhật dữ liệu đến 2024 và dự báo cho 2025:</p>
	<pre>hw.rev.a_2024 = HoltWinters(log(rev), seasonal = 'a') hw.rev.a_2024 ## Holt-Winters exponential smoothing with trend and additive s easonal component. ## ## Call: ## HoltWinters(x = log(rev), seasonal = "a") ## ## Smoothing parameters: ## alpha: 0.250636 ## beta : 0 ## gamma: 0.5168462 ## ## Coefficients: ## [,1] ## a 17.58006382 ## b 0.07079279 ## s1 -0.09784664 ## s2 0.24102779 ## s3 0.41791477 ## s4 0.30589349 pred = forecast(hw.rev.a_2024, 4) # Exponentiate forecast components pred_exp <- pred # Make a copy first pred_exp\$mean <- exp(pred\$mean) pred_exp\$lower <- exp(pred\$lower) pred_exp\$upper <- exp(pred\$upper) # Optionally, exponentiate fitted values too pred_exp\$fitted <- exp(pred\$fitted) # Print the exponentiated forecast print(pred_exp)</pre>

	<pre>## Point Forecast Lo 80 Hi 80 Lo 95 Hi 95 ## 2025 Q1 41992852 20339003 86700396 13856719 127259535 ## 2025 Q2 63254523 29957629 133559793 20168935 198381060 ## 2025 Q3 81032184 37550761 174862364 24991325 262739764 ## 2025 Q4 77759282 35279232 171389952 23217876 260424592 # Combine actual and forecasted mean into one ts object combined_series <- ts(c(rev, pred_exp\$mean), start = start(rev), frequency = frequency(rev)) # Get forecast start time forecast_start <- time(rev)[length(rev)] + 1/frequency(rev) forecast_time <- time(ts(pred_exp\$mean, start = forecast_start, frequency = frequency(rev))) # Calculate y-axis limits to include everything y_max <- max(c(rev, pred_exp\$mean, pred_exp\$upper[,1]), na.rm = TRUE) y_min <- min(c(rev, pred_exp\$mean, pred_exp\$lower[,1]), na.rm = TRUE) # Plot the actual + forecasted series plot.ts(combined_series, col = "blue", ylab = "Revenue", ylim = c(y_min, y_max)) # Add fitted values lines(pred_exp\$fitted, col = "red", lwd = 2) # Add forecasted mean lines(ts(pred_exp\$mean, start = forecast_start, frequency = frequency(rev)), col = "darkgreen", lwd = 2) # Add shaded confidence interval polygon(c(forecast_time, rev(forecast_time)), c(pred_exp\$lower[,1], rev(pred_exp\$upper[,1])), col = rgb(0.6, 0.6, 0.6, 0.4), border = NA) # Add Legend legend("topleft", legend = c("Actual", "Fitted", "Forecast", "80% CI"), col = c("blue", "red", "darkgreen", rgb(0.6, 0.6, 0.6, 0.4)), lty = c(1, 1, 1, NA), lwd = c(2, 2, 2, NA), pch = c(NA, NA, NA, 15), pt.cex = 2, bty = "n")</pre>
--	---

	
3.3	CÁC MÔ HÌNH VỚI GIÁ CỔ PHIẾU
	Tải dữ liệu
	<pre>VIC <- read_excel("Vingroup JSC Stock Price History.xlsx") r.actual = read_excel("Min_var.xlsx", sheet = "2025", range = "H3:H12", col_names = FALSE)[[1]] ## New names: ## • `` -> `...1` attach(VIC) `Log_return%` = `Log_return%`[-1]</pre>
3.3.1	CÁC MÔ HÌNH VỚI CHUỖI LOG RETURN
	Kiểm định ADF
	<pre>summary(ur.df(`Log_return%`, type = "trend", selectlags = "AIC")) ## ## ##### ## # Augmented Dickey-Fuller Test Unit Root Test # ## ##### ##</pre>

```
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.5146 -0.6823  0.0394  0.4692  7.3079
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7.833e-02  1.574e-01  -0.498    0.619
## z.lag.1      -8.997e-01  5.928e-02 -15.178 <2e-16 ***
## tt           7.592e-05  5.469e-04   0.139    0.890
## z.diff.lag    3.893e-02  4.502e-02   0.865    0.388
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.744 on 492 degrees of freedom
## Multiple R-squared:  0.4339, Adjusted R-squared:  0.4304
## F-statistic: 125.7 on 3 and 492 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -15.1781 76.7912 115.1868
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -3.98 -3.42 -3.13
## phi2  6.15  4.71  4.05
## phi3  8.34  6.30  5.36

summary(ur.df(`Log_return%`, type = "drift", selectlags = "AIC"))

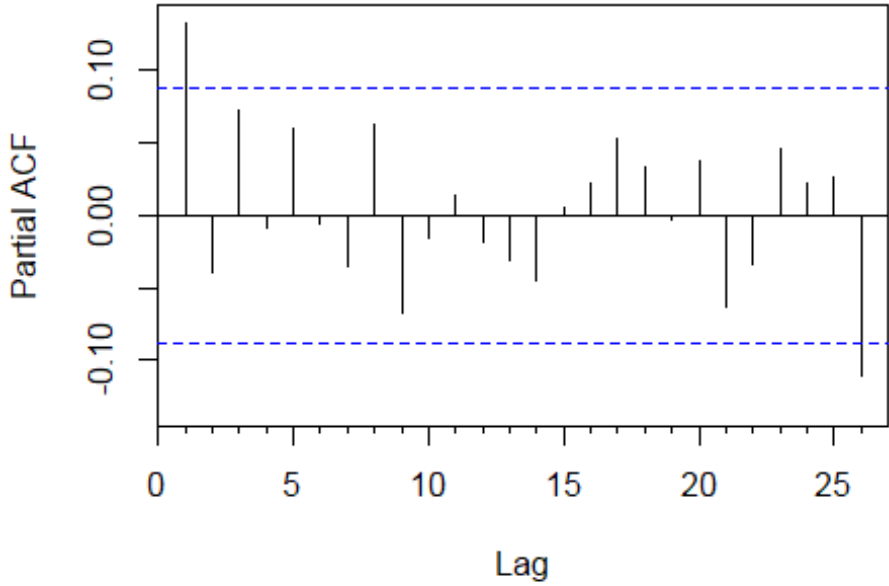
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.5184 -0.6773  0.0437  0.4758  7.3006
##
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.05938    0.07832  -0.758    0.449
## z.lag.1      -0.89966    0.05922 -15.193 <2e-16 ***
## z.diff.lag    0.03888    0.04497   0.865    0.388
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.742 on 493 degrees of freedom
## Multiple R-squared:  0.4339, Adjusted R-squared:  0.4316
## F-statistic: 188.9 on 2 and 493 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -15.1925 115.4068
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.44 -2.87 -2.57
## phi1  6.47  4.61  3.79

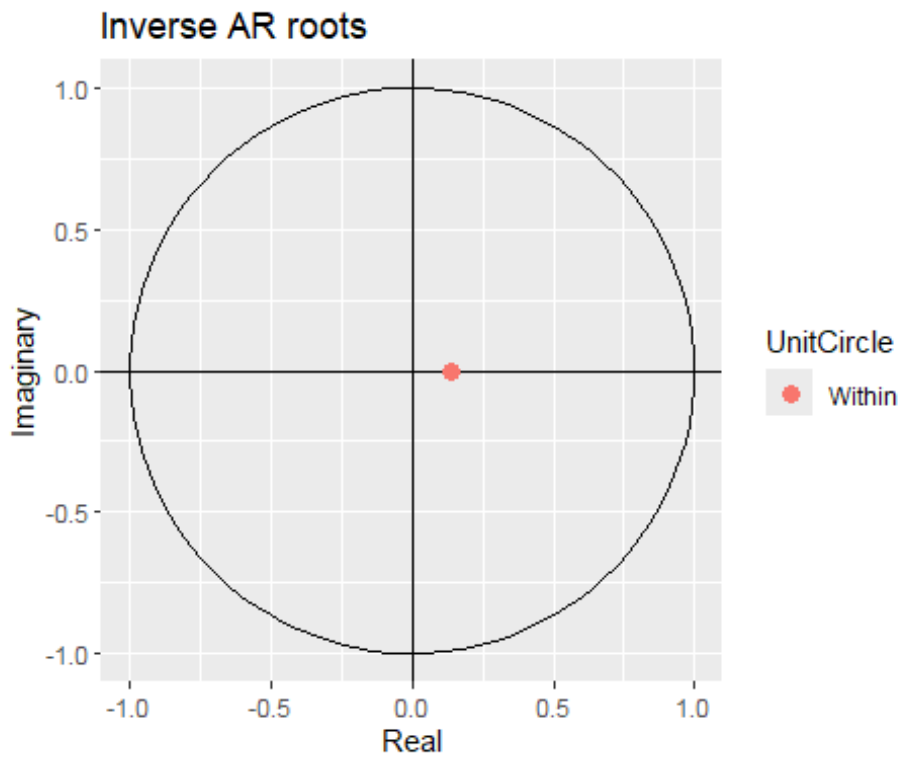
summary(ur.df(`Log_return%`,type = "none", selectlags = "AIC"))

##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.5810 -0.7365 -0.0157  0.4170  7.2441
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## z.lag.1    -0.89740    0.05912  -15.18 <2e-16 ***
## z.diff.lag  0.03772    0.04493   0.84    0.401
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.741 on 494 degrees of freedom
## Multiple R-squared:  0.4332, Adjusted R-squared:  0.4309
## F-statistic: 188.8 on 2 and 494 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -15.1802
```

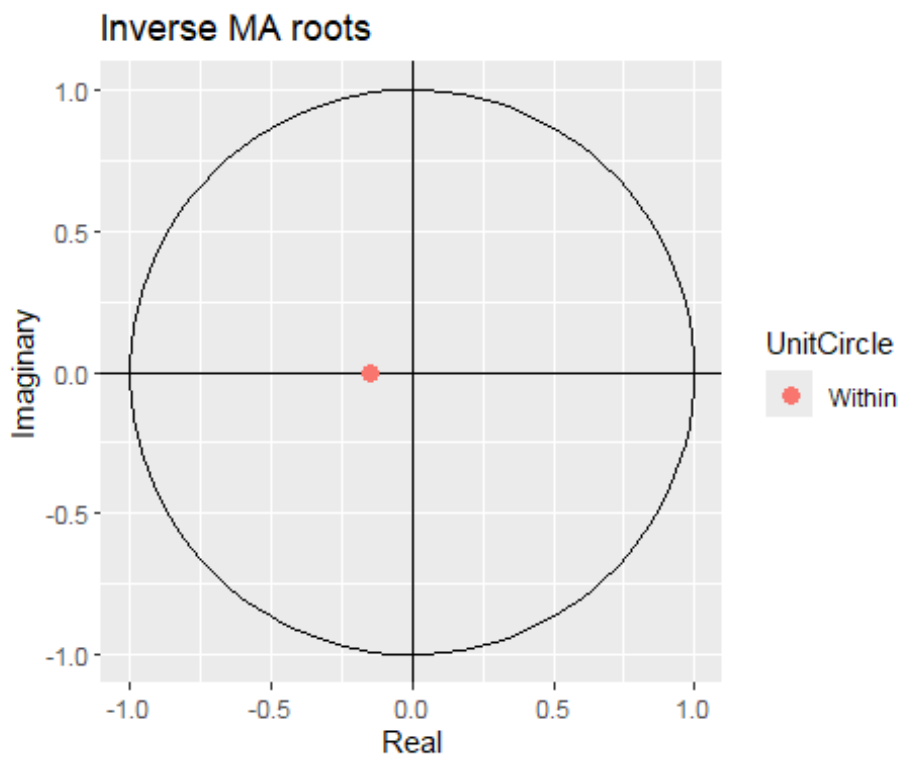
	<pre>## ## Critical values for test statistics: ## 1pct 5pct 10pct ## tau1 -2.58 -1.95 -1.62</pre>
	Lược đồ tương quan
	<pre>Acf(`Log_return%`)</pre> <p style="text-align: center;">Series Log_return%</p> <pre>Pacf(`Log_return%`)</pre>

	<p style="text-align: center;">Series Log_return%</p> 
	<p>ARIMA</p>
	<pre> r_ar1 = Arima(`Log_return%`, order = c(1,0,0), include.constant = 0) summary(r_ar1) ## Series: Log_return% ## ARIMA(1,0,0) with zero mean ## ## Coefficients: ## ar1 ## 0.1346 ## s.e. 0.0444 ## ## sigma^2 = 3.023: log likelihood = -981.58 ## AIC=1967.16 AICc=1967.18 BIC=1975.58 ## ## Training set error measures: ## ME RMSE MAE MPE MAPE MAS E ACF1 ## Training set -0.05853393 1.736868 1.090303 NaN Inf 0.719982 8 0.004087342 r_ma1 = Arima(`Log_return%`, order = c(0,0,1), include.constant = 0) summary(r_ma1) ## Series: Log_return% ## ARIMA(0,0,1) with zero mean ## ## Coefficients: </pre>

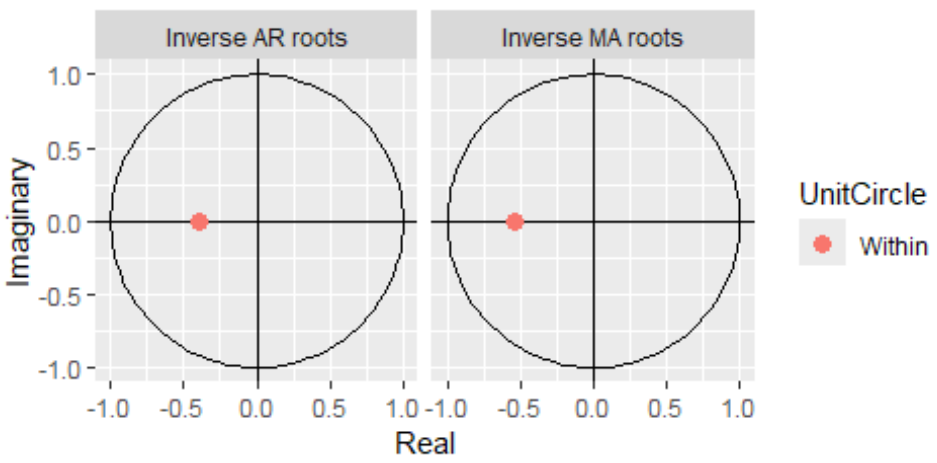
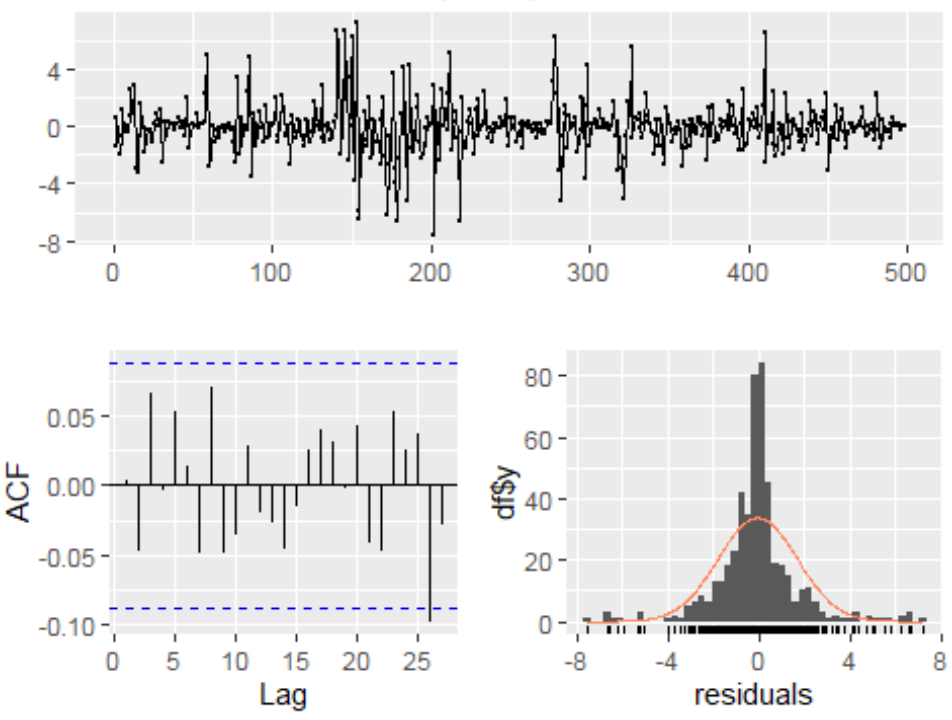
	<pre>## ma1 ## 0.1472 ## s.e. 0.0466 ## ## sigma^2 = 3.018: log likelihood = -981.2 ## AIC=1966.4 AICc=1966.42 BIC=1974.82 ## ## Training set error measures: ## ME RMSE MAE MPE MAPE MAS E ACF1 ## Training set -0.05896597 1.735543 1.091092 NaN Inf 0.720503 7 -0.006662006 r_ar1ma1 = Arima(`Log_return%`, order = c(1,0,1), include.constant = 0) summary(r_ar1ma1) ## Series: Log_return% ## ARIMA(1,0,1) with zero mean ## ## Coefficients: ## ar1 ma1 ## -0.3989 0.5397 ## s.e. 0.2064 0.1887 ## ## sigma^2 = 3.014: log likelihood = -980.33 ## AIC=1966.66 AICc=1966.71 BIC=1979.29 ## ## Training set error measures: ## ME RMSE MAE MPE MAPE MAS E ACF1 ## Training set -0.06148467 1.732497 1.092766 NaN Inf 0.721609 1 0.003661581</pre>
	Đồ thị nghiệm nghịch đảo
	<code>autoplot(r_ar1)</code>



```
autoplot(r_ma1)
```



