

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

Giảng viên	:Th.s Bùi Dương Hải
Nhóm thực hiện:	:Nhóm 2
Lớp:	:Toán kinh tế 64

Hà Nội, tháng 4 năm 2025

MỤC LỤC

THÔNG TIN VỀ NHÓM VÀ TỈ LỆ ĐÓNG GÓP	3
A. BÀI TẬP CÁ NHÂN.....	4
I. Nguyễn Bình An – CTCP Vật tư Kỹ thuật Nông nghiệp Cần Thơ.....	4
1. Giới thiệu công ty	4
2. Phân tích chuỗi doanh thu thuần (Financial series).....	4
3. Phân tích chuỗi giá và chuỗi log-return.....	6
II. VŨ PHẠM THÀNH PHƯƠNG – CTCP Vận tải Xăng dầu Vitaco.	10
1. Phân tích chuỗi doanh thu thuần.....	10
2. Phân tích chuỗi lịch sử giá.....	11
3. Phân tích chuỗi lợi suất của giá cổ phiếu.	13
III. Đỗ Quang Phước – CTCP Gemadept	16
1. Tổng quan về CTCP Gemadept	16
2. Phân tích chuỗi doanh thu thuần theo quý	16
3. Phân tích chuỗi giá cổ phiếu và log return theo ngày.....	17
IV. Đinh Ngọc Thảo - Tổng Công ty Phân bón và Hóa chất Dầu khí	26
1. Tổng quan về công ty	26
2. Phân tích chuỗi doanh thu thuần theo quý	26
3. Phân tích số liệu giá cổ phiếu theo ngày và log-return.....	28
B. BÀI TẬP NHÓM.....	32
1. Phân tích tương quan.....	32
2. Xây dựng danh mục tối ưu	32
3. Phân tích đồng tích hợp	33
4. Mô hình VAR	34
C. PHỤ LỤC	42
1. Nguyễn Bình An.....	42
2. Vũ Phạm Thành Phương	80
3. Đỗ Quang Phước	96
4. Đinh Ngọc Thảo	139

THÔNG TIN VỀ NHÓM VÀ TỈ LỆ ĐÓNG GÓP

STT	Họ và tên	Mã sinh viên	Đóng góp
1	Nguyễn Bình An	11220025	22%
2	Vũ Phạm Thành Phương	11225331	22%
3	Đỗ Quang Phước	11225151	30%
4	Đinh Ngọc Thảo	11225849	26%
Tổng			100%

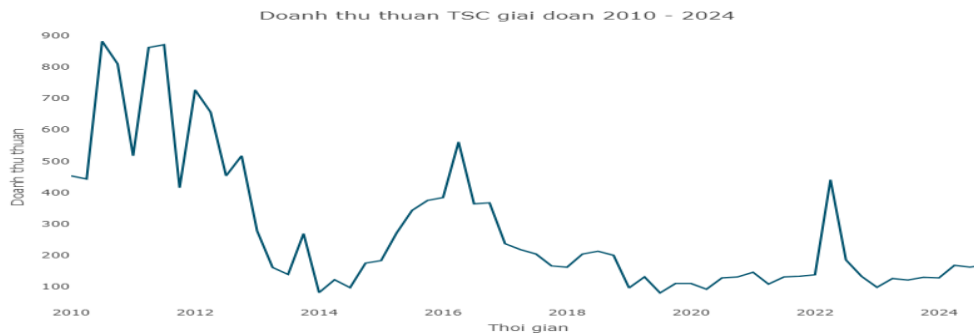
A. BÀI TẬP CÁ NHÂN

I. Nguyễn Bình An – CTCP Vật tư Kỹ thuật Nông nghiệp Cần Thơ

1. Giới thiệu công ty

Ra đời từ năm 1976, đến nay, với hàng ngàn đại lý, khách hàng trong và ngoài nước, CTCP Vật Tư Kỹ Thuật Nông Nghiệp Cần Thơ viết tắt là TSC là một trong những doanh nghiệp nông nghiệp hàng đầu Việt Nam. Cung ứng ra thị trường từ hạt giống, phân bón, nông dược, nông sản đến thực phẩm, chúng tôi xác định sứ mệnh là Cung cấp giải pháp toàn diện để phát triển ngành Nông nghiệp theo hướng công nghệ cao, bền vững.

2. Phân tích chuỗi doanh thu thuần (Financial series)



Hình 1. Biểu đồ Doanh thu thuần theo quý của mã cổ phiếu TSC giai đoạn 2010-2024.

Doanh thu của **CTCP Vật Tư Kỹ Thuật Nông Nghiệp Cần Thơ** chứng kiến các cuộc biến động doanh thu liên tục từ năm 2010 đến 2014. Đặc biệt là vào năm 2014, khi công ty phải chịu sự sụt giảm nghiêm trọng khi vào cuối năm 2014 đầu 2015, doanh thu thuần của công ty còn dưới mức 100. Mặc dù đã có những sự tăng trưởng mạnh mẽ trở lại từ 2015 đến 2019, công ty vẫn phải đối mặt với đợt khủng hoảng từ cuối 2019 đến đầu năm 2022 bởi tình hình đại hình. Cho tới giờ, công ty cũng đã có những bức tiến phát triển trở lại sau các biến cố trên.

2.1. Ước lượng mô hình Doanh thu thuần.

Training set: Từ Quý 1/2010 đến hết Quý 4/2023 (56 quan sát).

Validation set: Quý 1,2,3 và 4 năm 2024.

Mô hình	RMSE (training set)	MAPE (training set)	RMSE (validation set)	MAPE (validation set)	Dự báo Quý 1/2024 (Tỷ VND)	Dự báo Quý 2/2024 (Tỷ VND)	Dự báo Quý 3/2024 (Tỷ VND)	Dự báo Quý 4/2024 (Tỷ VND)
Lin-Lin	160.5813	0.5835454	150.5592	0.9392399	22.41	13.08	3.75	-5.58

Bài tập lớn Phân tích chuỗi thời gian và dự báo

Lin-Log	153.6373	0.4704595	48.64261	0.2725296	116.81	113.72	110.68	107.69
Log-Lin	159.8776	0.4213103	66.80729	0.3983233	96.94	94.12	91.39	88.73
Log-Log	186.8334	0.4177543	30.96587	0.1700909	133.32	132.07	130.86	129.67
Xu thế tuyến tính mùa vụ dạng nhân	159.1975	0.5993855	151.4999	0.9708991	-33.48	63.79	-1.12	0.89
Xu thế tuyến tính mùa vụ dạng cộng	158.4192	0.5913084	151.0263	0.9729269	-24.23	31.91	18.70	-1.30
Holt- Winters mùa vụ dạng cộng	147.3373	0.4556388	36.13883	0.2135265	109.18	201.08	129.08	117.11
Holt- Winters mùa vụ dạng nhân	146.5236	0.414646	35.27453	0.1965327	114.05	149.47	118.75	118.31

Bảng 1. Một số mô hình dự báo Doanh thu thuần VTO năm 2025.

Từ bảng trên quan sát được mô hình Holt-Winters dạng nhân có RMSE và MAPE nhỏ nhất ở cả hai tập training set và validation set. Trong khi đó, mô hình Holt-Winter dạng cộng đều có RMSE và MAPE lớn hơn so với mô hình dạng nhân. Từ đó, ta chọn mô hình Holt-Winter dạng nhân để dự báo doanh thu 4 quý năm 2025.

2.2. Lựa chọn mô hình.

Áp dụng mô hình Holt-Winters mùa vụ dạng nhân, dự báo được kết quả năm 2025 như sau:

Q1/2025	Q2/2025	Q3/2025	Q4/2025
151.37	197.99	166.93	165.39

Bảng 2. Kết quả dự báo năm 2025

3, Phân tích chuỗi giá và chuỗi log-return

3.1. Phân loại ADF

a. Chuỗi giá

```
## Test regression trend

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-403.23  -36.70   -0.18   35.15  398.39

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.04578     4.30726  -0.475  0.63502
z.lag.1      -0.93980     0.05769 -16.291 < 2e-16 ***
z.diff.lag    0.12502     0.04464   2.800  0.00531 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 95.9 on 493 degrees of freedom
Multiple R-squared:  0.4272,    Adjusted R-squared:  0.4249
F-statistic: 183.9 on 2 and 493 DF,  p-value: < 2.2e-16

Value of test-statistic is: -16.2905 132.6926

Critical values for test statistics:
      1pct  5pct 10pct
tau2  -3.44 -2.87 -2.57
phi1   6.47  4.61  3.79
```

Ở mô hình "Test regression trend", có sự hiện diện của thành phần xu thế với hệ số t có ý nghĩa (ít nhất trong phiên bản đầu tiên).

Mặc dù mô hình “drift” cho kết quả bác bỏ giả thiết đơn vị gốc rất mạnh ($z.lag.1$ với test statistic khoảng -16.29 , cực kỳ nhỏ so với các giá trị tới hạn), nhưng nếu dữ liệu có xu thế thì việc loại bỏ thành phần xu thế có thể tạo ra sự đánh giá thừa về mức độ dừng của chuỗi.

Từ đó, kiểm định ADF đúng và phù hợp với chuỗi giá là dạng dừng xu thế.

b, Chuỗi log-return

Mô hình này chỉ bao gồm hằng số (drift) và các thành phần sai phân.

Bài tập lớn Phân tích chuỗi thời gian và dự báo

Hệ số liên quan (đặc biệt $z.lag.1$) rất có ý nghĩa thống kê với giá trị test statistic khoảng -16.56, rất thấp so với các giá trị tới hạn (τ_2 : -3.44 ở mức 1%).

Điều này cho thấy ta có thể bác bỏ giả thiết đơn vị gốc trong mô hình dạng dừng sai phân.

```
Test regression drift

Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-7.3995 -1.1314 -0.0209  1.0526  7.0960

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.07163    0.10510  -0.682  0.49587
z.lag.1      -0.96980    0.05857 -16.557 < 2e-16 ***
z.diff.lag    0.12535    0.04463   2.808  0.00518 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.339 on 493 degrees of freedom
Multiple R-squared:  0.4404,    Adjusted R-squared:  0.4381
F-statistic: 194 on 2 and 493 DF,  p-value: < 2.2e-16

Value of test-statistic is: -16.5566 137.0646

Critical values for test statistics:
      1pct   5pct 10pct
tau2  -3.44 -2.87 -2.57
phi1   6.47  4.61  3.79
```

3.2. Xác định bậc AR,MA

a. Chuỗi giá

```
Best model: ARIMA(2,1,3)

Series: priceTSC
ARIMA(2,1,3)

Coefficients:
      ar1      ar2      ma1      ma2      ma3
    1.3963 -0.9140 -1.2455  0.6008  0.2081
s.e.  0.1014  0.0903  0.1046  0.1253  0.0478

sigma^2 = 8876: log likelihood = -2968.12
AIC=5948.24  AICc=5948.41  BIC=5973.5

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set -1.733657 93.64355 61.93823 -0.0840539 1.616478 0.9908525 0.001035072
```

Với chuỗi giá, ARIMA thích hợp nhất là ARIMA(2,1,3) với giá trị AIC = 5948.24.

b. Chuỗi lợi suất

Best model: ARIMA(3,0,2) with zero mean

Series: logTSC

ARIMA(3,0,2) with zero mean

Coefficients:

	ar1	ar2	ar3	ma1	ma2
	1.6436	-1.1169	0.2251	-1.5144	0.8014
s.e.	0.1353	0.0977	0.0465	0.1351	0.0810

sigma^2 = 5.343: log likelihood = -1121.54

AIC=2255.07 AICc=2255.24 BIC=2280.33

ARIMA(3,0,2) là thỏa mãn nhất với chuỗi log-return với AIC tính được bằng 2255.07.

Ngoài ra, với mức ý nghĩa 5%, $p_value = 0.2833 > 0.05$ nên chưa bác bỏ H_0 .

→ Không tồn tại hiện tượng tự tương quan trong phần dữ liệu. Qua đó thấy được mô hình phù hợp.

Mô hình dưới đây cho ta thấy được (p,d,q) tốt nhất của chuỗi giá và chuỗi lợi suất .

Mô hình	AIC	RMSE	Nghiệm nghịch đảo	P-value (Ljung-Box)
(2,1,3)	5948.24	93.64355	Có nghiệm nằm trên ĐTĐV	0.1257
(3,0,2)	2255.07	2.29921	Nằm trong ĐTĐV	0.2833

3.3. Dự báo 10 phiên giao dịch đầu năm 2025.

a, Với chuỗi giá

	(2,1,3)	Giá thực tế
1	2424.773	2430
2	2430.260	2430
3	2438.082	2390
4	2443.988	2400
5	2445.087	2390
6	2441.233	2400

Bài tập lớn Phân tích chuỗi thời gian và dự báo

7	2434.823	2360
8	2429.419	2340
9	2427.721	2340
10	2430.291	2340

Với dữ liệu bảng trên, mô hình ARIMA(2,1,3) là mô hình dự báo tốt nhất với sai số RMSE và MAPE nhỏ nhất với giá dự báo sát với giá thực tế của công ty TSC.

b, Chuỗi lợi suất

	(3,0,2)	Lợi suất thực tế
1	-0.04539154	0
2	0.28347536	-0.0165
3	0.22341308	0.004175
4	0.02286767	-0.004175
5	-0.16538216	0.004175
6	-0.26393635	-0.016807
7	-0.26057255	-0.0085106
8	-0.18733111	0
9	-0.0930473	0
10	-0.01934317	0

Bảng 9. So sánh mô hình với lợi suất thực tế.

Như vậy, với cách dự báo thông qua dự báo log return thì mô hình ARIMA(3,0,2) là phù hợp nhất với sai số dự báo RMSE nhỏ nhất.

II. VŨ PHẠM THÀNH PHƯƠNG – CTCP Vận tải Xăng dầu Vitaco.

1. Phân tích chuỗi doanh thu thuần.



Hình 2. Biểu đồ Doanh thu thuần theo quý của mã cổ phiếu VTO giai đoạn 2010-2024.

Doanh thu của **CTCP Vận tải Xăng dầu Vitaco** phụ thuộc đáng kể vào chu kỳ kinh tế và giá dầu thế giới. Từ nửa cuối năm 2014 đến đầu năm 2017 giảm mạnh theo giá dầu do thừa nguồn cung và kỳ vọng tăng trưởng kinh tế toàn cầu đạt mức thấp. Từ năm 2017 trở đi bắt đầu phục hồi dần tuy nhiên lại gặp cú sốc Covid-19 khiến doanh thu quý 4 năm 2021 đạt mức thấp nhất giai đoạn.

1.1. Ước lượng mô hình Doanh thu thuần.

Training set: Từ Quý 1/2010 đến hết Quý 4/2023 (56 quan sát).

Validation set: Quý 1,2,3 và 4 năm 2024.

Mô hình	RMSE (training set)	MAPE (training set)	RMSE (validation set)	MAPE (validation set)	Dự báo Quý 1/2024 (Tỷ VND)	Dự báo Quý 2/2024 (Tỷ VND)	Dự báo Quý 3/2024 (Tỷ VND)	Dự báo Quý 4/2024 (Tỷ VND)
Lin-Lin	47.3038	0.11984	21.16687	0.063815	272.700	270.363	268.026	265.689
Lin-Log	53.7249	0.14387	33.25875	0.106548	308.997	308.451	307.914	307.387
Log-Lin	47.9370	0.118416	21.48485	0.062204	270.917	268.946	266.989	265.046
Log-Log	55.0335	0.144276	28.54126	0.086377	303.340	302.820	302.311	301.811
Xu thế tuyến tính mùa vụ	47.176	0.120085	20.21466	0.057120	268.226	270.369	266.012	270.655

Bài tập lớn Phân tích chuỗi thời gian và dự báo

dạng cộng								
Xu thế tuyến tính mùa vụ dạng nhân	47.0693	0.118472	20.39319	0.052969	270.209	266.286	259.723	279.302
Holt-Winters mùa vụ dạng cộng	46.4913	0.116167	17.84044	0.052972	263.007	268.072	267.676	276.507
Holt-Winters mùa vụ dạng nhân	46.3834	0.115195	17.12226	0.050123	269.201	271.532	269.669	277.653

Bảng 3. Một số mô hình dự báo Doanh thu thuần VTO năm 2025.

Từ bảng trên quan sát được mô hình có RMSE và MAPE nhỏ nhất ở cả hai tập training set và validation set là:

$$\widehat{R}_{t+k} = (295.626 - 0.2129 * k) * s_Q.$$

1.2. Lựa chọn mô hình.

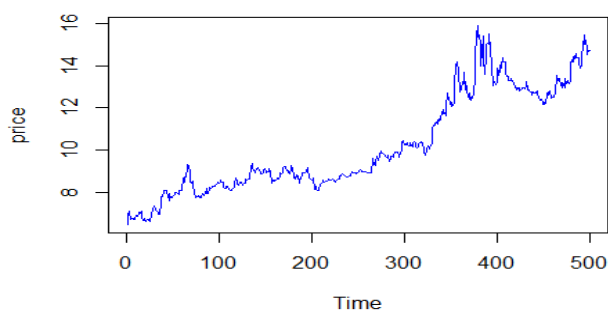
Áp dụng mô hình Holt-Winters mùa vụ dạng nhân, dự báo được kết quả năm 2025 như sau:

Q1/2025	Q2/2025	Q3/2025	Q4/2025
280.9046	283.1359	289.1929	295.3345

Bảng 4. Kết quả dự báo năm 2025

2. Phân tích chuỗi lịch sử giá.

Hình 3. Biểu đồ chuỗi giá theo từng phiên mã VTO giai đoạn 2023-2024



Đầu tiên, giá cổ phiếu của công ty phụ thuộc vào giá năng lượng thế giới. Tuy nhiên, đúng là giá dầu từ 2024 có tăng nhẹ do OPEC siết chặt nguồn cung, nhưng sự tăng trưởng lớn vào năm 2024 chủ yếu là sự phục hồi sau giai đoạn thiếu hụt nguồn cung xăng dầu trầm trọng ở Việt Nam từ cuối 2022 đến đầu năm 2023.

2.1. Kiểm định tính dừng của chuỗi giá.

Kiểm định	P-value	$ \tau_{statistic} $	$ \tau_{0.05} $
Trend	0.01547	3.0461	3.42
Drift	0.4012	0.8715	2.87
None	0.2866	1.1119	1.95

Bảng 5. Kết quả kiểm định ADF.

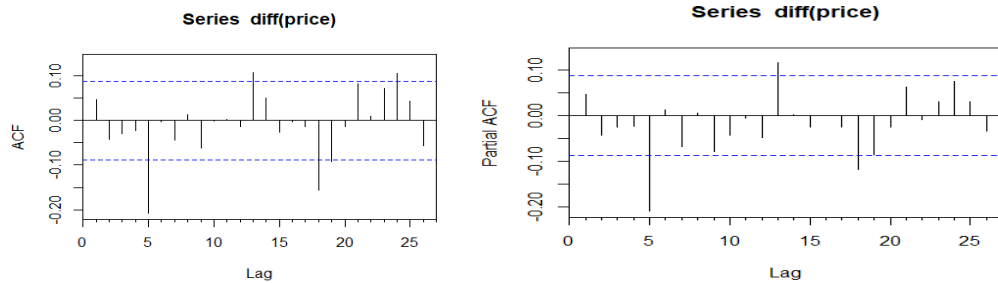
Trị tuyệt đối của giá trị quan sát luôn nhỏ hơn trị tuyệt đối của giá trị tới hạn ứng với mức ý nghĩa 5% nên chuỗi chưa dừng, tiến hành sai phân bậc 1.

Kiểm định	P-value	$ \tau_{statistic} $	$ \tau_{0.05} $
Trend	≈ 0	16.0628	3.42
Drift	≈ 0	16.0717	2.87
None	≈ 0	15.9978	1.95

Bảng 6. Kết quả kiểm định ADF cho sai phân bậc 1.

Trị tuyệt đối của giá trị quan sát lớn hơn trị tuyệt đối của giá trị tới hạn ứng với mức ý nghĩa 5% nên chuỗi đã dừng.

2.2. Xác định bậc của AR, MA.



Hình 4. Đồ thị ACF và PACF của sai phân bậc 1.

Theo đồ thị trên các lag 5 và lag 13 có spike lớn vượt quá ngưỡng nên các giá trị p,q có thể nhận là 5 và 13. Với d=1 ta xem xét các mô hình ARIMA sau:

Mô hình	AIC	RMSE	MAPE	Nghiem nghịch đảo	P-value (Ljung-Box)
(5,1,13)	10.77	0.23308	1.43959	Có nghiệm nằm trên ĐTĐV	0.04536
(5,1,5)	12.59	0.23936	1.44402	Nằm trong ĐTĐV	0.2531
(13,1,5)	12.27	0.23534	1.45426	Nằm trong ĐTĐV	0.0973

Bảng 5. So sánh và kiểm tra nhiễu trắng chuỗi giá.

Các giá trị có được trong bảng trên xem ở phần phụ lục. Mô hình (5,1,13) có AIC thấp nhưng giá trị P-value trong kiểm định Ljung-Box nhỏ hơn 0.05 nên Bác bỏ H_0 phần dư có tự tương quan, loại mô hình. Với hai mô hình còn lại ta thấy mô hình ARIMA(13,1,5) có AIC và RMSE nhỏ hơn nhưng

MAPE lớn hơn, nên ta phải tiếp tục tiến hành dự báo chuỗi giá và so sánh với mức giá thực tế để tìm ra mô hình phù hợp.

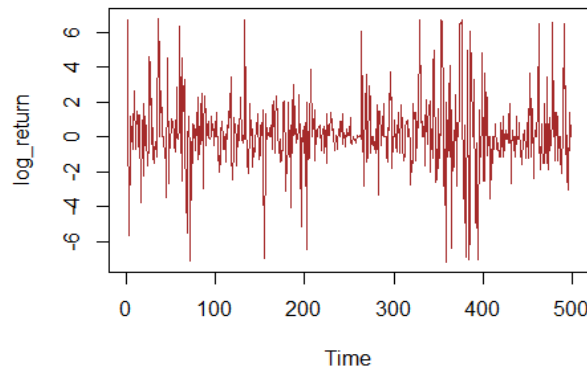
2.3. Dự báo 10 phiên giao dịch đầu năm 2025.

	(5,1,5)	(13,1,5)	Giá thực tế
1	14.85225	14.85859	14.8
2	14.78453	14.75645	14.45
3	14.86928	14.83313	13.9
4	14.79419	14.77616	13.8
5	14.78656	14.84584	14.15
6	14.77943	14.93313	14.05
7	14.78044	15.01064	13.8
8	14.79513	15.05530	13.85
9	14.78438	15.00995	13.8
10	14.78741	14.92340	14
RMSE	0.801461	0.9176132	
MAPE	0.053222	0.06034552	

Bảng 6. So sánh mô hình với giá thực tế.

Với dữ liệu bảng trên, mô hình ARIMA(5,1,5) là mô hình dự báo tốt nhất với sai số RMSE và MAPE nhỏ nhất trong 2 mô hình. Vậy nên mô hình ARIMA(13,1,5) dù có AIC nhỏ hơn nhưng không thể kết luận là mô hình tốt nhất.

3. Phân tích chuỗi lợi suất của giá cổ phiếu.



Hình 5. Đồ thị chuỗi lợi suất của cổ phiếu VTO.

3.1. Kiểm định tính dừng của chuỗi lợi suất.

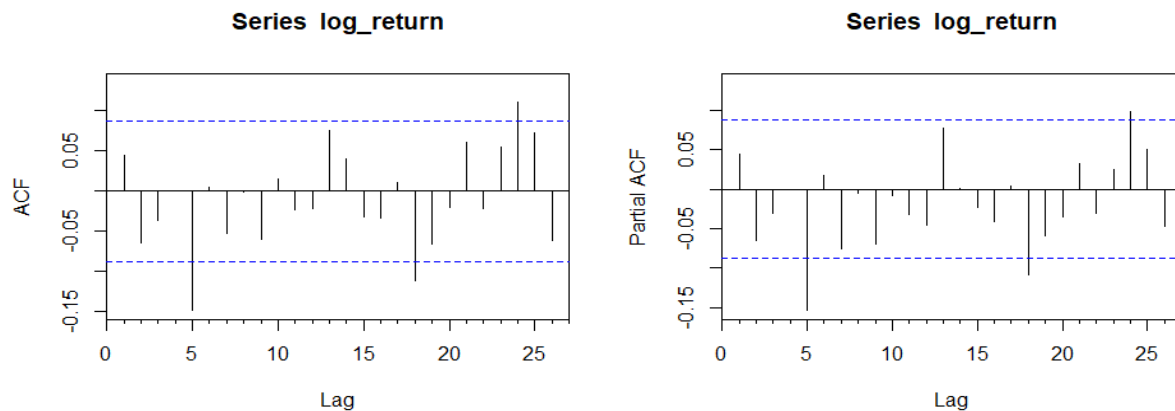
Kiểm định	P-value	$ \tau_{statistic} $	$ \tau_{0.05} $
Trend	≈ 0	16.5527	3.42

Drift	≈ 0	16.5696	2.87
None	≈ 0	16.4689	1.95

Bảng 7. Kết quả kiểm định ADF cho chuỗi lợi suất.

Trị tuyệt đối của giá trị quan sát lớn hơn trị tuyệt đối của giá trị tới hạn ứng với mức ý nghĩa 5% nên chuỗi đã dừng.

3.2. Xác định bậc AR, MA của chuỗi lợi suất.



Hình 6. Đồ thị ACF và PACF của chuỗi lợi suất.

Theo đồ thị trên các lag 5 và lag 18 có spike lớn vượt quá ngưỡng. Tuy giá trị 18 quá cao có thể làm sai lệch dự báo nhưng ta vẫn xem xét để đưa ra kết luận đúng nhất. Tiếp tục sử dụng hàm `auto.arima()` ta được mô hình ARIMA(5,0,0). Với các giá trị trên ta có bảng so sánh:

Mô hình	AIC	RMSE	Nghiệm nghịch đảo	P-value (Ljung-Box)
(18,0,5)	2174.34	2.031373	Có nghiệm nằm trên ĐĐV	0.1652
(5,0,18)	2177.96	2.039023	Có nghiệm nằm trên ĐĐV	0.1046
(5,0,5)	2171.48	2.089456	Nằm trong ĐĐV	0.4208
(5,0,0)	2168.33	2.104128	Nằm trong ĐĐV	0.2987

Bảng 8. So sánh và kiểm tra nhiều trắng chuỗi lợi suất.

Loại hai mô hình ARIMA(18,0,5) và (5,0,18) vì có nghiệm nghịch đảo nằm trên đường tròn đơn vị nên có thể chuỗi vẫn chưa dừng. Ta tiến hành dự báo với hai mô hình còn lại.

3.3. Dự báo lợi suất 10 phiên giao dịch đầu năm 2025.

So sánh sai số RMSE của hai mô hình ARIMA(5,0,5) và ARIMA(5,0,0) dựa trên lợi suất thực tế.

	(5,0,5)	(5,0,0)	Lợi suất thực tế
--	---------	---------	------------------

1	0.98019	0.63660	0.677968699
2	-0.38638	0.25692	-2.393276621
3	0.80653	0.66138	-3.880557442
4	-0.35845	-0.02992	-0.722024797
5	0.38646	0.14680	2.504603193
6	-0.07607	0.08357	-0.709222831
7	0.33562	0.15911	-1.795380362
8	0.09814	0.09070	0.361664047
9	0.20658	0.19556	-0.361664047
10	0.11301	0.17370	1.438873745
RMSE	1.944191	2.003145	

Bảng 9. So sánh mô hình với lợi suất thực tế.

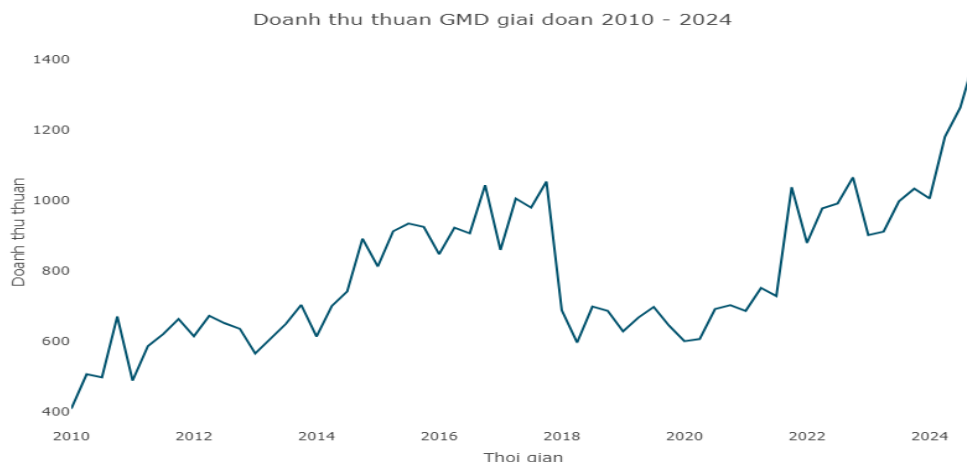
Như vậy, với cách dự báo thông qua dự báo log return thì mô hình ARIMA(5,0,5) là phù hợp nhất với sai số dự báo RMSE nhỏ nhất. Kết hợp cả hai cách dự báo bằng chuỗi giá và log return thì ta thấy mô hình ARIMA(5,1,5) là mô hình phù hợp nhất vì có sai số RMSE thấp hơn đáng kể và tương đối sát với chuỗi giá thực tế.

III. Đỗ Quang Phước – CTCP Gemadept

1. Tổng quan về CTCP Gemadept

Thành lập vào năm 1990, Công ty Cổ phần Gemadept cùng quốc gia khởi nghiệp, tiên phong đưa dịch vụ container vào Việt Nam và kết nối những tuyến hàng hải đầu tiên đến với thị trường quốc tế. Ngày nay, Gemadept là một thương hiệu hàng đầu trong ngành Khai thác Cảng và Logistics của Việt Nam.

2. Phân tích chuỗi doanh thu thuần theo quý



Hình 3.1 : Đồ thị chuỗi doanh thu thuần của GMD giai đoạn 2010 - 2024.

Các mô hình để dự báo cho doanh thu của CTCP Gemadept.

Dữ liệu về chuỗi doanh thu thuần của GMD gồm 60 quan sát được chia thành 2 tập là train set (từ quý 1 năm 2010 đến quý 4 năm 2023) và validation set (từ quý 1 năm 2024 đến quý 4 năm 2024). Ta sử dụng dữ liệu trong tập train set để dự báo 4 quý của năm 2024 cho tập validation set.

Mô hình	RMSE	MAPE	RMSE of 2024	MAPE of 2024
Linear - linear	138.8224	0.157557	309.149	0.2156405
Linear - log	132.3321	0.1479111	367.153	0.2669128
Log - linear	140.4618	0.1550435	300.1979	0.208309
Log log	133.0075	0.1448012	366.5778	0.2666646

Xu thế tuyến tính + mùa vụ dạng cộng	129.9238	0.1480687	296.3502	0.2257767
Xu thế tuyến tính + mùa vụ dạng nhân	132.2085	0.151284	288.0841	0.2271355
Holt – Winters dạng cộng	83.04816	0.08064433	226.2406	0.1669799
Holt – Winters dạng nhân	83.23577	0.07832165	239.3034	0.1808471

Bảng 3.1: Các mô hình dự báo cho chuỗi doanh thu của GMD

Từ bảng 2.1 cho thấy mô hình Holt – Winters dạng cộng có RMSE trên tập train set và validation set là nhỏ nhất. Trong khi đó, mô hình Holt – Winters dạng nhân có MAPE trên tập train set nhỏ nhất còn mô hình Holt – Winters dạng cộng có MAPE trên tập validation set nhỏ nhất. Vì vậy, ta có cơ sở để lựa chọn mô hình Holt – Winters dạng cộng để dự báo cho doanh thu của 4 quý trong năm 2025.

Kết hợp 2 tập train và validation lại và sử dụng mô hình Holt – Winters dạng cộng để dự báo doanh thu năm 2025 ta thu được kết quả:

Thời gian	Q1/2025	Q2/2025	Q3/2025	Q4/2025
Kết quả dự báo	1294.27	1365.50	1401.77	1487.05

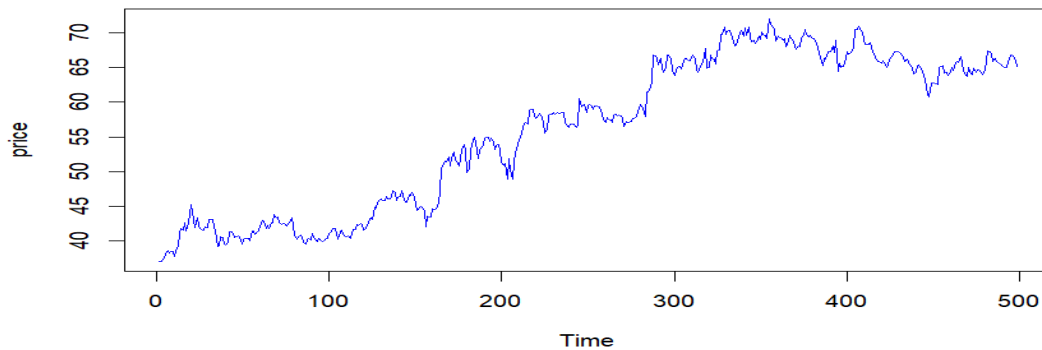
Bảng 3.2: Dự báo doanh thu thuần từ quý 1 đến quý 4 năm 2025 của GMD

3. Phân tích chuỗi giá cổ phiếu và log return theo ngày.

a. Mô hình với chuỗi giá cổ phiếu.

Kiểm định tính dừng của chuỗi giá cổ phiếu GMD.

Số liệu: Giá đóng cửa của mã chứng khoán GMD từ năm 2023 – 2024 (số liệu tính theo ngày, bao gồm 499 quan sát, đơn vị: nghìn VNĐ).



Hình 3.2: Đồ thị chuỗi giá cổ phiếu GMD.

Kiểm định ADF có xu thế (TREND):

Giả thuyết

H_0 : Chuỗi có nghiệm đơn vị (Chuỗi không dừng)

H_1 : Chuỗi không có nghiệm đơn vị (Chuỗi dừng vấp quanh xu thế)

Ta thu được kết quả:

	Estimate	Std. Error	t-value	p-value
Intercept	0.85074	0.38374	2.217	0.0271
z.lag.1	-0.01891	0.00972	-1.946	0.0523
tt	0.001072	0.000738	1.453	0.1468
z.diff.lag	-0.02480	0.04520	-0.549	0.5835
F-statistic: 1.711 on 3 and 493 DF, p-value: 0.1637				
Value of test-statistic is: -1.9457 2.199 2.278				
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				

Bảng 3.3: Kiểm định Trend

Với mức ý nghĩa 5%, $p\text{-value} = 0.0271 > 0.05$, cho thấy hệ số chặn không có ý nghĩa thống kê, do vậy, cần kiểm định NONE.

Giả thuyết:

H_0 : Chuỗi có nghiệm đơn vị (Chuỗi không dừng)

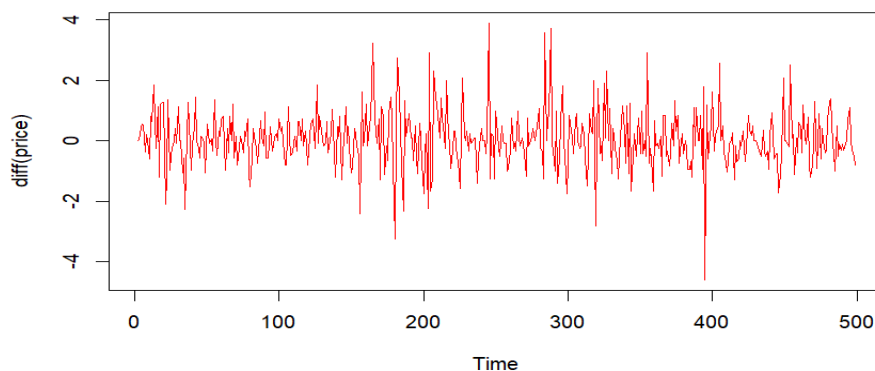
H_1 : Chuỗi không có nghiệm đơn vị (Chuỗi dừng vây quanh 0)

	Estimate	Std.Error	t-value	p-value
z.lag.1	0.0007945	0.0007222	1.100	0.272
z.diff.lag	-0.0332470	0.0450024	-0.739	0.460
Residual standard error: 0.9178 on 495 degrees of freedom				
Multiple R-squared: 0.003336, Adjusted R-squared: -0.0006911				
F-statistic: 0.8284 on 2 and 495 DF, p-value: 0.4374				
Value of test-statistic is: 1.1001				
Critical values for test statistics:				
1pct 5pct 10pct				
tau1 -2.58 -1.95 -1.62				

Bảng 3.4: Kiểm định NONE

Với mức ý nghĩa 5%, $|\tau_{qs}| = 0.9457 < |\tau_{0.05}| = 1.95$, nên chưa có cơ sở để bác bỏ H_0 , do đó có thể kết luận chuỗi giá GMD là không dừng.

Kiểm định tính dừng cho chuỗi sai phân của giá cổ phiếu GMD.



Hình 3.3: Đồ thị chuỗi sai phân giá cổ phiếu GMD.

Kiểm định tính dừng cho chuỗi sai phân trên ta được kết quả như sau:

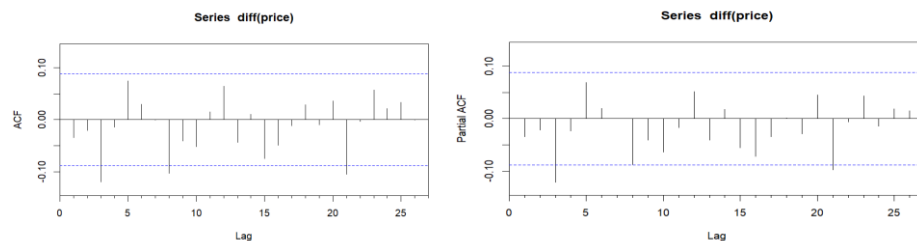
Value	Std.	Error	t-value	p-value
z.lag.1	-104.786	0.06466	-16.206	<2e-16
z.diff.lag	0.01732	0.04503	0.385	0.701
F-statistic: 262 on 2 and 494 DF, p-value: < 2.2e-16 Value of test-statistic is: -16.2063 Critical values for test statistics: 1pct 5pct 10pct tau1 -2.58 -1.95 -1.62				

Bảng 3.5 Kiểm định tính dừng cho chuỗi sai phân.