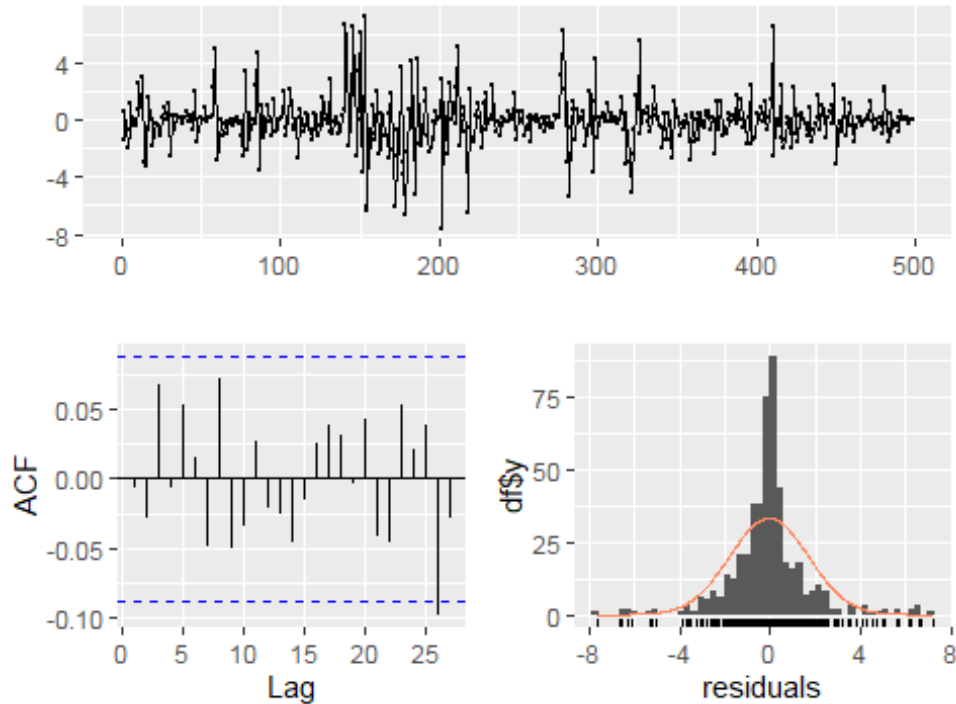
```
autoplot(r_ar1ma1)
```

	
	<p>Kiểm định tự tương quan của phần dư</p>
	<pre>checkresiduals(r_ar1)</pre> <p>Residuals from ARIMA(1,0,0) with zero mean</p>  <pre>## ## Ljung-Box test</pre>

```
##
## data: Residuals from ARIMA(1,0,0) with zero mean
## Q* = 10.518, df = 9, p-value = 0.3102
##
## Model df: 1. Total lags used: 10
```

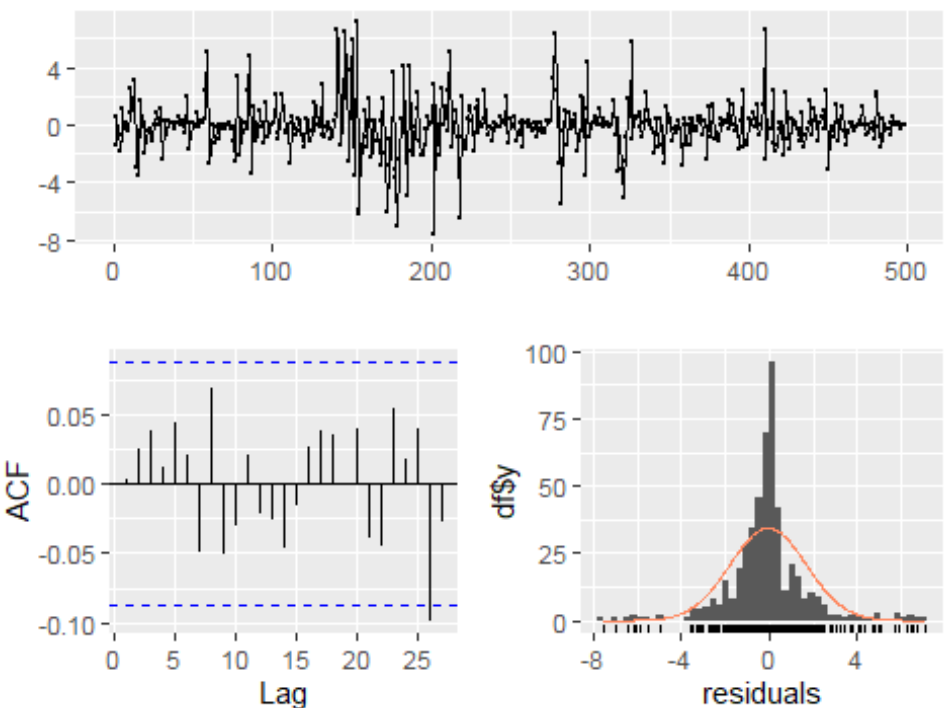
```
checkresiduals(r_ma1)
```

Residuals from ARIMA(0,0,1) with zero mean



```
##
## Ljung-Box test
##
## data: Residuals from ARIMA(0,0,1) with zero mean
## Q* = 10.061, df = 9, p-value = 0.3456
##
## Model df: 1. Total lags used: 10
```

```
checkresiduals(r_ar1ma1)
```

	<p>Residuals from ARIMA(1,0,1) with zero mean</p>  <pre>## ## Ljung-Box test ## ## data: Residuals from ARIMA(1,0,1) with zero mean ## Q* = 7.7899, df = 8, p-value = 0.4543 ## ## Model df: 2. Total lags used: 10</pre>
3.3.2	CÁC MÔ HÌNH VỚI CHUỖI GIÁ
	Kiểm định ADF chuỗi giá
	<pre>summary(ur.df(Price,type = "trend", selectlags = "AIC")) ## ## ##### ## # Augmented Dickey-Fuller Test Unit Root Test # ## ##### ## ## Test regression trend ## ## ## Call: ## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag) ## ## Residuals: ## Min 1Q Median 3Q Max ## -3895.6 -316.1 -5.3 222.0 5813.4</pre>

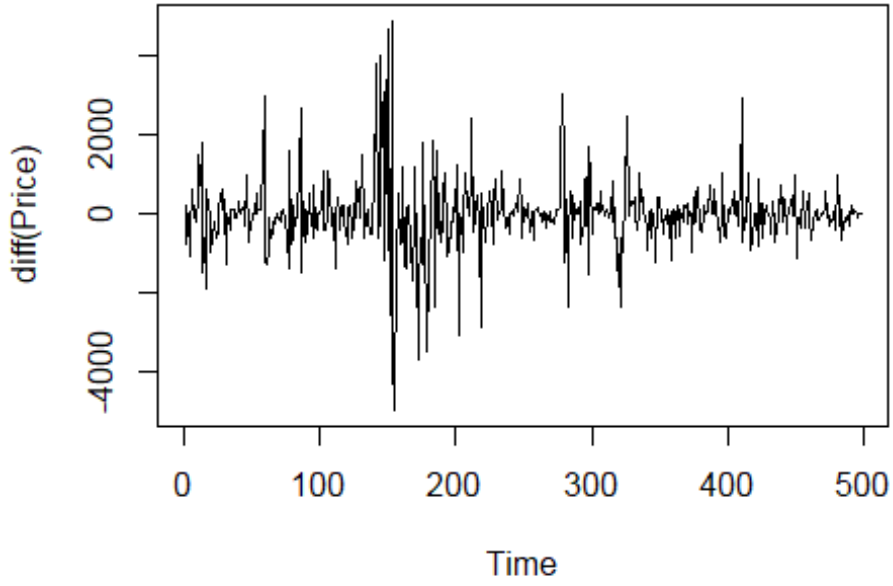
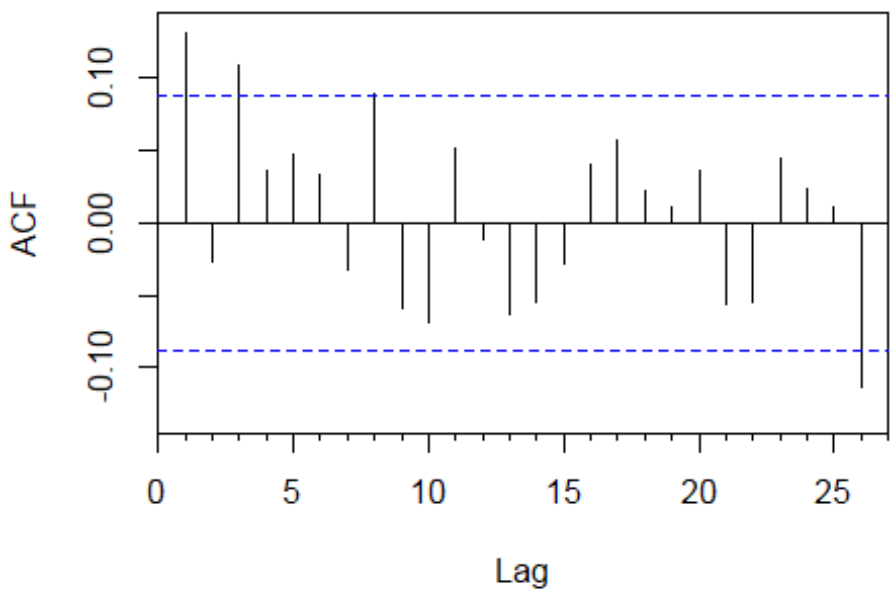
	<pre>## ## Coefficients: ## Estimate Std. Error t value Pr(> t) ## (Intercept) 1.393e+03 5.353e+02 2.603 0.00951 ** ## z.lag.1 -2.549e-02 9.434e-03 -2.702 0.00713 ** ## tt -8.467e-01 4.335e-01 -1.953 0.05137 . ## z.diff.lag 1.443e-01 4.454e-02 3.240 0.00128 ** ## --- ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ## ## Residual standard error: 912.3 on 493 degrees of freedom ## Multiple R-squared: 0.0317, Adjusted R-squared: 0.0258 ## F-statistic: 5.379 on 3 and 493 DF, p-value: 0.001191 ## ## ## Value of test-statistic is: -2.702 2.5814 3.6581 ## ## Critical values for test statistics: ## 1pct 5pct 10pct ## tau3 -3.98 -3.42 -3.13 ## phi2 6.15 4.71 4.05 ## phi3 8.34 6.30 5.36</pre>
	Kiểm định ADF chuỗi sai phân của chuỗi giá
	<pre>summary(ur.df(diff(Price),type = "trend", selectlags = "AIC")) ## ## ##### ## # Augmented Dickey-Fuller Test Unit Root Test # ## ##### ## ## Test regression trend ## ## ## Call: ## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag) ## ## Residuals: ## Min 1Q Median 3Q Max ## -4459.3 -312.2 16.8 227.1 5323.3 ## ## Coefficients: ## Estimate Std. Error t value Pr(> t) ## (Intercept) -40.21374 82.93813 -0.485 0.628 ## z.lag.1 -0.90674 0.05933 -15.284 <2e-16 *** ## tt 0.04607 0.28818 0.160 0.873 ## z.diff.lag 0.04473 0.04500 0.994 0.321 ## ---</pre>

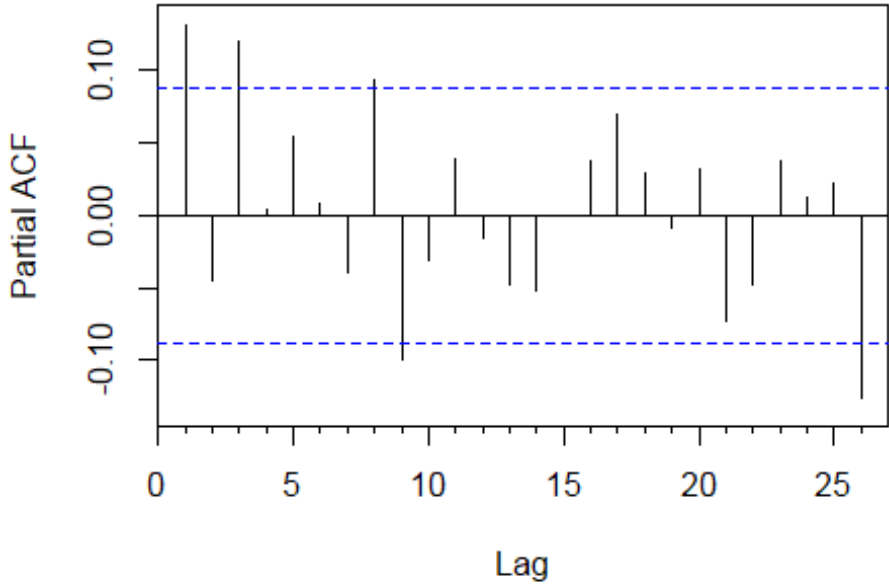
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 918.9 on 492 degrees of freedom
## Multiple R-squared:  0.4351, Adjusted R-squared:  0.4317
## F-statistic: 126.3 on 3 and 492 DF,  p-value: < 2.2e-16
##
## Value of test-statistic is: -15.2837 77.8638 116.7956
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -3.98 -3.42 -3.13
## phi2  6.15  4.71  4.05
## phi3  8.34  6.30  5.36

summary(ur.df(diff(Price),type = "drift",selectlags = "AIC"))