Như vậy có thể kết luận rằng chuỗi sai phân giá cổ phiếu GMD là chuỗi dừng.

Lựa chọn mô hình dự báo giá cổ phiếu GMD.

Để xác định bậc của mô hình ARMA ta cần xem xét đồ thị ACF và PACF của chuỗi sai phân giá cổ phiếu GMD.



Hình 3.4: Đồ thị ACF và PACF của chuỗi sai phân giá cổ phiếu GMD.

Qua 2 đồ thị trên ta có thể thấy tương quan giữa sai phân của giá GMD và các biến trễ ở thời kì thứ 3 và 8. Vì vậy, ta có thể đề xuất các mô hình ARIMA(1,1,1) (từ auto.arima); ARIMA(3,1,3) và ARIMA(8,1,8).

	ARIMA(3,1,3)	ARIMA(8,1,8)	ARIMA(1,1,1)
AIC	1324.61	1334.36	1326.99
RMSE	0.9008	0.8911	0.9087
MAPE(%)	1.1869	1.1788	1.1880

Xét theo tiêu chí AIC, ARIMA(3,1,3) là mô hình phù hợp nhất. Tuy nhiên mô hình ARIMA(8,1,8) lại có RMSE và MAPE nhỏ nhất cho thấy mô hình này có thể chính xác hơn về mặt dự báo.

Dự báo cho chuỗi giá cổ phiếu.

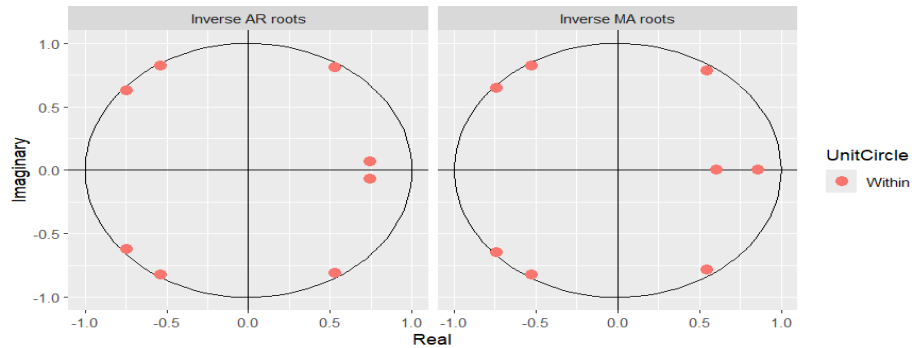
Sau khi sử dụng 2 mô hình là ARIMA(3,1,3) và ARIMA(8,1,8) để dự báo cho giá của cổ phiếu trong 10 phiên tiếp theo kết hợp với số liệu thực tế ta thu được kết quả như sau:

Phiên	Dự báo giá cổ phiếu		Giá cổ phiếu thực tế
	ARIMA(3,1,3)	ARIMA(8,1,8)	
1	65.47006	65.34552	66.2
2	65.56726	65.50256	65.8
3	65.46821	65.37566	64.1
4	65.23659	65.05748	62.7
5	65.10506	65.02492	63.5
6	65.17379	65.01357	63.5
7	65.37990	65.31354	62.0
8	65.52900	65.54113	61.8
9	65.49372	65.39196	60.0
10	65.31226	65.34124	60.2
RMSE	3.097038	3.051248	
MAPE	0.04193374	0.04120276	

Bảng 3.6: So sánh kết quả dự báo của các mô hình cho chuỗi giá cổ phiếu GMD

Từ bảng trên, ta thấy mô hình ARIMA (8,1,8) là mô hình dự báo tốt hơn với RMSE và MAPE nhỏ hơn. Như vậy mô hình với AIC thấp hơn chưa chắc hiệu quả hơn về mặt dự báo.

- Kiểm tra mô hình.



Hình 3.5: Kiểm tra nghiệm nghịch đảo mô hình ARIMA (8,1,8)

Nhận xét: Mô hình có nghiệm nghịch đảo nằm trong vòng tròn đơn vị. Quá trình đó không có xu hướng và không dao động quá mức, tức là rung bình và phương sai không đổi, điều này cho thấy mô hình là hợp lý.

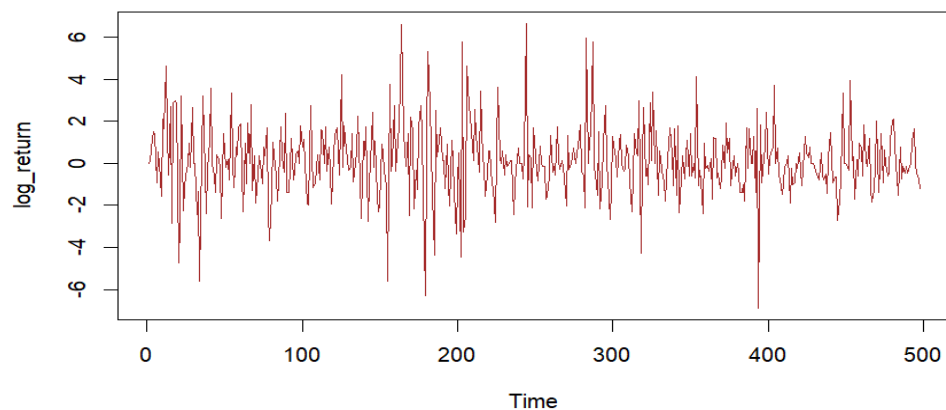
Ljung-Box test	
data: Residuals from ARIMA(8,1,8)	
$Q^* = 5.3448$, $df = 3$, $p\text{-value} = 0.1482$	
Model df: 16. Total lags used: 19	

Bảng 3.7: Kiểm định tự tương quan ARIMA (8,1,8).

Với mức ý nghĩa 5%, $pvalue = 0.1482 > 0.05 \rightarrow$ Chưa bác bỏ $H_0 \rightarrow$ không tồn tại hiện tượng tự tương quan trong phần dư. Điều này cho thấy mô hình là phù hợp

a. Mô hình với chuỗi lợi suất

Kiểm định tính dừng của chuỗi lợi suất.



Hình 3.6: Đồ thị chuỗi lợi suất(log return) của công ty GMD

Giả thuyết:

H_0 : Chuỗi có nghiệm đơn vị (Chuỗi không dừng)

H_1 : Chuỗi không có nghiệm đơn vị (Chuỗi dừng vây quanh 0)

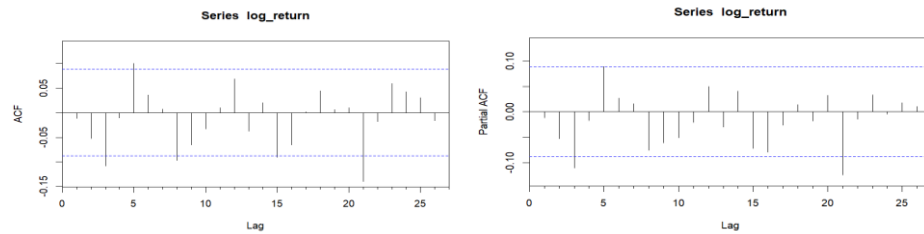
	Estimate	Std.Error	t-value	p-value
z.lag.1	-1.05428	0.06382	- 16.52	<2e-16
z.diff.lag	0.04765	0.04497	1.06	0.29
Residual standard error: 1.687 on 494 degrees of freedom Multiple R-squared: 0.504, Adjusted R-squared: 0.502 F-statistic: 251 on 2 and 494 DF, p-value: < 2.2e-16 Value of test-statistic is: -16.5204 Critical values for test statistics: 1pct 5pct 10pct tau1 -2.58 -1.95 -1.62				

Bảng 3.8: Kiểm định tính dừng cho chuỗi lợi suất.

Từ kết quả trên cho thấy chuỗi lợi suất(log return) của công ty GMD là chuỗi dừng.

Lựa chọn mô hình dự báo chuỗi lợi suất cổ phiếu GMD.

Để xác định bậc của mô hình ARMA ta cần xem xét đồ thị ACF và PACF của chuỗi lợi suất(log return) của GMD.



Hình 3.7: ACF và PACF cho chuỗi lợi suất.

Qua 2 đồ thị trên ta có thể thấy tương quan giữa sai phân của giá GMD và các biến trễ ở thời kì thứ 3 và 5. Vì vậy, ta có thể đề xuất các mô hình ARIMA(3,0,3); ARIMA(5,0,5)

	ARIMA(3,0,3)	ARIMA(5,0,5)
AIC	1923.11	1924.66
RMSE	1.640848	1.624614
SMAPE(%) (Symmetric MAPE)	1.608977	1.567315

Bảng 3.9: So sánh mô hình dự báo chuỗi lợi suất.

Ta thấy mô hình ARIMA(3,0,3) có chỉ số AIC nhỏ hơn còn mô hình ARIMA(5,0,5) có RMSE và MAPE nhỏ hơn.

Sau khi sử dụng 2 mô hình là ARIMA(3,0,3) và ARIMA(5,0,5) để dự báo cho lợi suất cổ phiếu trong 10 phiên tiếp theo kết hợp với số liệu thực tế ta thu được kết quả như sau:

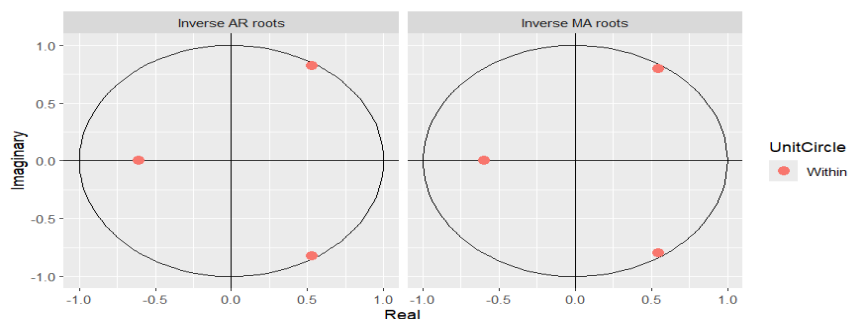
Phiên	Dự báo giá cổ phiếu		Lợi suất thực tế
	ARIMA(3,0,3)	ARIMA(5,0,5)	
1	0.5899	0.6241	1.5221
2	0.2717	0.3972	-0.6061
3	-0.1443	-0.4015	-2.6175
4	-0.3275	-0.1014	-2.2083
5	-0.0969	-0.2668	1.2678
6	0.3072	0.3137	0
7	0.5267	0.691	-2.3906

8	0.3673	0.1362	-0.3231
9	-0.0086	0.2062	-2.9559
10	-0.2584	-0.3873	0.3328
RMSE	1.768423	1.83957	
SMAPE	1.814551	1.812912	

Bảng 3.10: Dự báo chuỗi lợi suất .

Khi dự báo cho 10 phiên tiếp theo, mô hình ARIMA(3,0,3) tuy có SMAPE lớn hơn một chút nhưng RMSE lại thấp hơn khá đáng kể so với ARIMA(5,0,5) cùng với việc có AIC nhỏ hơn nên mô hình ARIMA(3,0,3) sẽ là mô hình phù hợp hơn để dự báo cho chuỗi lợi suất của cổ phiếu GMD.

Kiểm tra mô hình.



Hình 3.8 Kiểm tra nghiệm nghịch đảo mô hình ARIMA(3,0,3).

Nhận xét: Mô hình có nghiệm nghịch đảo nằm trong vòng tròn đơn vị. Quá trình đó không có xu hướng và không dao động quá mức, tức là rung bình và phương sai không đổi, điều này cho thấy mô hình là hợp lý.

<p>Ljung-Box test</p> <p>data: Residuals from ARIMA(3,0,3) with non-zero mean</p> <p>$Q^* = 3.5717$, $df = 4$, $p\text{-value} = 0.4671$</p> <p>Model df: 6. Total lags used: 10</p>
--

Với mức ý nghĩa 5%, $p\text{value} = 0.4671 > 0.05 \rightarrow$ Chưa bác bỏ $H_0 \Rightarrow$ không tồn tại hiện tượng tự tương quan trong phần dư. Điều này cho thấy mô hình là phù hợp.

IV. Đình Ngọc Thảo - Tổng Công ty Phân bón và Hóa chất Dầu khí

1. Tổng quan về công ty

Tổng Công ty Phân bón và Hóa chất Dầu khí là một trong những doanh nghiệp hàng đầu Việt Nam trong lĩnh vực sản xuất và kinh doanh phân bón và hóa chất. Là đơn vị thành viên của Tập đoàn Dầu khí Quốc gia Việt Nam, DPM sở hữu và vận hành Nhà máy Đạm Phú Mỹ – nhà máy sản xuất phân đạm urê lớn và hiện đại. Ngoài sản phẩm chủ lực là Đạm Phú Mỹ, công ty còn cung cấp các sản phẩm phân bón khác như NPK, Kali, SA, DAP, và các loại hóa chất phục vụ nông nghiệp và công nghiệp. Với mạng lưới phân phối rộng khắp cả nước và xuất khẩu ra thị trường quốc tế, DPM đóng vai trò quan trọng trong việc đảm bảo an ninh lương thực và phát triển bền vững ngành nông nghiệp Việt Nam.

2. Phân tích chuỗi doanh thu thuần theo quý

Sử dụng dữ liệu đến hết quý 3 năm 2024, dự báo doanh thu từ quý 4 năm 2024 đến quý 3 năm 2025

Sử dụng các mô hình là:

1. Mô hình linear-linear:

$$Revenue_t^F = 1936397 + 22921t$$

2. Mô hình linear-log:

$$Revenue_t^F = 1479478 + 365944 \ln t$$

3. Mô hình log-linear:

$$\ln(Revenue_t^F) = 14.494727 + 0.007787t$$

4. Mô hình log-log:

$$\ln(Revenue_t^F) = 14.30796 + 0.1344 \ln(t)$$

5. Mô hình xu thế tuyến tính dạng cộng:

$$Revenue_t^F = 1837073 + 23033t + 434201s_2 - 127544s_3 + 75883s_4$$

6. Mô hình xu thế tuyến tính dạng nhân:

$$Revenue_t^F = 1768776 + 25388t - 6129t * s_2 - 5255t * s_3 + 2960t * s_4$$

STT	R-squared	Toàn dữ liệu		4 quan sát cuối		Dự báo			
		RMSE	MAPE	RMSE	MAPE	Q4:2024	Q1:2025	Q2:2025	Q3:2025
1	0.1824	826366.8	0.2387738	367424.1	0.07722168	3311679	3334600	3357522	3380443

Bài tập lớn Phân tích chuỗi thời gian và dự báo

2	0.1276	853619.3	0.2450729	566041.7	0.1288417	2977779	2983828	2989778	2995633
3	0.1798	829035.4	0.2227881	471054.2	0.09960655	3146945	3171545	3196338	3221325
4	0.147	860204.4	0.2296787	687228.7	0.1704201	2836994	2843304	2849524	2855659
5	0.2354	799163.5	0.2269898	207848.9	0.04739009	3219064	3676298	3137586	3364046
6	0.24	796729.8	0.2308736	232963.1	0.04547852	3292071	3559310	3047695	3539442

Sử dụng dữ liệu đến hết quý 4 năm 2023, dự báo cho năm 2024:

1. Mô hình Holt-Winter dạng cộng:

$$Revenue_{T+k}^F = 3319137.28 + 69952.34k + S_Q$$

2. Mô hình Holt-Winter dạng nhân:

$$Revenue_{T+k}^F = (3413471 + 72495.34k)S_Q$$

STT	RMSE 2024	Dự báo cho 2024			
		Q1	Q2	Q3	Q4
1	280911.6	3363141	3734508	3216793	3661630
2	264541.3	3431617	4038474	3042989	3669114

Mô hình Holt-Winter dạng nhân có RMSE nhỏ hơn, suy ra mô hình dự báo sát với thực tế của năm 2024 hơn.

Sử dụng mô hình Holt-Winter dạng nhân và dữ liệu đến hết năm 2024, dự báo cho năm 2025

Q1	Q2	Q3	Q4
3224034	3802060	2870120	3387700

3. Phân tích số liệu giá cổ phiếu theo ngày và log-return

3.1. Phân tích chuỗi giá cổ phiếu theo ngày

a. Kiểm định tính dừng của chuỗi giá

Thực hiện kiểm định ADF với chuỗi giá thu được kết quả:

Test	τ_{stat}	$\tau_{0.05}$
With trend	-3.5982	-3.42
With drift	-2.7158	-2.87
None	-0.2543	-2.58

Bảng trên cho thấy kiểm định ADF with drift và none thu được $\tau_{stat} > \tau_{0.05}$, suy ra chưa bác bỏ giả thiết H_0 tức chuỗi giá không dừng ở mức ý nghĩa 5%.

Tiếp tục lấy sai phân bậc 1 chuỗi giá và thực hiện kiểm định ADF thu được:

Test	τ_{stat}	$\tau_{0.05}$
With trend	-16.0947	-3.42
With drift	-16.0925	-2.87
None	-16.1089	-1.95

Bảng trên cho thấy tất cả các giá trị τ_{stat} đều nhỏ hơn $\tau_{0.05}$, suy ra bác bỏ giả thiết ở H_0 ở mức 5% tức sai phân bậc 1 của chuỗi giá dừng.

b. Lựa chọn mô hình ARIMA

Kết quả đánh giá một số mô hình ARIMA cho sai phân bậc 1 của chuỗi giá

ARIMA	AIC	Log-likelihood	RMSE
1,0,1	897.29	-444.65	0.5909201
1,0,2	899.28	-444.64	0.5909126

2,0,2	891.72	-439.86	0.585119
2,0,3	892.52	-439.26	0.5844047
3,0,2	892.37	-439.19	0.5843154

Theo một số tiêu chí đánh giá ở trên, có thể thấy mô hình ARIMA(2,0,2) có AIC thấp nhất, log-likelihood và RMSE cao hơn 2 mô hình ARIMA(2,0,3) và ARIMA(3,0,2) nhưng không đáng kể. Có thể kết luận mô hình ARIMA(2,0,2) với chuỗi giá hay ARIMA(2,1,2) với sai phân bậc 1 của chuỗi giá là cân bằng giữa hiệu quả dự đoán, độ phù hợp với dữ liệu và mức độ đơn giản.

c. Đánh giá mô hình:

- Sử dụng vòng tròn đơn vị:

Tất cả nghiệm nghịch đảo đều nằm trong vòng tròn đơn vị, vì vậy kết luận rằng chuỗi sai phân là chuỗi dừng.

- Kiểm định tính nhiễu trắng của phần dư:

Thực hiện kiểm định Ljung-Box thu được kết quả là $p - value = 0.7171 > 0.05$ với số bậc tự do là 6 \Rightarrow Chưa đủ cơ sở bác bỏ giả thiết $H_0 \Rightarrow$ Phần dư là nhiễu trắng

d. Kết quả dự báo cho 10 giá trị tiếp theo của chuỗi giá:

Ngày	1,1,1	1,1,2	2,1,2	2,1,3	3,1,2	Giá thực tế
02/01/2025	35.01408	35.01471	34.95601	34.97010	34.97182	35.4
03/01/2025	35.01628	35.01879	34.95575	34.96994	34.97020	35.55
06/01/2025	35.01663	35.01864	34.99792	35.02169	35.02254	35
07/01/2025	35.01668	35.01865	35.00629	35.03103	35.03180	34.4
08/01/2025	35.01669	35.01865	34.96742	34.98320	34.98355	34.55
09/01/2025	35.01669	35.01865	34.95190	34.96580	34.96617	34.15
10/01/2025	35.01669	35.01865	34.98623	35.00843	35.00917	33.5
13/01/2025	35.01669	35.01865	35.00773	35.03264	35.03337	34.1
14/01/2025	35.01669	35.01865	34.97891	34.99617	34.99659	34
15/01/2025	35.01669	35.01865	34.95274	34.96656	34.96698	34

Tính được RMSE của các mô hình khi so với giá thực tế:

	1,1,1	1,1,2	2,1,2	2,1,3	3,1,2
RMSE	0.8361393	0.8374587	0.814492	0.8273381	0.8276651

Ta thấy được mô hình ARIMA(2,1,2) có RMSE nhỏ nhất \Rightarrow Mô hình tốt nhất để dự báo cho chuỗi giá.

3.2. Phân tích chuỗi lợi suất của cổ phiếu

Chuỗi lợi suất được tính từ chuỗi giá với công thức: $R_i = \ln\left(\frac{P_i}{P_{i-1}}\right) * 100$

a. Kiểm định tính dừng của chuỗi lợi suất

Test	τ_{stat}	$\tau_{0.05}$
With trend	-16.145	-3.42
With drift	-16.1425	-2.87
None	-16.1589	-1.95

Bảng trên cho thấy tất cả các giá trị τ_{stat} đều nhỏ hơn $\tau_{0.05}$, suy ra chưa đủ cơ sở bác bỏ giả thiết H_0
 \Rightarrow Chuỗi lợi suất dừng ở mức ý nghĩa 5%.

b. Lựa chọn mô hình ARIMA

Kết quả đánh giá một số mô hình ARIMA cho chuỗi lợi suất

ARIMA	AIC	Log-likelihood	RMSE
2,0,2	1997.69	-992.85	1.776398
2,0,3	1997.32	-991.66	1.77193
3,0,2	1999.62	-992.81	1.776269
3,0,4	1999.34	-990.67	1.768417
4,0,3	1999.59	-990.8	1.768873

Từ bảng trên có thể thấy mô hình ARIMA(2,0,3) có AIC thấp nhất, RMSE và log-likelihood ở mức giữa \Rightarrow Mô hình cân bằng giữa hiệu quả dự đoán, AIC thấp và độ đơn giản.

c. Đánh giá mô hình

Bài tập lớn Phân tích chuỗi thời gian và dự báo

- Sử dụng vòng tròn đơn vị:

Tất cả nghiệm nghịch đảo đều nằm trong vòng tròn đơn vị, vì vậy kết luận rằng chuỗi lợi suất là chuỗi dừng.

- Kiểm định tính nhiễu trắng của phần dư:

Thực hiện kiểm định Ljung-Box thu được kết quả là $p - value = 0.6819 > 0.05$ với số bậc tự do là 5 \Rightarrow Chưa đủ cơ sở bác bỏ giả thiết $H_0 \Rightarrow$ Phần dư là nhiễu trắng

d. Kết quả dự báo cho 10 giá trị tiếp theo của chuỗi lợi suất:

Ngày	2,0,2	2,0,3	3,0,2	3,0,4	4,0,3	Lợi suất thực tế
02/01/2025	0.0549	-0.0832	0.0438	-0.0869	-0.0952	1.1364
03/01/2025	0.1294	-0.0037	0.1343	-0.0667	-0.0578	0.4228
06/01/2025	0.0053	0.1471	0.0034	0.1205	0.1270	-1.5592
07/01/2025	-0.1124	0.0335	-0.1156	-0.0116	-0.0087	-1.5592
08/01/2025	-0.0482	-0.1316	-0.0465	-0.1490	-0.1437	0.4351
09/01/2025	0.0842	-0.0527	0.0881	-0.0287	-0.0325	-1.1645
10/01/2025	0.0800	0.1198	0.0808	0.1313	0.1244	-1.9217
13/01/2025	-0.0412	0.0755	-0.0425	0.0527	0.0535	1.7752
14/01/2025	-0.0867	-0.0979	-0.0861	-0.1110	-0.1059	-0.2937
15/01/2025	0.0042	-0.0868	0.0066	-0.0685	-0.0690	0.0000

Tính được RMSE của các mô hình khi so với lợi suất thực tế:

	2,0,2	2,0,3	3,0,2	3,0,4	4,0,3
RMSE	1.243331	1.278563	1.244213	1.278994	1.27899

Tuy đã nêu trên rằng mô hình ARIMA(2,0,3) là mô hình cân bằng nhất nhưng khi áp dụng với dữ liệu thực tế thì lại không hiệu quả bằng hai mô hình ARIMA(2,0,2) và ARIMA(3,0,2) tỏ ra có phần hạn chế hơn theo bảng trên.

B. BÀI TẬP NHÓM

1. Phân tích tương quan

	<i>TSC</i>	<i>VTO</i>	<i>GMD</i>	<i>DPM</i>
TSC	1			
VTO	-0.64154	1		
GMD	-0.63899	0.832392	1	
DPM	-0.20598	0.546636	0.520607	1

Bảng 3.1 tương quan giữa giá của 4 cổ phiếu

	<i>logTSC</i>	<i>logVTO</i>	<i>logGMD</i>	<i>logDPM</i>
logTSC	1			
logVTO	0.248494	1		
logGMD	0.281584	0.266947	1	
logDPM	0.312299	0.350616	0.425844	1

Bảng 3.2 tương quan giữa lợi suất các cổ phiếu

2. Xây dựng danh mục tối ưu

a. Xây dựng danh mục tối ưu với phương sai danh mục nhỏ nhất.

Bài toán: Xây dựng một danh mục gồm 4 cổ phiếu TSC,VTO,GMD,DPM sao cho phương sai của danh mục là nhỏ nhất với điều kiện lợi suất kỳ vọng phải lớn hơn hoặc bằng lợi suất trung bình thị trường.

Lợi suất kỳ vọng của một danh mục đầu tư được tính theo công thức:

$$E(R_p) = \sum_{i=1}^n w_i \cdot E(R_i)$$

Trong đó:

- $E(R_p)$: là lợi nhuận kỳ vọng của một danh mục đầu tư
- w_i : trọng số của cổ phiếu thứ i trong danh mục.
- $E(R_i)$: là lợi nhuận kỳ vọng của tài sản thứ i trong danh mục.
- Phương sai danh mục được tính theo công thức: $Var(R_p) = \sum_{i=1}^n \sum_{j=1}^n w_i \cdot w_j \cdot Cov(R_i, R_j)$
- $Cov(R_i, R_j)$: hiệp phương sai giữa lợi nhuận của cổ phiếu thứ i và thứ j .

Sau khi thực hiện tính toán ta thu được kết quả sau:

	TSC	VTO	GMD	DPM
TSC	5.639894	1.263144	1.123863	1.330435
VTO	1.263144	4.581437	0.960277	1.346234
GMD	1.123863	0.960277	2.824498	1.283834
DPM	1.330435	1.346234	1.283834	3.21792

Bảng 3.3: Ma trận hiệp phương sai của lợi suất 4 cổ phiếu.

wTSC	wVTO	wGMD	wDPM
0.135278878	0.196961	0.401603	0.266158

Bảng 3.4: Tỷ trọng tối ưu của danh mục đầu tư với phương sai nhỏ nhất.

Từ đó, ta có thể tính được phương sai danh mục tối ưu là $Var(R_p) = 1.817200212$ với lợi suất kì vọng $E(R_p) = 0.069749681$.

b. Xây dựng danh mục tối ưu với lợi suất kì vọng lớn nhất.

Bài toán: Xây dựng một danh mục gồm 4 cổ phiếu TSC,VTO,GMD,DPM sao cho lợi suất kỳ vọng của danh mục là lớn nhất nhất với điều kiện lợi phương sai danh mục nhỏ hơn hoặc phương sai trung bình của 4 cổ phiếu (tính theo trung bình nhân).

Sau khi thực hiện tính toán ta thu được kết quả như sau:

wTSC	wVTO	wGMD	wDPM
0	0.900428	0.099572	0

Bảng 3.5: Tỷ trọng tối ưu của danh mục đầu tư với lợi suất kì vọng lớn nhất.

Từ đây ta có thể tính được lợi suất kì vọng của danh mục tối ưu là $E(R_p) = 0.159712$ với phương sai là $Var(R_p) = 3.914689$.

3. Phân tích đồng tích hợp

Đồng tích hợp là hiện tượng hai hay nhiều chuỗi thời gian có tồn tại một số nguyên nhân gốc làm chúng có xu hướng biến động tương đồng nhau về dài hạn ngay cả khi nếu nhìn riêng lẻ thì chúng chỉ là bước ngẫu nhiên. Do các chuỗi mà nhóm sử dụng đều là chuỗi không dừng, trong khi đó lại dừng ở sai phân bậc nhất, nên cần kiểm định khả năng xảy ra các vector đồng tích hợp giữa các

chuỗi này. Kỹ thuật kiểm định đồng tích hợp được áp dụng là phương pháp của Johansen và Juselius được đề xuất vào năm 1990. Đây là kỹ thuật kiểm định đồng tích hợp phổ biến nhất trong việc áp dụng nguyên tắc cực đại hợp lý, nhằm xác định sự tồn tại của các vector đồng tích hợp giữa các chuỗi thời gian không dừng. Phương pháp này có ưu điểm là có thể giúp xác định số lượng các vector đồng tích hợp và cho phép nhà nghiên cứu có thể kiểm định nhiều giả thuyết khác nhau liên quan tới các phần tử vector. Số lượng mối quan hệ đồng tích hợp được tìm thấy sẽ tương ứng với số mối quan hệ dài hạn trong mô hình.

Kiểm đồng tích hợp của Johansen với tiêu chí “trace” với các cặp giả thuyết:

- $r=0$: $H_0:r=0$, không có mối quan hệ đồng tích hợp
 $H_1:r > 0$, có ít nhất 1 mối quan hệ đồng tích hợp
- $r \leq 1$: { $H_0:r \leq 1$, có 1 mối quan hệ đồng tích hợp
 $H_1:r > 1$, có nhiều hơn 1 mối quan hệ đồng tích hợp
- $r \leq 2$: { $H_0:r \leq 2$, có 2 mối quan hệ đồng tích hợp
 $H_1:r > 2$, có nhiều hơn 2 mối quan hệ đồng tích hợp

Sau khi thực hiện kiểm định thu được bảng kết quả:

Test type: trace statistic , without linear trend and constant in cointegration

Eigenvalues (lambda):

[1] 3.924193e-02 3.605806e-02 1.468083e-02 6.132501e-03 3.469447e-17

Values of teststatistic and critical values of test:

	test	10pct	5pct	1pct
$r \leq 3$		3.04	7.52	9.24 12.97
$r \leq 2$		10.37	17.85	19.96 24.60
$r \leq 1$		28.54	32.00	34.91 41.07
$r = 0$		48.36	49.65	53.12 60.16

Bảng 3.6 Kiểm định đồng tích hợp

Ở tất cả các mức r , giá trị thống kê kiểm định đều nhỏ hơn các ngưỡng tới hạn. Do đó, ta không bác bỏ giả thuyết H_0 ở bất kỳ mức nào. Như vậy không có mối quan hệ đồng tích hợp nào giữa các chuỗi giá.

4. Mô hình VAR

Mô hình VAR là một mô hình thống kê miêu tả mối quan hệ giữa nhiều chuỗi thời gian khi chúng thay đổi theo thời gian. Không như các mô hình Univariate như Arima, Sarima,... chỉ xét tác động của các trễ lên chính nó. VAR không những xét tác động của trễ chính nó mà còn thêm trễ của các chuỗi thời gian khác. Vậy so sánh với Arima và các mô hình Univariate VAR có điểm gì vượt trội? Vì là một mô hình hồi quy đa biến VAR đưa ra mối quan hệ giữa các biến thời gian với nhau giúp ta

đánh giá được tác động qua lại giữa chúng, điều mà mô hình đơn biến không làm được. Hơn nữa, theo đánh giá thực tế mô hình VAR thường đưa ra dự báo tốt hơn các mô hình đơn biến, tuy nhiên đòi hỏi quy mô mô hình VAR đòi hỏi nhiều nguồn lực hơn do nhu cầu về dữ liệu. Sau khi các bài cá nhân đã sử dụng mô hình đơn biến để dự báo chuỗi giá cổ phiếu, thì đến phần này chúng tôi sử dụng mô hình VAR với ba chuỗi giá để đánh giá tác động qua lại giữa các chuỗi, cũng như đánh giá khả năng dự báo của VAR so với các mô hình đơn biến trên. Ở phần này nhóm 2 sẽ sử dụng chuỗi log return của 4 mã cổ phiếu để xây dựng mô hình.

a. Ước lượng mô hình VAR

Sau khi sử dụng các phương pháp để xem xét nhóm tôi đã đi đến kết luận mô hình VAR có bậc 2 là phù hợp.

Sử dụng R để ước lượng mô hình với chuỗi log return của 4 mã cổ phiếu ta thu được kết quả như sau:

Estimation results for equation logTSC:

$$\begin{aligned} \log TSC = & \log TSC.11 + \log VTO.11 + \log GMD.11 + \log DPM.11 \\ & + \log TSC.12 + \log VTO.12 + \log GMD.12 + \log DPM.12 \end{aligned}$$

	Estimate	Std. Error	t value	Pr(> t)
logTSC.11	0.12604	0.04845	2.601	0.00957 **
logVTO.11	-0.07199	0.05380	-1.338	0.18152
logGMD.11	0.06621	0.07055	0.939	0.34841
logDPM.11	0.14096	0.06848	2.058	0.04009 *
logTSC.12	-0.12932	0.04781	-2.705	0.00707 **
logVTO.12	0.06600	0.05329	1.239	0.21612
logGMD.12	0.02122	0.07024	0.302	0.76272
logDPM.12	-0.02264	0.06869	-0.330	0.74185

Bảng 3.7a: Kết quả mô hình VAR với biến logTSC.

Estimation results for equation logVTO:

$$\begin{aligned} \log VTO = & \log TSC.11 + \log VTO.11 + \log GMD.11 + \log DPM.11 + \\ & \log TSC.12 + \log VTO.12 + \log GMD.12 + \log DPM.12 \end{aligned}$$

	Estimate	Std. Error	t value	Pr(> t)
logTSC.11	-0.02453	0.04419	-0.555	0.579
logVTO.11	0.02169	0.04907	0.442	0.659
logGMD.11	0.01516	0.06434	0.236	0.814

logDPM.11	0.09785	0.06246	1.567	0.118
logTSC.12	0.03342	0.04360	0.766	0.444
logVTO.12	-0.06631	0.04860	-1.364	0.173
logGMD.12	0.05451	0.06406	0.851	0.395
logDPM.12	-0.02212	0.06265	-0.353	0.724

Bảng 3.7b: Kết quả mô hình VAR với biến logVTO.

Estimation results for equation logGMD:

$$\log GMD = \log TSC.11 + \log VTO.11 + \log GMD.11 + \log DPM.11 + \log TSC.12 + \log VTO.12 + \log GMD.12 + \log DPM.12$$

	Estimate	Std. Error	t value	Pr(> t)
logTSC.11	-0.01706	0.03507	-0.486	0.6269
logVTO.11	-0.02776	0.03895	-0.713	0.4764
logGMD.11	0.02076	0.05107	0.406	0.6846
logDPM.11	-0.02715	0.04957	-0.548	0.5841
logTSC.12	-0.05924	0.03461	-1.712	0.0876
logVTO.12	0.04327	0.03858	1.121	0.2626
logGMD.12	-0.04677	0.05084	-0.920	0.3581
logDPM.12	0.02524	0.04972	0.508	0.6120

Bảng 3.7c: Kết quả mô hình VAR với biến logGMD.

Estimation results for equation logDPM:

$$\log DPM = \log TSC.11 + \log VTO.11 + \log GMD.11 + \log DPM.11 + \log TSC.12 + \log VTO.12 + \log GMD.12 + \log DPM.12$$

	Estimate	Std. Error	t value	Pr(> t)
logTSC.11	-0.004446	0.037266	-0.119	0.905
logVTO.11	-0.062197	0.041381	-1.503	0.133
logGMD.11	0.024308	0.054261	0.448	0.654
logDPM.11	-0.009396	0.052672	-0.178	0.858
logTSC.12	0.045094	0.036772	1.226	0.221
logVTO.12	0.003211	0.040990	0.078	0.938
logGMD.12	0.048544	0.054022	0.899	0.369

logDPM.12 -0.044088 0.052831 -0.835 0.404

Bảng 3.7d: Kết quả mô hình VAR với biến logDPM.

b. Kiểm định tự tương quan phần dư.

Thực hiện kiểm định Portmanteau Test trong R ta được kết quả như sau:

Portmanteau Test (asymptotic)

data: Residuals of VAR object var

Chi-squared = 275.58, df = 224, p-value = 0.01065

Bảng 3.8: Kiểm định tự tương quan phần dư.

$P\text{-value} = 0.01065 > 0.01$. Với mức ý nghĩa 1%, chưa bác bỏ giả thiết H_0 , các phần dư không xảy ra hiện tượng tự tương quan do đó mô hình VAR (2) là mô hình tốt để dự báo.

a. Dự báo

Kết quả dự báo được đưa ra như sau:

Kỳ	logTSC	logVTO	logGMD	logDPM
1	-0.2267977	-0.2630961	0.1621778	-0.06232195
2	0.1540306	-0.08300739	0.1137065	-0.036307
3	0.04461633	0.01267719	-0.004082665	0.007131611
4	-0.01671778	0.01746938	-0.02034126	0.01264663
5	-0.008110005	0.001988346	-0.002688958	-0.000085491
6	0.001241696	-0.002912634	0.003046292	-0.002394954
7	0.001355202	-0.000829379	0.00087798	-0.0002139
8	0.000024527	0.000394796	-0.000378547	0.00036902
9	-0.000204904	0.000191209	-0.000191878	0.000073166
10	-0.000035481	-0.00004073	0.000034863	-0.000048605

Bảng 3.9: Kết quả dự báo chuỗi log return của mô hình VAR

So sánh với mô hình ARIMA ở phần bài tập cá nhân ta có kết quả như sau:

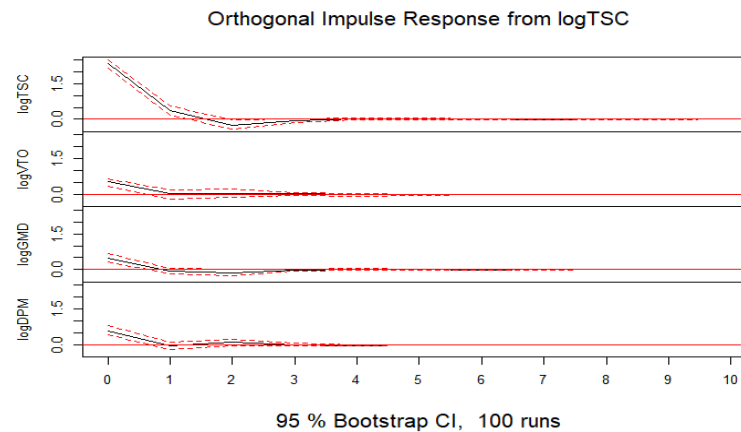
	logTSC		logVTO		logGMD		logDPM	
	arima	var	arima	var	arima	var	arima	var
RMSE	1.645	0.091	1.9441	0.5277	1.76842	1.7397	1.2785	1.2263
MAPE (%)	1.804	1.885	1.4414	1.9594	1.81455	1.9566	1.8664	1.9990

Bảng 3.10: So sánh RMSE và MAPE của mô hình ARIMA và VAR.

Nhìn chung mô hình VAR tuy sử dụng nhiều dữ liệu hơn nhưng lại chưa cho thấy hiệu quả vượt trội rõ rệt so với mô hình ARIMA, điều này có thể vì các công ty ở những nhóm ngành khác nhau hoặc các phân khúc khác nhau trong cùng ngành.

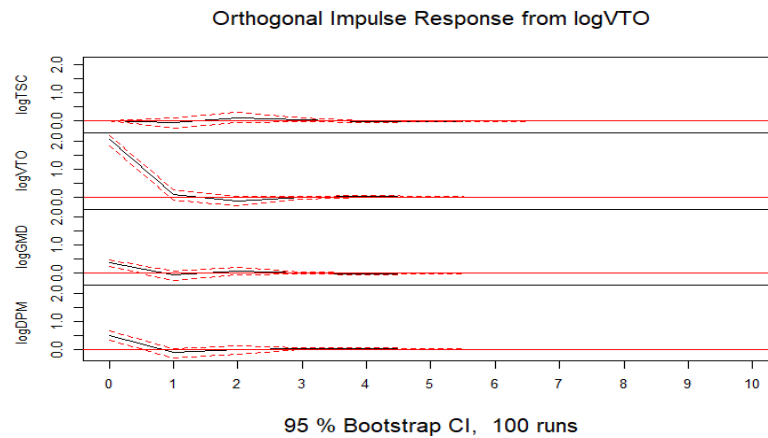
c. Hàm phản ứng

- Hàm phản ứng của TSC



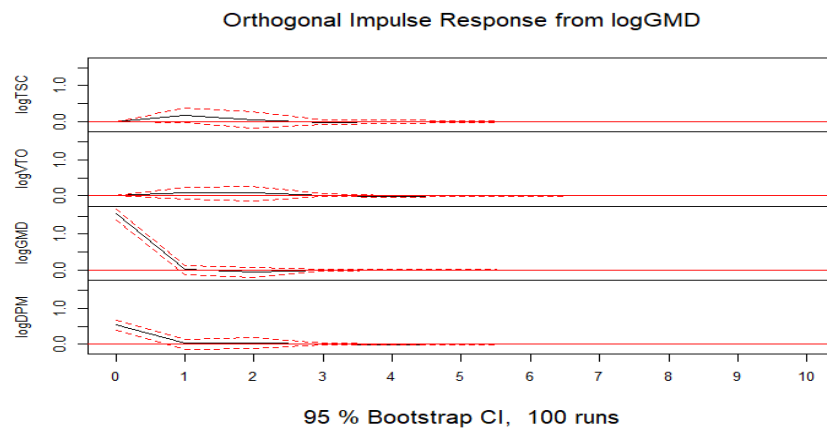
Nhận xét: Cú sốc vào logTSC làm chính nó tăng mạnh lúc đầu, sau đó giảm dần về 0. Ảnh hưởng từ logTSC đến 3 cổ phiếu còn lại đều có xu hướng giảm dần và tất hẳn cho thấy ảnh hưởng là tạm thời và rất yếu.

- Hàm phản ứng của VTO



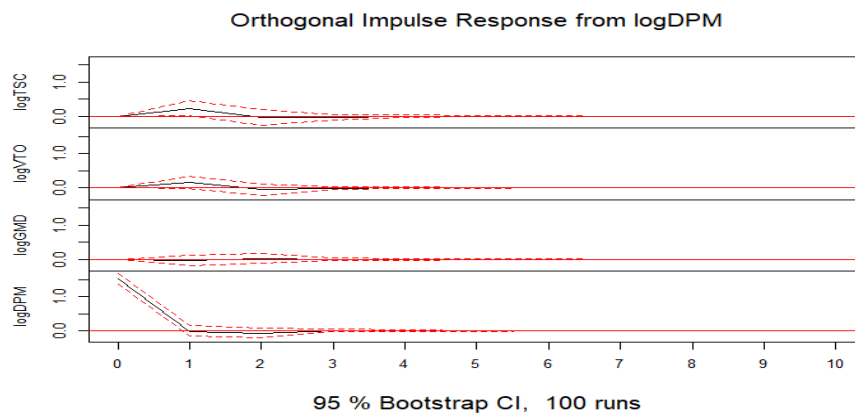
Nhận xét: Cú sốc vào logVTO tác động mạnh vào chính nó trong ngắn hạn, tiêu biến sau 3 chu kỳ. Ảnh hưởng từ logVTO đến 3 cổ phiếu còn lại đều rất yếu.

- Hàm phản ứng của GMD



Nhận xét: Sốc từ logGMD ảnh hưởng chính nó, ảnh hưởng lan tỏa không mạnh và ngắn hạn. Nó có hiệu ứng nhẹ lên logTSC và logVTO ở chu kỳ đầu tiên, nhưng CI nhanh chóng bao trùm 0.

- Hàm phản ứng của DPM



Nhận xét: LogDPM tự ảnh hưởng đến nó ở những chu kì đầu và nhanh chóng triệt tiêu. Shock từ logDPM hầu như không ảnh hưởng đáng kể đến các chuỗi còn lại.

d. Phân rã phương sai

LogTSC

- Gần như hoàn toàn tự giải thích (98.3%).
- Ba chuỗi log return còn lại có ảnh hưởng rất nhỏ (mỗi cái chỉ dưới 1%).
- \Rightarrow logTSC là biến chủ động, gần như không bị ảnh hưởng bởi những biến còn lại.

Log VTO

- Chủ yếu chịu ảnh hưởng rất lớn từ chính nó (93%).
- Chịu ảnh hưởng từ logTSC (6%) là đáng kể nhất trong các biến khác kinh.
- Ảnh hưởng từ giá đầu vào (logGMD) và chi phí (logDPM) rất thấp.

LogGMD

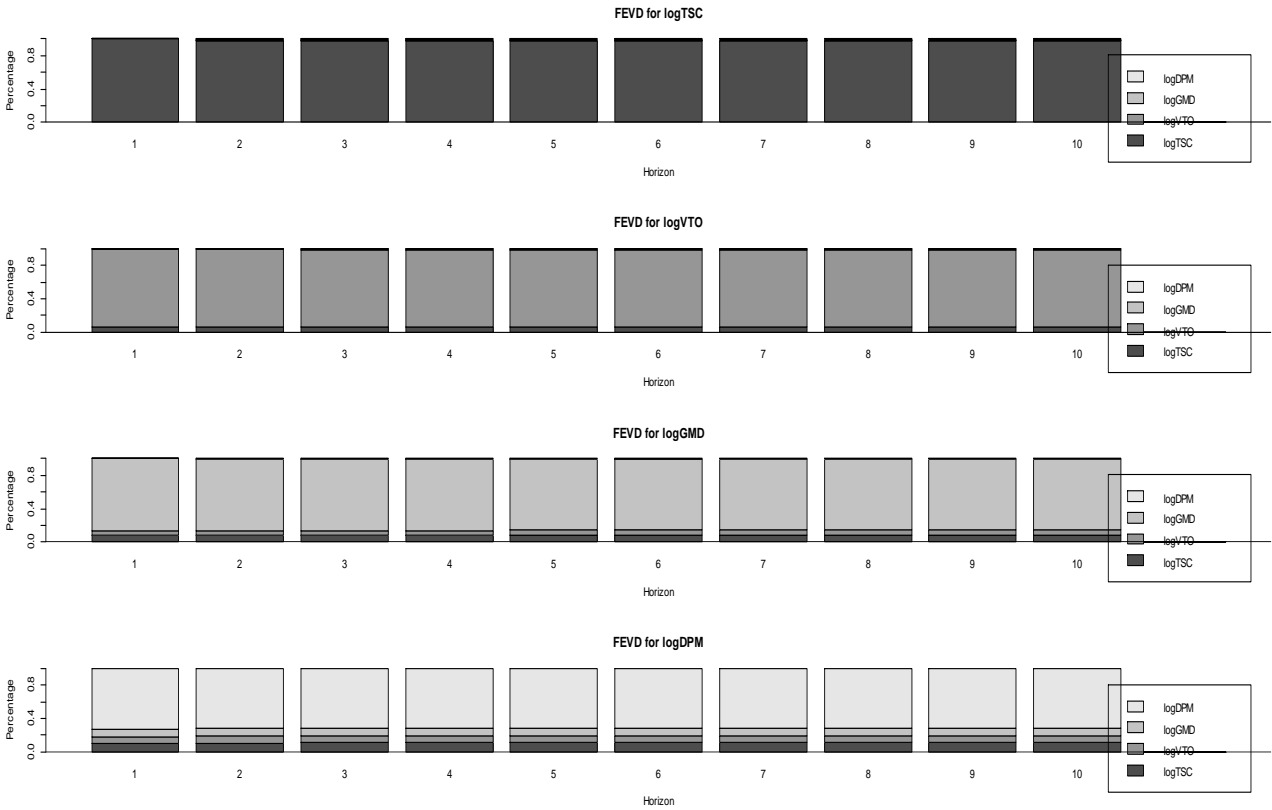
- logGMD chịu tác động nhiều nhất từ chính nó (85.7%), nhưng ảnh hưởng từ logTSC (8.7%) và logVTO (5.3%) cũng không nhỏ.
- logDPM hầu như không ảnh hưởng đến logGMD.

LogDPM

Đây là biến có sự phân rã sai số đa dạng nhất:

- logDPM: 71.9%
- logTSC: 11%
- logVTO: 8%
- logGMD: 9%

Bài tập lớn Phân tích chuỗi thời gian và dự báo



C. PHỤ LỤC

1. Nguyễn Bình An

Mo hình Lin Lin

##

Call:

lm(formula = rev1 ~ t1)

##

Residuals:

Min 1Q Median 3Q Max

-314.67 -85.70 -15.88 66.44 382.02

##

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 554.303 44.297 12.513 < 2e-16 ***

t1 -9.331 1.352 -6.902 5.94e-09 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##

Residual standard error: 163.5 on 54 degrees of freedom

Multiple R-squared: 0.4687, Adjusted R-squared: 0.4589

F-statistic: 47.64 on 1 and 54 DF, p-value: 5.942e-09

[1] 160.5813

[1] 0.5835454

1 2 3 4

22.41 13.08 3.75 -5.58

[1] 150.5592

[1] 0.9392399

Mo hình Lin Log

##

Bài tập lớn Phân tích chuỗi thời gian và dự báo

Call:

```
## lm(formula = rev1 ~ log(t1))
```

##

Residuals:

```
##   Min     1Q  Median     3Q      Max
```

```
## -382.31 -70.71 -14.67  52.55 381.51
```

##

Coefficients:

```
##           Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  835.31      75.38  11.081 1.59e-15 ***
```

```
## log(t1)     -177.71      23.53  -7.552 5.23e-10 ***
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

##

```
## Residual standard error: 156.5 on 54 degrees of freedom
```

```
## Multiple R-squared:  0.5137, Adjusted R-squared:  0.5047
```

```
## F-statistic: 57.03 on 1 and 54 DF, p-value: 5.232e-10
```

```
## [1] 153.6373
```

```
## [1] 0.4704595
```

```
##    1    2    3    4
```

```
## 116.81 113.72 110.68 107.69
```

```
## [1] 48.64261
```

```
## [1] 0.2725296
```

Mo hình Log Lin

##

Call:

```
## lm(formula = log(rev1) ~ t1)
```

##

Bài tập lớn Phân tích chuỗi thời gian và dự báo

Residuals:

```
##      Min       1Q   Median       3Q      Max
## -1.35963 -0.22345  0.03234  0.28458  1.30846
```

##

Coefficients:

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 6.255584  0.136056  45.978 < 2e-16 ***
## t1          -0.029500  0.004153  -7.104 2.79e-09 ***
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##

Residual standard error: 0.5023 on 54 degrees of freedom

Multiple R-squared: 0.4831, Adjusted R-squared: 0.4735

F-statistic: 50.47 on 1 and 54 DF, p-value: 2.792e-09

[1] 159.8776

[1] 0.4213103

```
##      1      2      3      4
```

```
## 96.94 94.12 91.39 88.73
```

[1] 66.80729

[1] 0.3983233

Mo hình Log Log

##

Call:

```
## lm(formula = log(rev1) ~ log(t1))
```

##

Residuals:

```
##      Min       1Q   Median       3Q      Max
## -1.15262 -0.30244 -0.04215  0.30609  1.12542
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

##

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 7.07936 0.24023 29.469 < 2e-16 ***

log(t1) -0.54083 0.07499 -7.212 1.87e-09 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##

Residual standard error: 0.4986 on 54 degrees of freedom

Multiple R-squared: 0.4906, Adjusted R-squared: 0.4812

F-statistic: 52.01 on 1 and 54 DF, p-value: 1.865e-09

[1] 186.8334

[1] 0.4177543

1 2 3 4

133.32 132.07 130.86 129.67

[1] 30.96587

[1] 0.1700909

Mo hình xu thế tuyến tính mùa vụ đang công

##

Call:

lm(formula = rev1 ~ t1 + s2 + s3 + s4)

##

Residuals:

Min 1Q Median 3Q Max

-297.82 -89.32 -17.59 67.07 363.35

##

Coefficients:

Estimate Std. Error t value Pr(>|t|)

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
## (Intercept) 511.739 57.863 8.844 7.15e-12 ***
```

```
## t1 -9.403 1.376 -6.835 9.92e-09 ***
```

```
## s2 65.546 62.758 1.044 0.301
```

```
## s3 61.735 62.804 0.983 0.330
```

```
## s4 51.138 62.879 0.813 0.420
```

```
## ---
```

```
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 166 on 51 degrees of freedom
```

```
## Multiple R-squared: 0.4829, Adjusted R-squared: 0.4424
```

```
## F-statistic: 11.91 on 4 and 51 DF, p-value: 6.609e-07
```

```
## [1] 158.4192
```

```
## [1] 0.5913084
```

```
## 1 2 3 4
```

```
## -24.23 31.91 18.70 -1.30
```

```
## [1] 151.0263
```

```
## [1] 0.9729269
```

Mo hình tuyến tính có mua vụ đang nhận

```
##
```

```
## Call:
```

```
## lm(formula = rev1 ~ t1 + tr2 + tr3 + tr4)
```

```
##
```

```
## Residuals:
```

```
## Min 1Q Median 3Q Max
```

```
## -298.39 -93.79 -18.07 75.80 382.10
```

```
##
```

```
## Coefficients:
```

```
## Estimate Std. Error t value Pr(>|t|)
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
## (Intercept) 554.8669 45.2167 12.271 < 2e-16 ***
```

```
## t1 -10.3219 1.8798 -5.491 1.27e-06 ***
```

```
## tr2 1.8551 1.9785 0.938 0.353
```

```
## tr3 0.8984 1.9538 0.460 0.648
```

```
## tr4 1.0889 1.9307 0.564 0.575
```

```
## ---
```

```
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 166.8 on 51 degrees of freedom
```

```
## Multiple R-squared: 0.4778, Adjusted R-squared: 0.4369
```

```
## F-statistic: 11.67 on 4 and 51 DF, p-value: 8.403e-07
```

```
## [1] 159.1975
```

```
## [1] 0.5993855
```

```
## 1 2 3 4
```

```
## -33.48 63.79 -1.12 0.89
```

```
## [1] 151.4999
```

```
## [1] 0.9708991
```

Mo hình Holt Winter dạng cong

```
## Holt-Winters exponential smoothing with trend and additive seasonal component.
```

```
##
```

```
## Call:
```

```
## HoltWinters(x = rev1, seasonal = "a")
```

```
##
```

```
## Smoothing parameters:
```

```
## alpha: 0.5093509
```

```
## beta : 0.04740083
```

```
## gamma: 0.7903639
```

```
##
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

Coefficients:

[1]

a 244.845762

b -1.875665

s1 -133.785100

s2 -40.015463

s3 -110.134718

s4 -120.231224

[1] 147.3373

[1] 0.4556388

Qtr1 Qtr2 Qtr3 Qtr4

2024 109.18 201.08 129.08 117.11

[1] 36.13883

[1] 0.2135265

Mo hình Holt Winter dạng nhân

Holt-Winters exponential smoothing with trend and multiplicative seasonal component.

##

Call:

HoltWinters(x = rev1, seasonal = "m")

##

Smoothing parameters:

alpha: 0.5470664

beta : 0.05120627

gamma: 0.6176897

##

Coefficients:

[1]

a 194.0026930

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
## b -3.2062959
```

```
## s1 0.5977771
```

```
## s2 0.7967804
```

```
## s3 0.6440516
```

```
## s4 0.6530369
```

```
## [1] 146.5236
```

```
## [1] 0.414646
```

```
##      Qtr1  Qtr2  Qtr3  Qtr4
```

```
## 2024 114.0537 149.4681 118.7527 118.3156
```

```
## [1] 35.27453
```

```
## [1] 0.1965327
```

Sung dung mo hinh Holt Winter dang nhan de du bao cho nam 2025

```
## Holt-Winters exponential smoothing with trend and multiplicative seasonal component.
```

```
##
```

```
## Call:
```

```
## HoltWinters(x = rev, seasonal = "m")
```

```
##
```

```
## Smoothing parameters:
```

```
## alpha: 0.5482734
```

```
## beta : 0.0503951
```

```
## gamma: 0.6185694
```

```
##
```

```
## Coefficients:
```

```
##      [,1]
```

```
## a 246.4354711
```

```
## b -0.1071959
```

```
## s1 0.6145240
```

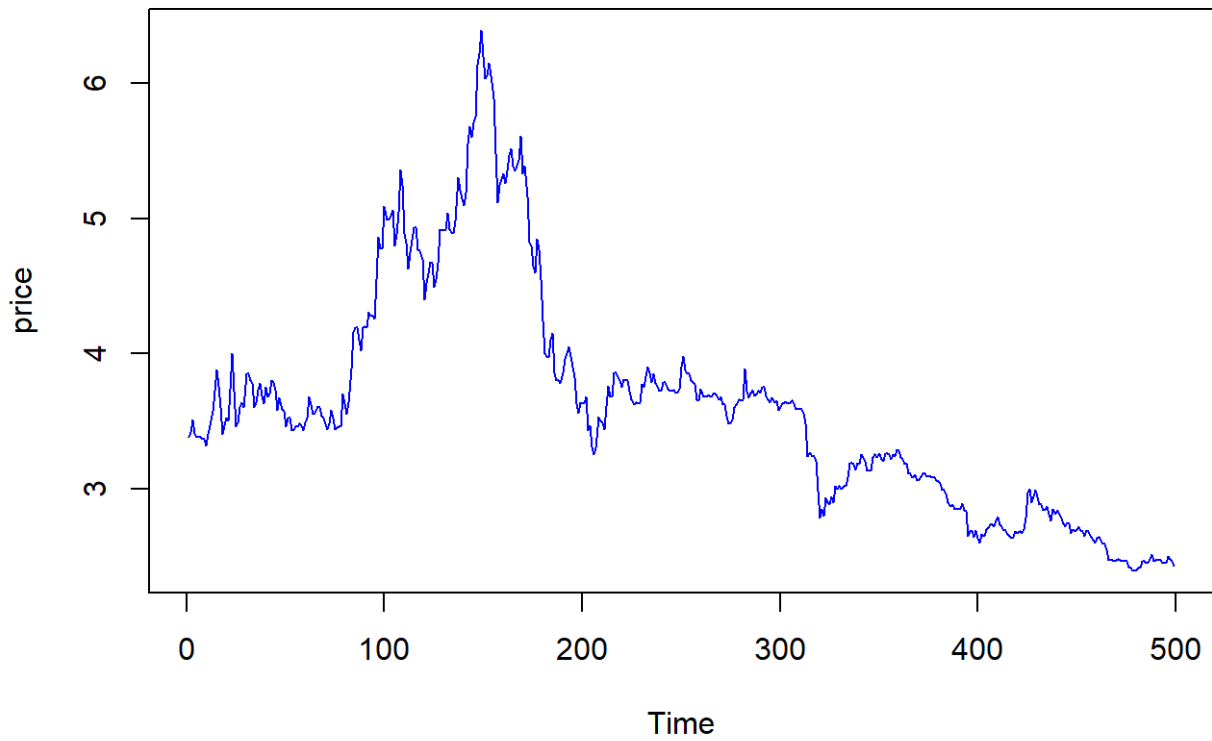
```
## s2 0.8041050
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

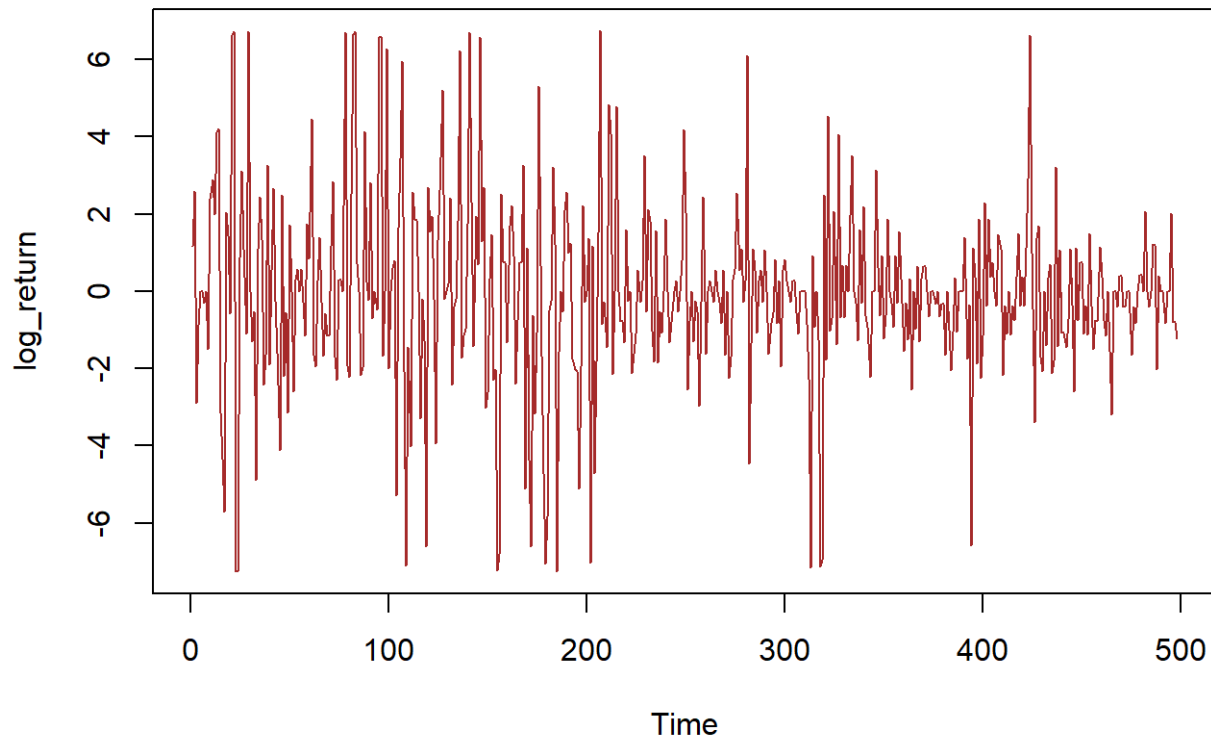
s3 0.6782669

s4 0.6722863

Qtr1 Qtr2 Qtr3 Qtr4



2025 151.37 197.99 166.93 165.39



Kiểm định nghiệm đơn vị cho chuỗi giá

```
##
```

```
## #####
```

```
## # Augmented Dickey-Fuller Test Unit Root Test #
```

```
## #####
```

```
##
```

```
## Test regression trend
```

```
##
```

```
##
```

```
## Call:
```

```
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
```

```
##
```

```
## Residuals:
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
##   Min    1Q  Median    3Q    Max
## -0.35862 -0.03684  0.00038  0.03860  0.41112
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.242e-02  3.265e-02   2.524 0.011906 *
## z.lag.1     -1.682e-02  6.868e-03  -2.449 0.014661 *
## tt          -9.305e-05  4.012e-05  -2.320 0.020773 *
## z.diff.lag   1.692e-01  4.423e-02   3.825 0.000147 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.09607 on 493 degrees of freedom
## Multiple R-squared:  0.04069,    Adjusted R-squared:  0.03486
## F-statistic: 6.971 on 3 and 493 DF,  p-value: 0.0001335
##
##
## Value of test-statistic is: -2.4493 2.34 3.4336
##
## 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
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
##  
## Test regression none  
##  
##  
## Call:  
## lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -0.38565 -0.03592  0.00022  0.03577  0.38775   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## z.lag.1    -0.0007614  0.0011660  -0.653  0.514070      
## z.diff.lag  0.1655838  0.0443035   3.737  0.000208 ***  
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 0.09651 on 495 degrees of freedom  
## Multiple R-squared:  0.02828,    Adjusted R-squared:  0.02435   
## F-statistic: 7.202 on 2 and 495 DF,  p-value: 0.0008259  
##  
##  
## Value of test-statistic is: -0.653  
##  
## Critical values for test statistics:  
##      1pct  5pct 10pct   
## tau1 -2.58 -1.95 -1.62
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
##
## #####
## # 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
## -0.40654 -0.03685  0.00155  0.03597  0.39557
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.360e-03  8.659e-03   0.619  0.53618
## z.lag.1      -9.434e-01  5.781e-02 -16.320 < 2e-16 ***
## tt           -2.971e-05  3.014e-05  -0.986  0.32466
## z.diff.lag    1.269e-01  4.469e-02   2.839  0.00471 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0959 on 492 degrees of freedom
## Multiple R-squared:  0.4284, Adjusted R-squared:  0.4249
## F-statistic: 122.9 on 3 and 492 DF, p-value: < 2.2e-16
##
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

##

Value of test-statistic is: -16.3198 88.7807 133.1689

##

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

##

#####

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

-0.40323 -0.03670 -0.00018 0.03515 0.39839

##

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.002046 0.004307 -0.475 0.63502

z.lag.1 -0.939802 0.057690 -16.291 < 2e-16 ***

z.diff.lag 0.125018 0.044645 2.800 0.00531 **

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0959 on 493 degrees of freedom
## Multiple R-squared:  0.4272, Adjusted R-squared:  0.4249
## F-statistic: 183.9 on 2 and 493 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -16.2905 132.6926
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.44 -2.87 -2.57
## phi1  6.47  4.61  3.79
##
## #####
## # 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
## -0.40519 -0.03875 -0.00222  0.03305  0.39629
```


Bài tập lớn Phân tích chuỗi thời gian và dự báo

##

Coefficients:

Estimate Std. Error t value Pr(>|t|)

z.lag.1 -0.93911 0.05763 -16.296 < 2e-16 ***

z.diff.lag 0.12469 0.04460 2.795 0.00538 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##

Residual standard error: 0.09582 on 494 degrees of freedom

Multiple R-squared: 0.427, Adjusted R-squared: 0.4247

F-statistic: 184 on 2 and 494 DF, p-value: < 2.2e-16

##

##

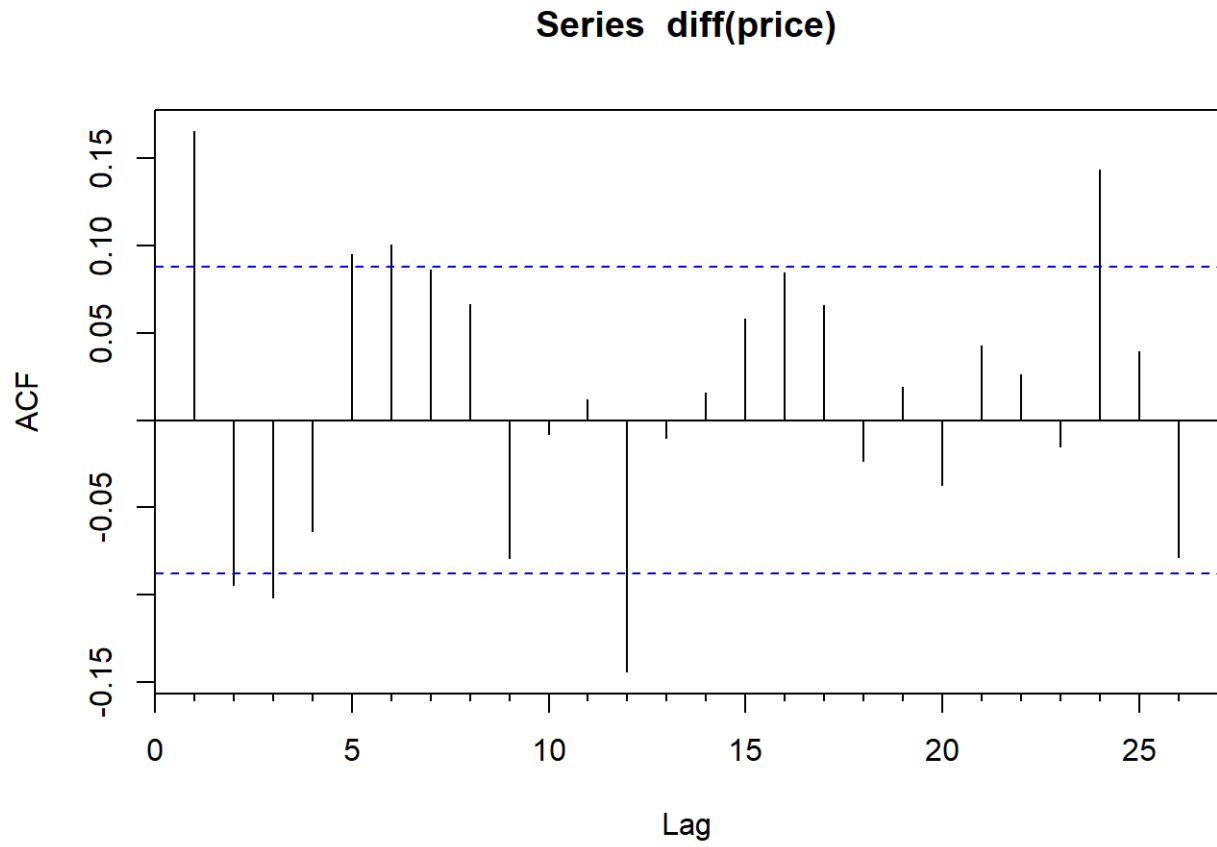
Value of test-statistic is: -16.2965

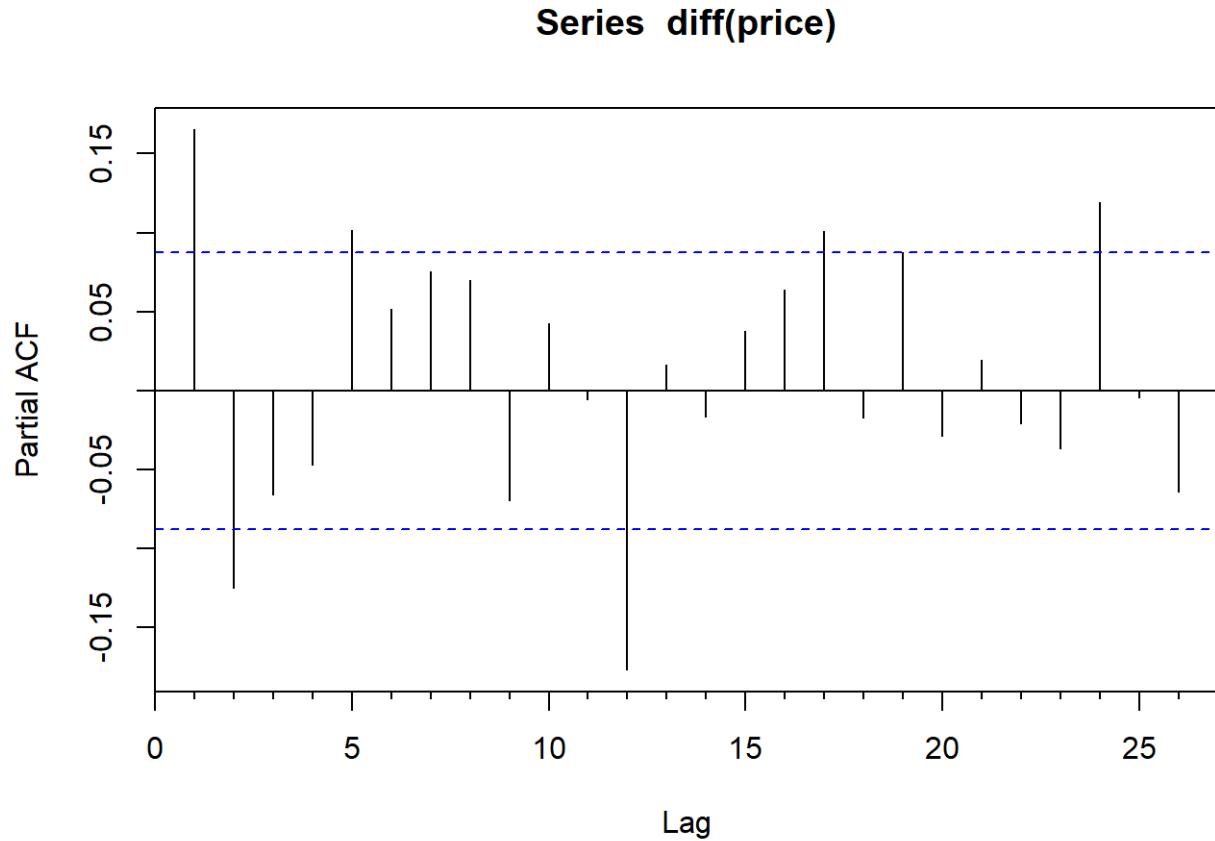
##

Critical values for test statistics:

1pct 5pct 10pct

tau1 -2.58 -1.95 -1.62





```
# Lua chon mo hinh ARIMA
```

```
##
```

```
## Call:
```

```
## arima(x = price, order = c(2, 1, 3))
```

```
##
```

```
## Coefficients:
```

```
##      ar1    ar2    ma1    ma2    ma3
```

```
##      1.3963 -0.9140 -1.2455 0.6008 0.2081
```

```
## s.e. 0.1014 0.0903 0.1046 0.1253 0.0478
```

```
##
```

```
## sigma^2 estimated as 0.008787: log likelihood = 471.94, aic = -931.89
```

```
##
```

```
## Training set error measures:
```

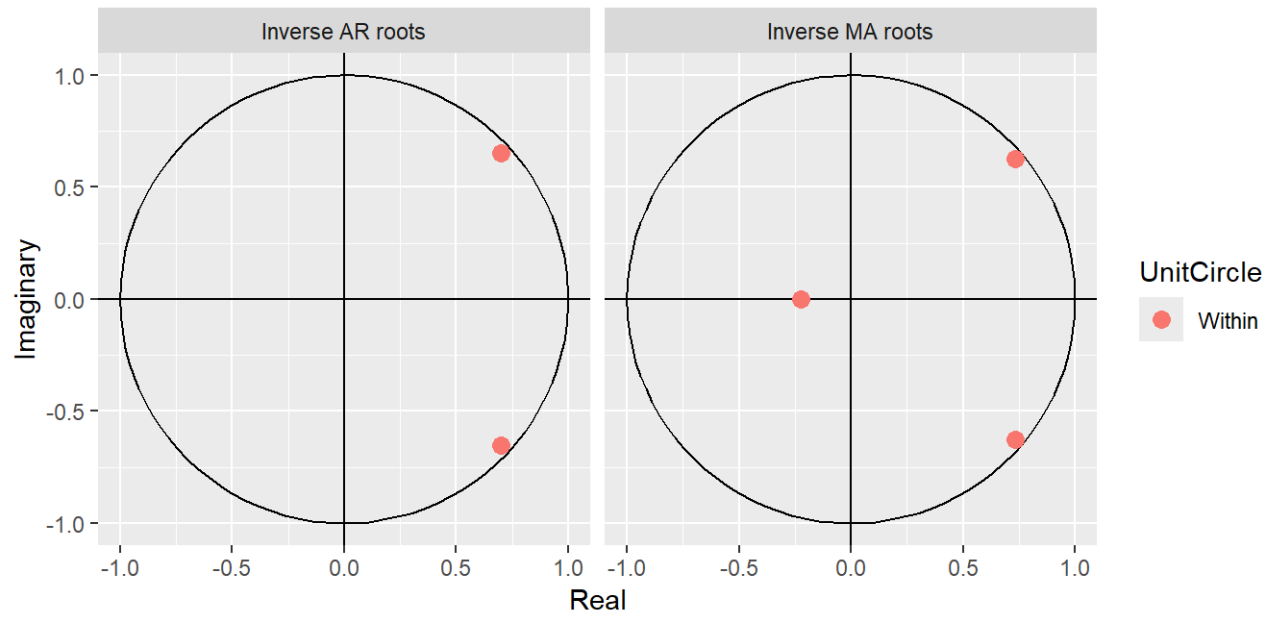
```
##           ME    RMSE    MAE    MPE    MAPE    MASE
```