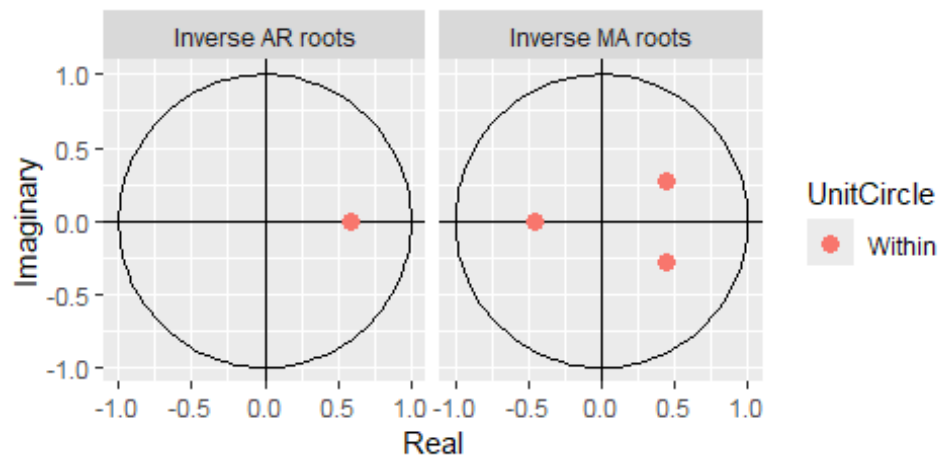
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression drift
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4463.8  -313.7    21.9   218.7  5318.9
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -28.71643   41.26291  -0.696   0.487
## z.lag.1      -0.90664    0.05927 -15.298 <2e-16 ***
## z.diff.lag    0.04467    0.04496   0.994   0.321
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 918 on 493 degrees of freedom
## Multiple R-squared:  0.4351, Adjusted R-squared:  0.4328
## F-statistic: 189.9 on 2 and 493 DF,  p-value: < 2.2e-16
##
## Value of test-statistic is: -15.298 117.0142
##
## Critical values for test statistics:
##      1pct  5pct 10pct
```


	<pre>## tau2 -3.44 -2.87 -2.57 ## phi1 6.47 4.61 3.79 summary(ur.df(diff(Price),type = "none",selectlags = "AIC")) ## ## ##### ## # Augmented Dickey-Fuller Test Unit Root Test # ## ##### ## ## Test regression none ## ## ## Call: ## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag) ## ## Residuals: ## Min 1Q Median 3Q Max ## -4494.4 -342.7 -6.8 190.4 5291.8 ## ## Coefficients: ## Estimate Std. Error t value Pr(> t) ## z.lag.1 -0.90475 0.05917 -15.290 <2e-16 *** ## z.diff.lag 0.04369 0.04491 0.973 0.331 ## --- ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ## ## Residual standard error: 917.5 on 494 degrees of freedom ## Multiple R-squared: 0.4345, Adjusted R-squared: 0.4322 ## F-statistic: 189.8 on 2 and 494 DF, p-value: < 2.2e-16 ## ## ## Value of test-statistic is: -15.2901 ## ## Critical values for test statistics: ## 1pct 5pct 10pct ## tau1 -2.58 -1.95 -1.62</pre>
	Đồ thị chuỗi sai phân
	<pre>plot(diff(Price), type = "l", xlab = "Time")</pre>

	
	<p>Lược đồ tương quan</p> <p><code>Acf(diff(Price))</code></p> <p>Series diff(Price)</p>  <p><code>Pacf(diff(Price))</code></p>

	<p style="text-align: center;">Series diff(Price)</p> 
	ARIMA
	<pre> p_arima113 = Arima(Price, order = c(1,1,3), include.constant = 0) summary(p_arima113) ## Series: Price ## ARIMA(1,1,3) ## ## Coefficients: ## ar1 ma1 ma2 ma3 ## 0.5804 -0.4380 -0.1302 0.1263 ## s.e. 0.2667 0.2671 0.0630 0.0436 ## ## sigma^2 = 829430: log likelihood = -4098.15 ## AIC=8206.31 AICc=8206.43 BIC=8227.36 ## ## Training set error measures: ## ME RMSE MAE MPE MAPE MASE ## Training set -24.3397 906.1564 544.3698 -0.06067848 1.098787 1.009481 ## ACF1 ## Training set -0.001870446 p_arima311 = Arima(Price, order = c(3,1,1), include.constant = 0) summary(p_arima311) ## Series: Price ## ARIMA(3,1,1) </pre>

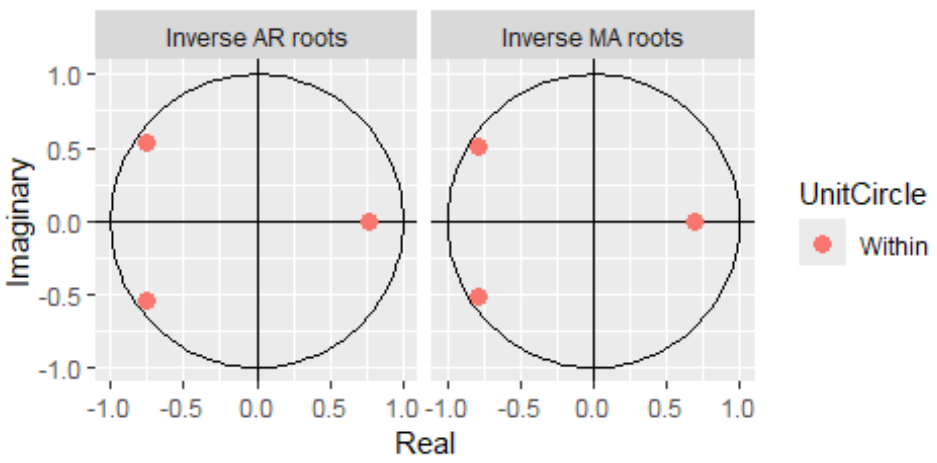
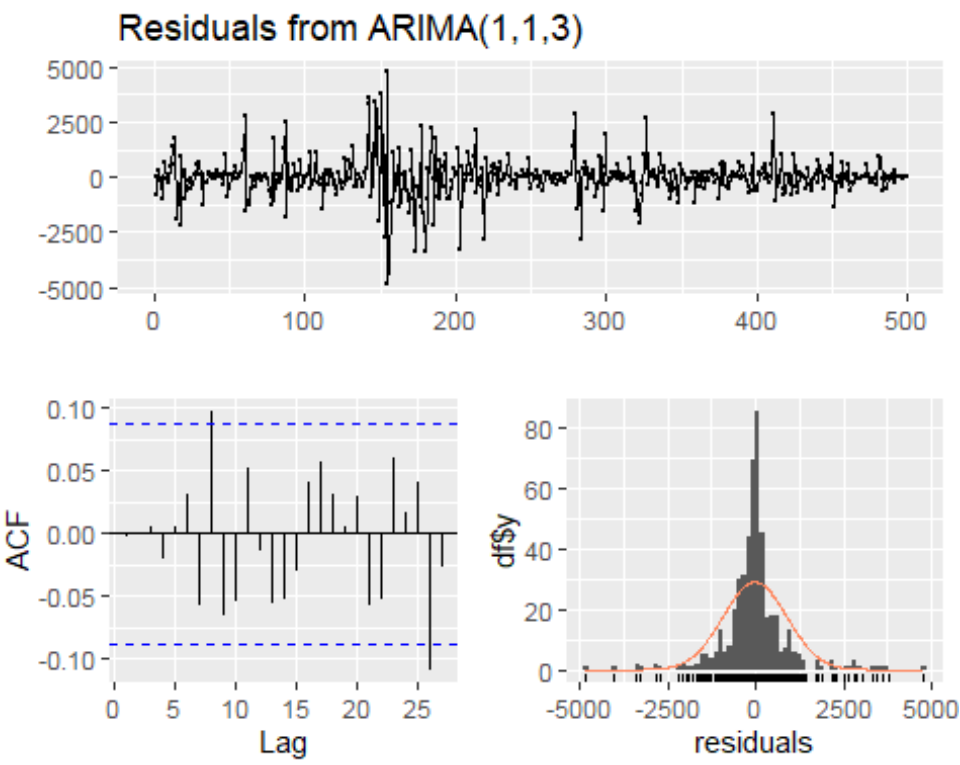
	<pre>## ## Coefficients: ## ar1 ar2 ar3 ma1 ## 0.3845 -0.0931 0.1288 -0.2460 ## s.e. 0.5305 0.0854 0.0448 0.5378 ## ## sigma^2 = 830644: log likelihood = -4098.51 ## AIC=8207.03 AICc=8207.15 BIC=8228.08 ## ## Training set error measures: ## ME RMSE MAE MPE MAP E MASE ## Training set -24.90583 906.8188 545.1821 -0.06225865 1.10003 9 1.010987 ## ACF1 ## Training set 0.002144244 p_arima313 = Arima(Price, order = c(3,1,3), include.constant = 0) summary(p_arima313) ## Series: Price ## ARIMA(3,1,3) ## ## Coefficients: ## ar1 ar2 ar3 ma1 ma2 ma3 ## -0.7364 0.2857 0.6491 0.8906 -0.2006 -0.6119 ## s.e. 0.1827 0.2425 0.1383 0.1884 0.2877 0.1571 ## ## sigma^2 = 820277: log likelihood = -4094.5 ## AIC=8202.99 AICc=8203.22 BIC=8232.47 ## ## Training set error measures: ## ME RMSE MAE MPE MAP E MASE ## Training set -24.01727 899.3165 546.7992 -0.05959358 1.10414 5 1.013986 ## ACF1 ## Training set 0.0017566</pre>
	Đồ thị nghiệm nghịch đảo
	<code>autoplot(p_arima113)</code>



```
autoplot(p_arima311)
```

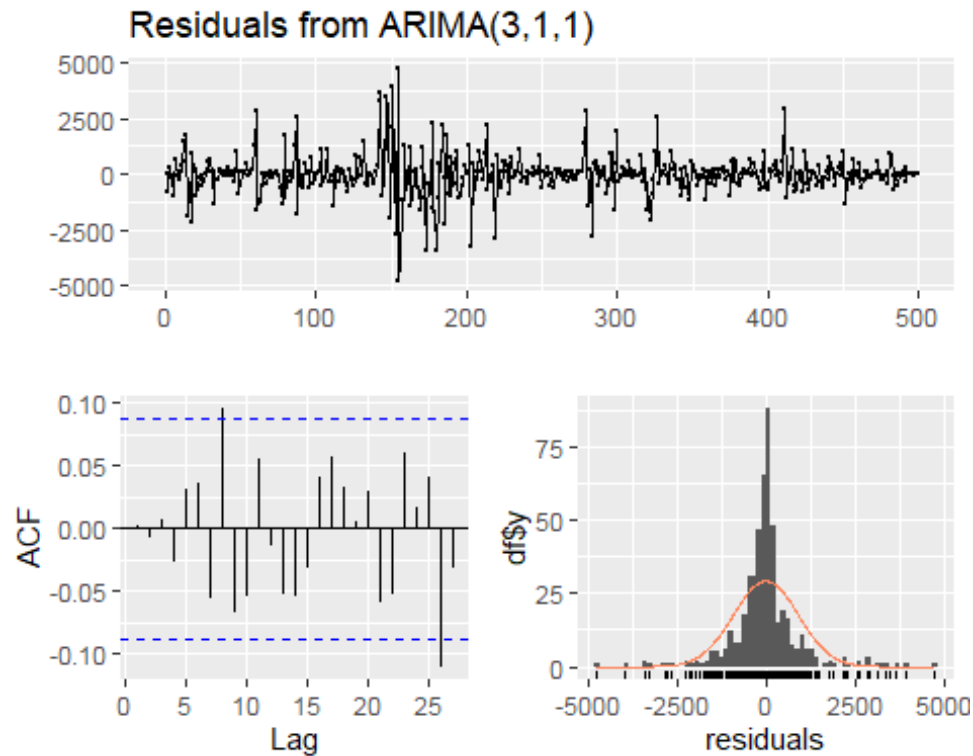


```
autoplot(p_arima313)
```

	
	<p>Kiểm định tự tương quan của phần dư</p>
	<pre>checkresiduals(p_arima113)</pre> <p>Residuals from ARIMA(1,1,3)</p>  <pre>## ## Ljung-Box test ##</pre>

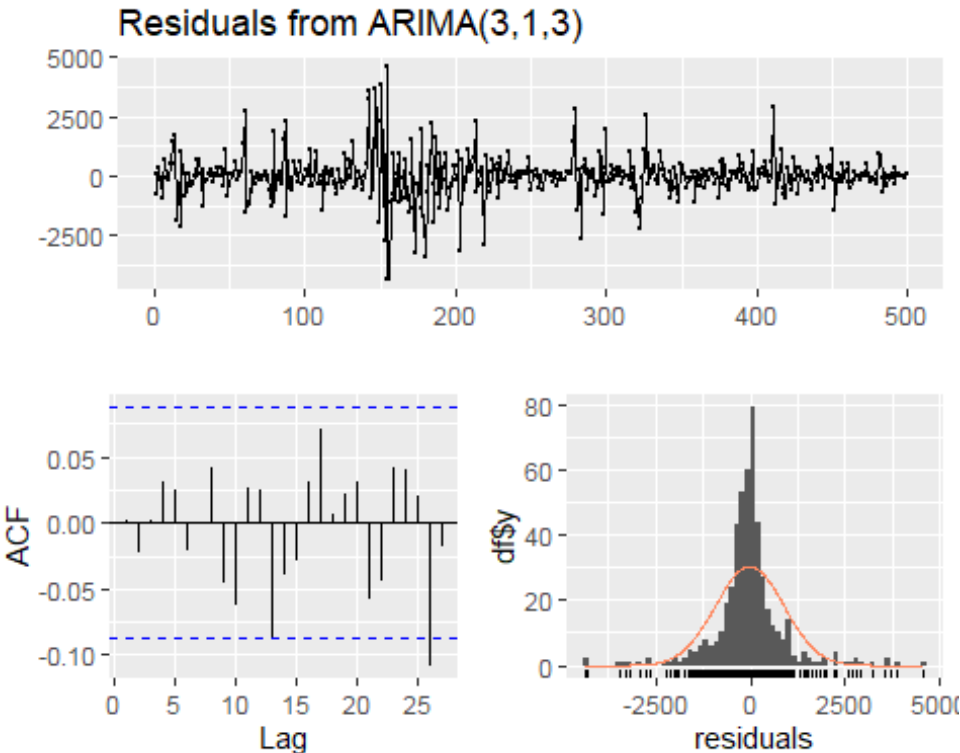
```
## data: Residuals from ARIMA(1,1,3)
## Q* = 10.858, df = 6, p-value = 0.09286
##
## Model df: 4. Total lags used: 10
```

```
checkresiduals(p_arma311)
```



```
##
## Ljung-Box test
##
## data: Residuals from ARIMA(3,1,1)
## Q* = 11.627, df = 6, p-value = 0.07082
##
## Model df: 4. Total lags used: 10
```

```
checkresiduals(p_arma313)
```

	 <pre>## ## Ljung-Box test ## ## data: Residuals from ARIMA(3,1,3) ## Q* = 5.2461, df = 4, p-value = 0.263 ## ## Model df: 6. Total lags used: 10</pre>
3.3.3	KẾT QUẢ
	Dự báo cho 10 phiên đầu năm 2025 và tính RMSE
	<pre>forecast(r_ar1, h=10)\$mean ## Time Series: ## Start = 499 ## End = 508 ## Frequency = 1 ## [1] 0 0 0 0 0 0 0 0 0 0 rmse(r.actual, forecast(r_ar1, h=10)\$mean) ## [1] 0.3916437 forecast(r_ma1, h=10)\$mean ## Time Series: ## Start = 499 ## End = 508 ## Frequency = 1</pre>


```
## [1] -0.0007770817 0.0000000000 0.0000000000 0.0000000000
0.0000000000
## [6] 0.0000000000 0.0000000000 0.0000000000 0.0000000000
0.0000000000

rmse(r.actual, forecast(r_ma1, h=10)$mean)

## [1] 0.3916683

forecast(r_ar1ma1, h=10)$mean

## Time Series:
## Start = 499
## End = 508
## Frequency = 1
## [1] -9.544781e-03 3.807087e-03 -1.518517e-03 6.056845e-04
-2.415868e-04
## [6] 9.636074e-05 -3.843501e-05 1.533041e-05 -6.114777e-06
2.438976e-06

rmse(r.actual, forecast(r_ar1ma1, h=10)$mean)

## [1] 0.3921733

p = c(tail(Price,1), forecast(p_arima113, h = 10)$mean)
r_fc = log(p[2:11]/p[1:10])*100
r_fc

## [1] -0.0239343568 -0.0091827203 -0.0083844668 -0.0048669985
-0.0028251070
## [6] -0.0016398404 -0.0009518405 -0.0005524900 -0.0003206885
-0.0001861407

rmse(r.actual, r_fc)

## [1] 0.3919672

p = c(tail(Price,1), forecast(p_arima311, h = 10)$mean)
r_fc = log(p[2:11]/p[1:10])*100
r_fc

## [1] -0.0272035674 -0.0104608680 -0.0014901513 -0.0031045418
-0.0024025943
## [6] -0.0008268027 -0.0004942334 -0.0004225728 -0.0002229803
-0.0001100715

rmse(r.actual, r_fc)