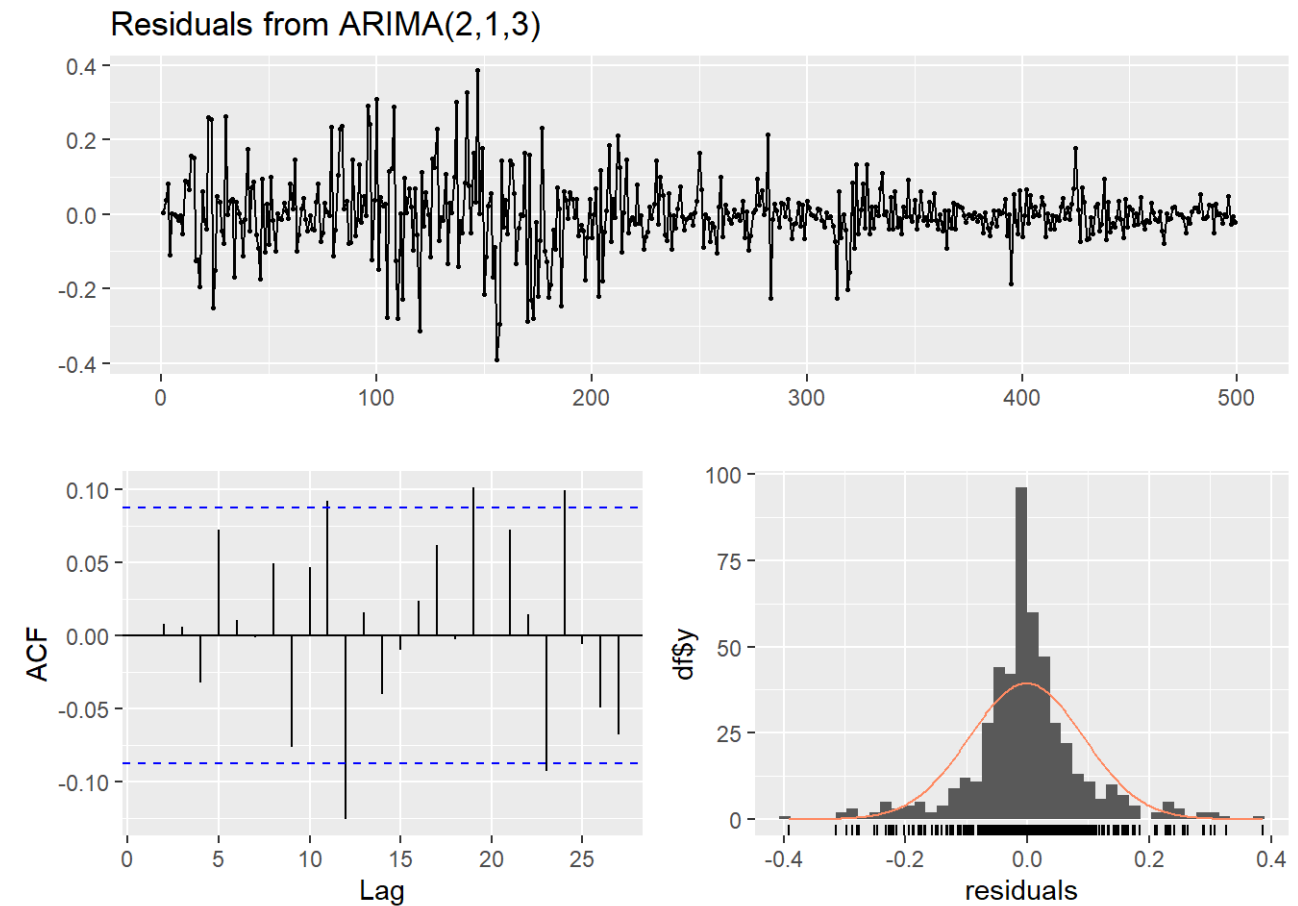
Bài tập lớn Phân tích chuỗi thời gian và dự báo

Training set -0.001733657 0.09364355 0.06193823 -0.08405391 1.616478 0.9908525

ACF1

Training set 0.001035073





```
##
```

```
## Ljung-Box test
```

```
##
```

```
## data: Residuals from ARIMA(2,1,3)
```

```
## Q* = 8.6088, df = 5, p-value = 0.1257
```

```
##
```

```
## Model df: 5. Total lags used: 10
```

```
##
```

```
## Call:
```

```
## arima(x = price, order = c(3, 1, 0))
```

```
##
```

```
## Coefficients:
```

```
##      ar1      ar2      ar3
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
##      0.1777 -0.1121 -0.0657
```

```
## s.e. 0.0447  0.0451  0.0447
```

```
##
```

```
## sigma^2 estimated as 0.009085: log likelihood = 463.91, aic = -919.82
```

```
##
```

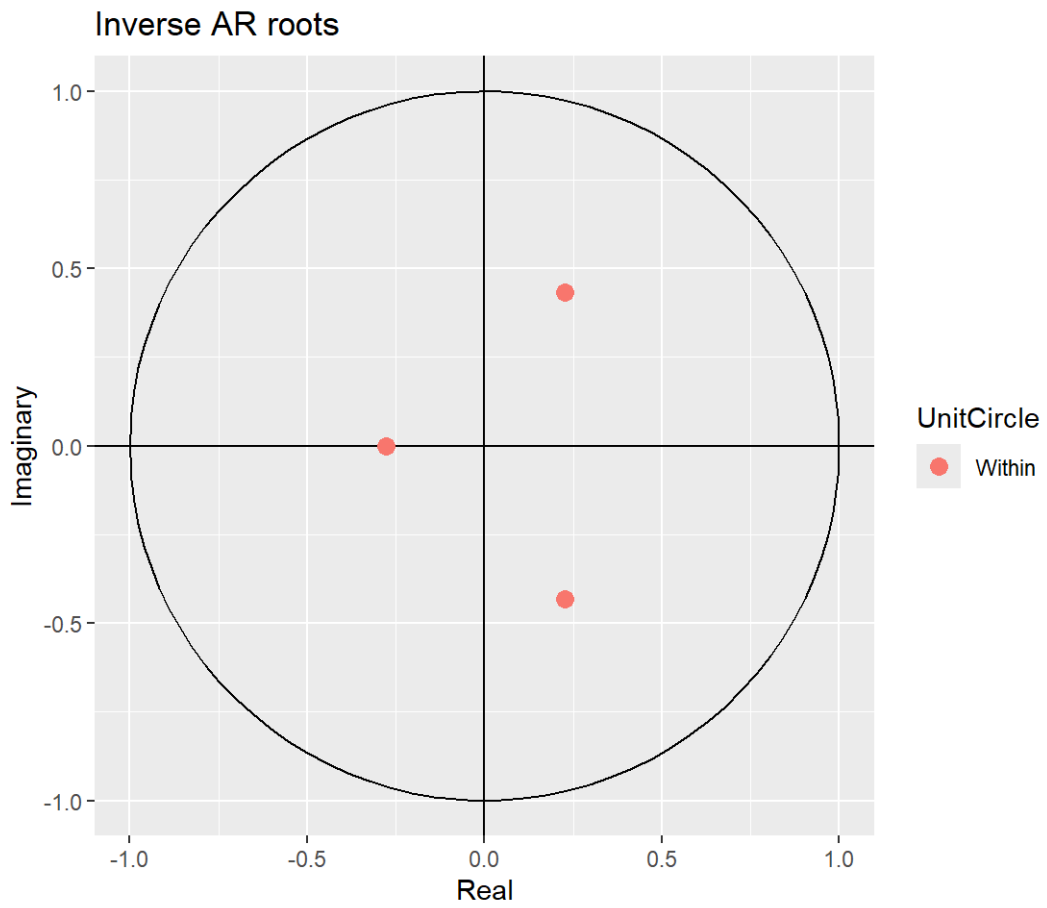
```
## Training set error measures:
```

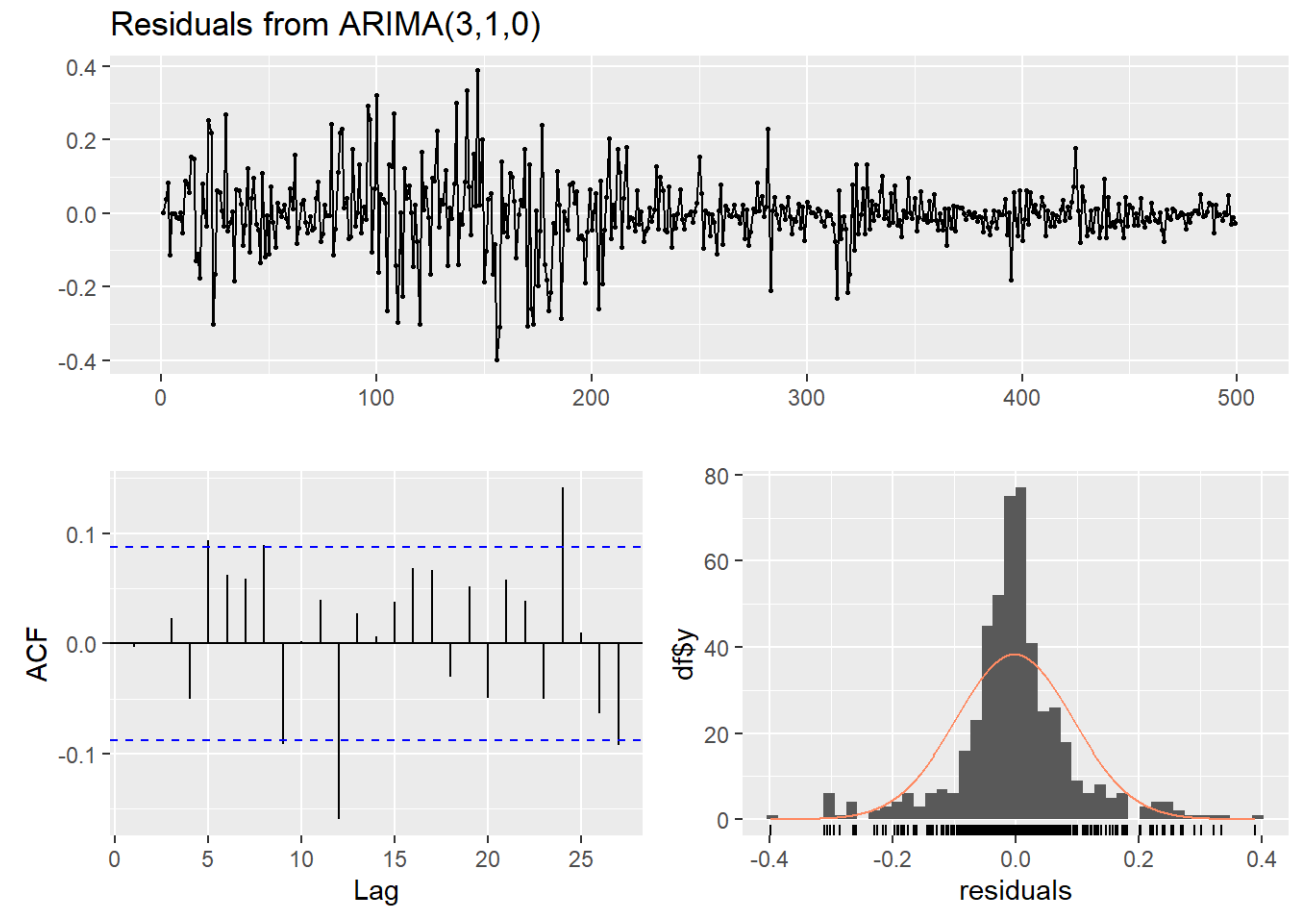
```
##           ME    RMSE    MAE    MPE    MAPE    MASE
```

```
## Training set -0.001889905 0.0952211 0.06239616 -0.09242793 1.625345 0.9981782
```

```
##           ACF1
```

```
## Training set -0.003092566
```





```
##
```

```
## Ljung-Box test
```

```
##
```

```
## data: Residuals from ARIMA(3,1,0)
```

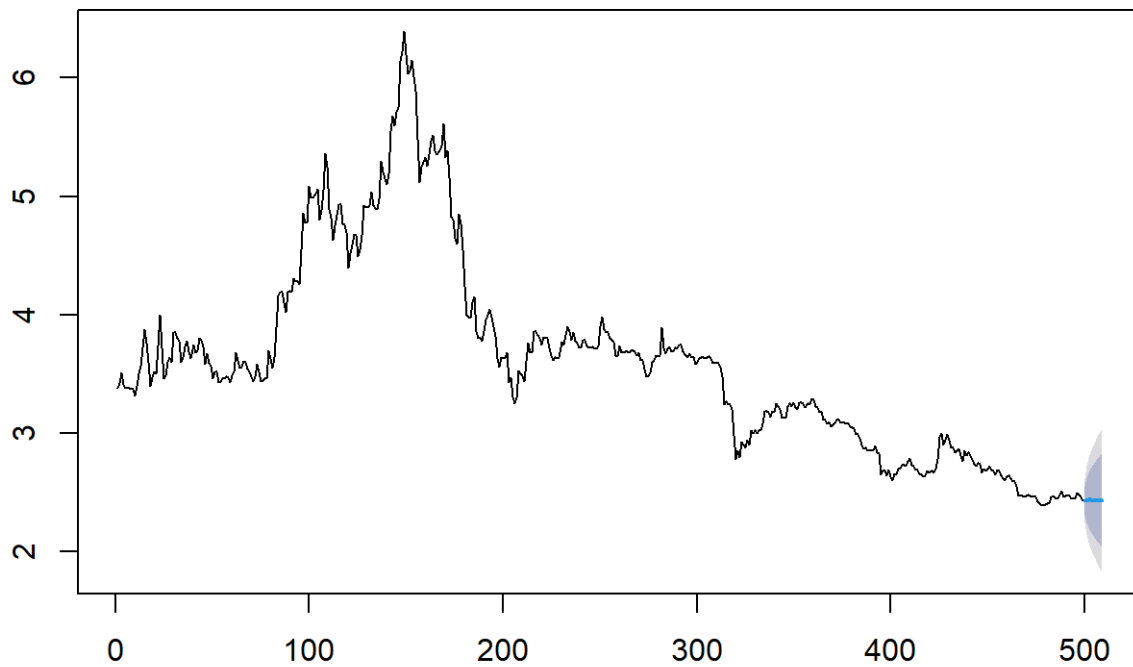
```
## Q* = 18.002, df = 7, p-value = 0.01196
```

```
##
```

```
## Model df: 3. Total lags used: 10
```

Su dụng mô hình đã chọn để dự báo cho chuỗi giá

Forecasts from ARIMA(3,1,0)



Kiểm định nghiệm đơn vị cho chuỗi Log return

```
##
```

```
## #####
```

```
## # Augmented Dickey-Fuller Test Unit Root Test #
```

```
## #####
```

```
##
```

```
## Test regression trend
```

```
##
```

```
##
```

```
## Call:
```

```
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
```

```
##
```

```
## Residuals:
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
##   Min   1Q Median   3Q   Max
## -7.5665 -1.0841 -0.0012  1.0116  6.8977
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.1518068  0.2111736   0.719   0.4726
## z.lag.1     -0.9752961  0.0587190 -16.610 <2e-16 ***
## tt          -0.0008970  0.0007354  -1.220   0.2232
## z.diff.lag   0.1281208  0.0446680   2.868   0.0043 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.338 on 492 degrees of freedom
## Multiple R-squared:  0.4421, Adjusted R-squared:  0.4387
## F-statistic: 130 on 3 and 492 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -16.6096 91.9627 137.9402
##
## 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
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

##

Test regression drift

##

##

Call:

lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

##

Residuals:

Min 1Q Median 3Q Max

-7.3995 -1.1314 -0.0209 1.0526 7.0960

##

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.07163 0.10510 -0.682 0.49587

z.lag.1 -0.96980 0.05857 -16.557 < 2e-16 ***

z.diff.lag 0.12535 0.04463 2.808 0.00518 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##

Residual standard error: 2.339 on 493 degrees of freedom

Multiple R-squared: 0.4404, Adjusted R-squared: 0.4381

F-statistic: 194 on 2 and 493 DF, p-value: < 2.2e-16

##

##

Value of test-statistic is: -16.5566 137.0646

##

Critical values for test statistics:

1pct 5pct 10pct

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
## tau2 -3.44 -2.87 -2.57
## phi1 6.47 4.61 3.79
##
## #####
## # 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
## -7.4808 -1.2019 -0.0915  0.9815  7.0244
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## z.lag.1    -0.9684     0.0585 -16.552 < 2e-16 ***
## z.diff.lag  0.1247     0.0446  2.795  0.00539 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.338 on 494 degrees of freedom
## Multiple R-squared:  0.4399, Adjusted R-squared:  0.4376
## F-statistic: 194 on 2 and 494 DF, p-value: < 2.2e-16
##
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

##

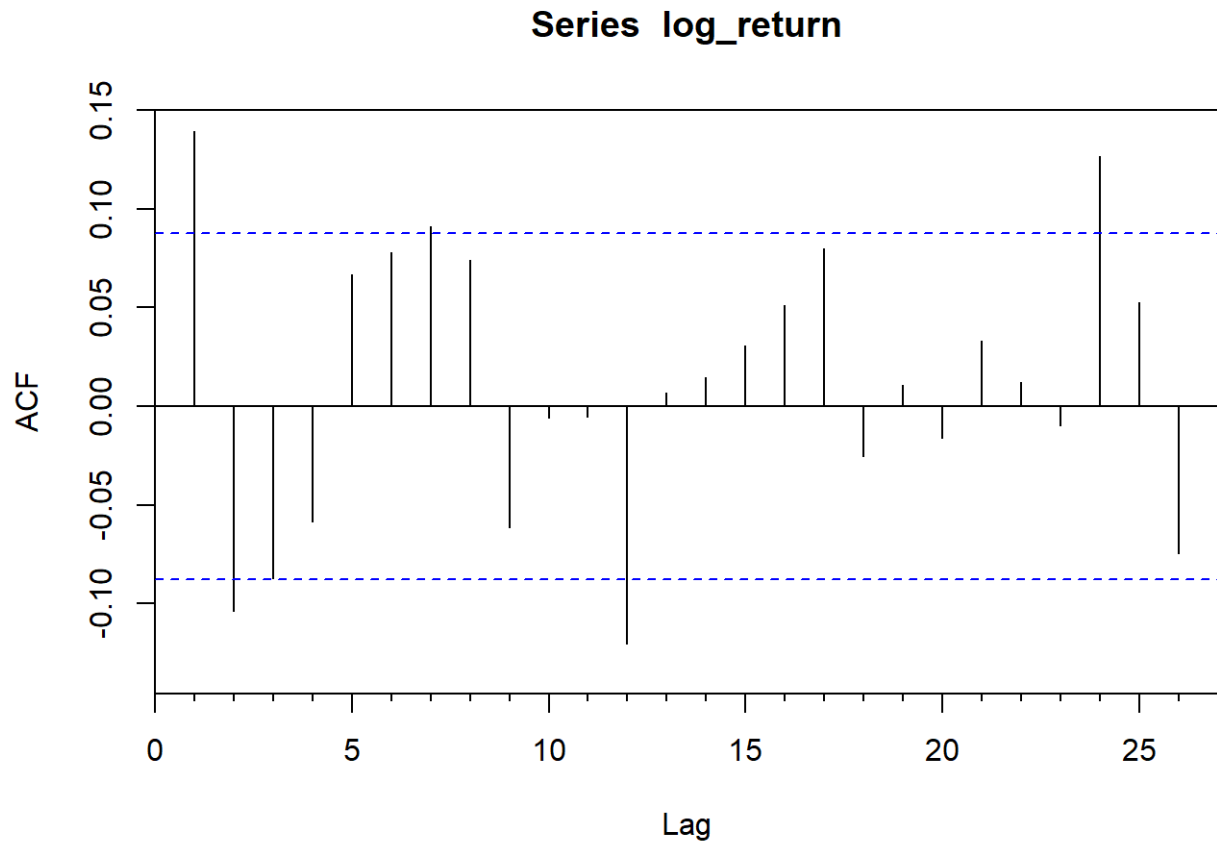
Value of test-statistic is: -16.5518

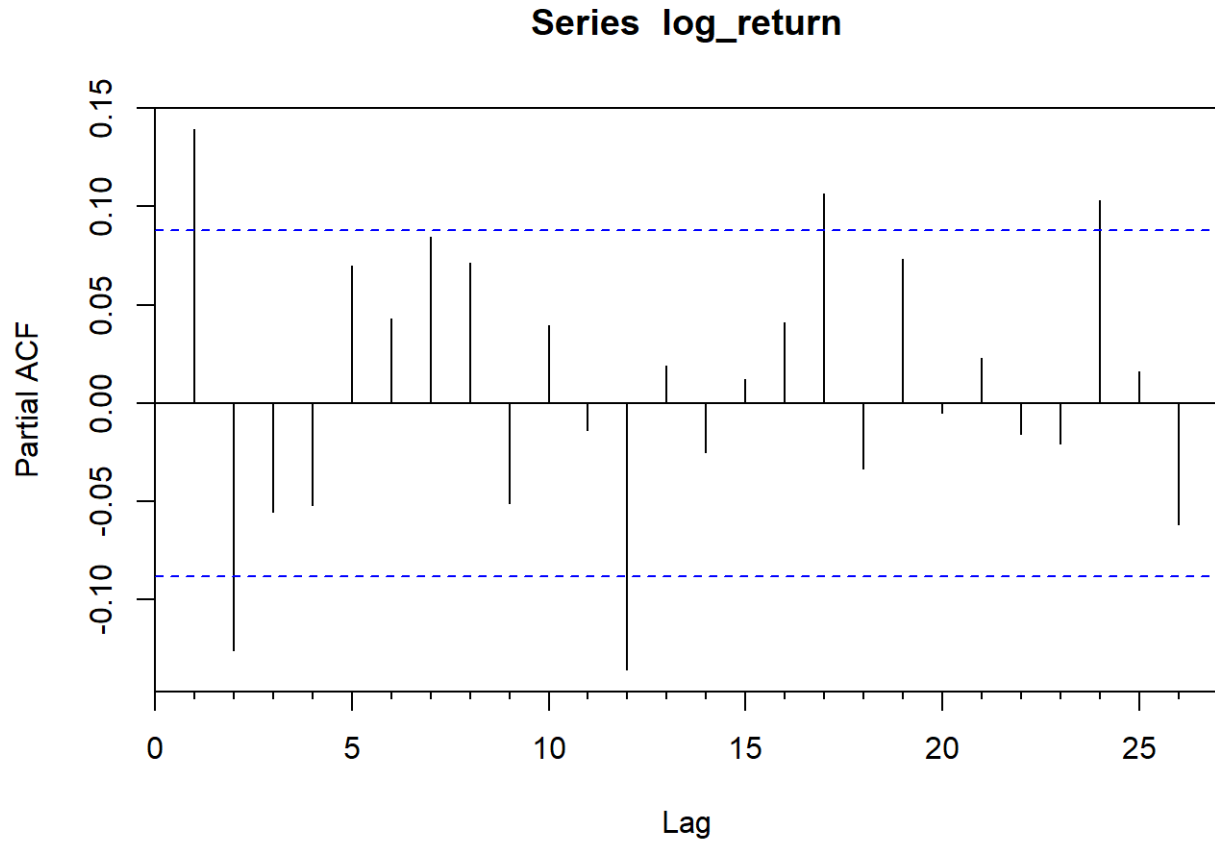
##

Critical values for test statistics:

1pct 5pct 10pct

tau1 -2.58 -1.95 -1.62





```
##
```

```
## Call:
```

```
## arima(x = log_return, order = c(3, 0, 2))
```

```
##
```

```
## Coefficients:
```

```
##      ar1    ar2    ar3    ma1    ma2 intercept
```

```
##      1.6399 -1.1150 0.2239 -1.5114 0.8003  -0.0688
```

```
## s.e. 0.1381  0.0977 0.0470  0.1378 0.0805   0.1184
```

```
##
```

```
## sigma^2 estimated as 5.286: log likelihood = -1121.37, aic = 2256.74
```

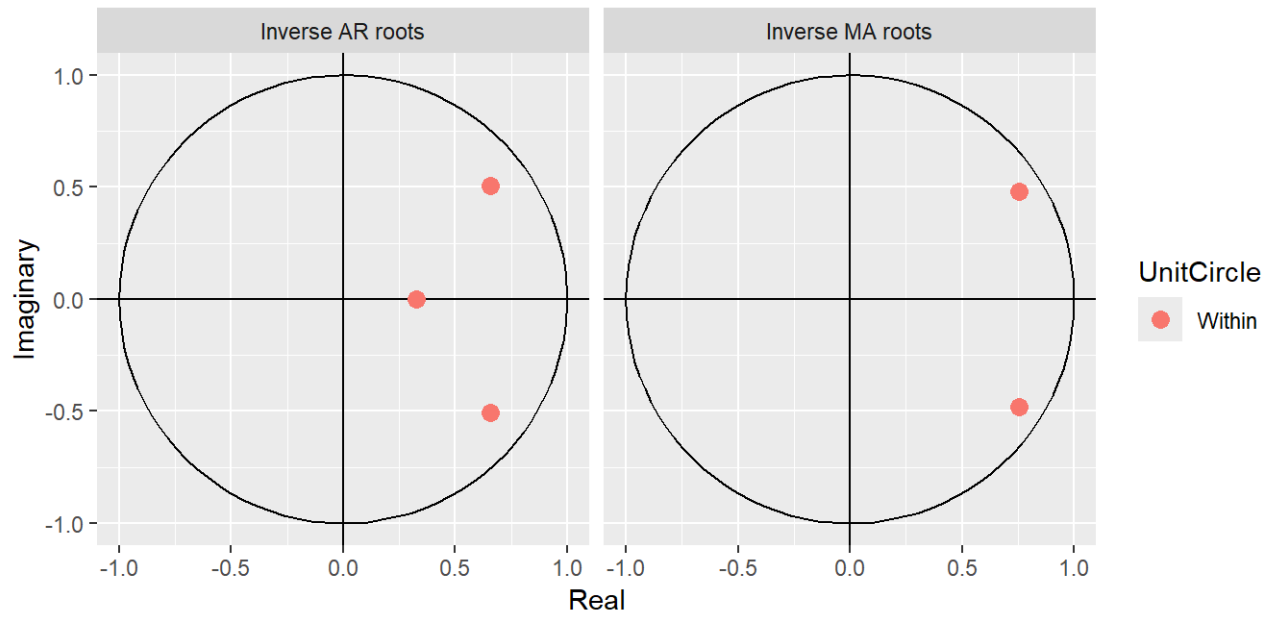
```
##
```

```
## Training set error measures:
```

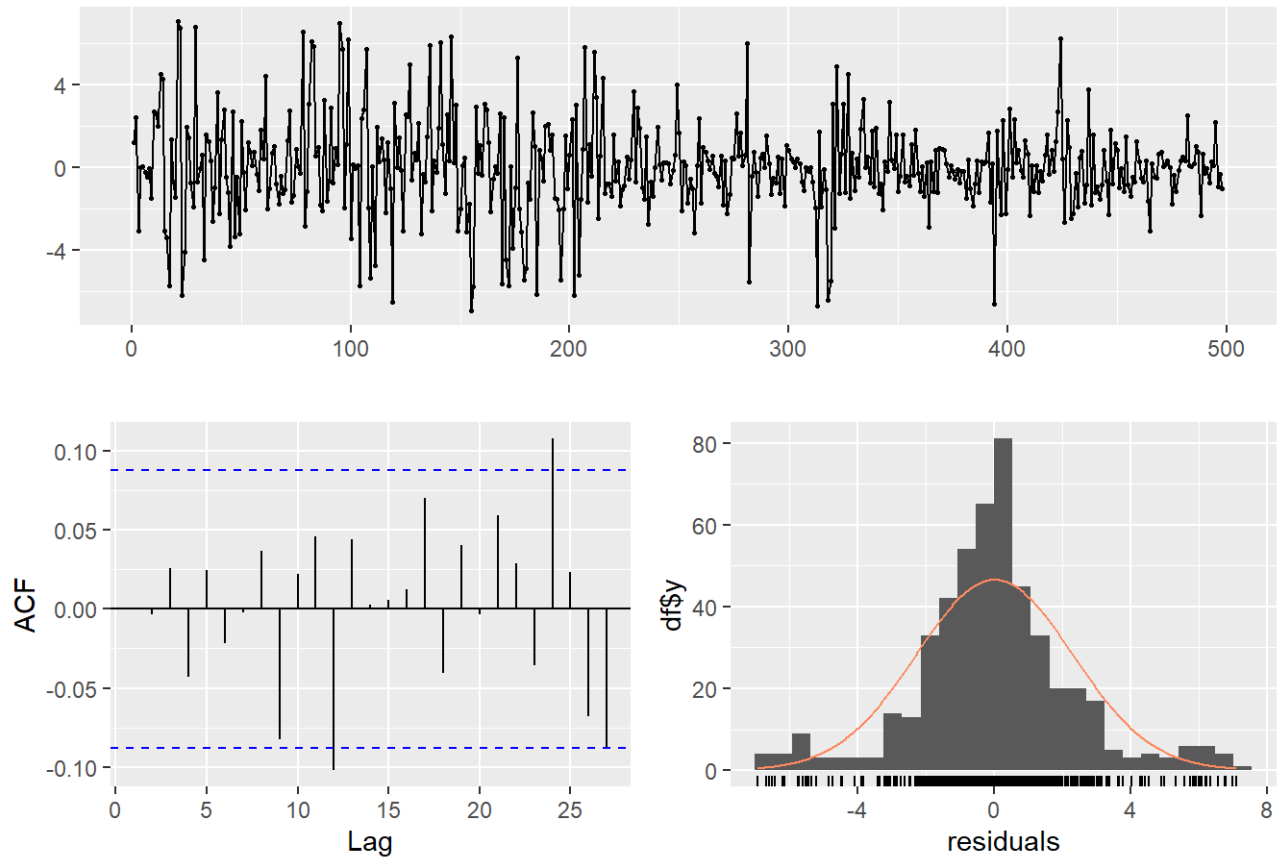
```
##           ME   RMSE    MAE  MPE  MAPE    MASE    ACF1
```


Bài tập lớn Phân tích chuỗi thời gian và dự báo

Training set 0.002703902 2.29915 1.621384 NaN Inf 0.728952 0.000280978



Residuals from ARIMA(3,0,2) with non-zero mean



```
##  
## Ljung-Box test  
##  
## data: Residuals from ARIMA(3,0,2) with non-zero mean  
## Q* = 6.2107, df = 5, p-value = 0.2863  
##  
## Model df: 5. Total lags used: 10  
##  
## Call:  
## arima(x = log_return, order = c(2, 0, 3))  
##  
## Coefficients:  
##      ar1      ar2      ma1      ma2      ma3 intercept
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
##      1.4543 -0.9637 -1.3275  0.7102  0.1884  -0.0644
```

```
## s.e. 0.0181  0.0168  0.0602  0.0855  0.0480   0.1147
```

```
##
```

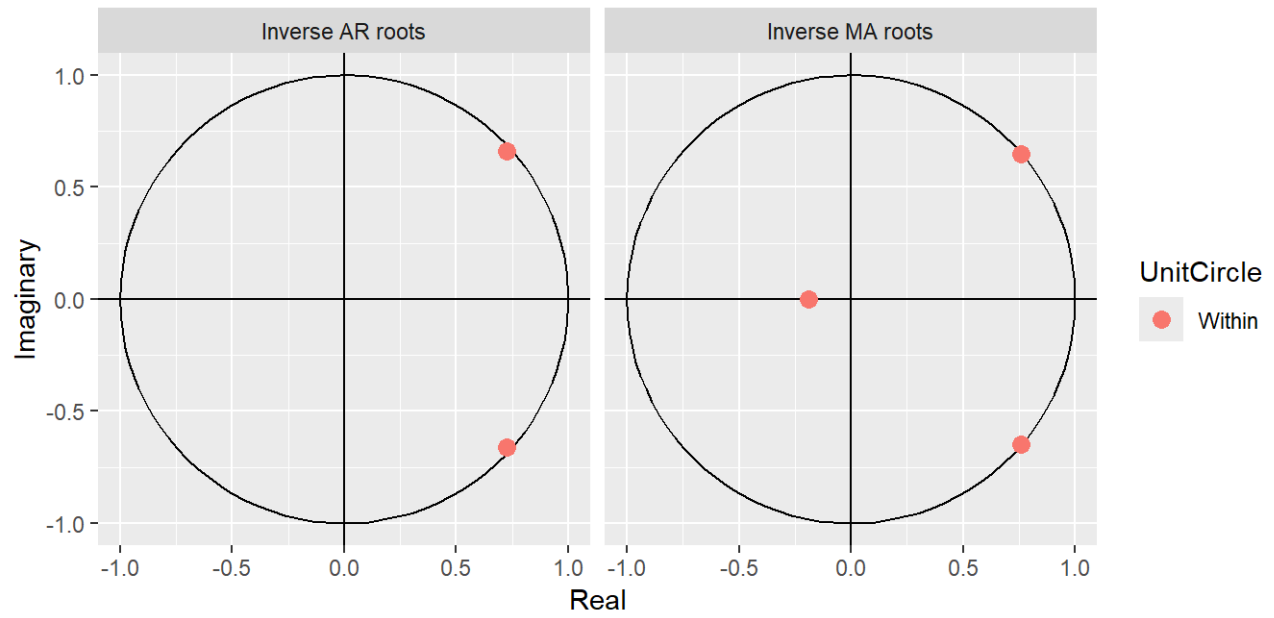
```
## sigma^2 estimated as 5.212: log likelihood = -1119.87, aic = 2253.75
```

```
##
```

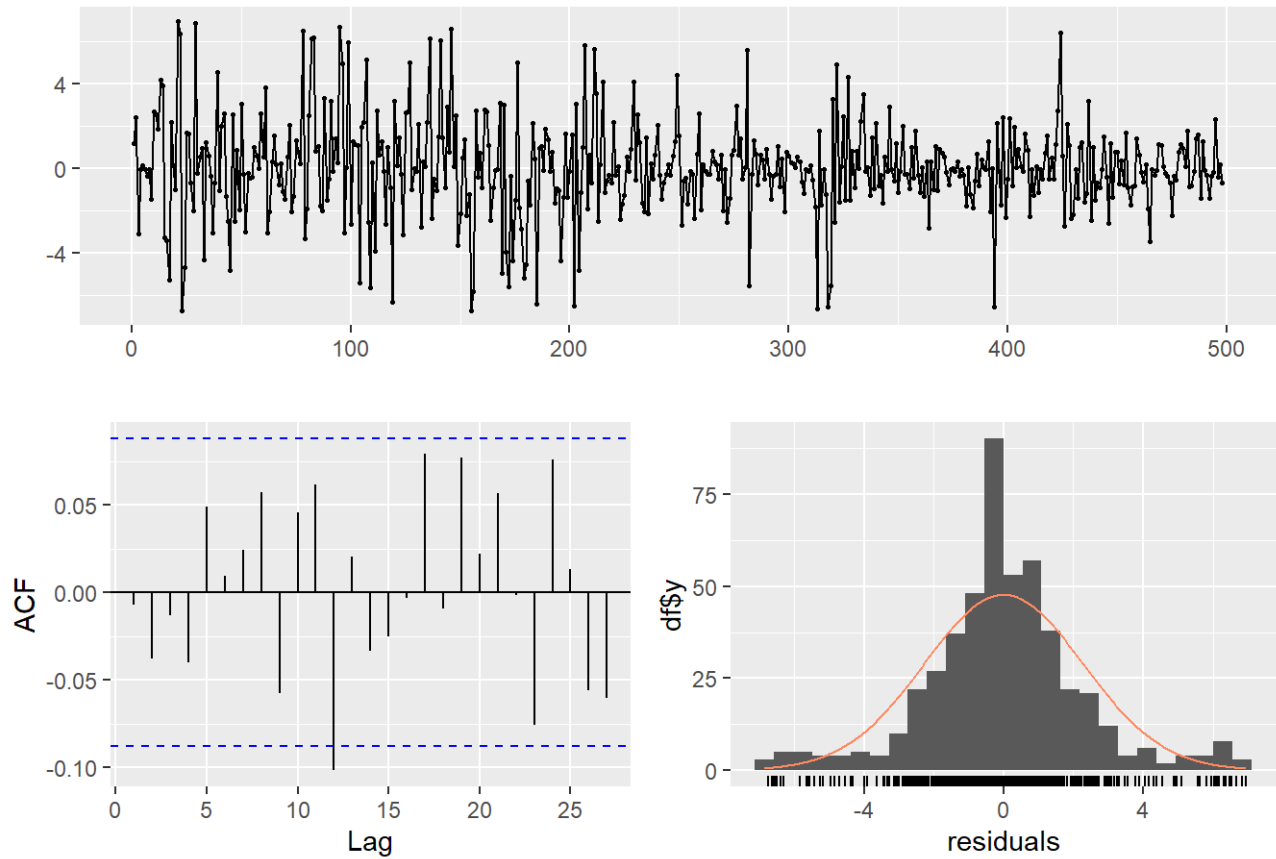
```
## Training set error measures:
```

```
##           ME   RMSE   MAE MPE MAPE   MASE   ACF1
```

```
## Training set 0.0001862634 2.283048 1.615877 NaN  Inf 0.7264762 -0.006901293
```



Residuals from ARIMA(2,0,3) with non-zero mean



```
##
```

```
## Ljung-Box test
```

```
##
```

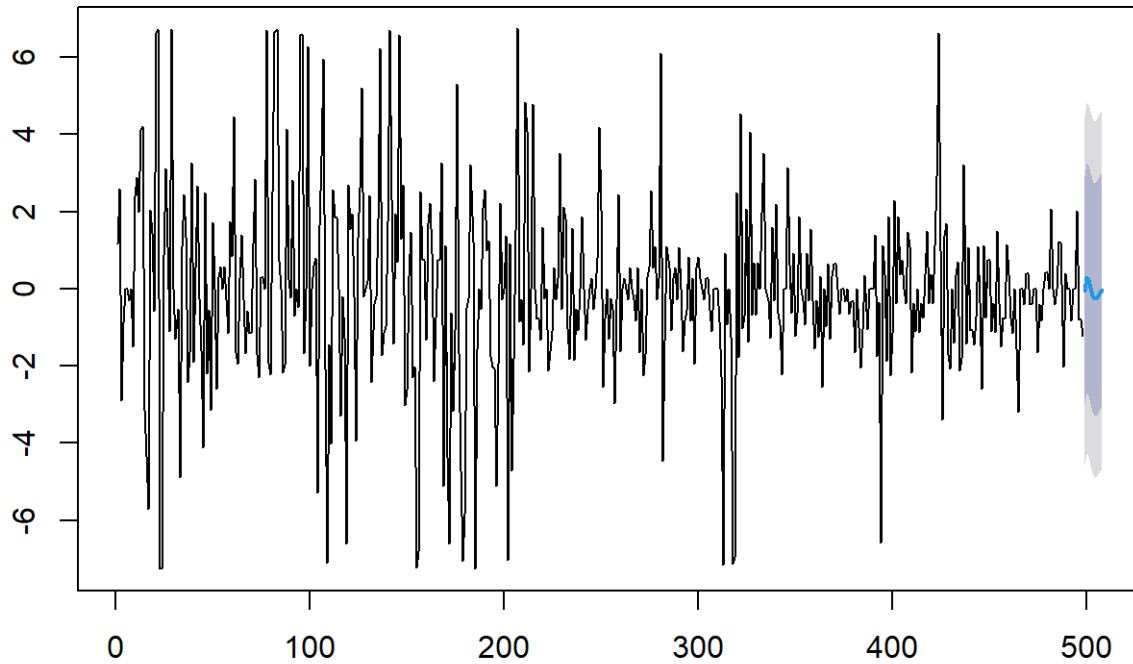
```
## data: Residuals from ARIMA(2,0,3) with non-zero mean
```

```
## Q* = 7.6332, df = 5, p-value = 0.1776
```

```
##
```

```
## Model df: 5. Total lags used: 10
```

Forecasts from ARIMA(3,0,2) with non-zero mean



Time Series:

Start = 499

End = 508

Frequency = 1

[1] -0.04539173 0.28347505 0.22341266 0.02286723 -0.16538250 -0.26393646

[7] -0.26057240 -0.18733078 -0.09304694 -0.01934293

2. Vũ Phạm Thành Phương

1. Mô hình lin-lin.

Multiple R-squared: 0.3894, Adjusted R-squared: 0.378

F-statistic: 34.43 on 1 and 54 DF, p-value: 2.776e-07

rmse(ts_train, **fitted**(model_lin_lin))

[1] 47.30389

mape(ts_train, **fitted**(model_lin_lin))

[1] 0.1198491

forecast_lin_lin

1 2 3 4

272.7006 270.3637 268.0268 265.6898

rmse(test_revenue, forecast_lin_lin)

[1] 21.16687

mape(test_revenue, forecast_lin_lin)

[1] 0.0638153

2. Mô hình lin-log.

Multiple R-squared: 0.2123, Adjusted R-squared: 0.1977

F-statistic: 14.56 on 1 and 54 DF, p-value: 0.0003519

rmse(ts_train, **fitted**(model_lin_log))

[1] 53.7249

mape(ts_train, **fitted**(model_lin_log))

[1] 0.1438712

forecast_lin_log

1 2 3 4

308.9974 308.4514 307.9147 307.3871

rmse(test_revenue, forecast_lin_log)

[1] 33.25875

mape(test_revenue, forecast_lin_log)

[1] 0.106548

3. Mô hình log-lin.

Bài tập lớn Phân tích chuỗi thời gian và dự báo

Multiple R-squared: 0.3971, Adjusted R-squared: 0.3859

F-statistic: 35.56 on 1 and 54 DF, p-value: 1.951e-07

rmse(ts_train, **exp**(**fitted**(model_log_lin)))

[1] 47.93705

mape(ts_train, **exp**(**fitted**(model_log_lin)))

[1] 0.1184165

forecast_log_lin

1 2 3 4

270.9177 268.9463 266.9892 265.0464

rmse(test_revenue, forecast_log_lin)

[1] 21.48485

mape(test_revenue, forecast_log_lin)

[1] 0.06220432

4. Mô hình log-log.

Multiple R-squared: 0.2183, Adjusted R-squared: 0.2039

F-statistic: 15.08 on 1 and 54 DF, p-value: 0.0002827

rmse(ts_train, **exp**(**fitted**(model_log_log)))

[1] 55.03356

mape(ts_train, **exp**(**fitted**(model_log_log)))

[1] 0.1442767

forecast_log_log

1 2 3 4

303.3401 302.8208 302.3113 301.8111

rmse(test_revenue, forecast_log_log)

[1] 28.54126

mape(test_revenue, forecast_log_log)

[1] 0.0863775

5. Mùa vụ tuyến tính dạng cộng.

Multiple R-squared: 0.3927, Adjusted R-squared: 0.345

F-statistic: 8.243 on 4 and 51 DF, p-value: 3.308e-05

```
rmse(ts_train, fitted(model_season_add))
```

```
## [1] 47.17603
```

```
mape(ts_train, fitted(model_season_add))
```

```
## [1] 0.1200855
```

```
forecast_season_add
```

```
##      1      2      3      4
```

```
## 268.2266 270.3695 266.0124 270.6552
```

```
rmse(test_revenue, forecast_season_add)
```

```
## [1] 20.21466
```

```
mape(test_revenue, forecast_season_add)
```

```
## [1] 0.05712073
```

6. Mùa vụ tuyến tính dạng nhân.