## [1] 0.3919538

p = c(tail(Price,1), forecast(p_arima313, h = 10)$mean)
r_fc = log(p[2:11]/p[1:10])*100
r_fc
```

[1] 0.05014658 0.02099961 -0.07877543 0.09655424 -0.07996630 0.03535084
[7] 0.01379944 -0.05198365 0.06516696 -0.05387824
rmse(r.actual, r_fc)
[1] 0.3929238

4. Chuỗi LSS – Phạm Phương Thảo

Bảng 4: Kết quả ước lượng mô hình dự báo doanh thu thuần sử dụng dữ liệu quý 1/2010-quý 4/2023

STT	Mô hình	Kết quả ước lượng
1	Lin-lin	<pre> #Mô hình 1: Lin-Lin df_train <- data.frame(LSS.train, time.train) linlin <- lm(LSS.train ~ time.train, data = df_train) summary(linlin) ## ## Call: ## lm(formula = LSS.train ~ time.train, data = df_train) ## ## Residuals: ## Min 1Q Median 3Q Max ## -336.18 -106.68 -21.18 86.07 392.48 ## ## Coefficients: ## Estimate Std. Error t value Pr(> t) ## (Intercept) 386.872 46.005 8.409 2.17e-11 *** ## time.train 2.333 1.404 1.661 0.102 ## --- ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ## ## Residual standard error: 169.8 on 54 degrees of freedom ## Multiple R-squared: 0.04863, Adjusted R-squared: 0.03101 ## F-statistic: 2.76 on 1 and 54 DF, p-value: 0.1024 #Lin-Lin: RMSE, MAPE toàn mẫu đến 2023Q4 rmse(LSS.train,fitted(linlin)) ## [1] 166.7757 mape(LSS.train,fitted(linlin)) ## [1] 0.3524753 #Lin-Lin: RMSE, MAPE cho forecast 2024Q1 đến 2024Q4 df_new <- data.frame(time.train = c(57, 58, 59, 60)) rmse(LSS.test,predict(linlin, newdata=df_new)) ## [1] 215.7434 mape(LSS.test,predict(linlin, newdata=df_new)) ## [1] 0.222638 </pre>

2	Lin-log	<pre> #Mô hình 2: Lin-Log linlog<-lm(LSS.train~log(time.train),data = df_train) summary(linlog) ## ## Call: ## lm(formula = LSS.train ~ log(time.train), data = df_train) ## ## Residuals: ## Min 1Q Median 3Q Max ## -339.38 -106.46 -22.65 82.52 377.96 ## ## Coefficients: ## Estimate Std. Error t value Pr(> t) ## (Intercept) 286.66 80.50 3.561 0.000781 *** ## log(time.train) 54.16 25.13 2.155 0.035612 * ## --- ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ## ## Residual standard error: 167.1 on 54 degrees of freedom ## Multiple R-squared: 0.07921, Adjusted R-squared: 0.06216 ## F-statistic: 4.645 on 1 and 54 DF, p-value: 0.03561 #Lin-Log: RMSE, MAPE toàn mẫu đến 2023Q4 rmse(LSS.train,fitted(linlog)) ## [1] 164.0733 mape(LSS.train,fitted(linlog)) ## [1] 0.3401541 #Lin-Log: RMSE, MAPE cho forecast 2024Q1 đến 2024Q4 rmse(LSS.test,predict(linlog, newdata=df_new)) ## [1] 226.6463 mape(LSS.test,predict(linlog, newdata=df_new)) ## [1] 0.2309166 </pre>
3	Log-lin	<pre> #Mô hình 3: Log-Lin loglin<-lm(log(LSS.train)~time.train) summary(loglin) ## ## Call: ## lm(formula = log(LSS.train) ~ time.train) ## ## Residuals: ## Min 1Q Median 3Q Max ## -1.09537 -0.20958 0.03165 0.23397 0.69589 ## ## Coefficients: ## Estimate Std. Error t value Pr(> t) ## (Intercept) 5.907365 0.105577 55.953 <2e-16 *** ## time.train 0.004774 0.003222 1.482 0.144 ## --- ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ## ## Residual standard error: 0.3898 on 54 degrees of freedom ## Multiple R-squared: 0.03906, Adjusted R-squared: 0.02127 ## F-statistic: 2.195 on 1 and 54 DF, p-value: 0.1443 #Log-Lin: RMSE, MAPE toàn mẫu đến 2023Q4 rmse(LSS.train,exp(fitted(loglin))) ## [1] 169.7508 mape(LSS.train,exp(fitted(loglin))) ## [1] 0.3278635 #Log-Lin: RMSE, MAPE cho forecast 2024Q1 đến 2024Q4 rmse(LSS.test,exp(predict(loglin, newdata=df_new))) ## [1] 242.8502 mape(LSS.test,exp(predict(loglin, newdata=df_new))) ## [1] 0.2423486 </pre>

4	Log-log	<pre> #Mô hình 4: Log-Log loglog<-lm(log(LSS.train)~log(time.train)) summary(loglog) ## ## Call: ## lm(formula = log(LSS.train) ~ log(time.train)) ## ## Residuals: ## Min 1Q Median 3Q Max ## -1.11015 -0.20911 0.02045 0.23994 0.66268 ## ## Coefficients: ## Estimate Std. Error t value Pr(> t) ## (Intercept) 5.66514 0.18392 30.803 <2e-16 *** ## log(time.train) 0.12291 0.05741 2.141 0.0368 * ## --- ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ## ## Residual standard error: 0.3817 on 54 degrees of freedom ## Multiple R-squared: 0.07823, Adjusted R-squared: 0.06116 ## F-statistic: 4.583 on 1 and 54 DF, p-value: 0.03682 #Log-Log: RMSE, MAPE toàn mẫu đến 2023Q4 rmse(LSS.train,exp(fitted(loglog))) ## [1] 167.1838 mape(LSS.train,exp(fitted(loglog))) ## [1] 0.3170051 #Log-Log: RMSE, MAPE cho forecast 2024Q1 đến 2024Q4 rmse(LSS.test,exp(predict(loglog, newdata=df_new))) ## [1] 250.2712 mape(LSS.test,exp(predict(loglog, newdata=df_new))) ## [1] 0.2532512 </pre>
---	---------	--

5	Linear trend + Additive	<pre> #Tạo biến giả mùa vụ theo quý s1 <- c(rep(c(1,0,0,0), 14)) s2 <- c(rep(c(0,1,0,0), 14)) s3 <- c(rep(c(0,0,1,0), 14)) s4 <- c(rep(c(0,0,0,1), 14)) #Mô hình 5: Linear trend + Additive df_train <- data.frame(LSS.train, time.train, s1, s2, s3, s4) linadd <- lm(LSS.train ~ time.train + s2 + s3 + s4, data = df_train) summary(linadd) ## ## Call: ## lm(formula = LSS.train ~ time.train + s2 + s3 + s4, data = df_train) ## ## Residuals: ## Min 1Q Median 3Q Max ## -252.61 -99.97 -3.44 90.89 438.99 ## ## Coefficients: ## Estimate Std. Error t value Pr(> t) ## (Intercept) 338.496 50.549 6.696 1.64e-08 *** ## time.train 2.397 1.202 1.994 0.051464 . ## s2 198.317 54.825 3.617 0.000683 *** ## s3 -50.294 54.865 -0.917 0.363621 ## s4 38.166 54.931 0.695 0.490330 ## --- ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ## ## Residual standard error: 145 on 51 degrees of freedom ## Multiple R-squared: 0.3449, Adjusted R-squared: 0.2935 ## F-statistic: 6.712 on 4 and 51 DF, p-value: 0.0002028 #Lin-add: RMSE, MAPE toàn mẫu đến 2023Q4 rmse(LSS.train,fitted(linadd)) ## [1] 138.394 mape(LSS.train,fitted(linadd)) ## [1] 0.2992807 #Lin-add: RMSE, MAPE cho forecast 2024Q1 đến 2024Q4 df_new2 <- data.frame(time.train = c(57, 58, 59, 60), s2 = c(0,1,0,0), s3 = c(0,0,1,0), s4 = c(0,0,0,1)) rmse(LSS.test,predict(linadd, newdata=df_new2)) ## [1] 171.5384 mape(LSS.test,predict(linadd, newdata=df_new2)) ## [1] 0.2028936 </pre>
---	-------------------------	--

6	Linear trend + Multiplicative	<pre> #Mô hình 6: Linear trend + Multiplicative linmul<-lm(LSS.train ~ time.train + time.train*s2+ time.train*s3+ time.train*s4, data = df_train) summary(linmul) ## ## Call: ## lm(formula = LSS.train ~ time.train + time.train * s2 + time.train * ## s3 + time.train * s4, data = df_train) ## ## Residuals: ## Min 1Q Median 3Q Max ## -255.32 -101.70 -8.19 85.18 435.75 ## ## Coefficients: ## Estimate Std. Error t value Pr(> t) ## (Intercept) 294.725 69.947 4.214 0.00011 *** ## time.train 4.018 2.224 1.807 0.07710 . ## s2 96.419 100.286 0.961 0.34115 ## s3 134.597 101.682 1.324 0.19187 ## s4 140.660 103.107 1.364 0.17886 ## time.train:s2 3.581 3.145 1.139 0.26054 ## time.train:s3 -6.487 3.145 -2.062 0.04460 * ## time.train:s4 -3.579 3.145 -1.138 0.26090 ## --- ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ## ## Residual standard error: 134.2 on 48 degrees of freedom ## Multiple R-squared: 0.4721, Adjusted R-squared: 0.3951 ## F-statistic: 6.132 on 7 and 48 DF, p-value: 3.809e-05 #Lin-mul: RMSE, MAPE toàn mẫu đến 2023Q4 rmse(LSS.train,fitted(linmul)) ## [1] 124.2359 mape(LSS.train,fitted(linmul)) ## [1] 0.2474361 #Lin-mul: RMSE, MAPE cho forecast 2024Q1 đến 2024Q4 rmse(LSS.test,predict(linmul, newdata=df_new2)) ## [1] 157.9477 mape(LSS.test,predict(linmul, newdata=df_new2)) ## [1] 0.2432298 </pre>
7	Holt-Winter + Additive	<pre> #Mô hình 7: Holt-Winter + Additive hw.lss.a <- HoltWinters(LSS.train, seasonal = "a") hw.lss.a ## Holt-Winters exponential smoothing with trend and additive seasonal ## component. ## ## Call: ## HoltWinters(x = LSS.train, seasonal = "a") ## ## Smoothing parameters: ## alpha: 0.2529007 ## beta : 0.1026824 ## gamma: 0.3477263 ## ## Coefficients: ## [,1] ## a 642.772957 ## b 7.900333 ## s1 -111.829106 ## s2 147.893398 ## s3 -228.556568 ## s4 -94.366600 #HW-add: RMSE toàn mẫu đến 2023Q4 rmse(LSS.train,fitted(hw.lss.a)) ## [1] 387.1592 #HW-add: RMSE cho forecast 2024Q1 đến 2024Q4 library(forecast) forecast_hw.a <- forecast(hw.lss.a, h = 4) rmse(LSS.test, forecast_hw.a\$mean) ## [1] 105.69 </pre>

8	Holt-Winter + Multiplicative	<pre> #Mô hình 8: Holt-Winter + Multiplicative hw.lss.m <- HoltWinters(LSS.train, seasonal = "m") # m = multiplicative hw.lss.m ## Holt-Winters exponential smoothing with trend and multiplicative seasonal ## component. ## Call: ## HoltWinters(x = LSS.train, seasonal = "m") ## ## Smoothing parameters: ## alpha: 0.2259463 ## beta : 0.1233019 ## gamma: 0.3912971 ## ## Coefficients: ## [,1] ## a 716.7293553 ## b 14.2255545 ## s1 0.7564842 ## s2 1.2162047 ## s3 0.5635833 ## s4 0.7821507 #HW-mul: RMSE toàn mẫu đến 2023Q4 rmse(LSS.train,fitted(hw.lss.m)) ## [1] 371.2349 #HW-mul: RMSE cho forecast 2024Q1 đến 2024Q4 library(forecast) forecast_hw.m <- forecast(hw.lss.m, h = 4) rmse(LSS.test, forecast_hw.m\$mean) ## [1] 91.29634 </pre>
---	------------------------------	--

Hình 4: Kết quả ước lượng mô hình Holt-Winter dạng nhân cho số liệu từ quý 1/2010-quý 4/2024

```

#Dựa vào RMSE cho 2024Q1 đến 2024Q4: Chọn mô hình Holt_Winter Multiplicative
#With data to 2024Q4, forecast for 2025
hw.lss.m.2 <- HoltWinters(LSS, seasonal = "m")      # m = multiplicative
hw.lss.m.2

```

```

## Holt-Winters exponential smoothing with trend and multiplicative seasonal
## component.
##
## Call:
## HoltWinters(x = LSS, seasonal = "m")
##
## Smoothing parameters:
## alpha: 0.2258967
## beta : 0.1191595
## gamma: 0.3847927
##
## Coefficients:
##      [,1]
## a 817.4024237
## b  17.2897618
## s1  0.8222935
## s2  1.1836681
## s3  0.5730897
## s4  0.7571990

```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

Bảng 5: Kết quả dự báo doanh thu thuần trong 4 quý năm 2025

```
forecast_hw.m <- forecast(hw.lss.m.2, h = 4)
forecast_hw.m
```

```
##          Point Forecast      Lo 80      Hi 80      Lo 95      Hi 95
## 2025 Q1      686.3620 576.4811 796.2428 518.3137 854.4103
## 2025 Q2     1008.4639 874.6632 1142.2645 803.8335 1213.0942
## 2025 Q3      498.1706 378.1455 618.1957 314.6081 681.7332
## 2025 Q4      671.3035 572.0434 770.5635 519.4983 823.1086
```

Bảng 6: Kết quả kiểm định Dickey-Fuller cho chuỗi giá

STT	Kiểm định	Kết quả
1	Kiểm định Dickey-Fuller: không có hệ số chặn (none)	<pre>#Chuỗi giá #Kiểm định Dickey-Fuller: không có hệ số chặn (none) summary(ur.df(LSS.pr,type = "none",lag = 0)) ## ## ##### ## # Augmented Dickey-Fuller Test Unit Root Test # ## ##### ## ## Test regression none ## ## Call: ## lm(formula = z.diff ~ z.lag.1 - 1) ## ## Residuals: ## Min 1Q Median 3Q Max ## -838.78 -95.94 -5.40 128.41 754.13 ## ## Coefficients: ## Estimate Std. Error t value Pr(> t) ## z.lag.1 0.0005574 0.0011305 0.493 0.622 ## ## Residual standard error: 248.3 on 497 degrees of freedom ## Multiple R-squared: 0.0004888, Adjusted R-squared: -0.001522 ## F-statistic: 0.243 on 1 and 497 DF, p-value: 0.6222 ## ## ## Value of test-statistic is: 0.493 ## ## Critical values for test statistics: ## 1pct 5pct 10pct ## tau1 -2.58 -1.95 -1.62</pre>

2	Kiểm định Dickey-Fuller: có hệ số chặn (drift)	<pre> #Chuỗi giá #Kiểm định Dickey-Fuller: có hệ số chặn (drift) summary(ur.df(LSS.pr,type = "drift",lag = 0)) ## ## ##### ## # Augmented Dickey-Fuller Test Unit Root Test # ## ##### ## ## Test regression drift ## ## Call: ## lm(formula = z.diff ~ z.lag.1 + 1) ## ## Residuals: ## Min 1Q Median 3Q Max ## -805.21 -114.60 -10.51 127.34 776.66 ## ## Coefficients: ## Estimate Std. Error t value Pr(> t) ## (Intercept) 175.685659 64.396450 2.728 0.00659 ** ## z.lag.1 -0.017028 0.006543 -2.602 0.00953 ** ## --- ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ## ## Residual standard error: 246.7 on 496 degrees of freedom ## Multiple R-squared: 0.01347, Adjusted R-squared: 0.01148 ## F-statistic: 6.773 on 1 and 496 DF, p-value: 0.009532 ## ## Value of test-statistic is: -2.6025 3.8446 ## ## Critical values for test statistics: ## 1pct 5pct 10pct ## tau2 -3.44 -2.87 -2.57 ## phi1 6.47 4.61 3.79 </pre>
3	Kiểm định Dickey-Fuller: có xu thế (trend)	<pre> #Chuỗi giá #Kiểm định Dickey-Fuller: có xu thế (trend) summary(ur.df(LSS.pr,type = "trend",lag = 0)) ## ## ##### ## # Augmented Dickey-Fuller Test Unit Root Test # ## ##### ## ## Test regression trend ## ## Call: ## lm(formula = z.diff ~ z.lag.1 + 1 + tt) ## ## Residuals: ## Min 1Q Median 3Q Max ## -788.15 -107.39 -14.77 127.76 775.42 ## ## Coefficients: ## Estimate Std. Error t value Pr(> t) ## (Intercept) 196.176018 69.906766 2.806 0.00521 ** ## z.lag.1 -0.021073 0.008458 -2.491 0.01305 * ## tt 0.075065 0.099418 0.755 0.45058 ## --- ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ## ## Residual standard error: 246.8 on 495 degrees of freedom ## Multiple R-squared: 0.01461, Adjusted R-squared: 0.01062 ## F-statistic: 3.669 on 2 and 495 DF, p-value: 0.02621 ## ## Value of test-statistic is: -2.4913 2.7509 3.6686 ## ## Critical values for test statistics: ## 1pct 5pct 10pct ## tau3 -3.98 -3.42 -3.13 ## phi2 6.15 4.71 4.05 ## phi3 8.34 6.30 5.36 </pre>

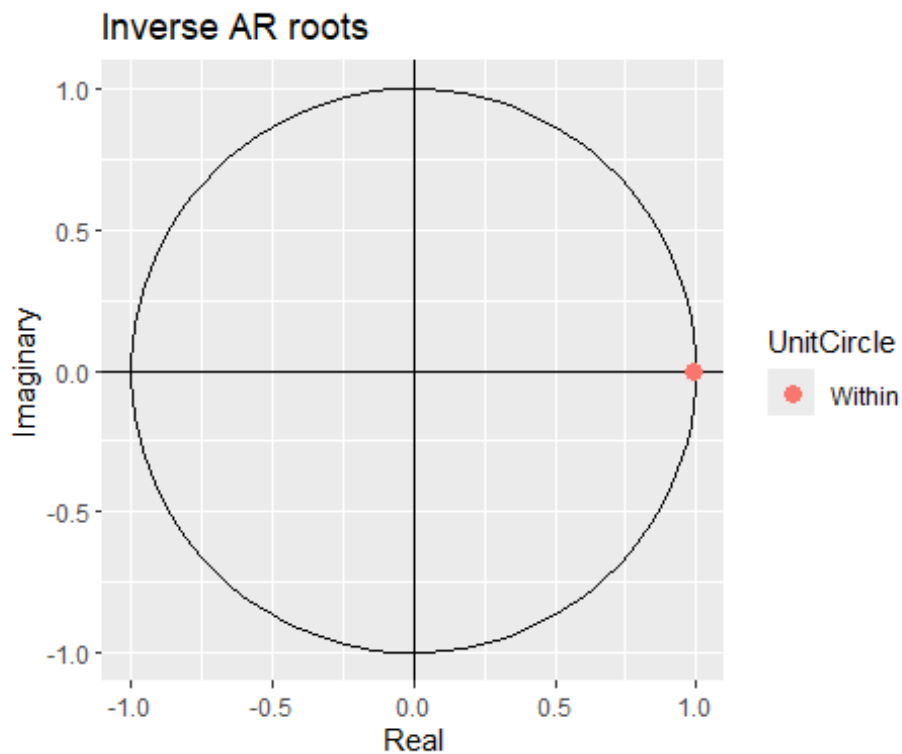
Hình 5: Kết quả ước lượng mô hình AR(1) cho chuỗi giá

```
#Chuỗi giá
#Mô hình AR(1)
reg.arima100<- Arima(LSS.pr,order = c(1,0,0))
reg.arima100

## Series: LSS.pr
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##          ar1      mean
##      0.9927  9191.073
## s.e.  0.0054  1235.708
##
## sigma^2 = 61647:  log likelihood = -3460.94
## AIC=6927.89  AICc=6927.94  BIC=6940.53
```

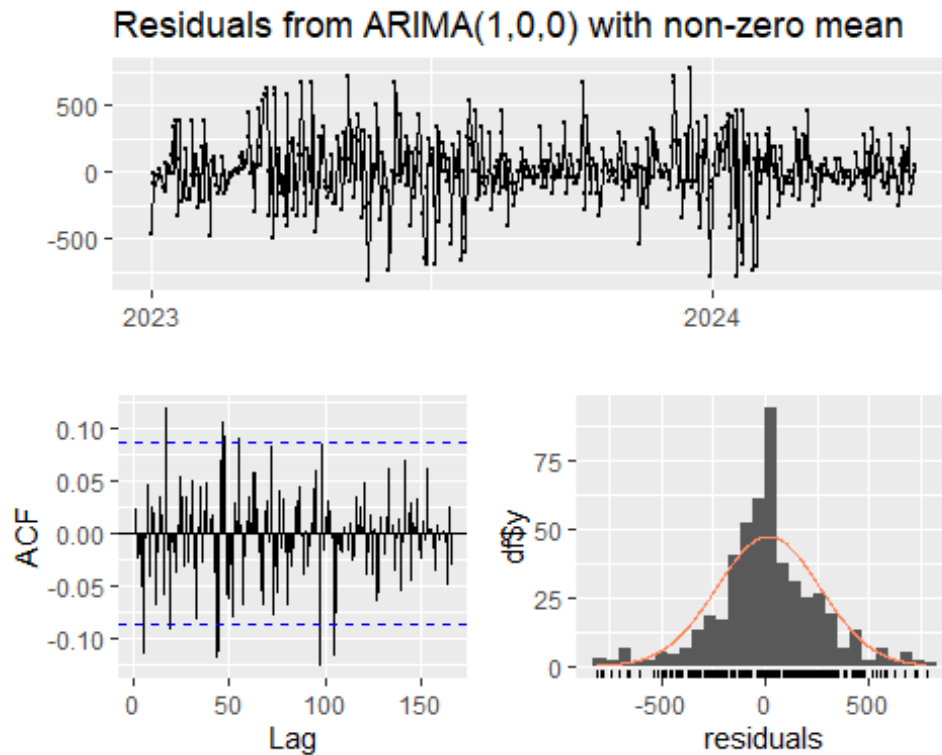
Hình 6: Vòng tròn đơn vị của mô hình AR(1) cho chuỗi giá

```
autoplot(reg.arima100)
```



Hình 7: Kết quả kiểm định phần dư cho mô hình $AR(1)$

```
checkresiduals(reg.arima100)
```



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,0) with non-zero mean
## Q* = 140.63, df = 99, p-value = 0.003806
##
## Model df: 1.   Total lags used: 100
```

Bảng 7: Kết quả kiểm định Dickey-Fuller cho chuỗi log-return

STT	Kiểm định	Kết quả
-----	-----------	---------

1	Kiểm định Dickey-Fuller: không có hệ số chặn (none)	<pre> #Chuỗi Log-return #Kiểm định Dickey-Fuller: không có hệ số chặn (none) summary(ur.df(rLSS,type = "none",lag = 0)) ## ## ##### ## # Augmented Dickey-Fuller Test Unit Root Test # ## ##### ## ## Test regression none ## ## ## Call: ## lm(formula = z.diff ~ z.lag.1 - 1) ## ## Residuals: ## Min 1Q Median 3Q Max ## -2.71436 -0.42211 0.00409 0.51979 2.54742 ## ## Coefficients: ## Estimate Std. Error t value Pr(> t) ## z.lag.1 -0.97497 0.04489 -21.72 <2e-16 *** ## --- ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ## ## Residual standard error: 0.9643 on 496 degrees of freedom ## Multiple R-squared: 0.4875, Adjusted R-squared: 0.4865 ## F-statistic: 471.8 on 1 and 496 DF, p-value: < 2.2e-16 ## ## Value of test-statistic is: -21.7211 ## ## Critical values for test statistics: ## 1pct 5pct 10pct ## tau1 -2.58 -1.95 -1.62 </pre>
2	Kiểm định Dickey-Fuller: có hệ số chặn (drift)	<pre> #Chuỗi Log-return #Kiểm định Dickey-Fuller: có hệ số chặn (drift) summary(ur.df(rLSS,type = "drift",lag = 0)) ## ## ##### ## # Augmented Dickey-Fuller Test Unit Root Test # ## ##### ## ## Test regression drift ## ## ## Call: ## lm(formula = z.diff ~ z.lag.1 + 1) ## ## Residuals: ## Min 1Q Median 3Q Max ## -2.76318 -0.47174 -0.04638 0.47163 2.49386 ## ## Coefficients: ## Estimate Std. Error t value Pr(> t) ## (Intercept) 0.05001 0.04330 1.155 0.249 ## z.lag.1 -0.97774 0.04493 -21.759 <2e-16 *** ## --- ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ## ## Residual standard error: 0.964 on 495 degrees of freedom ## Multiple R-squared: 0.4889, Adjusted R-squared: 0.4878 ## F-statistic: 473.5 on 1 and 495 DF, p-value: < 2.2e-16 ## ## Value of test-statistic is: -21.759 236.7279 ## ## Critical values for test statistics: ## 1pct 5pct 10pct ## tau2 -3.44 -2.87 -2.57 ## phi1 6.47 4.61 3.79 </pre>

3	Kiểm định Dickey-Fuller: có xu thế (trend)	<pre> #Chuỗi Log-return #Kiểm định Dickey-Fuller: có xu thế (trend) summary(ur.df(rLSS,type = "trend",lag = 0)) ## ## ##### ## # Augmented Dickey-Fuller Test Unit Root Test # ## ##### ## ## Test regression trend ## ## ## Call: ## lm(formula = z.diff ~ z.lag.1 + 1 + tt) ## ## Residuals: ## Min 1Q Median 3Q Max ## -2.84850 -0.44726 -0.01733 0.49901 2.46857 ## ## Coefficients: ## Estimate Std. Error t value Pr(> t) ## (Intercept) 0.1521096 0.0868241 1.752 0.0804 . ## z.lag.1 -0.9815493 0.0449846 -21.820 <2e-16 *** ## tt -0.0004092 0.0003017 -1.356 0.1756 ## --- ## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1 ## ## Residual standard error: 0.9632 on 494 degrees of freedom ## Multiple R-squared: 0.4908, Adjusted R-squared: 0.4887 ## F-statistic: 238 on 2 and 494 DF, p-value: < 2.2e-16 ## ## Value of test-statistic is: -21.8197 158.6995 238.0492 ## ## Critical values for test statistics: ## 1pct 5pct 10pct ## tau3 -3.98 -3.42 -3.13 ## phi2 6.15 4.71 4.05 ## phi3 8.34 6.30 5.36 </pre>
---	---	--

Hình 8: Kết quả kiểm định Box-test cho chuỗi log-return

```
Box.test(rLSS, lag = 20, type = "Ljung-Box")
```

```

##
## Box-Ljung test
##
## data:  rLSS
## X-squared = 25.38, df = 20, p-value = 0.1873

```

Hình 9: Kết quả chạy mô hình ARMA(1,1) cho chuỗi log-return

```

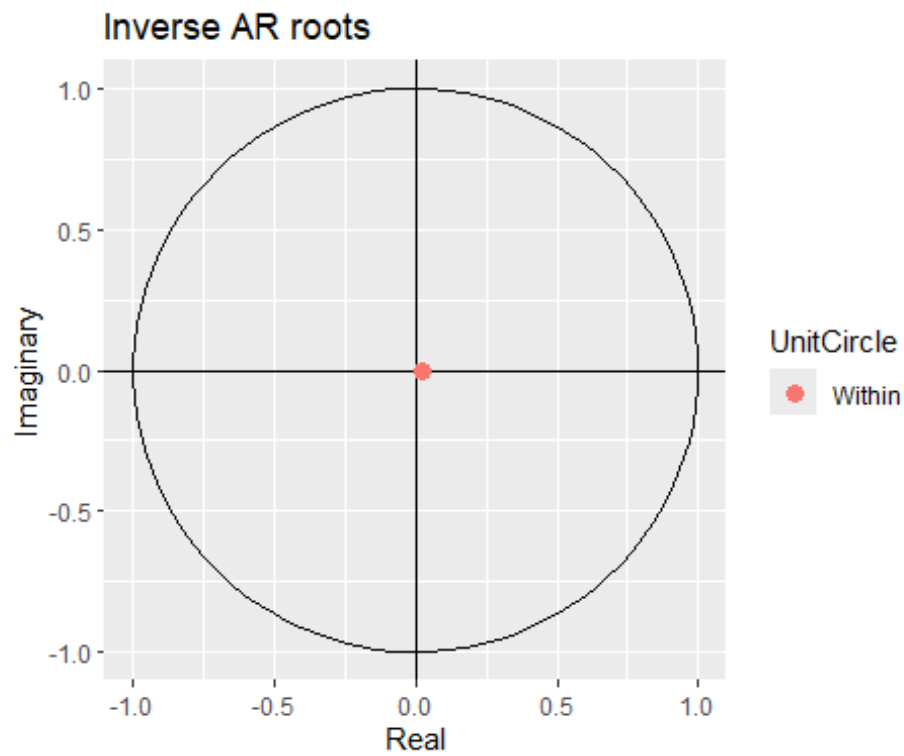
#Chuỗi Log-return
#Mô hình ARMA(1,1): hàm Arima
reg.arima101<- Arima(rLSS,order = c(1,0,0))
reg.arima101

## Series: rLSS
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##      ar1      mean
##    0.0222  0.0514
## s.e.  0.0448  0.0440
##
## sigma^2 = 0.9275: log likelihood = -686.88
## AIC=1379.76  AICc=1379.81  BIC=1392.4

```

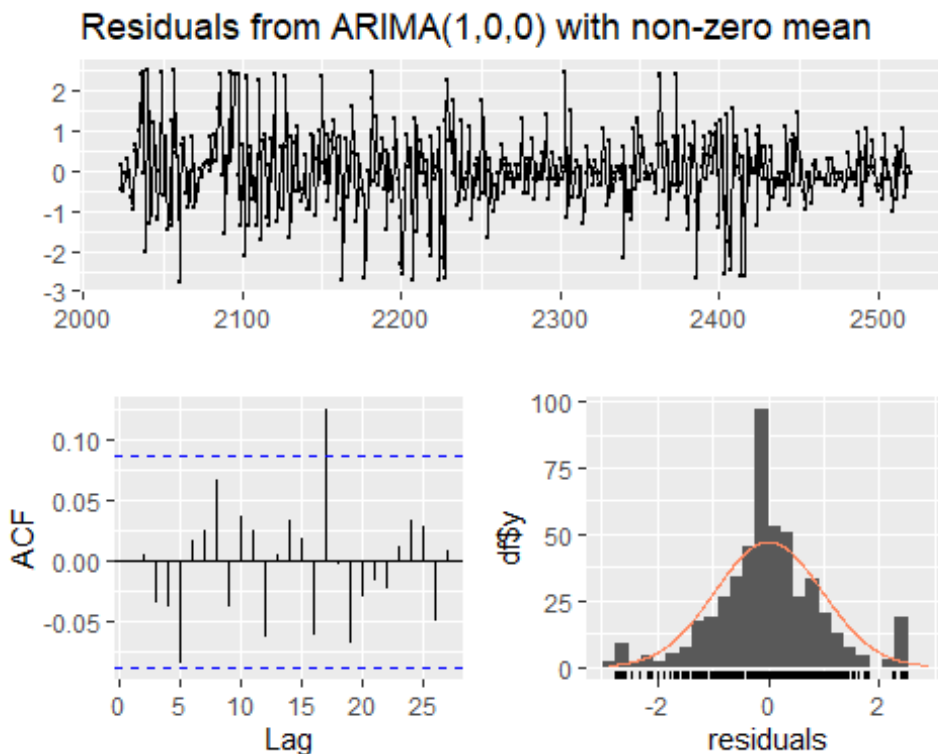
Hình 10: Vòng tròn nghiệm đơn vị cho mô hình $ARMA(1,1)$

```
library(ggplot2)
autoplot(reg.arima101)
```



Hình 11: Kiểm định phần dư cho mô hình ARMA(1,1)

```
library(forecast)
checkresiduals(reg.arima101)
```



```
##
##  Ljung-Box test
##
## data:  Residuals from ARIMA(1,0,0) with non-zero mean
## Q* = 9.3009, df = 9, p-value = 0.41
##
## Model df: 1.   Total lags used: 10
```

Hình 12: Kết quả dự báo giá cổ phiếu từ mô hình AR(1)

```
#Chuỗi giá
#Forecast từ mô hình AR(1)
usaf.arima100 <-forecast(reg.arima100)
usaf.arima100
```

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
## 2522	10635.12	10316.923	10953.31	10148.481	11121.75
## 2523	10624.57	10176.217	11072.93	9938.872	11310.27
## 2524	10614.10	10066.980	11161.23	9777.350	11450.86
## 2525	10603.71	9974.239	11233.18	9641.016	11566.41
## 2526	10593.40	9892.170	11294.62	9520.963	11665.83
## 2527	10583.16	9817.773	11348.54	9412.604	11753.71
## 2528	10572.99	9749.258	11396.72	9313.201	11832.78
## 2529	10562.90	9685.454	11440.34	9220.964	11904.83
## 2530	10552.88	9625.545	11480.22	9134.643	11971.12
## 2531	10542.94	9568.931	11516.94	9053.324	12032.55

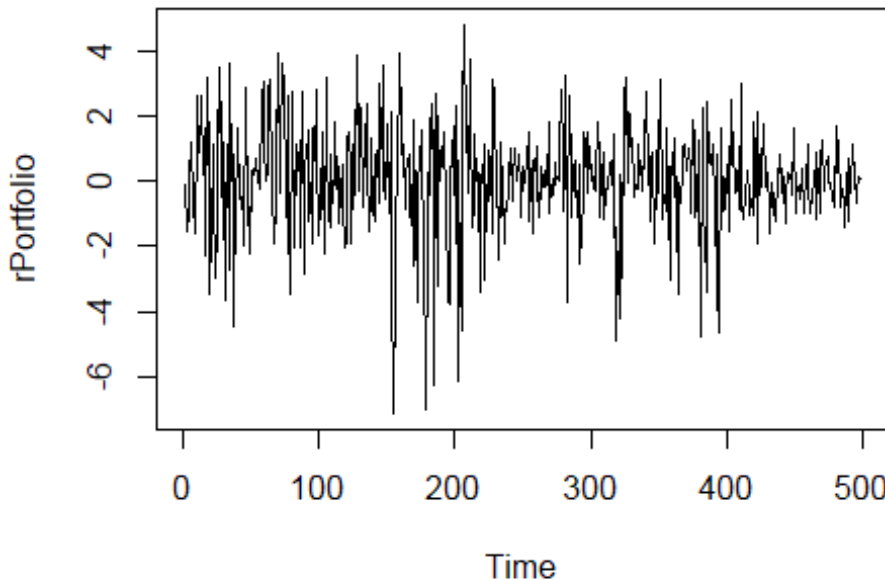
Hình 13: Kết quả dự báo log-return từ mô hình ARMA(1,1)

```
#Chuỗi Log-return
#Forecast từ mô hình ARMA(1,1)
usaf.arma101 <-forecast(reg.arma101)
usaf.arma101
```

##	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
## 2521	0.05026475	-1.183938	1.284468	-1.837286	1.937815
## 2522	0.05138140	-1.183126	1.285889	-1.836635	1.939398
## 2523	0.05140621	-1.183102	1.285914	-1.836610	1.939423
## 2524	0.05140676	-1.183101	1.285915	-1.836610	1.939423
## 2525	0.05140677	-1.183101	1.285915	-1.836610	1.939423
## 2526	0.05140678	-1.183101	1.285915	-1.836610	1.939423
## 2527	0.05140678	-1.183101	1.285915	-1.836610	1.939423
## 2528	0.05140678	-1.183101	1.285915	-1.836610	1.939423
## 2529	0.05140678	-1.183101	1.285915	-1.836610	1.939423
## 2530	0.05140678	-1.183101	1.285915	-1.836610	1.939423

PHỤ LỤC NHÓM

	TÀI CÁC THƯ VIỆN
	<pre> library(readxl) library(forecast) ## Registered S3 method overwritten by 'quantmod': ## method from ## as.zoo.data.frame zoo library(urca) library(tseries) library(vars) ## Loading required package: MASS ## Loading required package: strucchange ## Loading required package: zoo ## ## Attaching package: 'zoo' ## The following objects are masked from 'package:base': ## ## as.Date, as.Date.numeric ## Loading required package: sandwich ## Loading required package: lmtest library(Metrics) ## Warning: package 'Metrics' was built under R version 4.4.2 ## ## Attaching package: 'Metrics' ## The following object is masked from 'package:forecast': ## ## accuracy </pre>
1	PHÂN TÍCH TƯƠNG QUAN GIỮA CÁC CỔ PHIẾU
	Dữ liệu giá cổ phiếu (10 quan sát đầu tiên)

	<table><tr><th>Date</th><th>SBT</th><th>DIG</th><th>VIC</th><th>LSS</th></tr><tr><td>03/01/2023</td><td>12,231.65</td><td>15.3</td><td>56,800</td><td>5,375.73</td></tr><tr><td>04/01/2023</td><td>11,570.48</td><td>15.7</td><td>56,000</td><td>5,400.70</td></tr><tr><td>05/01/2023</td><td>11,694.45</td><td>15.85</td><td>56,200</td><td>5,342.45</td></tr><tr><td>06/01/2023</td><td>11,322.54</td><td>15.1</td><td>56,100</td><td>5,275.87</td></tr><tr><td>09/01/2023</td><td>11,198.57</td><td>14.8</td><td>55,000</td><td>5,275.87</td></tr><tr><td>10/01/2023</td><td>11,363.86</td><td>15.1</td><td>54,500</td><td>5,275.87</td></tr><tr><td>11/01/2023</td><td>11,487.83</td><td>15.5</td><td>55,100</td><td>5,325.80</td></tr><tr><td>12/01/2023</td><td>11,363.86</td><td>15.2</td><td>55,000</td><td>5,309.16</td></tr><tr><td>13/01/2023</td><td>11,363.86</td><td>15.25</td><td>55,000</td><td>5,225.95</td></tr><tr><td>16/01/2023</td><td>11,157.25</td><td>14.9</td><td>54,800</td><td>5,101.12</td></tr></table>	Date	SBT	DIG	VIC	LSS	03/01/2023	12,231.65	15.3	56,800	5,375.73	04/01/2023	11,570.48	15.7	56,000	5,400.70	05/01/2023	11,694.45	15.85	56,200	5,342.45	06/01/2023	11,322.54	15.1	56,100	5,275.87	09/01/2023	11,198.57	14.8	55,000	5,275.87	10/01/2023	11,363.86	15.1	54,500	5,275.87	11/01/2023	11,487.83	15.5	55,100	5,325.80	12/01/2023	11,363.86	15.2	55,000	5,309.16	13/01/2023	11,363.86	15.25	55,000	5,225.95	16/01/2023	11,157.25	14.9	54,800	5,101.12
Date	SBT	DIG	VIC	LSS																																																				
03/01/2023	12,231.65	15.3	56,800	5,375.73																																																				
04/01/2023	11,570.48	15.7	56,000	5,400.70																																																				
05/01/2023	11,694.45	15.85	56,200	5,342.45																																																				
06/01/2023	11,322.54	15.1	56,100	5,275.87																																																				
09/01/2023	11,198.57	14.8	55,000	5,275.87																																																				
10/01/2023	11,363.86	15.1	54,500	5,275.87																																																				
11/01/2023	11,487.83	15.5	55,100	5,325.80																																																				
12/01/2023	11,363.86	15.2	55,000	5,309.16																																																				
13/01/2023	11,363.86	15.25	55,000	5,225.95																																																				
16/01/2023	11,157.25	14.9	54,800	5,101.12																																																				
2	XÂY DỰNG DANH MỤC TỐI ƯU																																																							
3	DỰ BÁO LỢI SUẤT DANH MỤC																																																							
	CÁCH 2: DỰ BÁO LỢI SUẤT CHUỖI DANH MỤC DỰA TRÊN CHUỖI LỢI SUẤT DANH MỤC THỰC TẾ																																																							
	Tải dữ liệu và vẽ đồ thị																																																							
	<div><pre>rPortfolio <- read_excel("Min_var.xlsx", sheet = "Forecast", range = "I1:I499")[[1]] plot(rPortfolio, type="l", xlab = "Time")</pre></div> <div></div>																																																							
	Kiểm định ADF																																																							
	<pre>summary(ur.df(rPortfolio,type = "trend", selectlags = "AIC"))</pre>																																																							

```
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.2254 -0.8154  0.0626  0.9544  4.8406
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.1866937  0.1532398   1.218   0.224
## z.lag.1      -0.9460823  0.0629137 -15.038 <2e-16 ***
## tt           -0.0005944  0.0005324  -1.116   0.265
## z.diff.lag   -0.0295946  0.0450530  -0.657   0.512
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.693 on 492 degrees of freedom
## Multiple R-squared:  0.4879, Adjusted R-squared:  0.4848
## F-statistic: 156.3 on 3 and 492 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -15.0378 75.3786 113.0677
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau3 -3.98 -3.42 -3.13
## phi2  6.15  4.71  4.05
## phi3  8.34  6.30  5.36

summary(ur.df(rPortfolio,type = "drift", selectlags = "AIC"))

##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
```