```
## Multiple R-squared:  0.3954, Adjusted R-squared:  0.348
```

```
## F-statistic: 8.338 on 4 and 51 DF,  p-value: 2.966e-05
```

```
rmse(ts_train, fitted(model_season_mul))
```

```
## [1] 47.06931
```

```
mape(ts_train, fitted(model_season_mul))
```

```
## [1] 0.1184726
```

```
forecast_season_mul
```

```
##      1      2      3      4
```

```
## 270.2092 266.2865 259.7236 279.3023
```

```
rmse(test_revenue, forecast_season_mul)
```

```
## [1] 20.39319
```

```
mape(test_revenue, forecast_season_mul)
```

```
## [1] 0.05296969
```

7. Holt-Winters mùa vụ dạng cộng.

```
rmse(ts_train, fitted(model_hw_add)[,1])
```

```
## [1] 46.4913
```

```
mape(ts_train, fitted(model_hw_add)[,1])
```

```
## [1] 0.1161673
```

```
forecast_hw_add$mean
```

```
##      Qtr1  Qtr2  Qtr3  Qtr4
```

```
## 2024 263.0070 268.0720 267.6760 276.5073
```

```
rmse(test_revenue, forecast_hw_add$mean)
```

```
## [1] 17.84044
```

```
mape(test_revenue, forecast_hw_add$mean)
```

```
## [1] 0.05297274
```

8. Holt-Winters mùa vụ dạng nhân.

```
rmse(ts_train, fitted(model_hw_mul)[,1])
```

```
## [1] 46.38345
```

```
mape(ts_train, fitted(model_hw_mul)[,1])
```

```
## [1] 0.1151952
```

```
forecast_hw_mul$mean
```

```
##      Qtr1  Qtr2  Qtr3  Qtr4
```

```
## 2024 269.2018 271.5326 269.6692 277.6537
```

```
rmse(test_revenue, forecast_hw_mul$mean)
```

```
## [1] 17.12226
```

```
mape(test_revenue, forecast_hw_mul$mean)
```

```
## [1] 0.05012387
```

9. Dự báo bằng mô hình Holt-Winters dạng nhân.

```
##      Qtr1  Qtr2  Qtr3  Qtr4
```

```
## 2025 280.9046 283.1359 289.1929 295.3345
```

10. Kiểm định ADF cho chuỗi giá.

```
## Test regression trend
```

```
## Residuals:
```

```
##      Min      1Q  Median      3Q      Max
```

```
## -1.00409 -0.10273 -0.02009  0.08399  1.00524
```

```
##
```

```
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
## (Intercept) 0.2364029 0.0781692 3.024 0.00262 **
## z.lag.1 -0.0355467 0.0116694 -3.046 0.00244 **
## tt 0.0005649 0.0001922 2.939 0.00344 **
## z.diff.lag 0.0642412 0.0448251 1.433 0.15245
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2445 on 493 degrees of freedom
## Multiple R-squared: 0.02085, Adjusted R-squared: 0.01489
## F-statistic: 3.499 on 3 and 493 DF, p-value: 0.01547
##
##
## Value of test-statistic is: -3.0461 3.7485 4.7054
##
## 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

## Test regression drift
## Residuals:
## Min 1Q Median 3Q Max
## -1.08029 -0.10421 -0.01303 0.08552 0.95832
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.056477 0.048989 1.153 0.250
## z.lag.1 -0.004076 0.004677 -0.872 0.384
## z.diff.lag 0.048779 0.044858 1.087 0.277
##
## Residual standard error: 0.2464 on 494 degrees of freedom
## Multiple R-squared: 0.003691, Adjusted R-squared: -0.0003427
## F-statistic: 0.915 on 2 and 494 DF, p-value: 0.4012
##
##
## Value of test-statistic is: -0.8715 1.2831
##
## Critical values for test statistics:
## 1pct 5pct 10pct
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
## tau2 -3.44 -2.87 -2.57
```

```
## phi1 6.47 4.61 3.79
```

Test regression none

```
## Residuals:
```

```
##      Min      1Q  Median      3Q      Max
```

```
## -1.10291 -0.09768 -0.00872  0.09025  0.94131
```

```
##
```

```
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
```

```
## z.lag.1  0.001176  0.001058  1.112  0.267
```

```
## z.diff.lag 0.046391  0.044825  1.035  0.301
```

```
##
```

```
## Residual standard error: 0.2464 on 495 degrees of freedom
```

```
## Multiple R-squared:  0.005037, Adjusted R-squared:  0.001017
```

```
## F-statistic: 1.253 on 2 and 495 DF, p-value: 0.2866
```

```
##
```

```
##
```

```
## Value of test-statistic is: 1.1119
```

```
##
```

```
## Critical values for test statistics:
```

```
##      1pct  5pct 10pct
```

```
## tau1 -2.58 -1.95 -1.62
```

11. Kiểm định ADF cho chuỗi sai phân bậc 1.

Test regression trend

```
## Residuals:
```

```
##      Min      1Q  Median      3Q      Max
```

```
## -1.07582 -0.09841 -0.00927  0.08457  0.97624
```

```
##
```

```
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 7.465e-03  2.227e-02  0.335  0.738
```

```
## z.lag.1  -9.975e-01  6.210e-02 -16.063  <2e-16 ***
```

```
## tt      3.159e-05  7.738e-05  0.408  0.683
```

```
## z.diff.lag 4.374e-02  4.491e-02  0.974  0.330
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 0.2467 on 492 degrees of freedom
```

```
## Multiple R-squared:  0.4791, Adjusted R-squared:  0.4759
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
## F-statistic: 150.8 on 3 and 492 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -16.0628 86.0108 129.015
##
## 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

## Test regression drift
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.07192 -0.10156 -0.01263  0.08476  0.97982
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.01534    0.01111   1.380   0.168
## z.lag.1     -0.99705    0.06204 -16.072 <2e-16 ***
## z.diff.lag   0.04361    0.04487   0.972   0.332
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2465 on 493 degrees of freedom
## Multiple R-squared:  0.4789, Adjusted R-squared:  0.4768
## F-statistic: 226.5 on 2 and 493 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -16.0717 129.1513
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.44 -2.87 -2.57
## phi1  6.47  4.61  3.79

## Test regression none
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.06190 -0.08668  0.00307  0.10000  0.98850
##
## Coefficients:
```

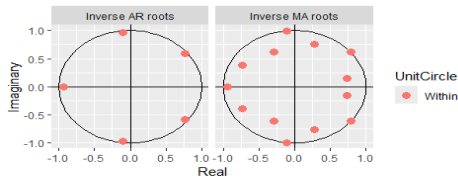
```
##      Estimate Std. Error t value Pr(>|t|)
## z.lag.1  -0.98943   0.06185 -15.998 <2e-16 ***
## z.diff.lag 0.03974   0.04482  0.887  0.376
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2467 on 494 degrees of freedom
## Multiple R-squared:  0.4769, Adjusted R-squared:  0.4748
## F-statistic: 225.2 on 2 and 494 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -15.9978
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

12. Chạy và kiểm định các mô hình ARIMA.

a) Mô hình ARIMA(5,1,13).

```
## arima(x = price, order = c(5, 1, 13))
##
## Coefficients:
##      ar1    ar2    ar3    ar4    ar5    ma1    ma2    ma3
##      0.4001 -0.3227 -0.1816 0.2864 -0.8311 -0.3567 0.2578 0.1894
## s.e. 0.1341 0.1760 0.2064 0.1690 0.1206 0.1405 0.1742 0.1963
##      ma4    ma5    ma6    ma7    ma8    ma9    ma10    ma11
##      -0.2968 0.6297 0.1674 -0.1924 0.0026 -0.0083 -0.1977 0.0334
## s.e. 0.1603 0.1154 0.0613 0.0706 0.0798 0.0709 0.0699 0.0555
##      ma12    ma13
##      -0.0399 0.1090
## s.e. 0.0518 0.0508
##
## sigma^2 estimated as 0.05444: log likelihood = 13.62, aic = 10.77
##
## Training set error measures:
##      ME    RMSE    MAE    MPE    MAPE    MASE
## Training set 0.02131319 0.2330877 0.15424 0.1855524 1.439599 1.00028
##      ACF1
## Training set -0.01195235
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo



Ljung-Box test

##

data: Residuals from ARIMA(5,1,13)

$Q^* = 8.0318$, $df = 3$, $p\text{-value} = 0.04536$

Model df: 18. Total lags used: 21

b) Mô hình ARIMA(5,1,5).

```
## arima(x = price, order = c(5, 1, 5))
```

##

Coefficients:

```
##      ar1   ar2   ar3   ar4   ar5   ma1   ma2   ma3
```

```
##    -0.0932 0.2145 -0.1939 -0.0292 0.1493 0.1511 -0.2674 0.1483
```

```
## s.e. 0.2087 0.2406 0.1913 0.2298 0.1835 0.1981 0.2354 0.1870
```

```
##      ma4   ma5
```

```
##    0.0305 -0.3767
```

```
## s.e. 0.2285 0.1755
```

##

```
## sigma^2 estimated as 0.05741: log likelihood = 4.7, aic = 12.59
```

##

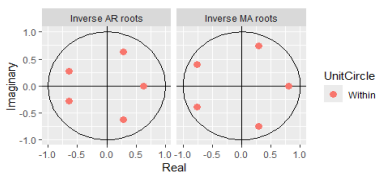
Training set error measures:

```
##           ME   RMSE   MAE   MPE   MAPE   MASE
```

```
## Training set 0.02311008 0.2393676 0.1551251 0.2014103 1.444021 1.00602
```

```
##           ACF1
```

```
## Training set -0.007405784
```



Ljung-Box test

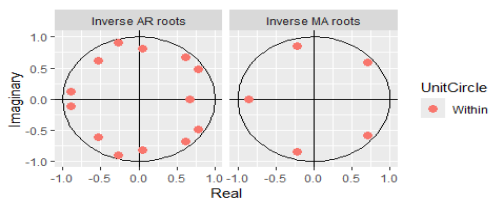
##

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
## data: Residuals from ARIMA(5,1,5)
##  $Q^* = 4.0782$ ,  $df = 3$ ,  $p\text{-value} = 0.2531$ 
## Model df: 10. Total lags used: 13
```

c) Mô hình ARIMA(13,1,5).

```
## arima(x = price, order = c(13, 1, 5))
##
## Coefficients:
##      ar1   ar2   ar3   ar4   ar5   ar6   ar7   ar8
##  0.1443 -0.2147 -0.1314 -0.0451 -0.7586 0.0782 -0.1469 0.0219
## s.e. 0.2214 0.1892 0.1623 0.2239 0.1471 0.0779 0.0690 0.0691
##      ar9  ar10  ar11  ar12  ar13   ma1   ma2   ma3   ma4
## -0.0816 -0.1398 0.0177 -0.087 0.1525 -0.0988 0.1520 0.1265 0.0431
## s.e. 0.0695 0.0604 0.0510 0.050 0.0532 0.2231 0.1857 0.1610 0.2230
##      ma5
##  0.5531
## s.e. 0.1395
##
## sigma^2 estimated as 0.0555: log likelihood = 12.86, aic = 12.27
##
## Training set error measures:
##              ME    RMSE    MAE    MPE    MAPE    MASE
## Training set 0.02076937 0.2353462 0.1559738 0.1801014 1.454261 1.011524
##              ACF1
## Training set -0.004320014
```



Ljung-Box test

```
##
## data: Residuals from ARIMA(13,1,5)
##  $Q^* = 6.3139$ ,  $df = 3$ ,  $p\text{-value} = 0.0973$ 
## Model df: 18. Total lags used: 21
```

13. So sánh dự báo của ARIMA(5,1,5) với giá thực tế và tính các sai số.

```
rmse(price_real, fc_price)
```

```
## [1] 0.801461
```

```
mape(price_real, fc_price)
```

```
## [1] 0.05322225
```

14. Kiểm định ADF chuỗi lợi suất.

```
## Test regression trend
```

```
## Residuals:
```

```
##   Min    1Q  Median    3Q   Max  
## -7.5193 -1.0200 -0.1129  0.9192  6.7045  
##
```

```
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)  
## (Intercept)  0.1260676  0.1917304   0.658   0.511  
## z.lag.1     -1.0235062  0.0618330 -16.553 <2e-16 ***  
## tt          0.0001022  0.0006656   0.154   0.878  
## z.diff.lag  0.0655977  0.0445370   1.473   0.141  
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##
```

```
## Residual standard error: 2.122 on 492 degrees of freedom
```

```
## Multiple R-squared:  0.4832, Adjusted R-squared:  0.4801
```

```
## F-statistic: 153.4 on 3 and 492 DF, p-value: < 2.2e-16
```

```
##
```

```
##
```

```
## Value of test-statistic is: -16.5527 91.3475 137.0164
```

```
##
```

```
## 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
```

```
## Test regression drift
```

```
## Residuals:
```

```
##   Min    1Q  Median    3Q   Max  
## -7.5049 -1.0182 -0.1156  0.9204  6.7175  
##
```

```
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)  
## (Intercept)  0.15158   0.09569   1.584   0.114
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

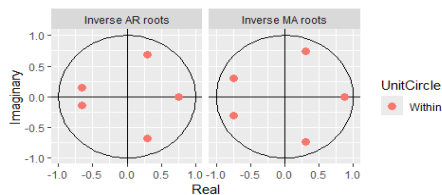
```
## z.lag.1   -1.02353   0.06177 -16.570 <2e-16 ***
## z.diff.lag 0.06566   0.04449  1.476  0.141
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.12 on 493 degrees of freedom
## Multiple R-squared:  0.4832, Adjusted R-squared:  0.4811
## F-statistic: 230.5 on 2 and 493 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -16.5696 137.2814
##
## Critical values for test statistics:
##      1pct 5pct 10pct
## tau2 -3.44 -2.87 -2.57
## phi1  6.47  4.61  3.79

## Test regression none
## Residuals:
##   Min     1Q  Median     3Q      Max
## -7.3253 -0.8714  0.0410  1.0683  6.8047
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## z.lag.1   -1.01360   0.06155 -16.469 <2e-16 ***
## z.diff.lag  0.06055   0.04444  1.362  0.174
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.123 on 494 degrees of freedom
## Multiple R-squared:  0.4806, Adjusted R-squared:  0.4785
## F-statistic: 228.5 on 2 and 494 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -16.4689
##
## Critical values for test statistics:
##      1pct 5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

15. Chạy và kiểm định mô hình ARIMA chuỗi lợi suất.

a) Mô hình ARIMA(5,0,5).

```
## arima(x = log_return, order = c(5, 0, 5))
##
## Coefficients:
##      ar1  ar2  ar3  ar4  ar5  ma1  ma2  ma3  ma4
##      0.0169 0.3053 -0.2824 0.0999 0.1873 0.0320 -0.3961 0.2187 -0.0664
## s.e. 0.4458 0.3576 0.6190 0.2565 0.3401 0.4376 0.3435 0.6454 0.2598
##      ma5 intercept
##      -0.3636 0.1628
## s.e. 0.3854 0.0595
##
## sigma^2 estimated as 4.366: log likelihood = -1073.74, aic = 2171.48
##
## Training set error measures:
##              ME  RMSE  MAE MPE MAPE  MASE  ACF1
## Training set 0.003570057 2.089456 1.433722 NaN Inf 0.6801143 0.0004672621
```



Ljung-Box test

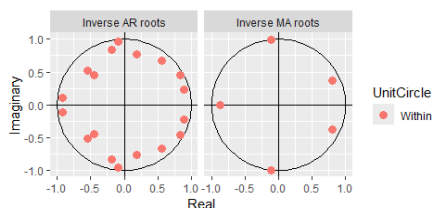
```
##
## data: Residuals from ARIMA(5,0,5) with non-zero mean
## Q* = 2.8165, df = 3, p-value = 0.4208
## Model df: 10. Total lags used: 13
```

b) Mô hình ARIMA(18,0,5).

```
## arima(x = log_return, order = c(18, 0, 5))
##
## Coefficients:
##      ar1  ar2  ar3  ar4  ar5  ar6  ar7  ar8
##      0.5553 -0.3288 0.1875 0.4344 -0.8178 0.1805 -0.1834 0.0672
## s.e. 0.1725 0.1881 0.1911 0.1947 0.1525 0.0752 0.0759 0.0840
##      ar9 ar10 ar11 ar12 ar13 ar14 ar15 ar16
##      -0.0473 -0.0635 -0.0019 -0.0865 0.1265 -0.1022 0.0270 -0.0747
## s.e. 0.0856 0.0834 0.0778 0.0735 0.0715 0.0679 0.0656 0.0629
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
##      ar17  ar18  ma1  ma2  ma3  ma4  ma5 intercept
##    -0.0483 -0.0481 -0.5128 0.2334 -0.1905 -0.4630 0.6710 0.1634
## s.e. 0.0608 0.0576 0.1681 0.1802 0.1717 0.1761 0.1385 0.0553
##
## sigma^2 estimated as 4.126: log likelihood = -1062.17, aic = 2174.34
##
## Training set error measures:
##           ME  RMSE  MAE MPE MAPE  MASE  ACF1
## Training set 0.003655175 2.031373 1.429586 NaN Inf 0.6781525 -0.0008185149
```



Ljung-Box test

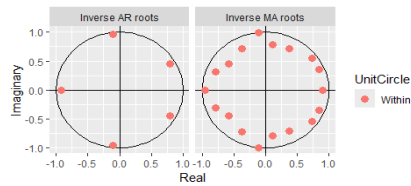
```
##
## data: Residuals from ARIMA(18,0,5) with non-zero mean
## Q* = 5.0913, df = 3, p-value = 0.1652
## Model df: 23. Total lags used: 26
```

c) Mô hình ARIMA(5,0,18).

```
## arima(x = log_return, order = c(5, 0, 18))
##
## Coefficients:
##      ar1  ar2  ar3  ar4  ar5  ma1  ma2  ma3
##    0.4585 -0.1509 -0.0217 0.4310 -0.7173 -0.4125 0.0653 0.0077
## s.e. 0.1700 0.1478 0.1446 0.1343 0.1417 0.1760 0.1536 0.1414
##      ma4  ma5  ma6  ma7  ma8  ma9  ma10  ma11
##   -0.4360 0.5706 0.1536 -0.1336 0.0075 -0.0057 -0.0704 -0.0018
## s.e. 0.1326 0.1484 0.0634 0.0622 0.0646 0.0640 0.0638 0.0623
##      ma12  ma13  ma14  ma15  ma16  ma17  ma18 intercept
##   -0.0412 0.0992 -0.0508 -0.0172 -0.0319 -0.0253 -0.0641 0.1618
## s.e. 0.0624 0.0669 0.0684 0.0543 0.0525 0.0615 0.0715 0.0567
##
## sigma^2 estimated as 4.158: log likelihood = -1063.98, aic = 2177.96
##
## Training set error measures:
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

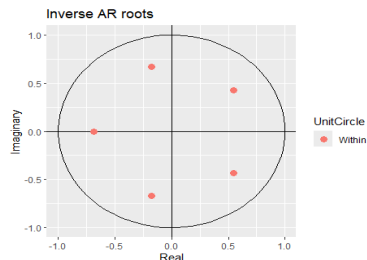
```
##          ME   RMSE   MAE MPE MAPE   MASE   ACF1
## Training set 0.00411781 2.039021 1.429853 NaN Inf 0.6782792 -0.002195243
```



```
##
## Ljung-Box test
##
## data: Residuals from ARIMA(5,0,18) with non-zero mean
## Q* = 6.147, df = 3, p-value = 0.1047
## Model df: 23. Total lags used: 26
```

d) Mô hình auto ARIMA(5,0,0).

```
## Coefficients:
##      ar1   ar2   ar3   ar4   ar5  mean
##    0.0460 -0.0701 -0.0401 0.0075 -0.1583 0.1647
## s.e. 0.0446  0.0446  0.0451 0.0450  0.0450 0.0777
##
## sigma^2 = 4.481: log likelihood = -1077.16
## AIC=2168.33 AICc=2168.56 BIC=2197.8
##
## Training set error measures:
##          ME   RMSE   MAE MPE MAPE   MASE   ACF1
## Training set 0.001733249 2.104128 1.449586 NaN Inf 0.6876398 0.002258355
```



```
## Ljung-Box test
##
## data: Residuals from ARIMA(5,0,0) with non-zero mean
## Q* = 6.0777, df = 5, p-value = 0.2987
## Model df: 5. Total lags used: 10
```

16. So sánh dự báo của ARIMA(5,0,5) với chuỗi lợi suất thực tế và tính các sai số.

```
rmse(log_return_real, fc_log_return)
```

```
## [1] 1.944191
```

3. Đỗ Quang Phước

Mo hình Lin Lin

##

Call:

lm(formula = rev1 ~ t1)

##

Residuals:

Min 1Q Median 3Q Max

-234.379 -114.594 -4.505 100.106 286.439

##

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 589.952 38.294 15.406 < 2e-16 ***

t1 5.986 1.169 5.122 4.16e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##

Residual standard error: 141.4 on 54 degrees of freedom

Multiple R-squared: 0.3269, Adjusted R-squared: 0.3145

F-statistic: 26.23 on 1 and 54 DF, p-value: 4.161e-06

[1] 138.8224

[1] 0.157557

1 2 3 4

931.16 937.14 943.13 949.11

[1] 309.149

[1] 0.2156405

Mo hình Lin Log

##

Call:

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
## lm(formula = rev1 ~ log(t1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -235.024 -118.998  -3.974  122.196  253.242
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  395.25     64.93   6.087 1.23e-07 ***
## log(t1)      118.69     20.27   5.856 2.90e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 134.8 on 54 degrees of freedom
## Multiple R-squared:  0.3884, Adjusted R-squared:  0.3771
## F-statistic: 34.29 on 1 and 54 DF, p-value: 2.898e-07
## [1] 132.3321
## [1] 0.1479111
##      1      2      3      4
## 875.13 877.20 879.22 881.22
## [1] 367.153
## [1] 0.2669128
Mo hình Log Lin
##
## Call:
## lm(formula = log(rev1) ~ t1)
##
## Residuals:
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
##      Min      1Q  Median      3Q      Max
## -0.36907 -0.15087 -0.00184  0.14633  0.34604
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.374562   0.050404 126.469 < 2e-16 ***
## t1           0.008222   0.001538   5.344 1.87e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1861 on 54 degrees of freedom
## Multiple R-squared:  0.346, Adjusted R-squared:  0.3338
## F-statistic: 28.56 on 1 and 54 DF, p-value: 1.871e-06
## [1] 140.4618
## [1] 0.1550435
##      1      2      3      4
## 937.49 945.23 953.03 960.90
## [1] 300.1979
## [1] 0.208309
Mo hình Log Log
##
## Call:
## lm(formula = log(rev1) ~ log(t1))
##
## Residuals:
##      Min      1Q  Median      3Q      Max
## -0.31897 -0.14763  0.00797  0.14964  0.29844
##
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 6.08284 0.08208 74.108 < 2e-16 ***

log(t1) 0.17092 0.02562 6.671 1.41e-08 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##

Residual standard error: 0.1704 on 54 degrees of freedom

Multiple R-squared: 0.4518, Adjusted R-squared: 0.4416

F-statistic: 44.5 on 1 and 54 DF, p-value: 1.412e-08

[1] 133.0075

[1] 0.1448012

1 2 3 4

874.69 877.30 879.86 882.40

[1] 366.5778

[1] 0.2666646

Mo hình xu thế tuyến tính mùa vụ đang công

##

Call:

lm(formula = rev1 ~ t1 + s2 + s3 + s4)

##

Residuals:

Min 1Q Median 3Q Max

-253.01 -103.82 -16.92 100.38 249.23

##

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 529.802 47.455 11.164 2.64e-15 ***

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
## t1      5.780    1.128  5.123 4.66e-06 ***
## s2      53.577   51.470  1.041  0.3028
## s3      73.512   51.507  1.427  0.1596
## s4     137.018   51.569  2.657  0.0105 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 136.1 on 51 degrees of freedom
## Multiple R-squared:  0.4105, Adjusted R-squared:  0.3642
## F-statistic: 8.877 on 4 and 51 DF, p-value: 1.613e-05
## [1] 129.9238
## [1] 0.1480687
##      1      2      3      4
## 859.25 918.61 944.32 1013.61
## [1] 296.3502
## [1] 0.2257767
```

Mo hình tuyến tính có mua vụ đang nhan

```
##
## Call:
## lm(formula = rev1 ~ t1 + tr2 + tr3 + tr4)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -257.01 -108.31 -10.20   96.24  252.63
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  592.969    37.551  15.791  <2e-16 ***
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
## t1      4.171    1.561  2.672  0.0101 *
```

```
## tr2     1.176    1.643  0.716  0.4775
```

```
## tr3     1.907    1.623  1.175  0.2454
```

```
## tr4     3.555    1.603  2.217  0.0311 *
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 138.5 on 51 degrees of freedom
```

```
## Multiple R-squared:  0.3895, Adjusted R-squared:  0.3417
```

```
## F-statistic: 8.136 on 4 and 51 DF, p-value: 3.74e-05
```

```
## [1] 132.2085
```

```
## [1] 0.151284
```

```
##      1      2      3      4
```

```
## 830.70 903.07 951.54 1056.52
```

```
## [1] 288.0841
```

```
## [1] 0.2271355
```

Mo hình Holt Winter dạng cong

```
## Holt-Winters exponential smoothing with trend and additive seasonal component.
```

```
##
```

```
## Call:
```

```
## HoltWinters(x = rev1, seasonal = "a")
```

```
##
```

```
## Smoothing parameters:
```

```
## alpha: 0.7367552
```

```
## beta : 0.01296075
```

```
## gamma: 0.4912907
```

```
##
```

```
## Coefficients:
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
##      [,1]
```

```
## a 987.98838
```

```
## b 14.74969
```

```
## s1 -80.91137
```

```
## s2 -36.71584
```

```
## s3 -11.95649
```

```
## s4 51.70186
```

```
## [1] 83.04816
```

```
## [1] 0.08064433
```

```
##      Qtr1  Qtr2  Qtr3  Qtr4
```

```
## 2024 921.83 980.77 1020.28 1098.69
```

```
## [1] 226.2406
```

```
## [1] 0.1669799
```

Mo hình Holt Winter dạng nhân

```
## Holt-Winters exponential smoothing with trend and multiplicative seasonal component.
```

```
##
```

```
## Call:
```

```
## HoltWinters(x = rev1, seasonal = "m")
```

```
##
```

```
## Smoothing parameters:
```

```
## alpha: 0.7500857
```

```
## beta : 0.0145502
```

```
## gamma: 0.8659548
```

```
##
```

```
## Coefficients:
```

```
##      [,1]
```

```
## a 1012.6973404
```

```
## b 14.4776496
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
## s1 0.8793330
```

```
## s2 0.9130726
```

```
## s3 0.9546391
```

```
## s4 1.0231669
```

```
## [1] 83.23577
```

```
## [1] 0.07832165
```

```
##      Qtr1  Qtr2  Qtr3  Qtr4
```

```
## 2024 903.2288 951.1045 1008.2233 1095.4106
```

```
## [1] 239.3034
```

```
## [1] 0.1808471
```

Sung dung mo hinh Holt Winter dang cong de du bao cho nam 2025

```
## Holt-Winters exponential smoothing with trend and additive seasonal component.
```

```
##
```

```
## Call:
```

```
## HoltWinters(x = rev, seasonal = "a")
```

```
##
```

```
## Smoothing parameters:
```

```
## alpha: 0.7902633
```

```
## beta : 0
```

```
## gamma: 0.5749062
```

```
##
```

```
## Coefficients:
```

```
##      [,1]
```

```
## a 1347.778811
```

```
## b 20.500000
```

```
## s1 -74.009101
```

```
## s2 -23.280092
```

```
## s3 -7.511381
```

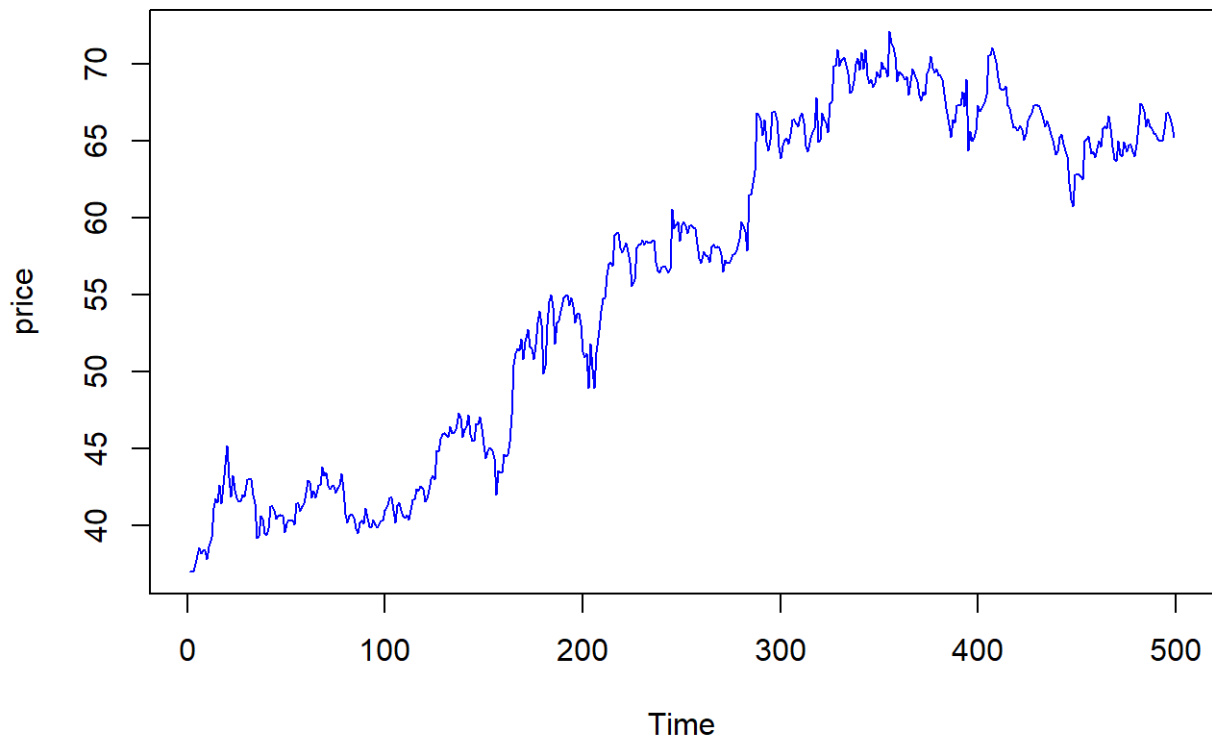
Bài tập lớn Phân tích chuỗi thời gian và dự báo