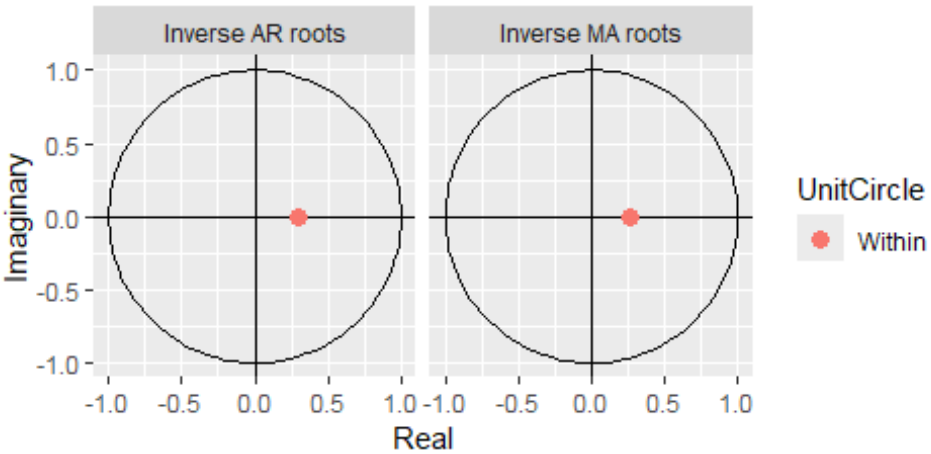
```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.1668 -0.7737  0.0287  0.9416  4.8732
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.03818    0.07609   0.502   0.616
## z.lag.1      -0.94095    0.06276 -14.993 <2e-16 ***
## z.diff.lag   -0.03213    0.04501  -0.714   0.476
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.694 on 493 degrees of freedom
## Multiple R-squared:  0.4866, Adjusted R-squared:  0.4845
## F-statistic: 233.7 on 2 and 493 DF, p-value: < 2.2e-16
##
## Value of test-statistic is: -14.9926 112.3884
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.44 -2.87 -2.57
## phi1  6.47  4.61  3.79

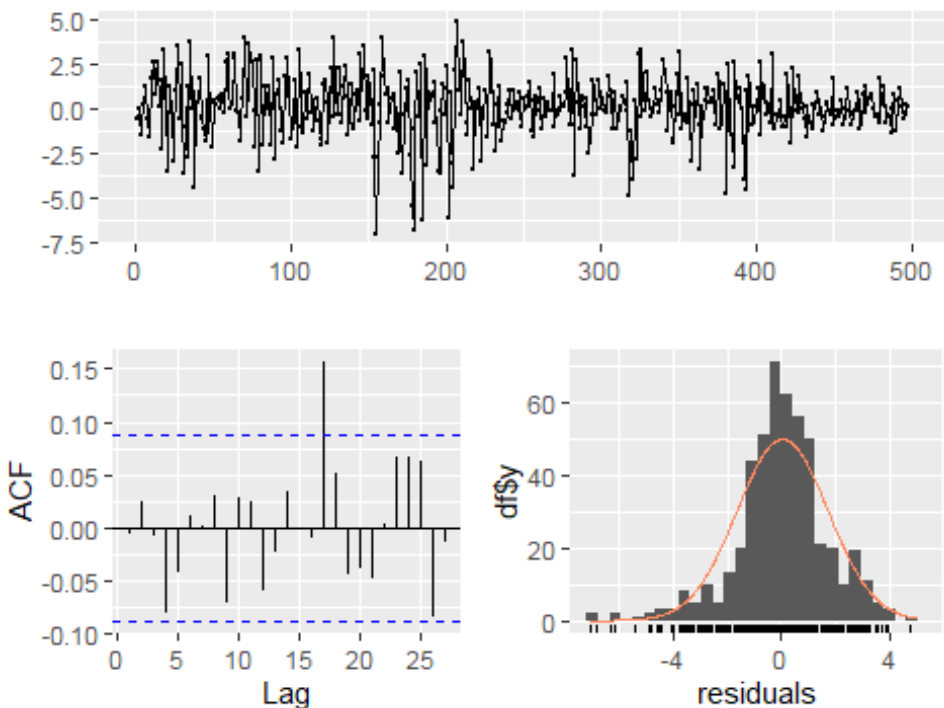
summary(ur.df(rPortfolio,type = "none", selectlags = "AIC"))

##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
##
## Test regression none
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.1283 -0.7349  0.0670  0.9787  4.9126
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## z.lag.1      -0.93993    0.06268 -14.996 <2e-16 ***
## z.diff.lag   -0.03263    0.04496  -0.726   0.468
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.692 on 494 degrees of freedom
```

	<pre>## Multiple R-squared: 0.4864, Adjusted R-squared: 0.4843 ## F-statistic: 233.9 on 2 and 494 DF, p-value: < 2.2e-16 ## ## ## Value of test-statistic is: -14.9955 ## ## Critical values for test statistics: ## 1pct 5pct 10pct ## tau1 -2.58 -1.95 -1.62</pre>
	Lược đồ tương quan
	<p><code>Acf(rPortfolio)</code></p> <p>Series rPortfolio</p> <p><code>Pacf(rPortfolio)</code></p>

	<p style="text-align: center;">Series rPortfolio</p>
	Kiểm tra nhiễu trắng
	<pre>Box.test(rPortfolio, lag = 20, type = "Ljung-Box") ## ## Box-Ljung test ## ## data: rPortfolio ## X-squared = 27.769, df = 20, p-value = 0.115</pre>
	Mô hình ARIMA(1,0,1)
	<pre>arma11 = Arima(rPortfolio, order = c(1,0,1), include.constant = 0) summary(arma11) ## Series: rPortfolio ## ARIMA(1,0,1) with zero mean ## ## Coefficients: ## ar1 ma1 ## 0.2911 -0.2587 ## s.e. 0.5404 0.5448 ## ## sigma^2 = 2.856: log likelihood = -966.92 ## AIC=1939.84 AICc=1939.89 BIC=1952.48 ## ## Training set error measures: ## ME RMSE MAE MPE MAPE</pre>

	<pre> MASE ## Training set 0.03716724 1.686526 1.221315 108.5861 121.1539 0.6839225 ## ACF1 ## Training set -0.005310784 </pre>
	Đồ thị nghiệm nghịch đảo và kiểm tra tự tương quan của phần dư
	<pre>autoplot(arma11)</pre>  <pre>checkresiduals(arma11)</pre>

	<p>Residuals from ARIMA(1,0,1) with zero mean</p>  <pre>## ## Ljung-Box test ## ## data: Residuals from ARIMA(1,0,1) with zero mean ## Q* = 8.1313, df = 8, p-value = 0.4207 ## ## Model df: 2. Total lags used: 10</pre>
	<p>Dự báo</p> <pre>forecast(arma11, h=10)\$mean ## Time Series: ## Start = 499 ## End = 508 ## Frequency = 1 ## [1] 2.814024e-03 8.192551e-04 2.385122e-04 6.943876e-05 2.0 21591e-05 ## [6] 5.885518e-06 1.713468e-06 4.988472e-07 1.452309e-07 4.2 28149e-08</pre>
4.	<p>PHÂN TÍCH ĐỒNG TÍCH HỢP</p> <pre>stock <- read_excel("D:/Time Series/Min_var.xlsx", range = "A1:D500") attach(stock) stock\$DIG = stock\$DIG*1000</pre>


```
summary(ca.jo(data.frame(SBT, DIG, VIC),type = "trace"))

##
## #####
## # Johansen-Procedure #
## #####
##
## Test type: trace statistic , with linear trend
##
## Eigenvalues (lambda):
## [1] 0.024388596 0.012005375 0.006917714
##
## Values of teststatistic and critical values of test:
##
##          test 10pct  5pct  1pct
## r <= 2 |   3.45   6.50   8.18 11.65
## r <= 1 |   9.45  15.66  17.95 23.52
## r = 0  |  21.72  28.71  31.52 37.22
##
## Eigenvectors, normalised to first column:
## (These are the cointegration relations)
##
##          SBT.l2      DIG.l2      VIC.l2
## SBT.l2  1.0000000    1.000000    1.0000000
## DIG.l2 64.2200068 -449.732180 -535.85825582
## VIC.l2 -0.1050132   -1.008579   -0.04191106
##
## Weights W:
## (This is the loading matrix)
##
##          SBT.l2      DIG.l2      VIC.l2
## SBT.d -3.253897e-02 7.330767e-04 2.941868e-03
## DIG.d  2.621027e-06 5.746781e-07 1.855059e-05
## VIC.d  1.335419e-02 1.640308e-02 7.717818e-03

summary(ca.jo(data.frame(SBT, DIG, VIC),type = "eigen"))

##
## #####
## # Johansen-Procedure #
## #####
##
## Test type: maximal eigenvalue statistic (lambda max) , with
linear trend
##
## Eigenvalues (lambda):
## [1] 0.024388596 0.012005375 0.006917714
##
## Values of teststatistic and critical values of test:
##
##          test 10pct  5pct  1pct
```

```
## r <= 2 | 3.45 6.50 8.18 11.65
## r <= 1 | 6.00 12.91 14.90 19.19
## r = 0 | 12.27 18.90 21.07 25.75
##
## Eigenvectors, normalised to first column:
## (These are the cointegration relations)
##
##           SBT.l2      DIG.l2      VIC.l2
## SBT.l2  1.0000000    1.000000    1.00000000
## DIG.l2  64.2200068 -449.732180 -535.85825582
## VIC.l2  -0.1050132  -1.008579  -0.04191106
##
## Weights W:
## (This is the loading matrix)
##
##           SBT.l2      DIG.l2      VIC.l2
## SBT.d  -3.253897e-02  7.330767e-04  2.941868e-03
## DIG.d   2.621027e-06  5.746781e-07  1.855059e-05
## VIC.d   1.335419e-02  1.640308e-02  7.717818e-03

summary(ca.jo(data.frame(DIG, VIC),type = "trace"))

##
## #####
## # Johansen-Procedure #
## #####
##
## Test type: trace statistic , with linear trend
##
## Eigenvalues (lambda):
## [1] 0.01117509 0.00576203
##
## Values of teststatistic and critical values of test:
##
##           test 10pct  5pct  1pct
## r <= 1 | 2.87  6.50  8.18 11.65
## r = 0 | 8.46 15.66 17.95 23.52
##
## Eigenvectors, normalised to first column:
## (These are the cointegration relations)
##
##           DIG.l2      VIC.l2
## DIG.l2  1.000000000  1.000000e+00
## VIC.l2  0.004570645  4.696036e-05
##
## Weights W:
## (This is the loading matrix)
##
##           DIG.l2      VIC.l2
```

	<pre>## DIG.d 0.0003257741 -0.009957458 ## VIC.d -2.9076701612 -6.805969983 summary(ca.jo(data.frame(DIG, VIC),type = "eigen")) ## ## ##### ## # Johansen-Procedure # ## ##### ## ## Test type: maximal eigenvalue statistic (lambda max) , with linear trend ## ## Eigenvalues (lambda): ## [1] 0.01117509 0.00576203 ## ## Values of teststatistic and critical values of test: ## ## test 10pct 5pct 1pct ## r <= 1 2.87 6.50 8.18 11.65 ## r = 0 5.59 12.91 14.90 19.19 ## ## Eigenvectors, normalised to first column: ## (These are the cointegration relations) ## ## DIG.12 VIC.12 ## DIG.12 1.000000000 1.000000e+00 ## VIC.12 0.004570645 4.696036e-05 ## ## Weights W: ## (This is the loading matrix) ## ## DIG.12 VIC.12 ## DIG.d 0.0003257741 -0.009957458 ## VIC.d -2.9076701612 -6.805969983</pre>
5.	PHÂN TÍCH VAR
5.1.	ƯỚC LƯỢNG VAR
	Tải dữ liệu
	<pre>stock1 <- read_excel("D:/Time Series/Min_var.xlsx", range = "F2:I500") attach(stock1) r2025 <- read_excel("Min_var.xlsx", sheet = "2025", range = "F1:I12") [-1,]</pre>
	Chọn bậc của trễ
	VARselect(stock1)

<pre>## \$selection ## AIC(n) HQ(n) SC(n) FPE(n) ## 1 1 1 1 ## ## \$criteria ## 1 2 3 4 5 6 ## AIC(n) 5.491762 5.524037 5.541843 5.543148 5.57649 0 5.602050 ## HQ(n) 5.559220 5.645462 5.717234 5.772504 5.85981 3 5.939339 ## SC(n) 5.663496 5.833159 5.988353 6.127044 6.29777 4 6.460721 ## FPE(n) 242.685157 250.649146 255.160795 255.509727 264.19912 4 271.078608 ## 7 8 9 10 ## AIC(n) 5.640046 5.630439 5.659625 5.687859 ## HQ(n) 6.031302 6.075661 6.158814 6.241014 ## SC(n) 6.636105 6.763885 6.930459 7.096080 ## FPE(n) 281.633621 279.015194 287.376094 295.730118</pre>	
VAR(1)	
<pre>var1 = VAR(stock1, p=1, type = "const") summary(var1) ## ## VAR Estimation Results: ## ===== ## Endogenous variables: rSBT, rDIG, rVIC, rLSS ## Deterministic variables: const ## Sample size: 497 ## Log Likelihood: -4155.371 ## Roots of the characteristic polynomial: ## 0.1331 0.05967 0.05967 0.0292 ## Call: ## VAR(y = stock1, p = 1, type = "const") ## ## ## Estimation results for equation rSBT: ## ===== ## rSBT = rSBT.l1 + rDIG.l1 + rVIC.l1 + rLSS.l1 + const ## ## Estimate Std. Error t value Pr(> t) ## rSBT.l1 -0.03104 0.06277 -0.495 0.621 ## rDIG.l1 0.01127 0.03465 0.325 0.745 ## rVIC.l1 0.06562 0.05254 1.249 0.212 ## rLSS.l1 -0.02297 0.04690 -0.490 0.624 ## const 0.01272 0.08698 0.146 0.884 ##</pre>	

```
##
## Residual standard error: 1.931 on 492 degrees of freedom
## Multiple R-Squared: 0.00523, Adjusted R-squared: -0.002858
## F-statistic: 0.6466 on 4 and 492 DF, p-value: 0.6295
##
##
## Estimation results for equation rDIG:
## =====
## rDIG = rSBT.l1 + rDIG.l1 + rVIC.l1 + rLSS.l1 + const
##
##           Estimate Std. Error t value Pr(>|t|)
## rSBT.l1  0.02842    0.09257   0.307   0.759
## rDIG.l1 -0.01758    0.05110  -0.344   0.731
## rVIC.l1 -0.02468    0.07748  -0.319   0.750
## rLSS.l1 -0.06951    0.06916  -1.005   0.315
## const    0.04615    0.12827   0.360   0.719
##
##
## Residual standard error: 2.847 on 492 degrees of freedom
## Multiple R-Squared: 0.004221, Adjusted R-squared: -0.0038
74
## F-statistic: 0.5214 on 4 and 492 DF, p-value: 0.72
##
##
## Estimation results for equation rVIC:
## =====
## rVIC = rSBT.l1 + rDIG.l1 + rVIC.l1 + rLSS.l1 + const
##
##           Estimate Std. Error t value Pr(>|t|)
## rSBT.l1  0.03373    0.05649   0.597  0.55067
## rDIG.l1  0.04572    0.03118   1.466  0.14327
## rVIC.l1  0.12534    0.04728   2.651  0.00828 **
## rLSS.l1 -0.07053    0.04220  -1.671  0.09532 .
## const   -0.04847    0.07827  -0.619  0.53605
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 1.738 on 492 degrees of freedom
## Multiple R-Squared: 0.0266, Adjusted R-squared: 0.01869
## F-statistic: 3.361 on 4 and 492 DF, p-value: 0.009937
##
##
## Estimation results for equation rLSS:
## =====
## rLSS = rSBT.l1 + rDIG.l1 + rVIC.l1 + rLSS.l1 + const
##
##           Estimate Std. Error t value Pr(>|t|)
## rSBT.l1  0.09139    0.08370   1.092   0.275
```

	<pre>## rDIG.l1 0.03996 0.04620 0.865 0.388 ## rVIC.l1 -0.02947 0.07005 -0.421 0.674 ## rLSS.l1 -0.03686 0.06253 -0.589 0.556 ## const 0.13855 0.11598 1.195 0.233 ## ## ## Residual standard error: 2.575 on 492 degrees of freedom ## Multiple R-Squared: 0.005396, Adjusted R-squared: -0.0026 9 ## F-statistic: 0.6673 on 4 and 492 DF, p-value: 0.6149 ## ## ## ## Covariance matrix of residuals: ## rSBT rDIG rVIC rLSS ## rSBT 3.7283 2.261 0.7261 3.463 ## rDIG 2.2607 8.108 1.6063 2.741 ## rVIC 0.7261 1.606 3.0192 1.025 ## rLSS 3.4628 2.741 1.0254 6.628 ## ## Correlation matrix of residuals: ## rSBT rDIG rVIC rLSS ## rSBT 1.0000 0.4112 0.2164 0.6966 ## rDIG 0.4112 1.0000 0.3247 0.3739 ## rVIC 0.2164 0.3247 1.0000 0.2292 ## rLSS 0.6966 0.3739 0.2292 1.0000</pre>
	Kiểm định tự tương quan của phần dư
	<pre>serial.test(var1) ## ## Portmanteau Test (asymptotic) ## ## data: Residuals of VAR object var1 ## Chi-squared = 273.57, df = 240, p-value = 0.0673</pre>
	Dự báo và vẽ đồ thị dự báo
	<pre>forecast1 = predict(var1, n.ahead = 10) layout(matrix(1:5, ncol = 1), heights = c(1, 1, 1, 1, 0.3)) # Tăng khoảng trắng phía trên (top margin) par(mar = c(4, 4, 4.5, 2)) # top margin = 4.5 (default thường là 3-4) vars_to_plot <- c("rSBT", "rDIG", "rVIC", "rLSS") for (varname in vars_to_plot) {</pre>

```

fc <- forecast1$fcst[[varname]]

# Actual data + r2025
actual <- c(tail(stock1[[varname]], 120), head(r2025[[varname]], 10
))
total_len <- length(actual)

time_index <- 380:(380 + total_len - 1) # 380 to 509

forecast_values <- fc[, "fcst"]
lower <- fc[, "lower"]
upper <- fc[, "upper"]

# Plot
plot(x = time_index, y = actual, type = "l", col = "black", lwd = 2
,
      ylim = range(c(actual, lower, upper)),
      ylab = varname, main = paste("Forecast vs Actual -", varname),
      xlab = "Time")

# Forecast line
lines(time_index[(total_len - 9):total_len], forecast_values, col =
"blue", lty = 2, lwd = 2)

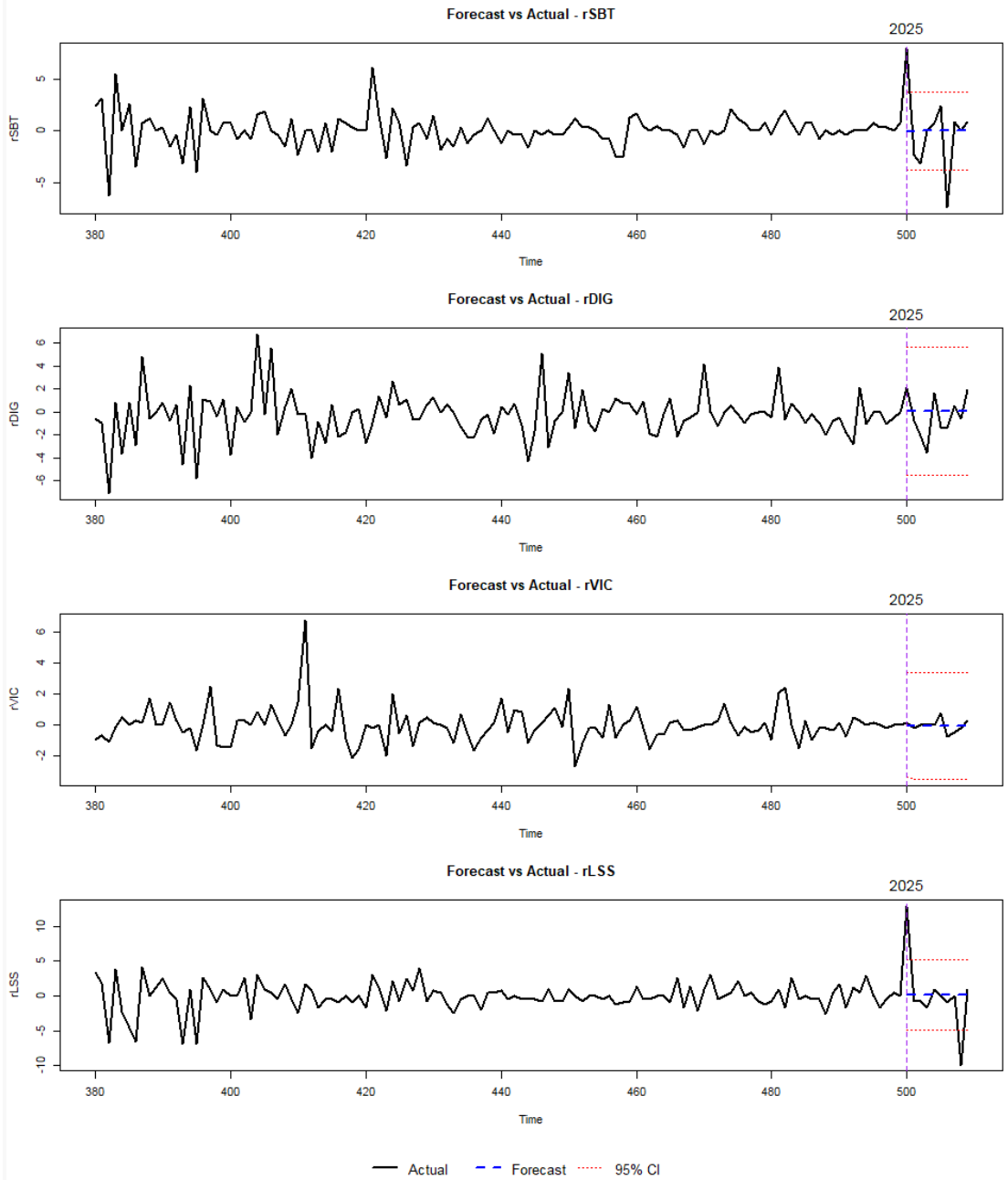
# Confidence intervals
lines(time_index[(total_len - 9):total_len], lower, col = "red", lty
= 3)
lines(time_index[(total_len - 9):total_len], upper, col = "red", lty
= 3)

# Vertical line
forecast_start_time <- time_index[total_len - 9] # = 500
abline(v = forecast_start_time, col = "purple", lty = 2)

# "2025" Label in margin top (side 3)
mtext("2025", side = 3, line = 0.5, at = forecast_start_time, cex =
0.9, col = "black")
}

# Legend
par(mar = c(0, 0, 0, 0))
plot.new()
legend("center", legend = c("Actual", "Forecast", "95% CI"),
      col = c("black", "blue", "red"),
      lty = c(1, 2, 3), lwd = c(2, 2, 1),
      horiz = TRUE, bty = "n", cex = 1.2)

```

	 <p>Forecast vs Actual - rSBT</p> <p>Forecast vs Actual - rDIG</p> <p>Forecast vs Actual - rVIC</p> <p>Forecast vs Actual - rLSS</p> <p>— Actual - - Forecast 95% CI</p>
	<p>So sánh giá trị dự báo với thực tế</p>
	<pre>forecast1\$fcst\$rSBT[,1] ## [1] -0.011067188 0.007552743 0.005618088 0.005609363 0.005596272 ## [6] 0.005594750 0.005594565 0.005594539 0.005594536 0.005594535 rmse(r2025\$rSBT, forecast1\$fcst\$rSBT[,1]) ## [1] 3.780792 forecast1\$fcst\$rDIG[,1]</pre>

	<pre>## [1] 0.06792788 0.03070166 0.03812674 0.03767113 0.03768271 0.03768407 ## [7] 0.03768403 0.03768404 0.03768404 0.03768404 rmse(r2025\$rDIG, forecast1\$fcst\$rDIG[,1]) ## [1] 1.814161 forecast1\$fcst\$rVIC[,1] ## [1] -0.02261872 -0.06328185 -0.06413807 -0.06426672 -0.0643 0331 -0.06430676 ## [7] -0.06430725 -0.06430732 -0.06430732 -0.06430733 rmse(r2025\$rVIC, forecast1\$fcst\$rVIC[,1]) ## [1] 0.384982 forecast1\$fcst\$rLSS[,1] ## [1] 0.2085823 0.1332291 0.1374191 0.1374098 0.1373949 0.137 3958 0.1373958 ## [8] 0.1373958 0.1373958 0.1373958 rmse(r2025\$rLSS, forecast1\$fcst\$rLSS[,1]) ## [1] 5.161905</pre>
5.2	HÀM PHẢN ỨNG IRF
	IRF
	<pre>irf_result <- irf(var1) irf_result ## ## Impulse response coefficients ## \$rSBT ## rSBT rDIG rVIC rLS S ## [1,] 1.930883e+00 1.170823e+00 3.760600e-01 1.793379e+0 0 ## [2,] -6.326814e-02 -9.964880e-02 3.931648e-02 1.460746e-0 1 ## [3,] 6.538066e-05 -1.117068e-02 -1.206379e-02 -1.630704e-0 2 ## [4,] -5.448942e-04 1.629500e-03 -8.705257e-04 5.161934e-0 4 ## [5,] -3.370720e-05 -5.852285e-05 -8.940855e-05 2.194333e-0 5 ## [6,] -5.984091e-06 7.522815e-07 -1.656697e-05 -3.592927e-0 6 ## [7,] -8.103414e-07 4.753763e-07 -1.990660e-06 1.038359e-0 7</pre>

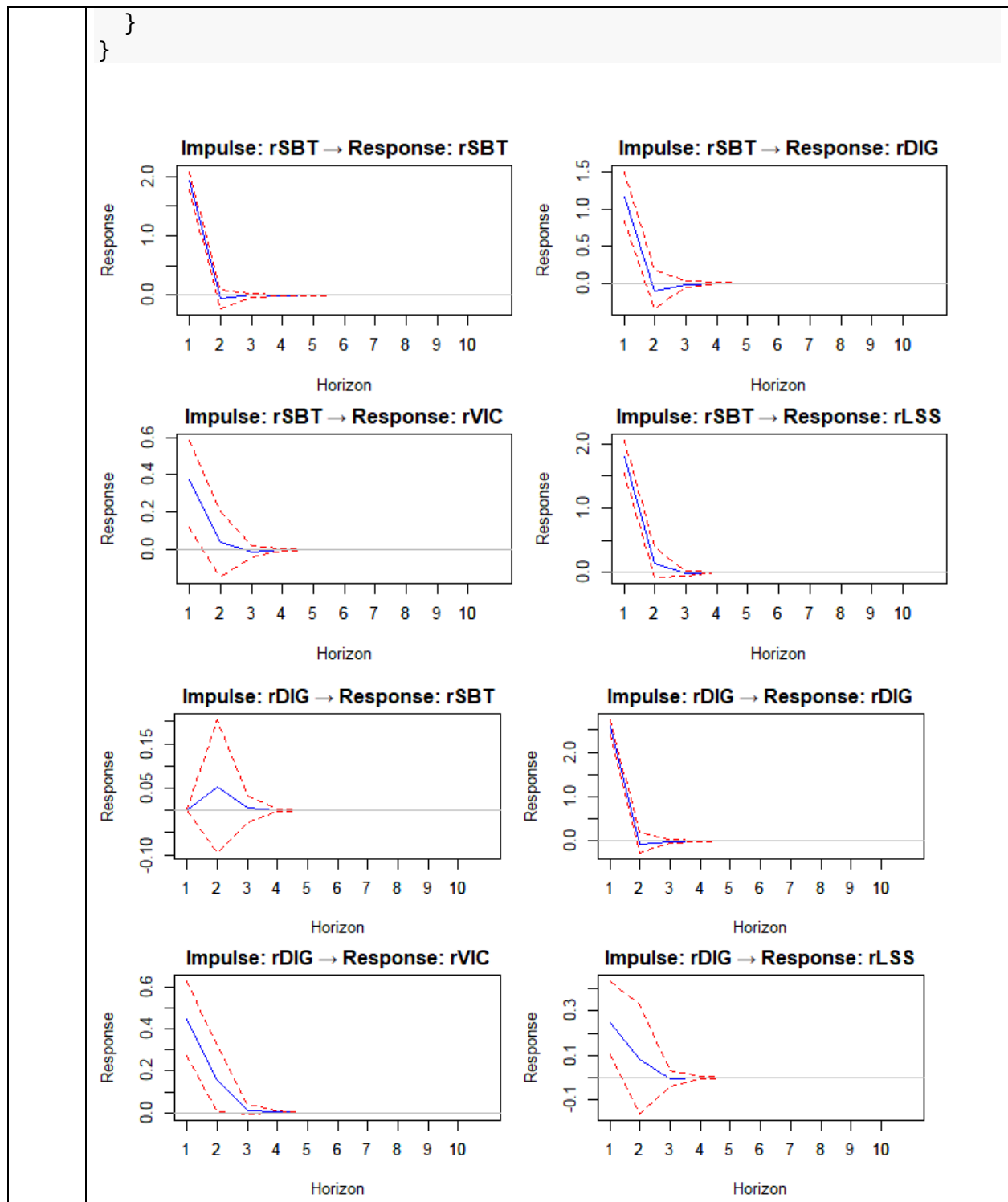
##	[8,]	-1.025006e-07	1.053137e-08	-2.624455e-07	-2.237458e-10
##	[9,]	-1.391591e-08	3.395257e-09	-3.585666e-08	-1.204147e-09
##	[10,]	-1.854999e-09	5.135733e-10	-4.723737e-09	-3.501898e-11
##	[11,]	-2.457951e-10	5.728274e-11	-6.287233e-10	-8.506423e-12
##	\$rDIG				
##		rSBT	rDIG	rVIC	rLSS
##	[1,]	0.000000e+00	2.595576e+00	4.492098e-01	2.470578e-01
##	[2,]	5.304634e-02	-7.388519e-02	1.575408e-01	8.137202e-02
##	[3,]	5.989252e-03	-6.738504e-03	1.241975e-02	-5.746492e-03
##	[4,]	6.851470e-04	3.815530e-04	1.856014e-03	1.238903e-04
##	[5,]	1.019752e-04	-4.165830e-05	2.644596e-04	1.859859e-05
##	[6,]	1.329154e-05	-4.190018e-06	3.337248e-05	-8.242924e-07
##	[7,]	1.749014e-06	-3.150322e-07	4.498022e-06	9.417503e-08
##	[8,]	2.351518e-07	-6.232545e-08	6.017603e-07	1.122543e-08
##	[9,]	3.122748e-08	-7.854866e-09	7.971908e-08	8.523364e-10
##	[10,]	4.153675e-09	-1.001381e-09	1.062658e-08	1.592570e-10
##	[11,]	5.534306e-10	-1.377144e-10	1.415094e-09	2.055315e-11
##	\$rVIC				
##		rSBT	rDIG	rVIC	rLSS
##	[1,]	0.000000e+00	0.000000e+00	1.635836e+00	1.466965e-01
##	[2,]	1.039725e-01	-5.057343e-02	1.946976e-01	-5.361844e-02
##	[3,]	1.021038e-02	2.765244e-03	2.938121e-02	3.719812e-03
##	[4,]	1.556730e-03	-7.422058e-04	3.891284e-03	4.064368e-05
##	[5,]	1.977235e-04	-4.158442e-05	5.034692e-04	-3.562936e-06
##	[6,]	2.651294e-05	-5.829101e-06	6.812738e-05	1.702288e-06
##	[7,]	3.542682e-06	-9.439420e-07	9.047250e-06	1.196243e-07
##	[8,]	4.703198e-07	-1.143513e-07	1.201943e-06	1.501365e-08
##	[9,]	6.263784e-08	-1.533441e-08	1.602364e-07	2.438331e-09
##	[10,]	8.341425e-09	-2.074861e-09	2.132480e-08	2.996607e-10
##	[11,]	1.110125e-09	-2.736508e-10	2.838350e-09	3.992320e-11
##	\$rLSS				
##		rSBT	rDIG	rVIC	rLS
##	[1,]	0.000000e+00	0.000000e+00	0.000000e+00	1.824736e+00
##	[2,]	-4.191915e-02	-1.268398e-01	-1.286924e-01	-6.725611e-02
##	[3,]	-7.027552e-03	8.889752e-03	-1.860024e-02	-2.627861e-03
##	[4,]	-8.418561e-04	2.857873e-04	-1.976772e-03	3.579904e-04
##	[5,]	-1.085858e-04	-5.041093e-06	-2.883597e-04	-2.045678e-05