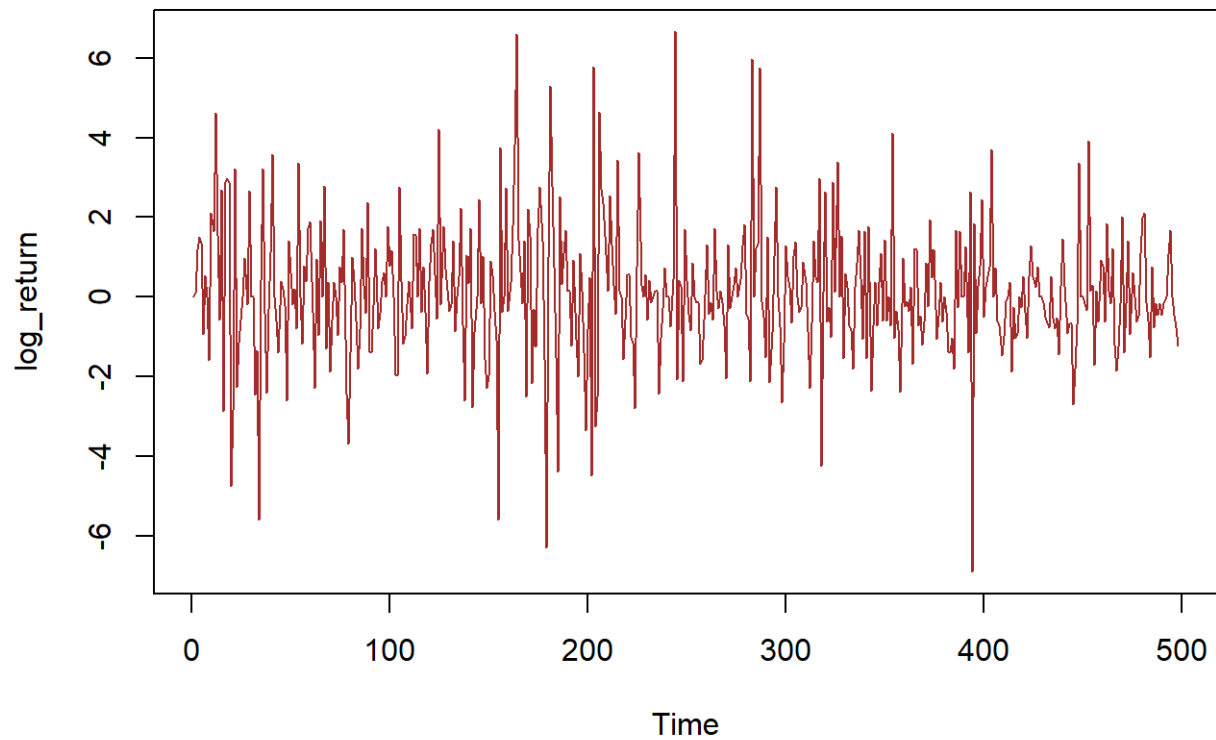
s4 57.266541

Qtr1 Qtr2 Qtr3 Qtr4

2025 1294.27 1365.50 1401.77 1487.05

PHAN 2





Kiểm định nghiệm đơn vị cho chuỗi giá

```
##
```

```
## #####
```

```
## # Augmented Dickey-Fuller Test Unit Root Test #
```

```
## #####
```

```
##
```

```
## Test regression trend
```

```
##
```

```
##
```

```
## Call:
```

```
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
```

```
##
```

```
## Residuals:
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
##   Min    1Q  Median    3Q    Max
## -4.5231 -0.4963 -0.0538  0.4372  3.8641
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.850740   0.383741   2.217  0.0271 *
## z.lag.1     -0.018914   0.009721  -1.946  0.0523 .
## tt          0.001072   0.000738   1.453  0.1468
## z.diff.lag  -0.024801   0.045204  -0.549  0.5835
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9147 on 493 degrees of freedom
## Multiple R-squared:  0.01031,   Adjusted R-squared:  0.004284
## F-statistic: 1.711 on 3 and 493 DF,  p-value: 0.1637
##
##
## Value of test-statistic is: -1.9457 2.199 2.278
##
## 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
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

##

Test regression none

##

##

Call:

lm(formula = z.diff ~ z.lag.1 - 1 + z.diff.lag)

##

Residuals:

Min 1Q Median 3Q Max

-4.5944 -0.4905 -0.0320 0.4636 3.8622

##

Coefficients:

Estimate Std. Error t value Pr(>|t|)

z.lag.1 0.0007945 0.0007222 1.100 0.272

z.diff.lag -0.0332470 0.0450024 -0.739 0.460

##

Residual standard error: 0.9178 on 495 degrees of freedom

Multiple R-squared: 0.003336, Adjusted R-squared: -0.0006911

F-statistic: 0.8284 on 2 and 495 DF, p-value: 0.4374

##

##

Value of test-statistic is: 1.1001

##

Critical values for test statistics:

1pct 5pct 10pct

tau1 -2.58 -1.95 -1.62

##

#####

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
## # 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
## -4.5789 -0.4877 -0.0383  0.4547  3.8365
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.1246812  0.0832147   1.498   0.135
## z.lag.1     -1.0593832  0.0649529 -16.310 <2e-16 ***
## tt          -0.0002581  0.0002885  -0.895   0.371
## z.diff.lag   0.0230682  0.0451206   0.511   0.609
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9189 on 492 degrees of freedom
## Multiple R-squared:  0.5176, Adjusted R-squared:  0.5147
## F-statistic: 176 on 3 and 492 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -16.31 88.6767 133.0136
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

##

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

##

#####

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

-4.6171 -0.5007 -0.0374 0.4643 3.8386

##

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.06011 0.04143 1.451 0.147

z.lag.1 -1.05655 0.06486 -16.289 <2e-16 ***

z.diff.lag 0.02170 0.04509 0.481 0.631

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Bài tập lớn Phân tích chuỗi thời gian và dự báo

##

Residual standard error: 0.9187 on 493 degrees of freedom

Multiple R-squared: 0.5168, Adjusted R-squared: 0.5149

F-statistic: 263.7 on 2 and 493 DF, p-value: < 2.2e-16

##

##

Value of test-statistic is: -16.2891 132.6684

##

Critical values for test statistics:

1pct 5pct 10pct

tau2 -3.44 -2.87 -2.57

phi1 6.47 4.61 3.79

##

#####

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

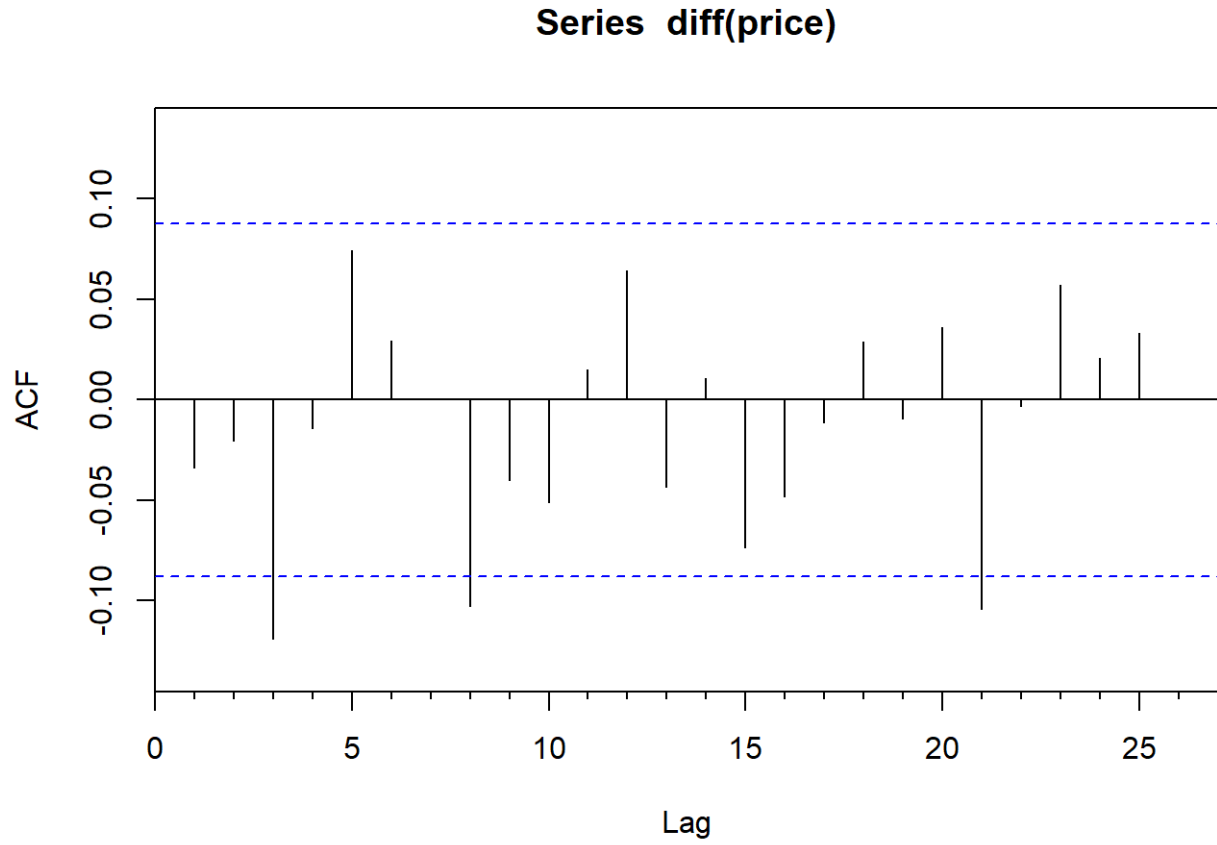
-4.5606 -0.4387 0.0189 0.5141 3.8994

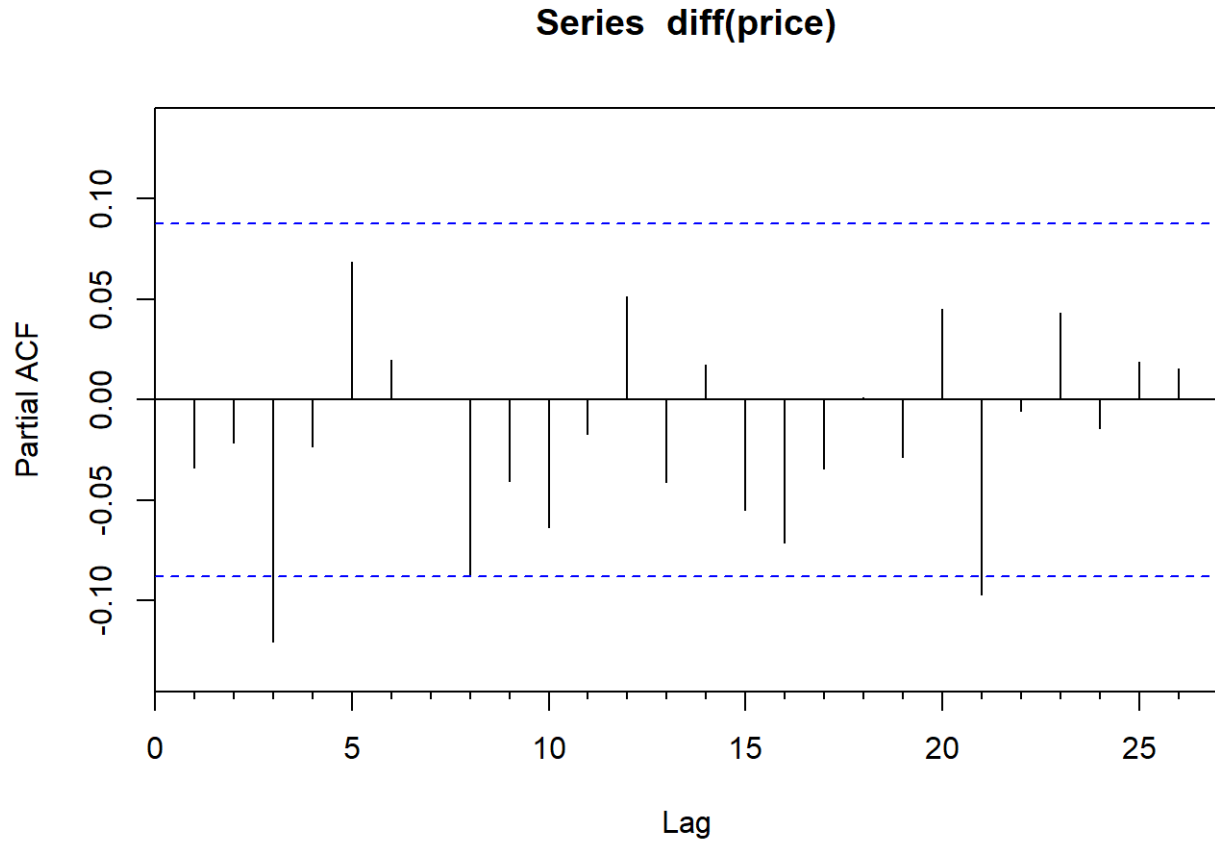
##

Coefficients:

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
##      Estimate Std. Error t value Pr(>|t|)
## z.lag.1  -1.04786   0.06466 -16.206  <2e-16 ***
## z.diff.lag 0.01732   0.04503  0.385   0.701
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9197 on 494 degrees of freedom
## Multiple R-squared:  0.5148, Adjusted R-squared:  0.5128
## F-statistic: 262 on 2 and 494 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -16.2063
##
## Critical values for test statistics:
##      1pct 5pct 10pct
## tau1 -2.58 -1.95 -1.62
```



```
# Lua chon mo hinh ARIMA
```

```
##
```

```
## Call:
```

```
## arima(x = price, order = c(3, 1, 3))
```

```
##
```

```
## Coefficients:
```

```
##      ar1    ar2    ar3    ma1    ma2    ma3
```

```
##      0.4566 -0.3083 -0.5792 -0.5196 0.3102 0.5244
```

```
## s.e. 0.3223 0.3435 0.3050 0.3310 0.3654 0.3082
```

```
##
```

```
## sigma^2 estimated as 0.8131: log likelihood = -655.3, aic = 1324.61
```

```
##
```

```
## Training set error measures:
```

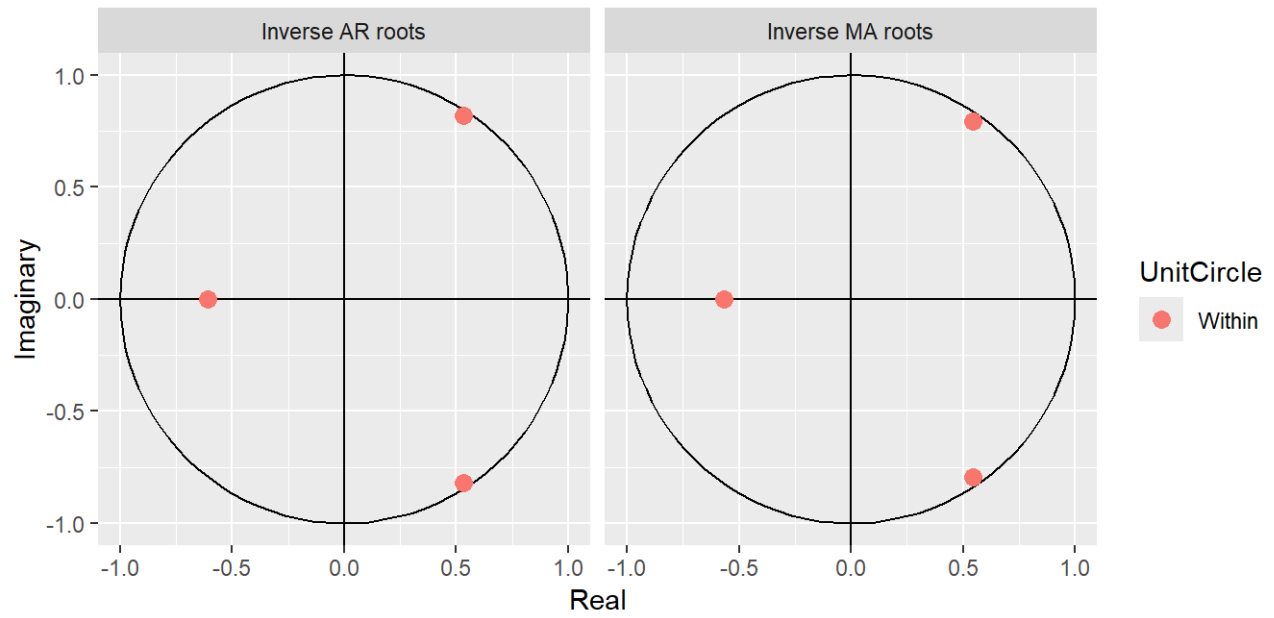
```
##           ME    RMSE    MAE    MPE    MAPE    MASE
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

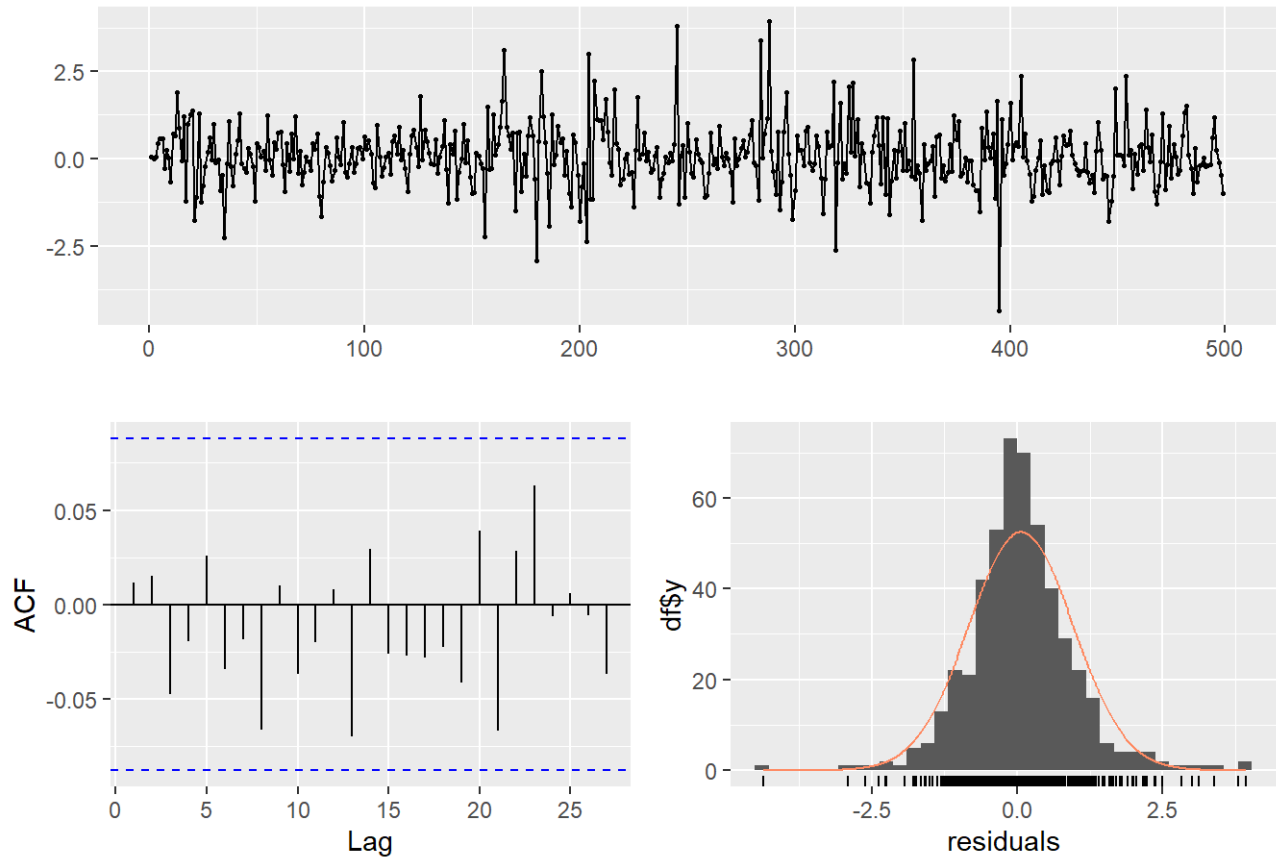
Training set 0.06184381 0.9007989 0.6495164 0.109501 1.186906 1.005015

ACF1

Training set 0.01168256



Residuals from ARIMA(3,1,3)



```
##
```

```
## Ljung-Box test
```

```
##
```

```
## data: Residuals from ARIMA(3,1,3)
```

```
## Q* = 5.6005, df = 4, p-value = 0.231
```

```
##
```

```
## Model df: 6. Total lags used: 10
```

```
## Series: price
```

```
## ARIMA(1,1,1) with drift
```

```
##
```

```
## Coefficients:
```

```
##      ar1      ma1      drift
```

```
##      0.8802 -0.9308  0.0560
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
## s.e. 0.0559 0.0416 0.0238
```

```
##
```

```
## sigma^2 = 0.8324: log likelihood = -659.49
```

```
## AIC=1326.99 AICc=1327.07 BIC=1343.83
```

```
##
```

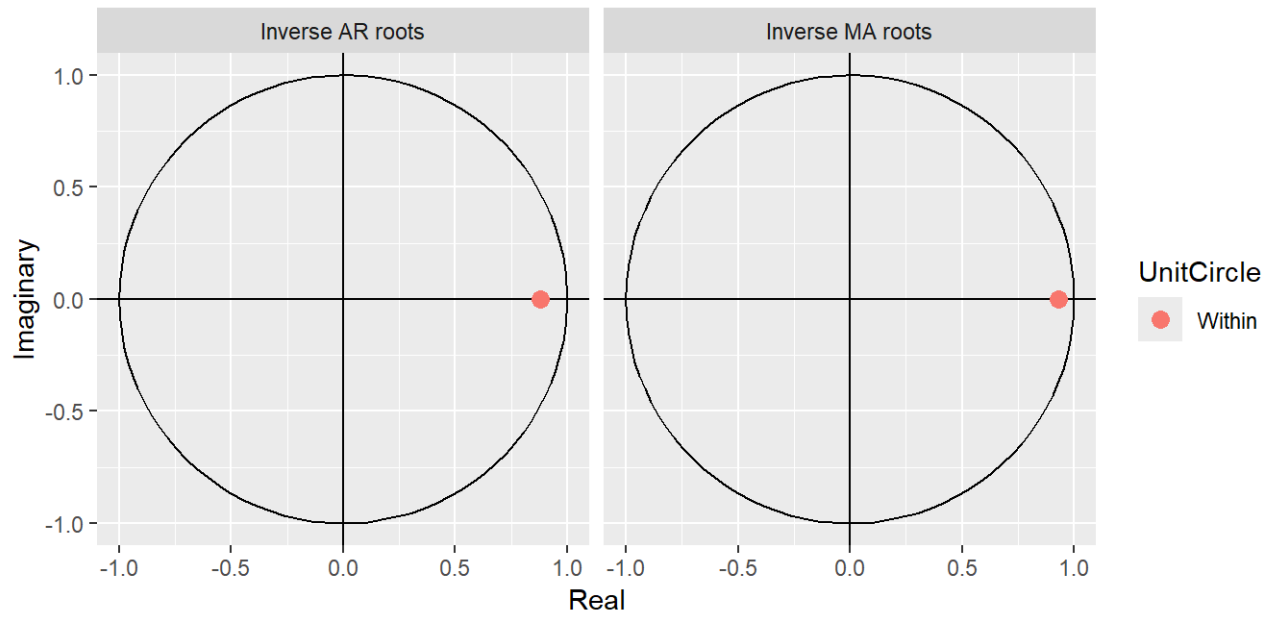
```
## Training set error measures:
```

```
##           ME    RMSE    MAE    MPE    MAPE    MASE
```

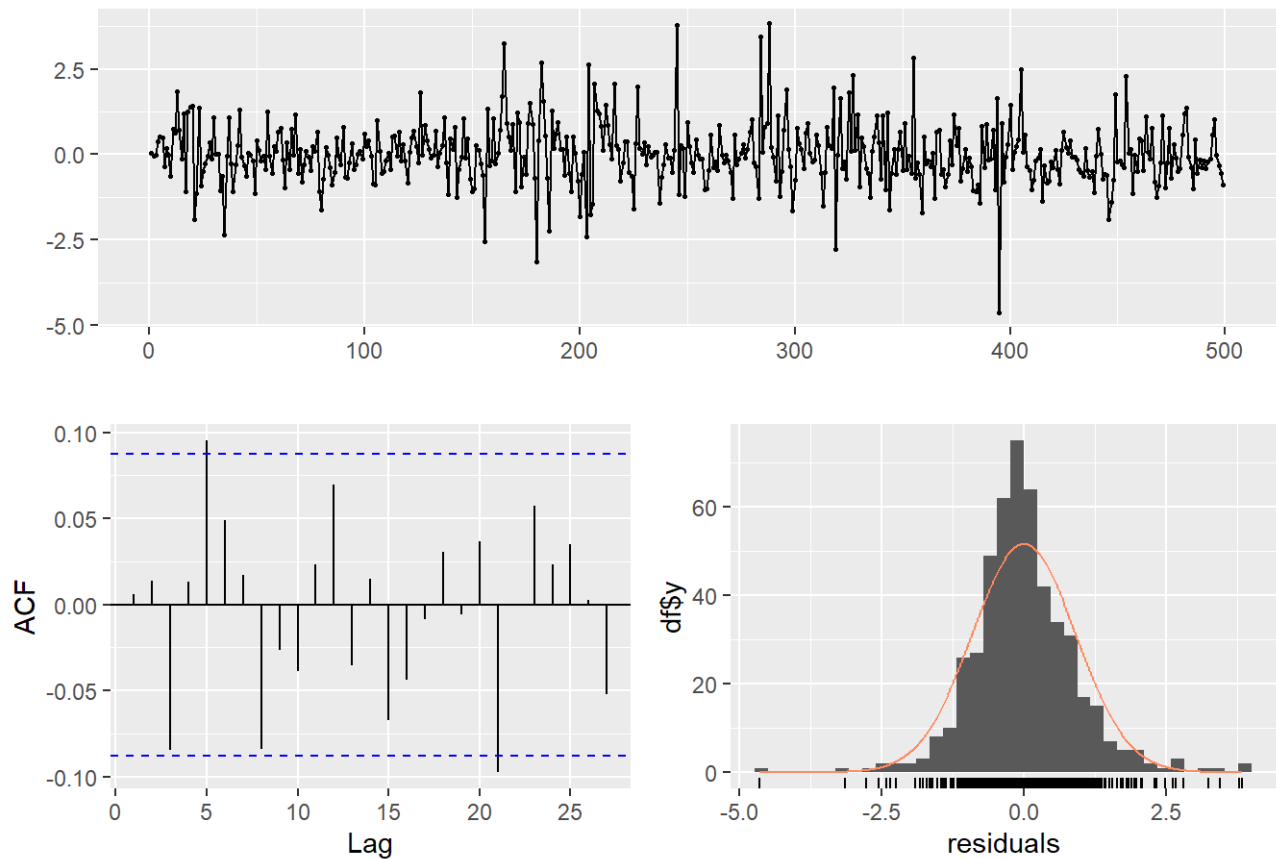
```
## Training set 0.002487233 0.9087019 0.6496175 -0.001517884 1.188033 1.005172
```

```
##           ACF1
```

```
## Training set 0.005695038
```



Residuals from ARIMA(1,1,1) with drift



```
##
```

```
## Ljung-Box test
```

```
##
```

```
## data: Residuals from ARIMA(1,1,1) with drift
```

```
## Q* = 14.459, df = 8, p-value = 0.07056
```

```
##
```

```
## Model df: 2. Total lags used: 10
```

```
## Warning in arima(price, c(8, 1, 8)): possible convergence problem: optim gave
```

```
## code = 1
```

```
##
```

```
## Call:
```

```
## arima(x = price, order = c(8, 1, 8))
```

```
##
```


Bài tập lớn Phân tích chuỗi thời gian và dự báo

Coefficients:

ar1 ar2 ar3 ar4 ar5 ar6 ar7 ar8

-0.0053 -0.0533 0.5963 -0.8744 0.4291 0.2321 0.5184 -0.4748

s.e. 0.5548 0.3061 0.1980 0.2868 0.3718 0.0979 0.1802 0.4560

ma1 ma2 ma3 ma4 ma5 ma6 ma7 ma8

-0.0293 0.0314 -0.6870 0.9107 -0.4124 -0.1788 -0.6045 0.4382

s.e. 0.5578 0.2723 0.1645 0.3489 0.3955 0.0838 0.1589 0.4937

##

sigma^2 estimated as 0.7956: log likelihood = -650.18, aic = 1334.36

##

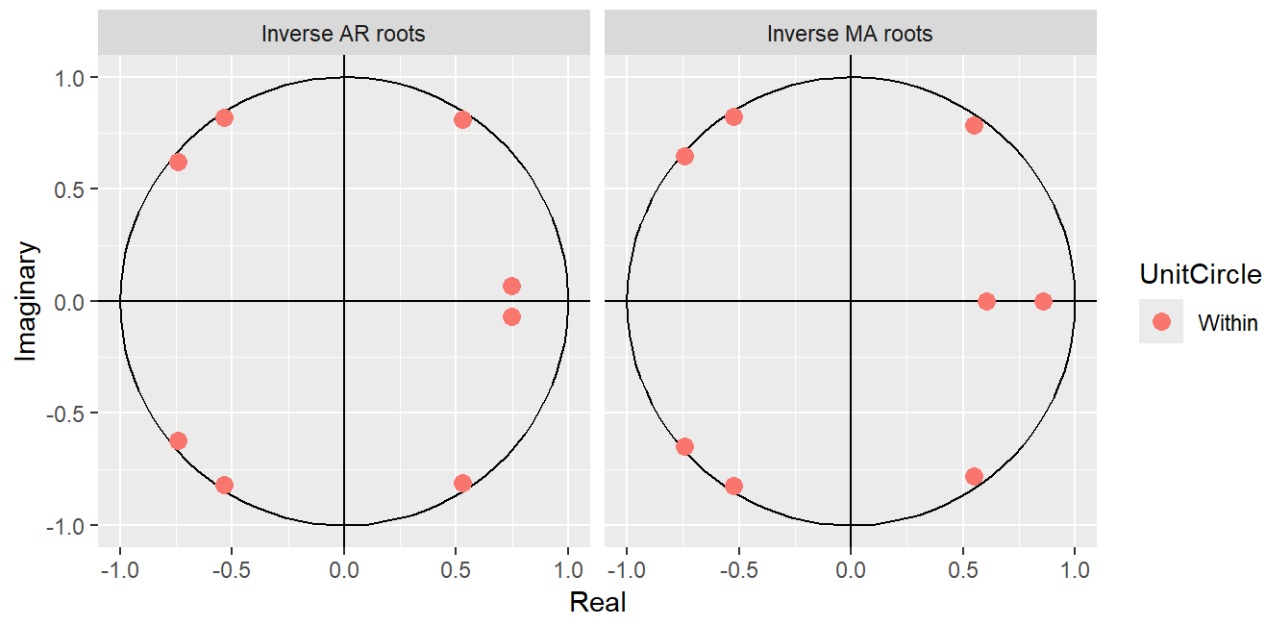
Training set error measures:

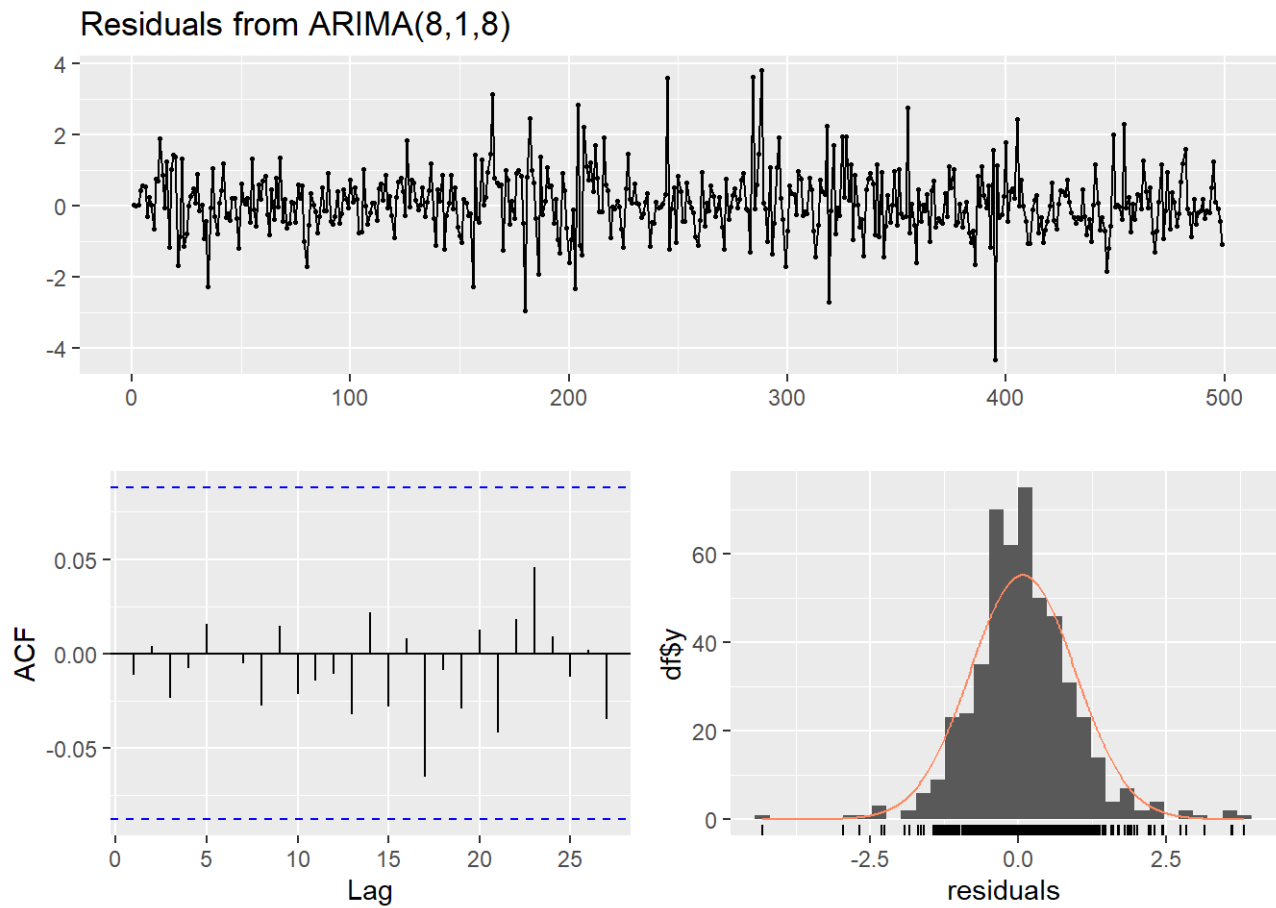
ME RMSE MAE MPE MAPE MASE

Training set 0.07666024 0.8910569 0.643906 0.1368482 1.178841 0.9963342

ACF1

Training set -0.01144687





```
##
```

```
## Ljung-Box test
```

```
##
```

```
## data: Residuals from ARIMA(8,1,8)
```

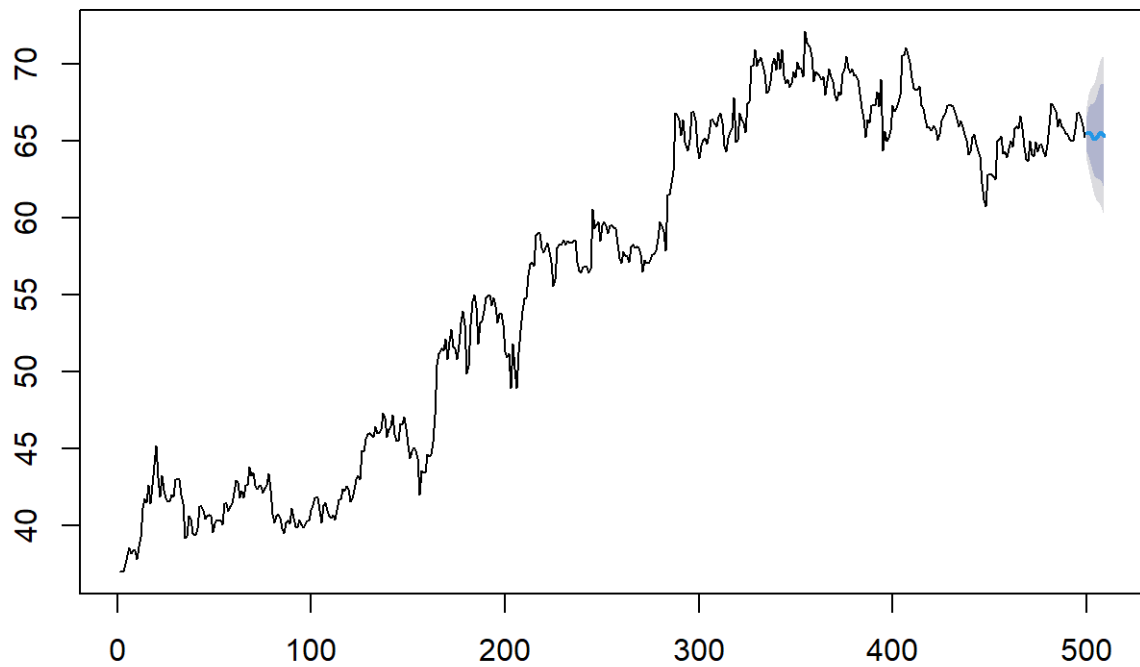
```
## Q* = 5.3448, df = 3, p-value = 0.1482
```

```
##
```

```
## Model df: 16. Total lags used: 19
```

Su dụng mô hình đã chọn để dự báo cho chuỗi giá

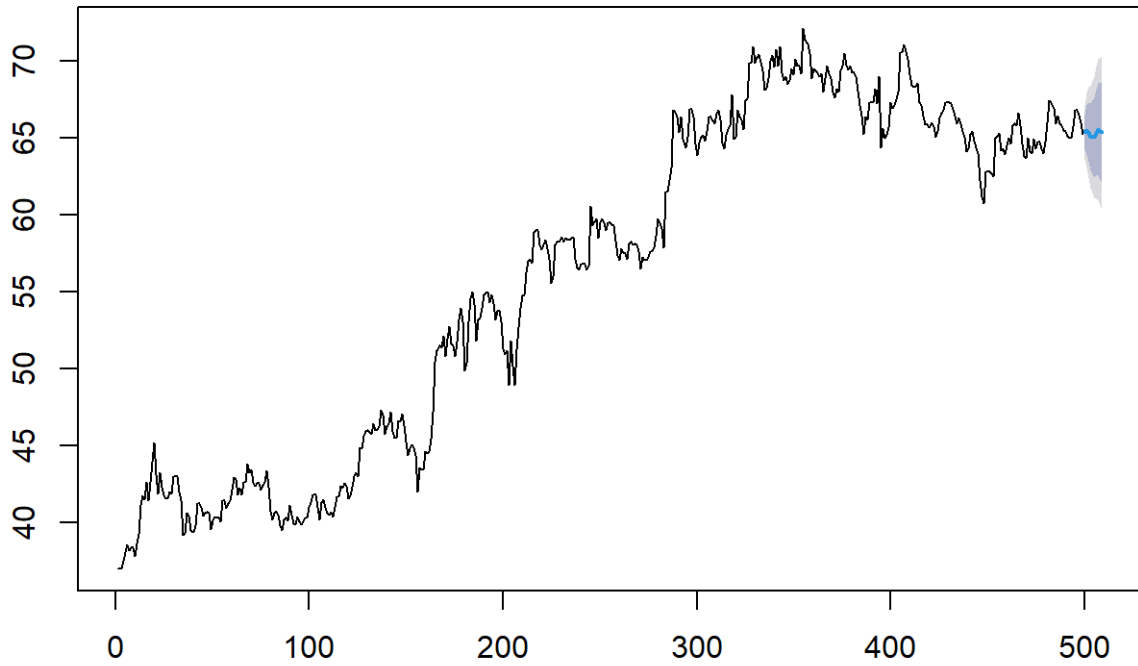
Forecasts from ARIMA(3,1,3)