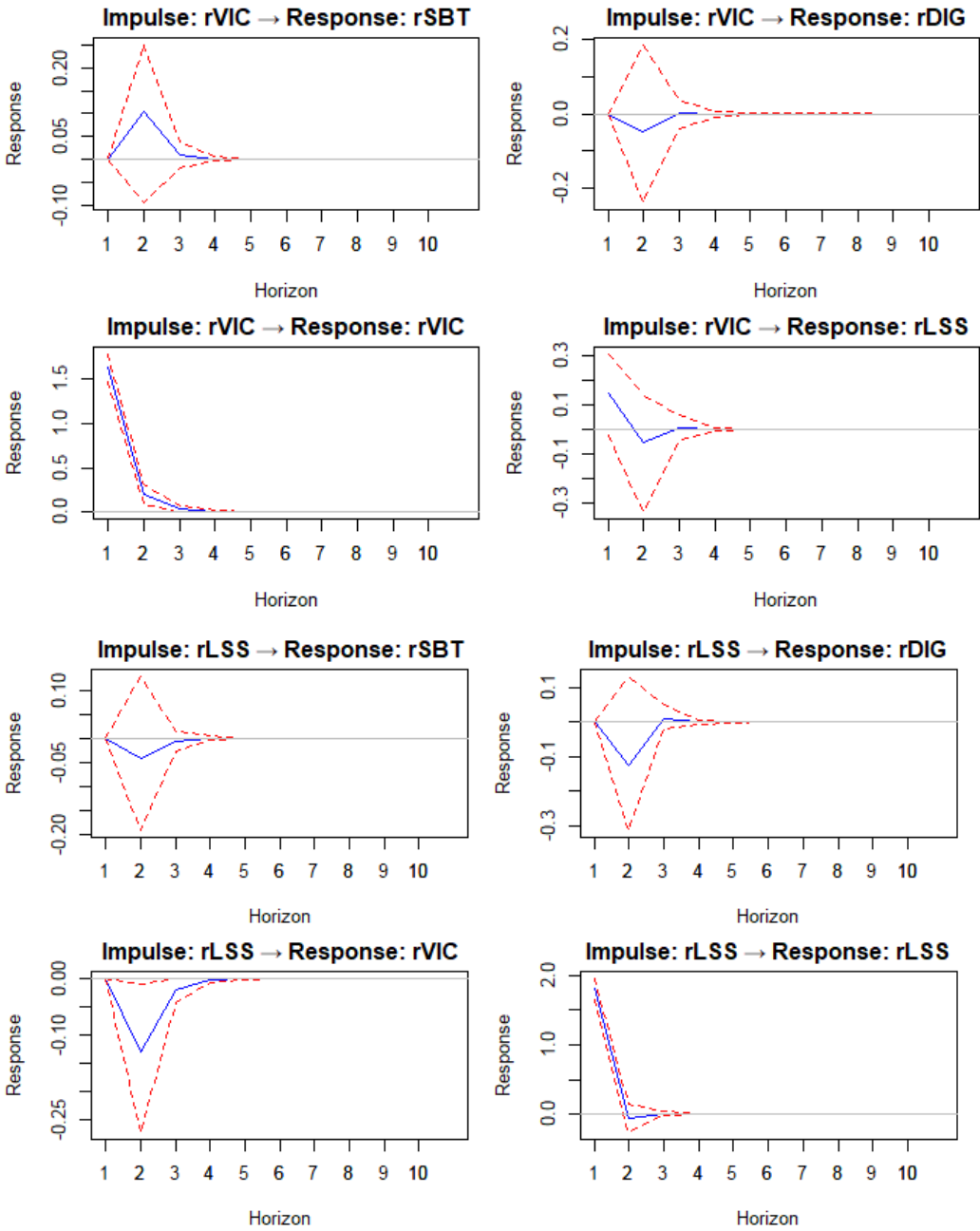
##	[6,]	-1.513814e-05	5.542100e-06	-3.859509e-05	-8.730731e-07
##	[7,]	-1.980169e-06	4.856798e-07	-5.033420e-06	7.565986e-09
##	[8,]	-2.635234e-07	5.889942e-08	-6.760415e-07	-1.350357e-08
##	[9,]	-3.520733e-08	9.100592e-09	-8.998287e-08	-1.309065e-09
##	[10,]	-4.679109e-09	1.151464e-09	-1.195819e-08	-1.538875e-10
##	[11,]	-6.229317e-10	1.526380e-10	-1.593246e-09	-2.353064e-11
##					
##					
##	Lower Band, CI= 0.95				
##	\$rSBT				
##		rSBT	rDIG	rVIC	rLS
S					
##	[1,]	1.726452e+00	8.446627e-01	1.939117e-01	1.518559e+00
##	[2,]	-2.548954e-01	-3.705207e-01	-9.485949e-02	-6.219125e-02
##	[3,]	-2.763482e-02	-4.935538e-02	-4.494724e-02	-6.549756e-02
##	[4,]	-6.129372e-03	-4.027461e-03	-4.483426e-03	-3.515353e-03
##	[5,]	-4.153444e-04	-8.577921e-04	-8.280742e-04	-6.349941e-04
##	[6,]	-2.216734e-04	-2.373775e-04	-1.551096e-04	-3.473384e-04
##	[7,]	-1.732590e-05	-1.758326e-05	-3.229016e-05	-3.238238e-05
##	[8,]	-6.365143e-06	-3.653819e-06	-6.890811e-06	-2.960042e-06
##	[9,]	-7.034238e-07	-1.289489e-06	-1.385778e-06	-9.610110e-07
##	[10,]	-2.391580e-07	-2.009282e-07	-3.680009e-07	-1.516235e-07
##	[11,]	-2.611366e-08	-6.684281e-08	-5.993635e-08	-1.928733e-08
##					
##	\$rDIG				
##		rSBT	rDIG	rVIC	rLS
S					
##	[1,]	0.000000e+00	2.414883e+00	2.436536e-01	6.322426e-02
##	[2,]	-1.055625e-01	-3.451819e-01	1.018012e-02	-1.578716e-01
##	[3,]	-3.104931e-02	-4.824438e-02	-1.275260e-02	-4.887567e-02

##	[4,]	-3.051345e-03	-8.018476e-03	-1.974295e-03	-6.265026e-03
3					
##	[5,]	-4.812638e-04	-8.806352e-04	-2.938987e-04	-9.894578e-04
4					
##	[6,]	-9.618971e-05	-3.384123e-04	-6.155815e-05	-2.378207e-04
4					
##	[7,]	-2.324320e-05	-3.072605e-05	-1.025163e-05	-6.658270e-05
5					
##	[8,]	-2.709374e-06	-7.975225e-06	-1.207911e-06	-1.039499e-05
5					
##	[9,]	-1.160658e-06	-1.100552e-06	-2.473869e-07	-2.382192e-06
6					
##	[10,]	-1.831531e-07	-3.492714e-07	-6.490429e-08	-5.263855e-07
7					
##	[11,]	-3.382931e-08	-2.466845e-08	-7.521649e-09	-1.131709e-07
7					
##					
##	\$rVIC				
##		rSBT	rDIG	rVIC	rLS
S					
##	[1,]	0.000000e+00	0.000000e+00	1.446849e+00	-1.340319e-02
2					
##	[2,]	-7.997534e-02	-3.268536e-01	6.545063e-02	-2.279720e-01
1					
##	[3,]	-1.756060e-02	-3.672514e-02	-3.522644e-03	-4.423807e-02
2					
##	[4,]	-1.485212e-03	-6.409404e-03	-2.061678e-03	-7.069697e-03
3					
##	[5,]	-6.156791e-04	-1.310546e-03	-3.808582e-04	-1.469351e-03
3					
##	[6,]	-9.168933e-05	-1.813575e-04	-7.488187e-05	-3.795342e-04
4					
##	[7,]	-2.120610e-05	-3.139996e-05	-6.608251e-06	-6.883336e-05
5					
##	[8,]	-1.659115e-06	-3.895859e-06	-1.135188e-06	-1.602095e-05
5					
##	[9,]	-7.396248e-07	-2.559463e-06	-1.535176e-07	-3.745930e-06
6					
##	[10,]	-4.412007e-08	-1.143980e-07	-3.935037e-08	-8.809767e-07
7					
##	[11,]	-2.056666e-08	-3.676025e-08	-4.045372e-09	-2.071399e-07
7					
##					
##	\$rLSS				
##		rSBT	rDIG	rVIC	rLS
S					
##	[1,]	0.000000e+00	0.000000e+00	0.000000e+00	1.696410e+00
0					
##	[2,]	-2.315905e-01	-3.465266e-01	-2.757627e-01	-3.313688e-01
1					

##	[3,]	-2.953403e-02	-1.800223e-02	-4.590341e-02	-2.934136e-02
2					
##	[4,]	-4.093691e-03	-5.782902e-03	-1.075072e-02	-4.357867e-03
3					
##	[5,]	-1.290840e-03	-1.587869e-03	-1.714998e-03	-1.260248e-03
3					
##	[6,]	-1.290640e-04	-2.052982e-04	-4.623798e-04	-2.346183e-04
4					
##	[7,]	-3.518785e-05	-2.688756e-05	-7.856770e-05	-2.993879e-05
5					
##	[8,]	-6.917460e-06	-1.476821e-05	-1.929341e-05	-7.022209e-06
6					
##	[9,]	-1.381914e-06	-2.620465e-06	-3.770364e-06	-9.801607e-07
7					
##	[10,]	-2.853890e-07	-2.259874e-07	-8.616069e-07	-2.076863e-07
7					
##	[11,]	-6.530366e-08	-9.396293e-08	-1.824985e-07	-3.454201e-08
8					
##					
##					
##	Upper Band, CI= 0.95				
##	\$rSBT				
##		rSBT	rDIG	rVIC	rLSS
##	[1,]	2.073504e+00	1.427413e+00	5.604571e-01	1.990308e+00
##	[2,]	1.051334e-01	1.325948e-01	2.209802e-01	3.101499e-01
##	[3,]	4.081538e-02	3.419959e-02	2.090048e-02	1.585142e-02
##	[4,]	7.580857e-03	1.269171e-02	5.606724e-03	1.019904e-02
##	[5,]	1.116549e-03	9.186525e-04	1.298413e-03	1.665278e-03
##	[6,]	1.495398e-04	1.037127e-04	2.083184e-04	2.178917e-04
##	[7,]	4.407940e-05	5.723493e-05	3.585960e-05	3.523863e-05
##	[8,]	7.544875e-06	3.048024e-06	8.088521e-06	6.438916e-06
##	[9,]	1.180347e-06	6.692748e-07	1.143676e-06	7.410656e-07
##	[10,]	2.193941e-07	1.281673e-07	2.836109e-07	2.457726e-07
##	[11,]	5.412572e-08	3.644001e-08	4.714946e-08	4.474513e-08
##					
##	\$rDIG				
##		rSBT	rDIG	rVIC	rLSS
##	[1,]	0.000000e+00	2.726822e+00	5.920601e-01	3.816506e-01
##	[2,]	2.317090e-01	1.475381e-01	3.397268e-01	2.960281e-01
##	[3,]	2.801752e-02	5.401371e-02	4.000573e-02	2.238626e-02
##	[4,]	5.093664e-03	7.730038e-03	9.801986e-03	5.512056e-03
##	[5,]	8.113724e-04	2.153905e-03	1.900597e-03	1.047074e-03
##	[6,]	1.970650e-04	1.969499e-04	3.973773e-04	1.562522e-04
##	[7,]	2.576866e-05	5.063801e-05	8.098067e-05	2.695603e-05
##	[8,]	7.972586e-06	1.397427e-05	1.750547e-05	4.531094e-06
##	[9,]	1.504499e-06	2.221913e-06	3.738535e-06	8.705552e-07
##	[10,]	3.570039e-07	5.779314e-07	8.031115e-07	1.783134e-07
##	[11,]	8.830835e-08	1.346531e-07	1.720125e-07	4.187001e-08
##					
##	\$rVIC				

	<pre>## rSBT rDIG rVIC rLSS ## [1,] 0.000000e+00 0.000000e+00 1.794864e+00 3.535994e-01 ## [2,] 2.694470e-01 1.690754e-01 3.356712e-01 1.463928e-01 ## [3,] 3.485543e-02 4.386660e-02 8.111093e-02 4.207109e-02 ## [4,] 7.520327e-03 6.573600e-03 1.671135e-02 5.896330e-03 ## [5,] 1.507146e-03 1.120943e-03 3.702178e-03 1.038451e-03 ## [6,] 3.250244e-04 3.288746e-04 7.785031e-04 1.902159e-04 ## [7,] 5.810769e-05 7.472508e-05 1.681787e-04 3.715405e-05 ## [8,] 1.388251e-05 1.281666e-05 3.577774e-05 8.005657e-06 ## [9,] 2.674179e-06 3.028846e-06 7.774141e-06 1.541773e-06 ## [10,] 6.368219e-07 5.955859e-07 1.709849e-06 2.772202e-07 ## [11,] 1.458213e-07 1.307563e-07 3.771073e-07 6.639504e-08 ## ## \$rLSS ## rSBT rDIG rVIC rLSS ## [1,] 0.000000e+00 0.000000e+00 0.000000e+00 1.913224e+00 ## [2,] 1.280245e-01 1.043118e-01 2.283652e-02 1.902333e-01 ## [3,] 2.137722e-02 5.056665e-02 7.122941e-03 3.244741e-02 ## [4,] 5.512056e-03 8.083529e-03 2.925634e-03 1.060624e-02 ## [5,] 5.253215e-04 9.401724e-04 4.324839e-04 1.494516e-03 ## [6,] 8.795446e-05 2.668576e-04 5.319326e-05 2.685220e-04 ## [7,] 2.085208e-05 6.184808e-05 1.761026e-05 8.375518e-05 ## [8,] 3.983649e-06 6.118315e-06 3.857739e-06 1.275299e-05 ## [9,] 9.844954e-07 1.796026e-06 5.391314e-07 2.567732e-06 ## [10,] 2.115016e-07 2.503266e-07 1.312462e-07 5.494045e-07 ## [11,] 3.486939e-08 4.109160e-08 2.741295e-08 1.324609e-07</pre>
	Vẽ đồ thị IRF
	<pre>impulses <- irf_result\$impulse responses <- irf_result\$response horizon <- nrow(irf_result\$irf[[1]]) par(mfrow = c(2, 2), mar = c(4, 4, 2, 1)) # 2x2 grid, nice margins for (imp in impulses) { for (resp in responses) { main_title <- paste("Impulse:", imp, "→ Response:", resp) irf_vals <- irf_result\$irf[[imp]][, resp] lower <- irf_result\$Lower[[imp]][, resp] upper <- irf_result\$Upper[[imp]][, resp] ts.plot(irf_vals, col = "blue", ylim = range(c(lower, upper)), main = main_title, ylab = "Response", xlab = "Horizon") axis(1, at = 1:10) lines(lower, col = "red", lty = 2) lines(upper, col = "red", lty = 2) abline(h = 0, col = "gray") } }</pre>



	
5.3	PHÂN RÃ PHƯƠNG SAI
	FEVD
	<pre>fevd(var1) ## \$rSBT ## rSBT rDIG rVIC rLSS ## [1,] 1.0000000 0.0000000000 0.000000000 0.000000000 ## [2,] 0.9958958 0.0007508387 0.002884515 0.0004688791 ## [3,] 0.9958454 0.0007603717 0.002912185 0.0004820325 ## [4,] 0.9958445 0.0007604962 0.002912829 0.0004822211 ## [5,] 0.9958444 0.0007604990 0.002912839 0.0004822243 ## [6,] 0.9958444 0.0007604990 0.002912839 0.0004822243 ## [7,] 0.9958444 0.0007604990 0.002912839 0.0004822243</pre>

	## [8,] 0.9958444 0.0007604990 0.002912839 0.0004822243
	## [9,] 0.9958444 0.0007604990 0.002912839 0.0004822243
	## [10,] 0.9958444 0.0007604990 0.002912839 0.0004822243
	##
	## \$rDIG
	## rSBT rDIG rVIC rLSS
##	[1,] 0.1690740 0.8309260 0.0000000000 0.0000000000
##	[2,] 0.1695869 0.8281230 0.0003141379 0.001975998
##	[3,] 0.1695968 0.8281024 0.0003150671 0.001985642
##	[4,] 0.1695971 0.8281021 0.0003151346 0.001985651
##	[5,] 0.1695971 0.8281021 0.0003151348 0.001985651
##	[6,] 0.1695971 0.8281021 0.0003151349 0.001985651
##	[7,] 0.1695971 0.8281021 0.0003151349 0.001985651
##	[8,] 0.1695971 0.8281021 0.0003151349 0.001985651
##	[9,] 0.1695971 0.8281021 0.0003151349 0.001985651
##	[10,] 0.1695971 0.8281021 0.0003151349 0.001985651
##	##
	## \$rVIC
	## rSBT rDIG rVIC rLSS
##	[1,] 0.04684108 0.06683609 0.8863228 0.0000000000
##	[2,] 0.04611833 0.07309947 0.8754397 0.005342492
##	[3,] 0.04614281 0.07311364 0.8752921 0.005451442
##	[4,] 0.04614271 0.07311420 0.8752904 0.005452661
##	[5,] 0.04614271 0.07311421 0.8752904 0.005452687
##	[6,] 0.04614271 0.07311421 0.8752904 0.005452687
##	[7,] 0.04614271 0.07311421 0.8752904 0.005452687
##	[8,] 0.04614271 0.07311421 0.8752904 0.005452687
##	[9,] 0.04614271 0.07311421 0.8752904 0.005452687
##	[10,] 0.04614271 0.07311421 0.8752904 0.005452687
##	##
	## \$rLSS
	## rSBT rDIG rVIC rLSS
##	[1,] 0.4852145 0.009208448 0.003246600 0.5023305
##	[2,] 0.4858420 0.010153229 0.003660801 0.5003440
##	[3,] 0.4858586 0.010157697 0.003662702 0.5003210
##	[4,] 0.4858586 0.010157698 0.003662702 0.5003210
##	[5,] 0.4858586 0.010157698 0.003662702 0.5003210
##	[6,] 0.4858586 0.010157698 0.003662702 0.5003210
##	[7,] 0.4858586 0.010157698 0.003662702 0.5003210
##	[8,] 0.4858586 0.010157698 0.003662702 0.5003210
##	[9,] 0.4858586 0.010157698 0.003662702 0.5003210
##	[10,] 0.4858586 0.010157698 0.003662702 0.5003210
	Vẽ đồ thị FEVD
	fevd_result <- fevd(var1)
	contrib_names <- colnames(stock1)
	colors <- c("lightblue", "lightgreen", "lightpink", "lightyellow")
	<i># Vẽ từng đồ thị riêng, Legend nằm bên ngoài</i>
	for (v in names(fevd_result)) {
	mat <- t(fevd_result[[v]])

```

rownames(mat) <- contrib_names[1:nrow(mat)]

# Tăng lề bên phải để chứa Legend
par(mar = c(5, 4, 4, 8), xpd = TRUE) # xpd = TRUE cho phép vẽ Legend ngoài vùng plot

barpos <- barplot(mat,
  beside = FALSE,
  col = colors[1:nrow(mat)],
  main = paste("FEVD for", v),
  xlab = "Horizon",
  ylab = "Percentage",
  xaxt = "n")

axis(1, at = barpos, labels = 1:ncol(mat))

# Legend bên phải, ra ngoài vùng plot
legend("topright",
  inset = c(-0.25, 0),
  legend = rownames(mat),
  fill = colors[1:nrow(mat)],
  bty = "n")
}

```