```
## [1] 3.097038
```

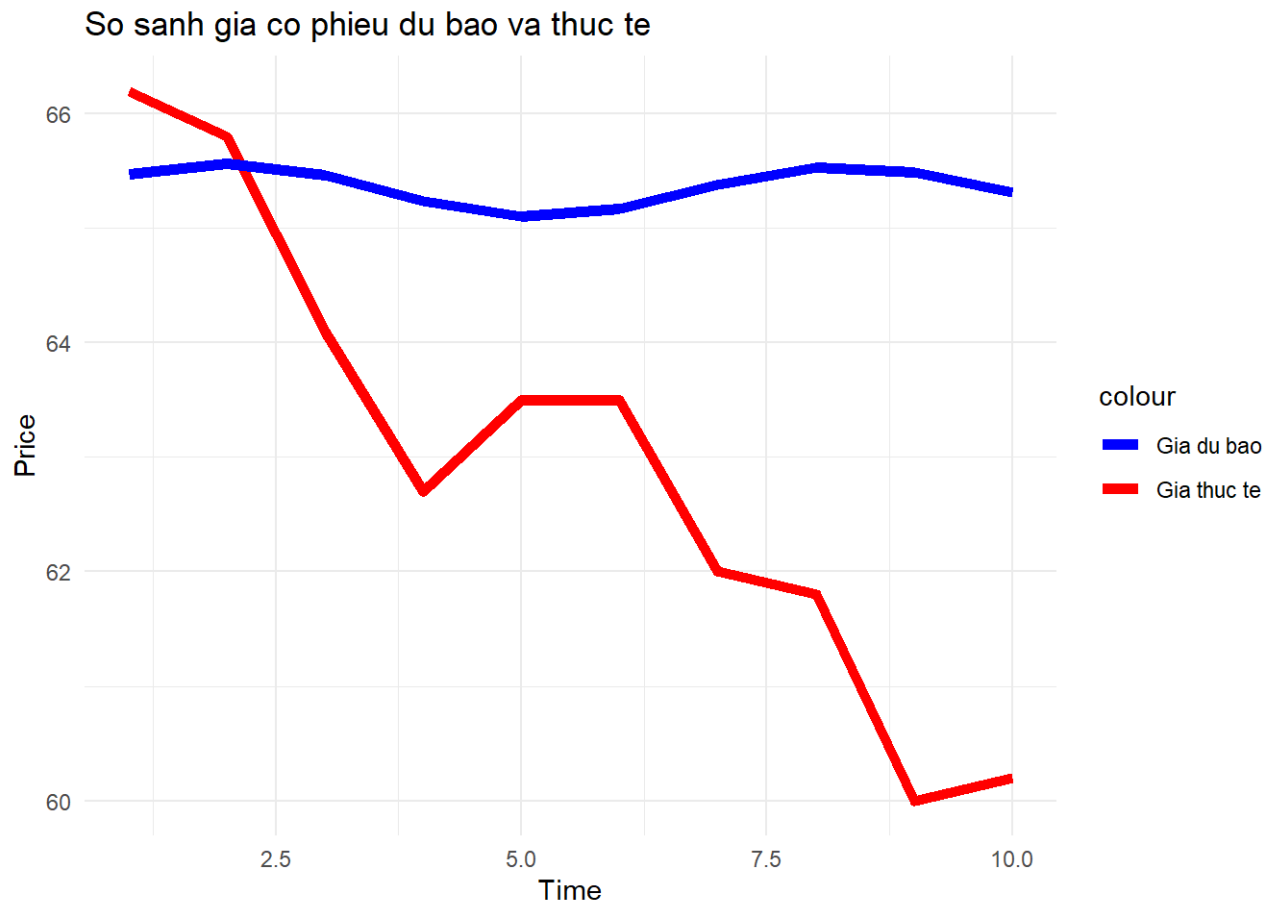
```
## [1] 0.04193374
```

Forecasts from ARIMA(8,1,8)



[1] 3.051248

[1] 0.04120276



Kiem dinh nghiem don vi cho chuoai Log renturn

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

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
##   Min   1Q Median   3Q   Max
## -6.9766 -0.8144 -0.0642  0.8042  6.5722
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.2685949  0.1526758   1.759  0.0792 .
## z.lag.1     -1.0688104  0.0641376 -16.664 <2e-16 ***
## tt          -0.0005870  0.0005289  -1.110  0.2676
## z.diff.lag   0.0548988  0.0450469   1.219  0.2235
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.684 on 492 degrees of freedom
## Multiple R-squared:  0.5078, Adjusted R-squared:  0.5048
## F-statistic: 169.2 on 3 and 492 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -16.6643 92.5698 138.8537
##
## 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
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

##

Test regression drift

##

##

Call:

lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

##

Residuals:

Min 1Q Median 3Q Max

-7.0632 -0.8723 -0.0956 0.8127 6.6111

##

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.12162 0.07598 1.601 0.110

z.lag.1 -1.06443 0.06403 -16.624 <2e-16 ***

z.diff.lag 0.05276 0.04502 1.172 0.242

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##

Residual standard error: 1.684 on 493 degrees of freedom

Multiple R-squared: 0.5066, Adjusted R-squared: 0.5046

F-statistic: 253.1 on 2 and 493 DF, p-value: < 2.2e-16

##

##

Value of test-statistic is: -16.6237 138.174

##

Critical values for test statistics:

1pct 5pct 10pct

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
## tau2 -3.44 -2.87 -2.57
## phi1 6.47 4.61 3.79
##
## #####
## # 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
## -6.9478 -0.7479  0.0213  0.9371  6.7043
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## z.lag.1   -1.05428    0.06382  -16.52  <2e-16 ***
## z.diff.lag  0.04765    0.04497   1.06   0.29
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.687 on 494 degrees of freedom
## Multiple R-squared:  0.504, Adjusted R-squared:  0.502
## F-statistic: 251 on 2 and 494 DF, p-value: < 2.2e-16
##
```

```
##
```

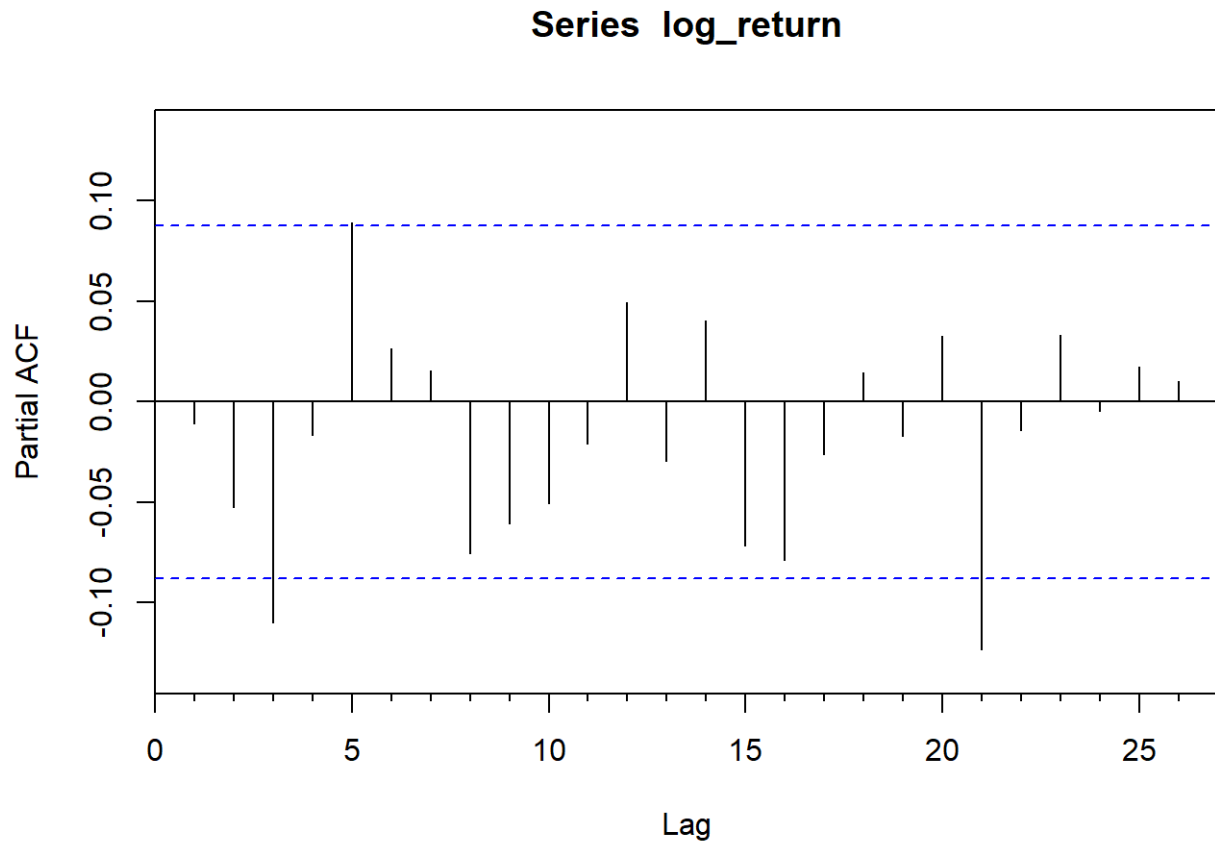
```
## Value of test-statistic is: -16.5204
```

```
##
```

```
## Critical values for test statistics:
```

```
##      1pct  5pct 10pct
```

```
## tau1 -2.58 -1.95 -1.62
```



```
##
```

```
## Call:
```

```
## arima(x = log_return, order = c(3, 0, 3))
```

```
##
```

```
## Coefficients:
```

```
##      ar1    ar2    ar3    ma1    ma2    ma3 intercept
```

```
##      0.4554 -0.3091 -0.5827 -0.5035  0.2770  0.5575   0.1155
```

```
## s.e. 0.7310  0.7792  0.6979  0.7447  0.8238  0.6973   0.0681
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

##

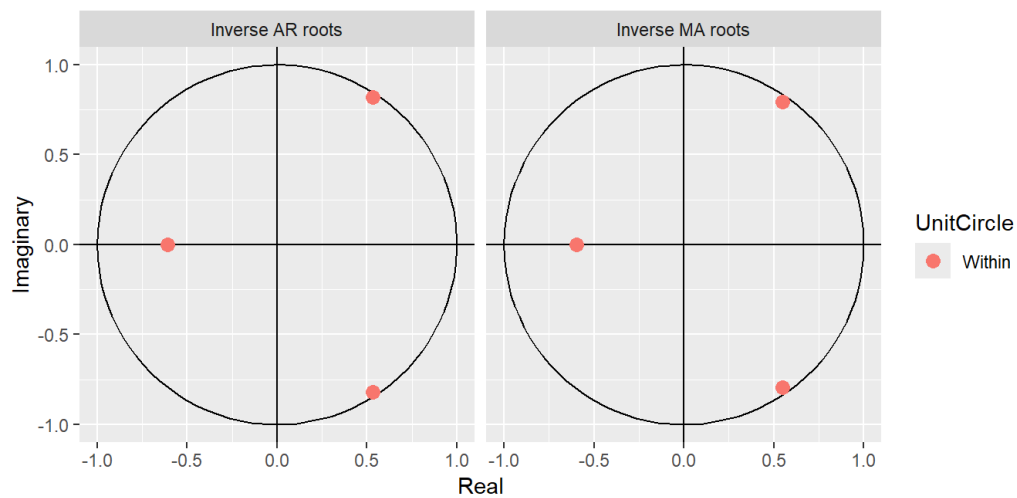
σ^2 estimated as 2.692: log likelihood = -953.56, aic = 1923.11

##

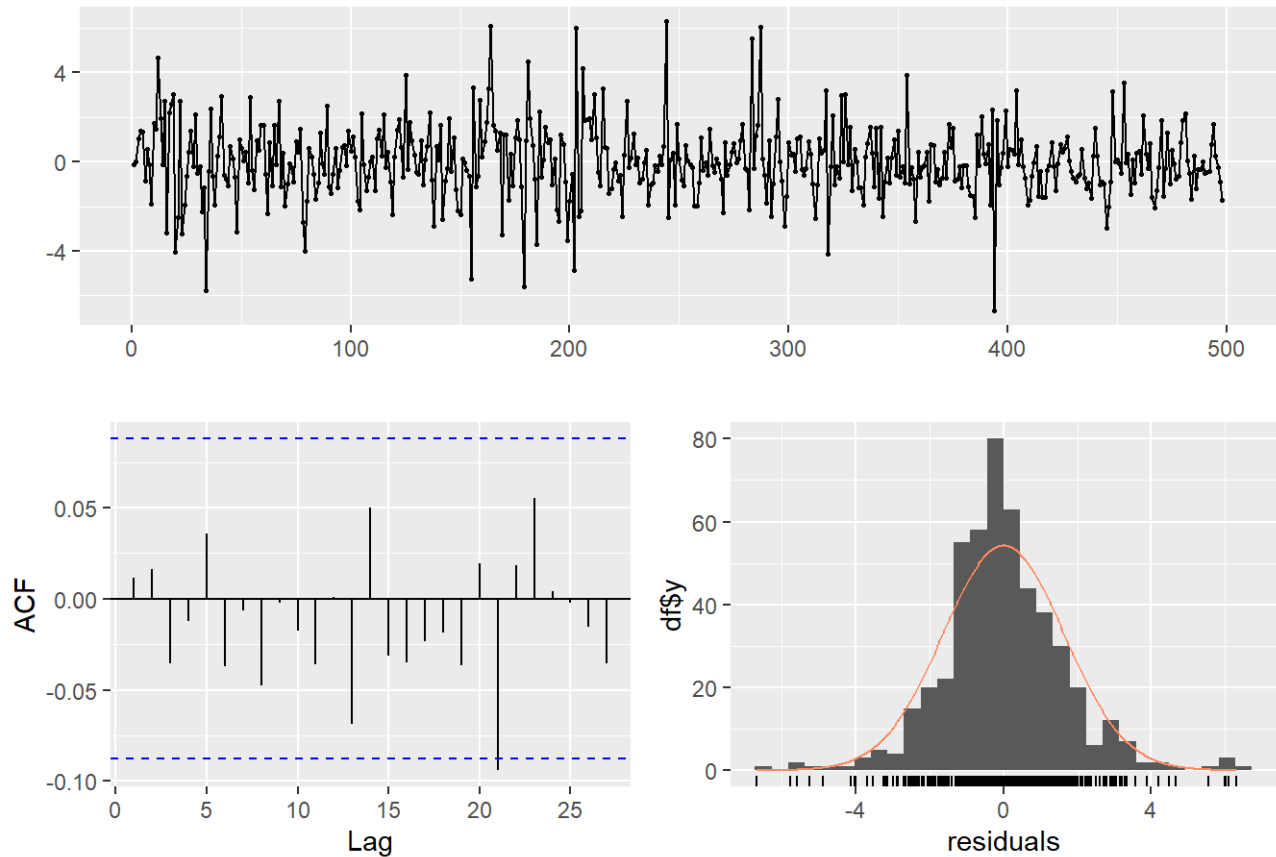
Training set error measures:

ME RMSE MAE MPE MAPE MASE ACF1

Training set -0.001585658 1.640848 1.196468 NaN Inf 0.6950206 0.01131949



Residuals from ARIMA(3,0,3) with non-zero mean



```
##
```

```
## Ljung-Box test
```

```
##
```

```
## data: Residuals from ARIMA(3,0,3) with non-zero mean
```

```
## Q* = 3.5717, df = 4, p-value = 0.4671
```

```
##
```

```
## Model df: 6. Total lags used: 10
```

```
## Warning in arima(log_return, c(5, 0, 5)): possible convergence problem: optim
```

```
## gave code = 1
```

```
##
```

```
## Call:
```

```
## arima(x = log_return, order = c(5, 0, 5))
```

```
##
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

Coefficients:

ar1 ar2 ar3 ar4 ar5 ma1 ma2 ma3

0.0038 -0.0261 -0.3067 -0.7137 0.4727 -0.0316 -0.0205 0.2518

s.e. 0.7820 0.3898 0.2178 0.3468 0.7347 0.8111 0.3724 0.1714

ma4 ma5 intercept

0.7555 -0.4602 0.1130

s.e. 0.2925 0.7603 0.0693

##

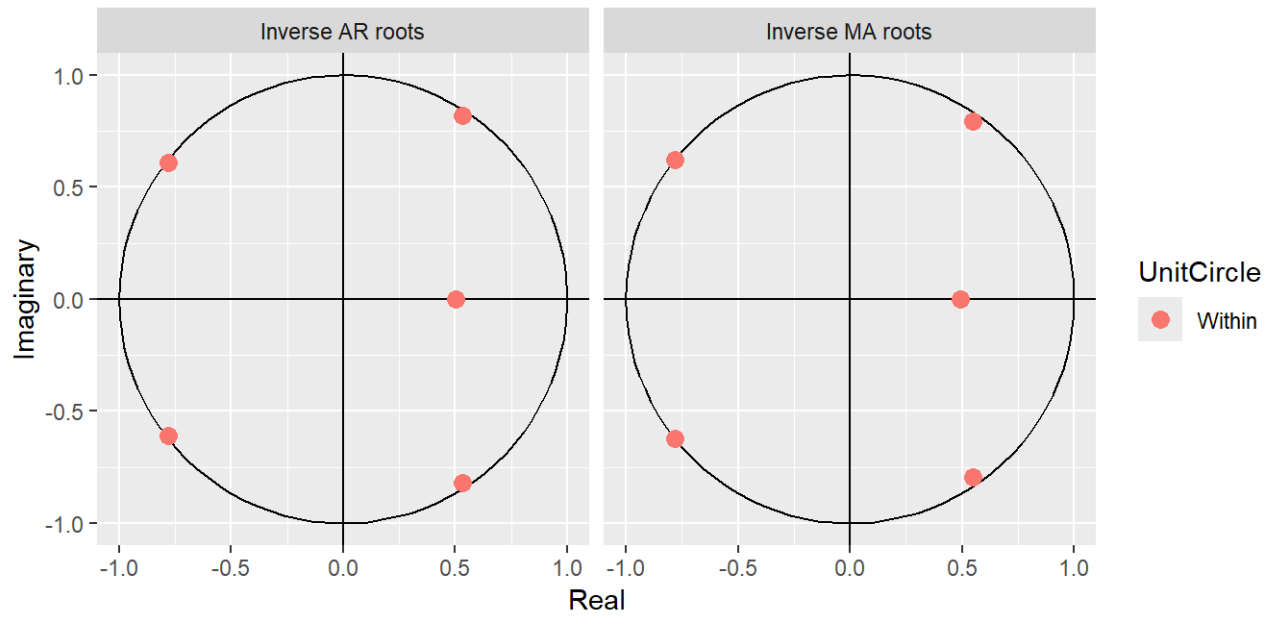
sigma^2 estimated as 2.639: log likelihood = -950.33, aic = 1924.66

##

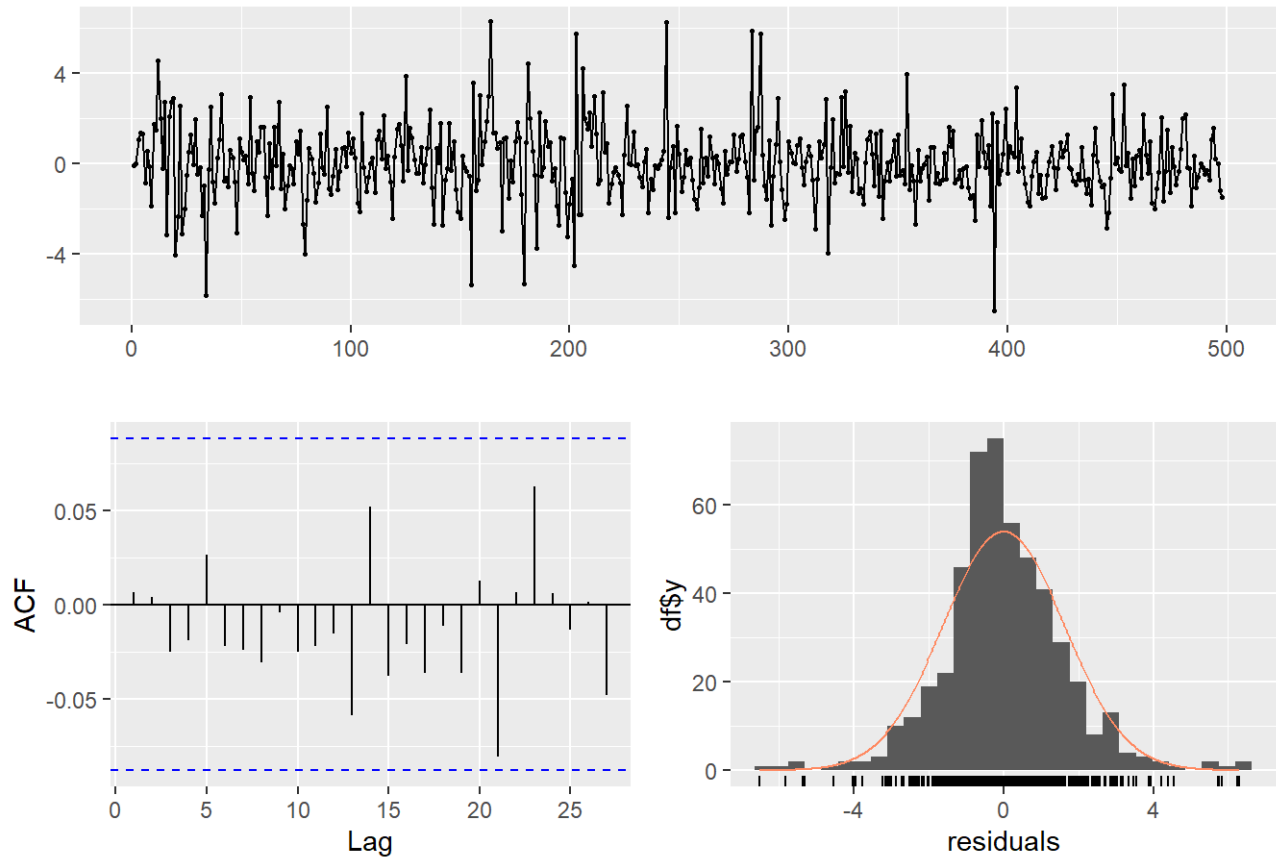
Training set error measures:

ME RMSE MAE MPE MAPE MASE ACF1

Training set 0.001186833 1.624614 1.187036 NaN Inf 0.6895416 0.006299839



Residuals from ARIMA(5,0,5) with non-zero mean



```
##
```

```
## Ljung-Box test
```

```
##
```

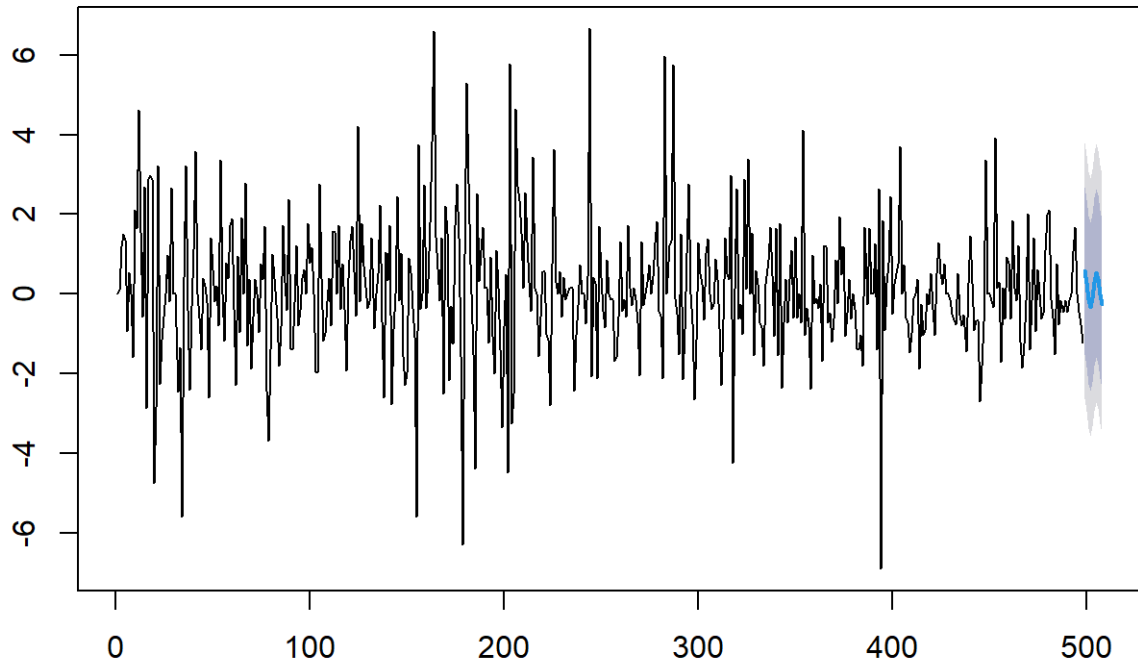
```
## data: Residuals from ARIMA(5,0,5) with non-zero mean
```

```
## Q* = 4.3464, df = 3, p-value = 0.2264
```

```
##
```

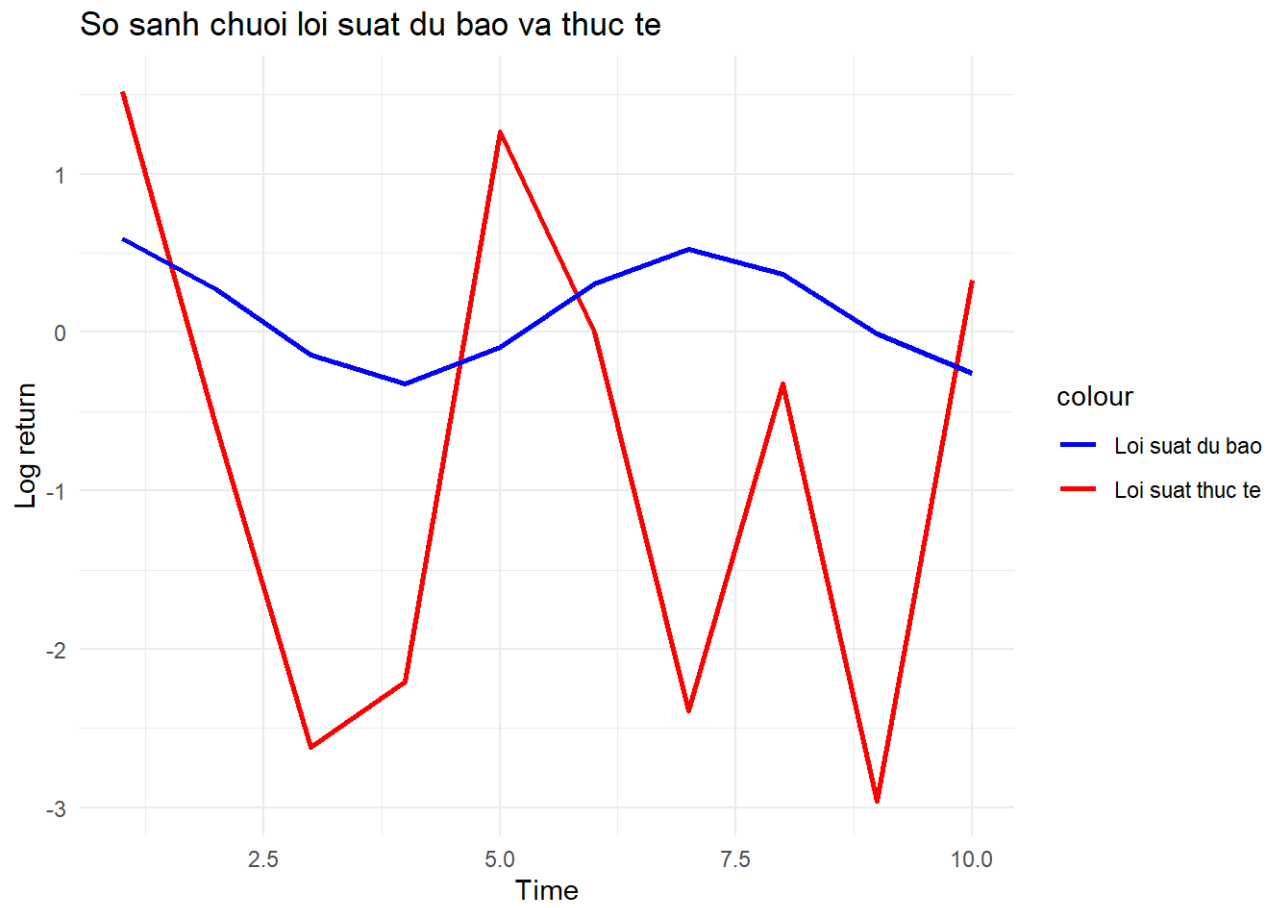
```
## Model df: 10. Total lags used: 13
```

Forecasts from ARIMA(3,0,3) with non-zero mean

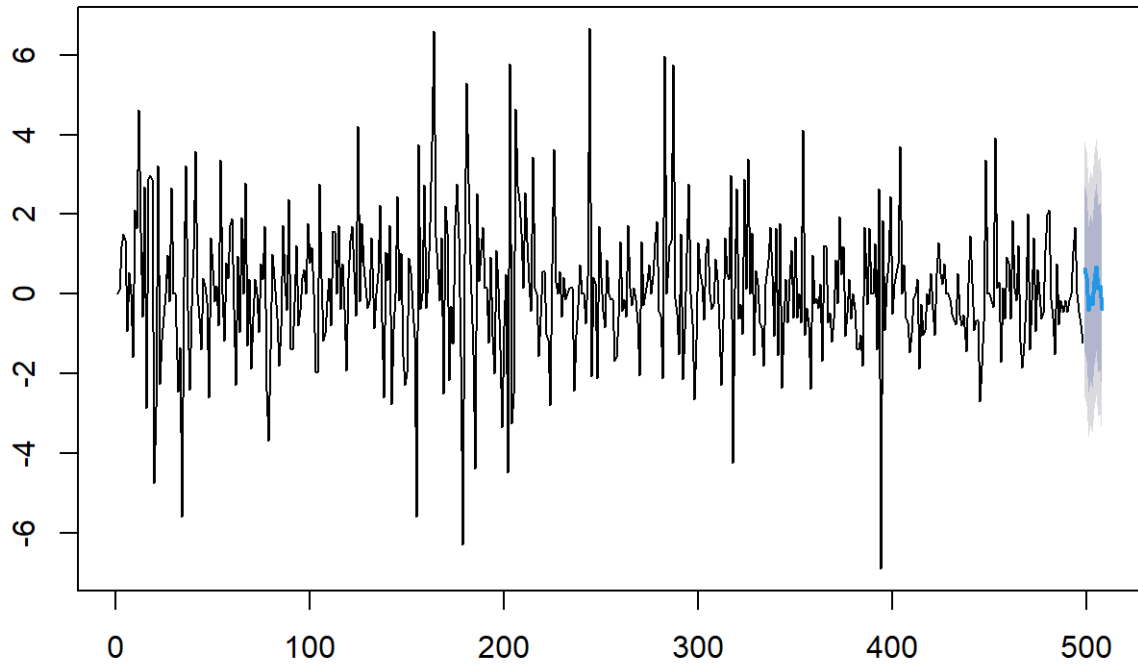


```
## [1] 1.768423
```

```
## [1] 1.814551
```

Forecasts from ARIMA(5,0,5) with non-zero mean



```
## [1] 1.83957
```

```
## [1] 1.812912
```

4. Đinh Ngọc Thảo

#Mô hình linear-linear

```
##  
## Call:  
## lm(formula = revenue1 ~ t1)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -1210532 -556947  -99612   310433  2769568   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept) 1936397    221723   8.733 4.22e-12 ***   
## t1          22921      6427   3.566 0.000742 ***   
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 840700 on 57 degrees of freedom  
## Multiple R-squared:  0.1824, Adjusted R-squared:  0.1681   
## F-statistic: 12.72 on 1 and 57 DF, p-value: 0.0007418
```

```
rmse(revenue1,fitted(reg_linlin))
```

```
## [1] 826366.8
```

```
mape(revenue1,fitted(reg_linlin))
```

```
## [1] 0.2387738
```

```
rmse(tail(revenue1,4), tail(fitted(reg_linlin),4))
```

```
## [1] 367424.1
```

```
mape(tail(revenue1,4), tail(fitted(reg_linlin),4))
```

```
## [1] 0.07722168
```

```
new1<-data.frame(t1=c(60:63))
```

```
pre1<-predict(reg_linlin,new1)
```

```
pre1
```

```
##      1      2      3      4
```

```
## 3311679 3334600 3357522 3380443
```

#Mô hình linear-log

```
##
## Call:
## lm(formula = revenue1 ~ log(t1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1226916 -605126 -64393  358877 2925446
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1479478    412206   3.589 0.000691 ***
## log(t1)      365944    126738   2.887 0.005480 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 868500 on 57 degrees of freedom
## Multiple R-squared:  0.1276, Adjusted R-squared:  0.1123
## F-statistic: 8.337 on 1 and 57 DF, p-value: 0.00548

rmse(revenue1,fitted(reg_linlog))
## [1] 853619.3

mape(revenue1,fitted(reg_linlog))
## [1] 0.2450729

rmse(tail(revenue1,4), tail(fitted(reg_linlog),4))
## [1] 566041.7

mape(tail(revenue1,4), tail(fitted(reg_linlog),4))
## [1] 0.1288417

pre2<-predict(reg_linlog,new1)
pre2
##      1      2      3      4
## 2977779 2983828 2989778 2995633

#Mô hình log-linear
##
## Call:
## lm(formula = log(revenue1) ~ t1)
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.51374 -0.19351 -0.00428  0.15249  0.71481
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.494727  0.075999 190.722 < 2e-16 ***
## t1           0.007787  0.002203   3.535 0.000819 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2882 on 57 degrees of freedom
## Multiple R-squared:  0.1798, Adjusted R-squared:  0.1654
## F-statistic: 12.49 on 1 and 57 DF, p-value: 0.0008185

rmse(revenue1,exp(fitted(reg_loglin)))

## [1] 829035.4

mape(revenue1,exp(fitted(reg_loglin)))

## [1] 0.2227881

rmse(tail(revenue1,4), tail(exp(fitted(reg_loglin)),4))

## [1] 471054.2

mape(tail(revenue1,4), tail(exp(fitted(reg_loglin)),4))

## [1] 0.09960655

pre3<-exp(predict(reg_loglin,new1))
pre3

##      1      2      3      4
## 3146945 3171545 3196338 3221325

#Mô hình log-log

##
## Call:
## lm(formula = log(revenue1) ~ log(t1))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.53143 -0.21481  0.00027  0.17257  0.74734
##
```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 14.30796   0.13949 102.575 < 2e-16 ***
## log(t1)     0.13440   0.04289   3.134 0.00272 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2939 on 57 degrees of freedom
## Multiple R-squared:  0.147, Adjusted R-squared:  0.132
## F-statistic: 9.821 on 1 and 57 DF, p-value: 0.002725
```

```
rmse(revenue1,exp(fitted(reg_loglog)))
```

```
## [1] 860204.4
```

```
mape(revenue1,exp(fitted(reg_loglog)))
```

```
## [1] 0.2296787
```

```
rmse(tail(revenue1,4), tail(exp(fitted(reg_loglog)),4))
```

```
## [1] 687228.7
```

```
mape(tail(revenue1,4), tail(exp(fitted(reg_loglog)),4))
```

```
## [1] 0.1704201
```

```
pre4<-exp(predict(reg_loglog,new1))
```

```
pre4
```

```
##      1      2      3      4
```

```
## 2836994 2843304 2849524 2855659
```

#Tạo biến giả mùa vụ theo quý:

```
s1 <- c(rep(c(1,0,0,0), 15))
```

```
s2 <- c(rep(c(0,1,0,0), 15))
```

```
s3 <- c(rep(c(0,0,1,0), 15))
```

```
s4 <- c(rep(c(0,0,0,1), 15))
```

```
s1 = s1[1:59]
```

```
s2 = s2[1:59]
```

```
s3 = s3[1:59]
```

```
s4 = s4[1:59]
```

#Mô hình xu thế tuyến tính dạng cộng

```
##
```

```
## Call:
```

```
## lm(formula = revenue1 ~ t1 + s2 + s3 + s4)
```

```
##
## Residuals:
##   Min     1Q   Median     3Q      Max
## -1215293 -481522 -27395  256517 2863413
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1837073   284392   6.460 3.1e-08 ***
## t1           23033     6392   3.604 0.000684 ***
## s2           434201    305091   1.423 0.160434
## s3          -127544    305292  -0.418 0.677767
## s4           75883     310489   0.244 0.807850
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 835300 on 54 degrees of freedom
## Multiple R-squared:  0.2354, Adjusted R-squared:  0.1787
## F-statistic: 4.155 on 4 and 54 DF, p-value: 0.005247

rmse(revenue1, fitted(reg_add))

## [1] 799163.5

rmse(tail(revenue1, 4), tail(fitted(reg_add),4))

## [1] 207848.9

mape(revenue1, fitted(reg_add))

## [1] 0.2269898

mape(tail(revenue1, 4), tail(fitted(reg_add),4))

## [1] 0.04739009

new_add<-data.frame(t1=c(60:63),s2=c(0,1,0,0),s3=c(0,0,1,0),s4=c(0,0,0,1))
pre_reg_add<-predict(reg_add,new_add)
pre_reg_add

##      1      2      3      4
## 3219064 3676298 3137586 3364046

#Mô hình xu thế tuyến tính dạng nhân

##
## Call:
## lm(formula = revenue1 ~ t1 + t1 * s2 + t1 * s3 + t1 * s4)
```

```
##
## Residuals:
##   Min     1Q   Median     3Q      Max
## -1185103 -521715  -4247  305444 2816312
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1768776   432217  4.092 0.000152 ***
## t1           25388    12803  1.983 0.052767 .
## s2           615713    619108  0.995 0.324666
## s3           30645     627131  0.049 0.961217
## s4          -15271     648719 -0.024 0.981311
## t1:s2        -6129     18106 -0.338 0.736378
## t1:s3        -5255     18106 -0.290 0.772824
## t1:s4         2960     19122  0.155 0.877598
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 856900 on 51 degrees of freedom
## Multiple R-squared:  0.24, Adjusted R-squared:  0.1357
## F-statistic: 2.301 on 7 and 51 DF, p-value: 0.04065

rmse(revenue1, fitted(reg_mul))

## [1] 796729.8

rmse(tail(revenue1, 4), tail(fitted(reg_mul), 4))

## [1] 232963.1

mape(revenue1, fitted(reg_mul))

## [1] 0.2308736

mape(tail(revenue1, 4), tail(fitted(reg_mul), 4))

## [1] 0.04547852

new_mul<-data.frame(t1=c(60:63),s2=c(0,1,0,0),s3=c(0,0,1,0),s4=c(0,0,0,1))
pre_reg_mul<-predict(reg_mul,new_mul)
pre_reg_mul

##      1      2      3      4
## 3292071 3559310 3047695 3539442

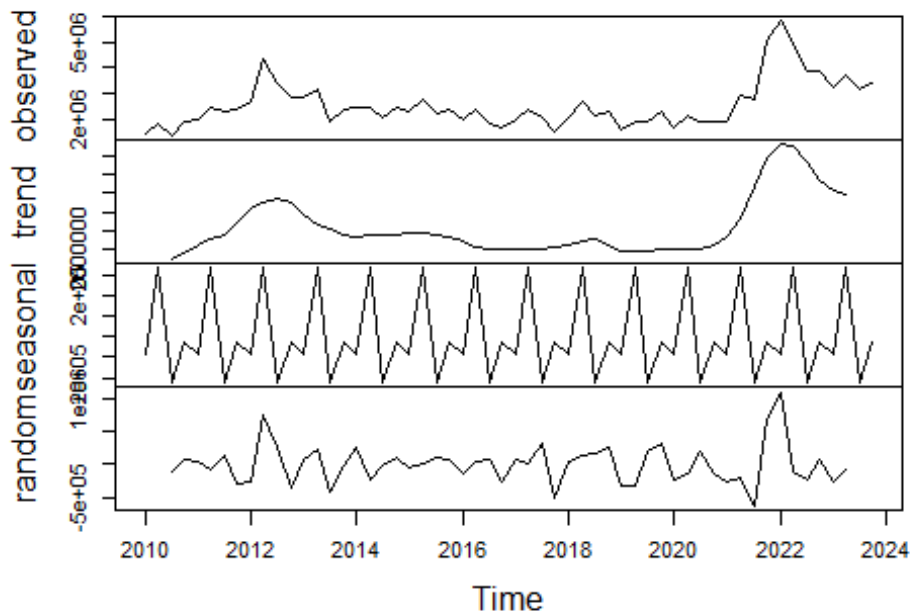
time<-seq_along(Revenue)
t2<-head(time,-4)
```



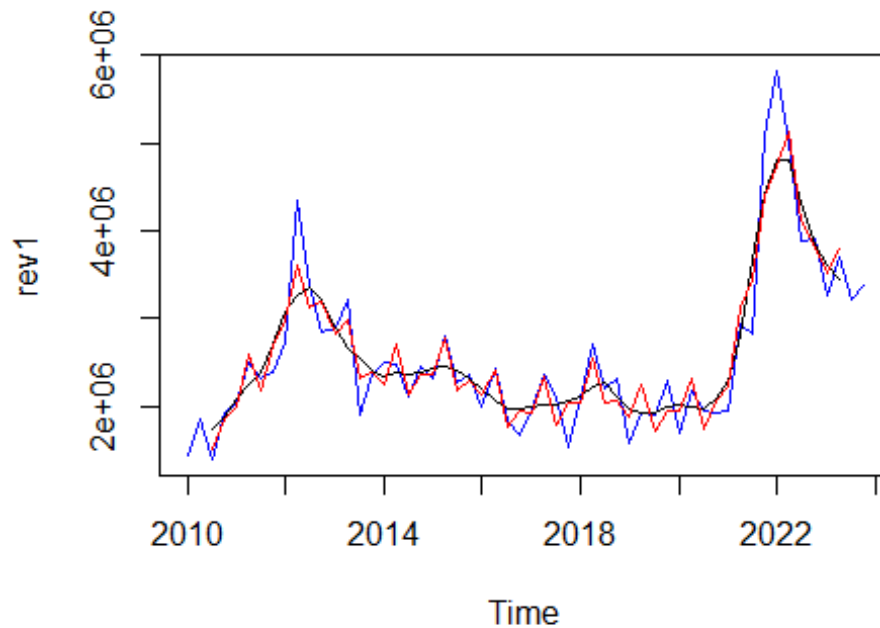
```
t3<-tail(time,4)
t3<-as.data.frame(t3)
train.set<-Revenue[1:56]
valid.set<-Revenue[57:60]
rev1<-ts(train.set,start = c(2010,1),frequency = 4)

#Phân tích cấu trúc chuỗi rev1
decom.rev1.a = decompose(rev1, type = 'additive')
plot(decom.rev1.a)
```

Decomposition of additive time series



```
rev1.trend.a = decom.rev1.a$trend
rev1.seas.a = decom.rev1.a$seasonal
rev1.rand.a = decom.rev1.a$random
rev1.trendseas.a = rev1 - rev1.rand.a
plot(rev1, col = 'blue')
lines(rev1.trend.a)
lines(rev1.trendseas.a, col = 'red')
```



```
rev1.weight.a = rev1.trendseas.a - rev1.trend.a
```

```
rev1.weight.a
```

##	Qtr1	Qtr2	Qtr3	Qtr4
## 2010	NA	NA	-223952.58	-27539.95
## 2011	-81154.57	332647.10	-223952.58	-27539.95
## 2012	-81154.57	332647.10	-223952.58	-27539.95
## 2013	-81154.57	332647.10	-223952.58	-27539.95
## 2014	-81154.57	332647.10	-223952.58	-27539.95
## 2015	-81154.57	332647.10	-223952.58	-27539.95
## 2016	-81154.57	332647.10	-223952.58	-27539.95
## 2017	-81154.57	332647.10	-223952.58	-27539.95
## 2018	-81154.57	332647.10	-223952.58	-27539.95
## 2019	-81154.57	332647.10	-223952.58	-27539.95
## 2020	-81154.57	332647.10	-223952.58	-27539.95
## 2021	-81154.57	332647.10	-223952.58	-27539.95
## 2022	-81154.57	332647.10	-223952.58	-27539.95
## 2023	-81154.57	332647.10	NA	NA

#Mô hình Holt-Winters dạng cộng

```
## Holt-Winters exponential smoothing with trend and additive seasonal component.
```

```
##
```

```
## Call:
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
## HoltWinters(x = rev1, seasonal = "a")
##
## Smoothing parameters:
## alpha: 1
## beta : 0.02470859
## gamma: 0
##
## Coefficients:
##      [,1]
## a 3319137.28
## b  69952.34
## s1 -25948.16
## s2 275465.97
## s3 -312201.53
## s4  62683.72

f_add<-forecast(hw.add,h=4)
f_add$mean

##      Qtr1  Qtr2  Qtr3  Qtr4
## 2024 3363141 3734508 3216793 3661630

accuracy(f_add,valid.set)

##              ME  RMSE  MAE  MPE  MAPE  MASE
## Training set -82292.92 544380.3 394008.4 -4.963368 15.286670 0.8489270
## Test set    -119982.64 280911.6 226533.7 -4.138446  6.837573 0.4880874
##              ACF1
## Training set 0.1319903
## Test set      NA

rmse(valid.set,f_add$mean)

## [1] 280911.6

#Mô hình Holt-Winters dạng nhân

## Holt-Winters exponential smoothing with trend and multiplicative seasonal component.
##
## Call:
## HoltWinters(x = rev1, seasonal = "m")
##
## Smoothing parameters:
## alpha: 0.9828025
## beta : 0.02427243
```

```
## gamma: 1
##
## Coefficients:
##      [,1]
## a 3.413471e+06
## b 7.249534e+04
## s1 9.844091e-01
## s2 1.134893e+00
## s3 8.380682e-01
## s4 9.907280e-01

f_mul<-forecast(hw.mul,h=4)
f_mul

##      Point Forecast  Lo 80  Hi 80  Lo 95  Hi 95
## 2024 Q1      3431617 2652418 4210815 2239935 4623298
## 2024 Q2      4038474 2844784 5232164 2212882 5864066
## 2024 Q3      3042989 1855564 4230414 1226979 4858999
## 2024 Q4      3669114 2242634 5095593 1487502 5850726

accuracy(f_mul,valid.set)

##           ME  RMSE  MAE  MPE  MAPE  MASE
## Training set -97712.67 610013.9 442015.7 -5.429761 17.108694 0.9523632
## Test set    -171512.88 264541.3 188530.3 -5.228046  5.781093 0.4062056
##           ACF1
## Training set 0.01262003
## Test set      NA

rmse(valid.set,f_mul$mean)

## [1] 264541.3

#Dự báo cho năm 2025 sử dụng mô hình Holt-Winters dạng nhân

## Holt-Winters exponential smoothing with trend and multiplicative seasonal component.
##
## Call:
## HoltWinters(x = revenue, seasonal = "m")
##
## Smoothing parameters:
## alpha: 0.9785257
## beta : 0.02594431
## gamma: 1
##
```

```
## Coefficients:
##      [,1]
## a 3.235957e+06
## b 5.718656e+04
## s1 9.790142e-01
## s2 1.134831e+00
## s3 8.422909e-01
## s4 9.777751e-01

f_mul2<-forecast(hw.mul2,h=4)
f_mul2$mean

##      Qtr1  Qtr2  Qtr3  Qtr4
## 2025 3224034 3802060 2870120 3387700

#Chuỗi giá:

Price<-ts(Price,start = 1,frequency = 1)
Return<-ts(Return[2:499],start = 1,frequency = 1)

#Kiểm định ADF chuỗi giá

##
## #####
## # 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
## -2.23909 -0.29031  0.01124  0.30019  2.63741
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.3747121  0.3944235   3.485 0.000535 ***
## z.lag.1     -0.0457098  0.0127034  -3.598 0.000353 ***
## tt          0.0005208  0.0002137   2.437 0.015162 *
## z.diff.lag  -0.0086966  0.0446699  -0.195 0.845718
## ---
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5838 on 493 degrees of freedom
## Multiple R-squared:  0.02726,   Adjusted R-squared:  0.02134
## F-statistic: 4.605 on 3 and 493 DF,  p-value: 0.003442
##
##
## Value of test-statistic is: -3.5982 4.4639 6.6941
##
## 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
##
## #####
## # 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
## -2.17742 -0.28313  0.01889  0.30257  2.54784
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.97476    0.36045   2.704  0.00708 **
## z.lag.1     -0.02962    0.01091  -2.716  0.00684 **
## z.diff.lag  -0.01380    0.04484  -0.308  0.75849
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5867 on 494 degrees of freedom
## Multiple R-squared:  0.01554,   Adjusted R-squared:  0.01155
## F-statistic: 3.899 on 2 and 494 DF,  p-value: 0.0209
```

```
##
##
## Value of test-statistic is: -2.7158 3.6894
##
## Critical values for test statistics:
##      1pct 5pct 10pct
## tau2 -3.44 -2.87 -2.57
## phi1  6.47  4.61  3.79
##
## #####
## # 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
## -2.16639 -0.28347  0.01183  0.31984  2.46731
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## z.lag.1   -0.0002039  0.0008015  -0.254   0.799
## z.diff.lag -0.0287701  0.0447831  -0.642   0.521
##
## Residual standard error: 0.5904 on 495 degrees of freedom
## Multiple R-squared:  0.0009715, Adjusted R-squared: -0.003065
## F-statistic: 0.2407 on 2 and 495 DF, p-value: 0.7862
##
##
## Value of test-statistic is: -0.2543
##
## Critical values for test statistics:
##      1pct 5pct 10pct
## tau1 -2.58 -1.95 -1.62
```

#Kiểm định ADF sai phân bậc 1 chuỗi giá

```
##
## #####
## # 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
## -2.19021 -0.28460  0.01389  0.31120  2.47615
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.0299667  0.0534260  -0.561   0.575
## z.lag.1      -1.0399795  0.0646161 -16.095 <2e-16 ***
## tt           0.0001163  0.0001858   0.626   0.531
## z.diff.lag    0.0116812  0.0449268   0.260   0.795
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5919 on 492 degrees of freedom
## Multiple R-squared:  0.5138, Adjusted R-squared:  0.5109
## F-statistic: 173.3 on 3 and 492 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -16.0947 86.3484 129.5208
##
## 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
##
## #####
## # Augmented Dickey-Fuller Test Unit Root Test #
## #####
```

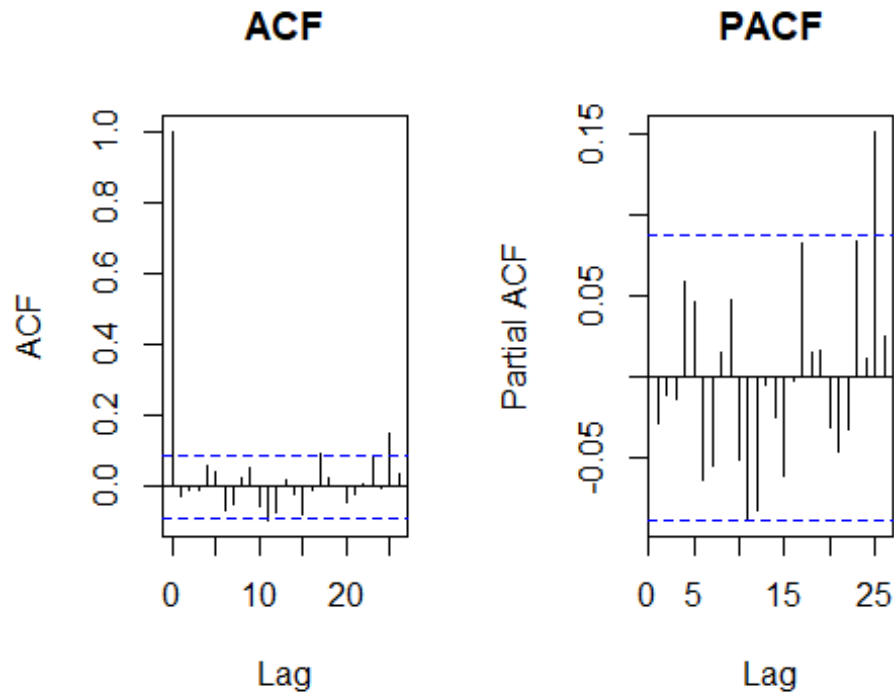


```
##
## Test regression drift
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.17511 -0.29075  0.00589  0.31954  2.46619
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.0009444  0.0265596  -0.036   0.972
## z.lag.1      -1.0383218  0.0645221 -16.093 <2e-16 ***
## z.diff.lag    0.0109362  0.0448833   0.244   0.808
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5915 on 493 degrees of freedom
## Multiple R-squared:  0.5134, Adjusted R-squared:  0.5115
## F-statistic: 260.1 on 2 and 493 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -16.0925 129.4862
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.44 -2.87 -2.57
## phi1  6.47  4.61  3.79
##
## #####
## # 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
## -2.17605 -0.29169  0.00495  0.31860  2.46525
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## z.lag.1    -1.03833    0.06446 -16.109  <2e-16 ***
## z.diff.lag  0.01094    0.04484   0.244   0.807
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5909 on 494 degrees of freedom
## Multiple R-squared:  0.5134, Adjusted R-squared:  0.5115
## F-statistic: 260.6 on 2 and 494 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -16.1089
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau1 -2.58 -1.95 -1.62

#Chuỗi giá dừng sai phân bậc 1

#Vẽ ACF, PACF của sai phân bậc 1 chuỗi giá
par(mfrow = c(1,2))
acf(diff(Price), main = "ACF")
pacf(diff(Price), main = "PACF")
```



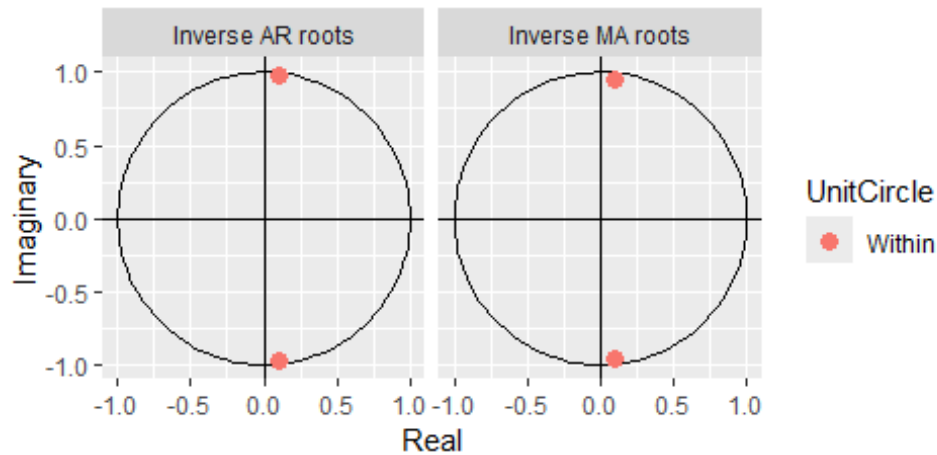
```
##
## Call:
## arima(x = diff(Price), order = c(1, 0, 1))
##
## Coefficients:
##      ar1    ma1 intercept
##  0.1561 -0.1864  0.0007
## s.e. 0.8751  0.8721  0.0255
##
## sigma^2 estimated as 0.3492: log likelihood = -444.65, aic = 897.29
##
## Training set error measures:
##              ME    RMSE    MAE MPE MAPE    MASE    ACF1
## Training set 5.223763e-05 0.5909201 0.4195312 NaN Inf 0.6606685 0.001087262
##
## Call:
## arima(x = diff(Price), order = c(1, 0, 2))
##
## Coefficients:
##      ar1    ma1    ma2 intercept
## -0.0253 -0.0043 -0.0101  0.0007
## s.e. 1.3867  1.3853  0.0510  0.0255
```

```
##
## sigma^2 estimated as 0.3492: log likelihood = -444.64, aic = 899.28
##
## Training set error measures:
##           ME    RMSE    MAE MPE MAPE    MASE    ACF1
## Training set 9.025431e-05 0.5909126 0.4196201 NaN Inf 0.6608085 0.0002360952
##
## Call:
## arima(x = diff(Price), order = c(2, 0, 2))
##
## Coefficients:
##      ar1    ar2    ma1    ma2 intercept
##  0.1924 -0.9602 -0.1804 0.9210  0.0007
## s.e. 0.0383 0.0331 0.0547 0.0471  0.0258
##
## sigma^2 estimated as 0.3424: log likelihood = -439.86, aic = 891.72
##
## Training set error measures:
##           ME    RMSE    MAE MPE MAPE    MASE    ACF1
## Training set 2.954685e-05 0.585119 0.4181477 NaN Inf 0.6584898 -0.04502682
##
## Call:
## arima(x = diff(Price), order = c(2, 0, 3))
##
## Coefficients:
##      ar1    ar2    ma1    ma2    ma3 intercept
##  0.1775 -0.9560 -0.2048 0.9260 -0.0495  0.0006
## s.e. 0.0430 0.0336 0.0615 0.0469 0.0446  0.0246
##
## sigma^2 estimated as 0.3415: log likelihood = -439.26, aic = 892.52
##
## Training set error measures:
##           ME    RMSE    MAE MPE MAPE    MASE    ACF1
## Training set 0.000102899 0.5844047 0.4159007 NaN Inf 0.6549514 -0.003110328
##
## Call:
## arima(x = diff(Price), order = c(3, 0, 2))
##
## Coefficients:
```

Bài tập lớn Phân tích chuỗi thời gian và dự báo

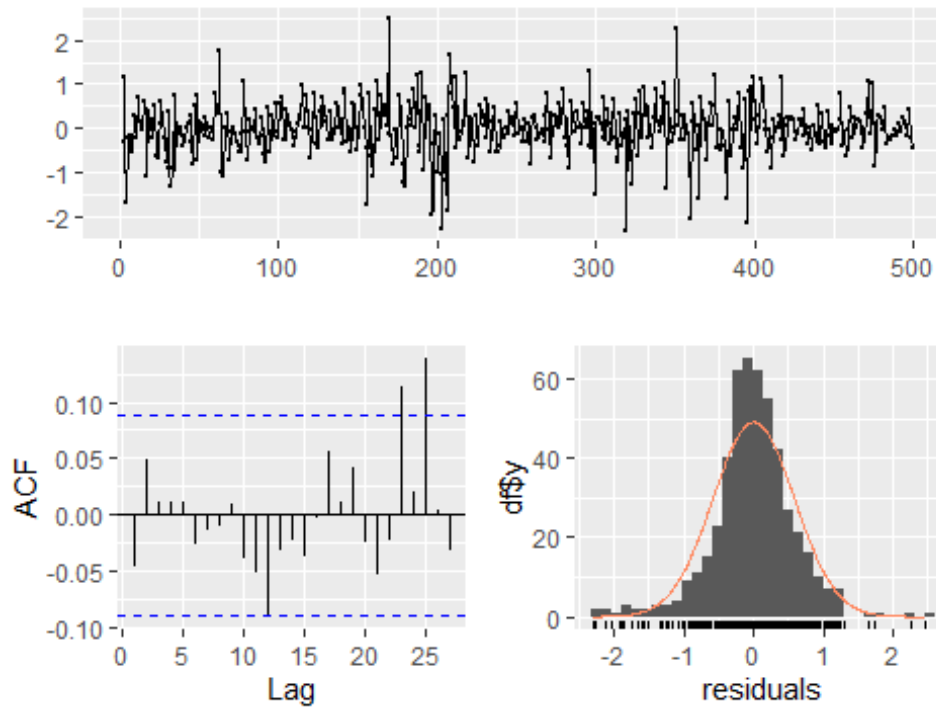
```
##      ar1   ar2   ar3   ma1   ma2 intercept
##      0.1165 -0.9443 -0.0576 -0.1489 0.9160  0.0006
## s.e. 0.0733  0.0373  0.0487  0.0591 0.0486  0.0245
##
## sigma^2 estimated as 0.3414: log likelihood = -439.19, aic = 892.37
##
## Training set error measures:
##              ME    RMSE    MAE MPE MAPE    MASE    ACF1
## Training set 0.0001194972 0.5843154 0.4156231 NaN Inf 0.6545143 0.002107827

autoplot(reg3)
```



```
checkresiduals(reg3)
```

Residuals from ARIMA(2,0,2) with non-zero mean



```
##
## Ljung-Box test
##
## data: Residuals from ARIMA(2,0,2) with non-zero mean
## Q* = 3.7006, df = 6, p-value = 0.7171
##
## Model df: 4. Total lags used: 10
#Mô hình phù hợp
#Dự báo price các mô hình
## Time Series:
## Start = 500
## End = 509
## Frequency = 1
## [1] 35.01408 35.01628 35.01663 35.01668 35.01669 35.01669 35.01669 35.01669
## [9] 35.01669 35.01669
## Time Series:
## Start = 500
## End = 509
## Frequency = 1
## [1] 35.01471 35.01879 35.01864 35.01865 35.01865 35.01865 35.01865 35.01865
## [9] 35.01865 35.01865
```

```
## Time Series:
## Start = 500
## End = 509
## Frequency = 1
## [1] 34.95601 34.95575 34.99792 35.00629 34.96742 34.95190 34.98623 35.00773
## [9] 34.97891 34.95274
```

```
## Time Series:
## Start = 500
## End = 509
## Frequency = 1
## [1] 34.97010 34.96994 35.02169 35.03103 34.98320 34.96580 35.00843 35.03264
## [9] 34.99617 34.96656
```

```
## Time Series:
## Start = 500
## End = 509
## Frequency = 1
## [1] 34.97182 34.97020 35.02254 35.03180 34.98355 34.96617 35.00917 35.03337
## [9] 34.99659 34.96698
```

```
#Tính RMSE cho dự báo so với giá thực tế
valid1<-c(35.4,35.55,35,34.4,34.55,34.15,33.5,34.1,34,34)
rmse(valid1,f1$mean)
```

```
## [1] 0.8361393
```

```
rmse(valid1,f2$mean)
```

```
## [1] 0.8374587
```

```
rmse(valid1,f3$mean)
```

```
## [1] 0.814492
```

```
rmse(valid1,f4$mean)
```

```
## [1] 0.8273381
```

```
rmse(valid1,f5$mean)
```

```
## [1] 0.8276651
```

```
#Kiểm định ADF chuỗi log-return
```

```
##
```

```
## #####
```

```
## # Augmented Dickey-Fuller Test Unit Root Test #
```

```
## #####
```

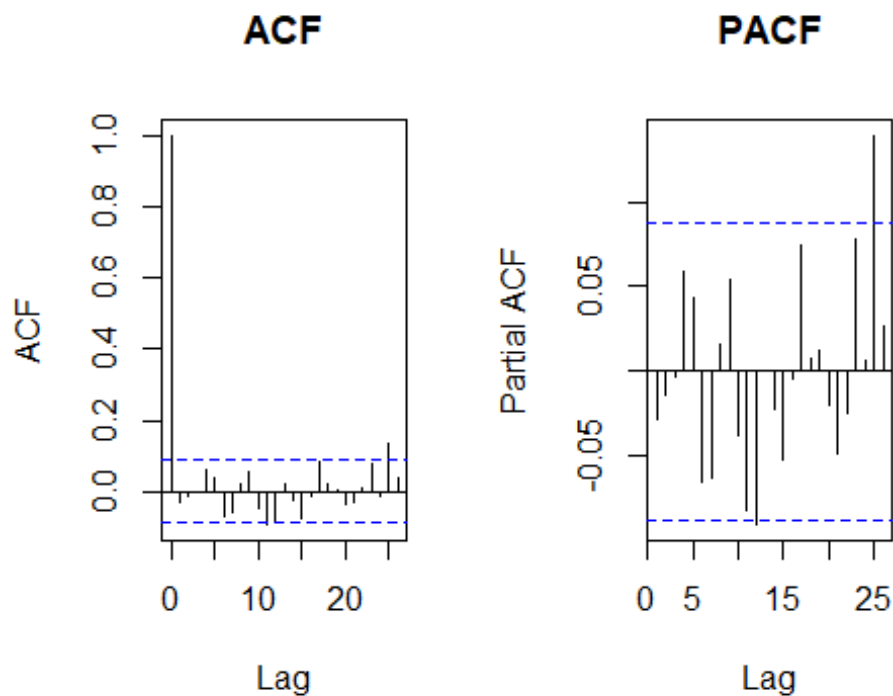
```
##
## Test regression trend
##
##
## Call:
## lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
##
## Residuals:
##   Min     1Q  Median     3Q    Max
## -7.2284 -0.8646  0.0381  0.9671  6.6853
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.0915416  0.1620222  -0.565   0.572
## z.lag.1      -1.0431629  0.0646119 -16.145 <2e-16 ***
## tt           0.0003562  0.0005634   0.632   0.527
## z.diff.lag    0.0148863  0.0449436   0.331   0.741
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.795 on 492 degrees of freedom
## Multiple R-squared:  0.5138, Adjusted R-squared:  0.5109
## F-statistic: 173.3 on 3 and 492 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -16.145 86.8889 130.3318
##
## 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
##
## #####
## # 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
## -7.2412 -0.8850  0.0234  0.9696  6.7044
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.002662  0.080546  -0.033   0.974
## z.lag.1     -1.041485  0.064518 -16.143 <2e-16 ***
## z.diff.lag   0.014127  0.044900   0.315   0.753
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.794 on 493 degrees of freedom
## Multiple R-squared:  0.5134, Adjusted R-squared:  0.5115
## F-statistic: 260.1 on 2 and 493 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -16.1425 130.292
##
## Critical values for test statistics:
##      1pct  5pct 10pct
## tau2 -3.44 -2.87 -2.57
## phi1  6.47  4.61  3.79
##
## #####
## # 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
## -7.2438 -0.8876  0.0208  0.9670  6.7017
```

```
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## z.lag.1   -1.04149   0.06445 -16.159 <2e-16 ***
## z.diff.lag 0.01413   0.04485  0.315  0.753
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.792 on 494 degrees of freedom
## Multiple R-squared:  0.5134, Adjusted R-squared:  0.5115
## F-statistic: 260.6 on 2 and 494 DF, p-value: < 2.2e-16
##
##
## Value of test-statistic is: -16.1589
##
## Critical values for test statistics:
##      1pct 5pct 10pct
## tau1 -2.58 -1.95 -1.62

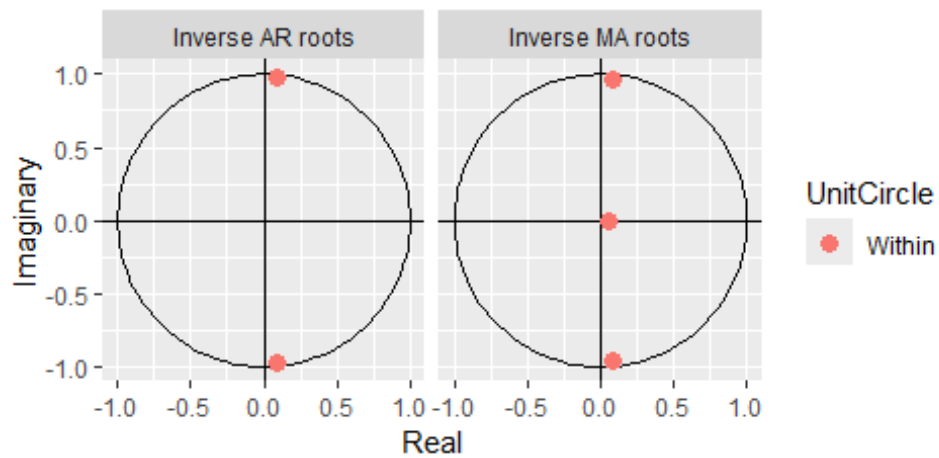
#Chuỗi dừng
#Vẽ ACF, PACF chuỗi log-return
```



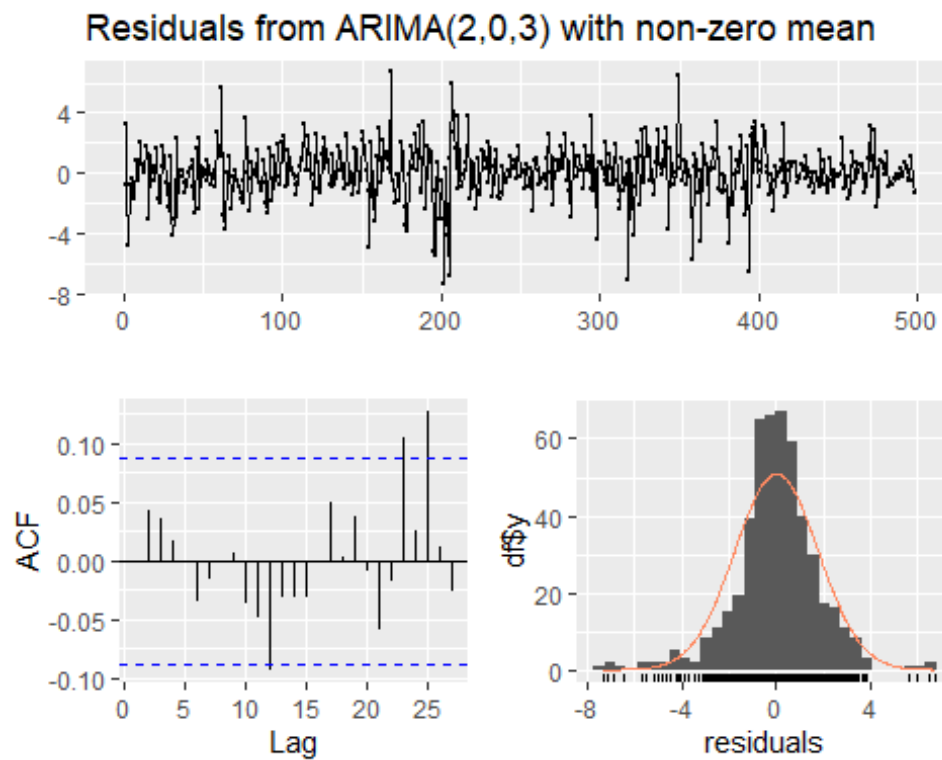
#Chọn mô hình arima

```
##
## Call:
## arima(x = Return, order = c(2, 0, 2))
##
## Coefficients:
##      ar1    ar2    ma1    ma2 intercept
##    0.4065 -0.9028 -0.4496 0.8740  0.0015
## s.e. 0.1069 0.0949 0.1093 0.1217  0.0758
##
## sigma^2 estimated as 3.156: log likelihood = -992.85, aic = 1997.69
##
## Training set error measures:
##              ME   RMSE   MAE MPE MAPE   MASE   ACF1
## Training set 0.0008831109 1.776398 1.271074 NaN Inf 0.6575227 0.009951026
##
## Call:
## arima(x = Return, order = c(2, 0, 3))
##
## Coefficients:
##      ar1    ar2    ma1    ma2    ma3 intercept
##    0.1814 -0.9583 -0.2089 0.9299 -0.0485  0.0018
## s.e. 0.0429 0.0321 0.0613 0.0461 0.0449  0.0747
##
## sigma^2 estimated as 3.14: log likelihood = -991.66, aic = 1997.32
##
## Training set error measures:
##              ME   RMSE   MAE MPE MAPE   MASE   ACF1
## Training set 0.000291098 1.77193 1.264024 NaN Inf 0.6538759 -0.0027747
##
## Call:
## arima(x = Return, order = c(3, 0, 2))
##
## Coefficients:
##      ar1    ar2    ar3    ma1    ma2 intercept
##    0.4208 -0.9012 0.0134 -0.4541 0.8671  0.0023
## s.e. 0.1052 0.0872 0.0499 0.0954 0.1078  0.0767
##
## sigma^2 estimated as 3.155: log likelihood = -992.81, aic = 1999.62
```

```
##
## Training set error measures:
##           ME   RMSE   MAE MPE MAPE   MASE   ACF1
## Training set -1.410238e-05 1.776269 1.27217 NaN Inf 0.6580896 -0.0006130149
##
## Call:
## arima(x = Return, order = c(3, 0, 4))
##
## Coefficients:
##      ar1   ar2   ar3   ma1   ma2   ma3   ma4 intercept
##    0.5583 -1.0057 0.3423 -0.5918 1.0401 -0.3924 0.0845  0.0014
## s.e. 0.3998  0.1084 0.3896  0.3982 0.1232  0.3761 0.0504  0.0817
##
## sigma^2 estimated as 3.127: log likelihood = -990.67, aic = 1999.34
##
## Training set error measures:
##           ME   RMSE   MAE MPE MAPE   MASE   ACF1
## Training set 0.0004881115 1.768417 1.263172 NaN Inf 0.6534353 0.0008937259
##
## Call:
## arima(x = Return, order = c(4, 0, 3))
##
## Coefficients:
##      ar1   ar2   ar3   ar4   ma1   ma2   ma3 intercept
##    0.5131 -0.9183 0.2854 0.0793 -0.5441 0.9467 -0.3348  0.0015
## s.e. 0.4119  0.1211 0.3975 0.0505  0.4117 0.1188  0.3841  0.0813
##
## sigma^2 estimated as 3.129: log likelihood = -990.8, aic = 1999.59
##
## Training set error measures:
##           ME   RMSE   MAE MPE MAPE   MASE   ACF1
## Training set 0.0003938122 1.768873 1.262792 NaN Inf 0.6532388 -0.0014392
autoplot(reg7)
```



```
checkresiduals(reg7)
```



```
##  
## Ljung-Box test  
##
```

```
## data: Residuals from ARIMA(2,0,3) with non-zero mean
```

```
## Q* = 3.1172, df = 5, p-value = 0.6819
```

```
##
```

```
## Model df: 5. Total lags used: 10
```

#Mô hình đáng tin cậy

#Dự báo chuỗi lợi suất

```
## Time Series:
```

```
## Start = 499
```

```
## End = 508
```

```
## Frequency = 1
```

```
## [1] 0.0549 0.1294 0.0053 -0.1124 -0.0482 0.0842 0.0800 -0.0412 -0.0867
```

```
## [10] 0.0042
```

```
## Time Series:
```

```
## Start = 499
```

```
## End = 508
```

```
## Frequency = 1
```

```
## [1] -0.0832 -0.0037 0.1471 0.0335 -0.1316 -0.0527 0.1198 0.0755 -0.0979
```

```
## [10] -0.0868
```

```
## Time Series:
```

```
## Start = 499
```

```
## End = 508
```

```
## Frequency = 1
```

```
## [1] 0.0438 0.1343 0.0034 -0.1156 -0.0465 0.0881 0.0808 -0.0425 -0.0861
```

```
## [10] 0.0066
```

```
## Time Series:
```

```
## Start = 499
```

```
## End = 508
```

```
## Frequency = 1
```

```
## [1] -0.0869 -0.0667 0.1205 -0.0116 -0.1490 -0.0287 0.1313 0.0527 -0.1110
```

```
## [10] -0.0685
```

```
## Time Series:
```

```
## Start = 499
```

```
## End = 508
```

```
## Frequency = 1
```

```
## [1] -0.0952 -0.0578 0.1270 -0.0087 -0.1437 -0.0325 0.1244 0.0535 -0.1059
```

```
## [10] -0.0690
```

#Tính RMSE của dự báo so với lợi suất thực tế

Bài tập lớn Phân tích chuỗi thời gian và dự báo

```
## [1] 1.243331
```

```
rmse(valid2,f7$mean)
```

```
## [1] 1.278563
```

```
rmse(valid2,f8$mean)
```

```
## [1] 1.244213
```

```
rmse(valid2,f9$mean)
```

```
## [1] 1.278994
```

```
rmse(valid2,f10$mean)
```

```
## [1] 1.27899
```