

VŨ PHẠM THÀNH PHƯƠNG – REPORT – 11225331

1. Phân tích chuỗi doanh thu thuần.



Hình 1. 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	47.176	0.120085	20.21466	0.057120	268.226	270.369	266.012	270.655

mùa vụ 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 1. 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.

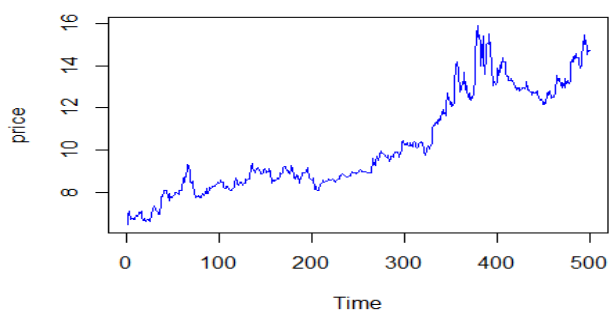
Áp dụng mô hình Holt-Winters mùa vụ dạng nhân, dự báo được kết quả năm 2025 như sau:

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

Bảng 2. Kết quả dự báo năm 2025

2. Phân tích chuỗi lịch sử giá.

Hình 2. Biểu đồ chuỗi giá theo từng phiên mã VTO giai đoạn 2023-2034



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 3. Kết quả kiểm định ADF.

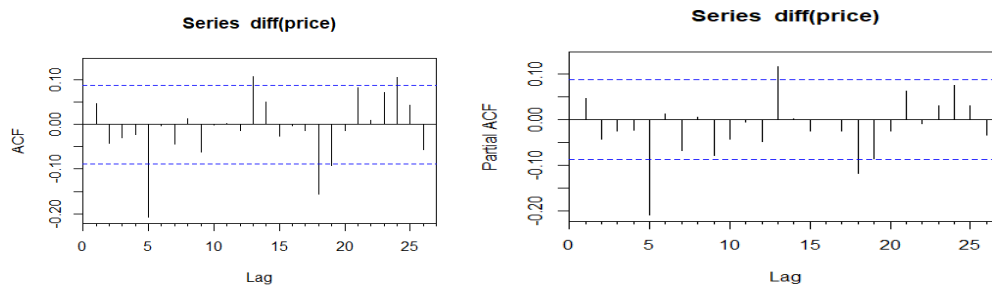
Trị tuyệt đối của giá trị quan sát luôn nhỏ hơn trị tuyệt đối của giá trị tới hạn ứng với mức ý nghĩa 5% nên chuỗi chưa dừng, tiến hành sai phân bậc 1.

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 4. 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 3. 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.

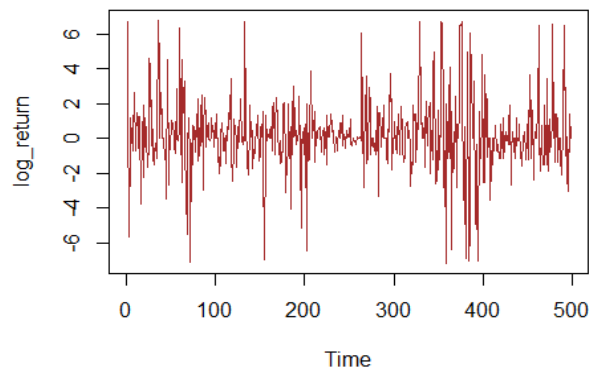
2.3. Dự báo 10 phiên giao dịch đầu năm 2025.

	(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 4. Đồ thị chuỗi lợi suất của cổ phiếu VTO.

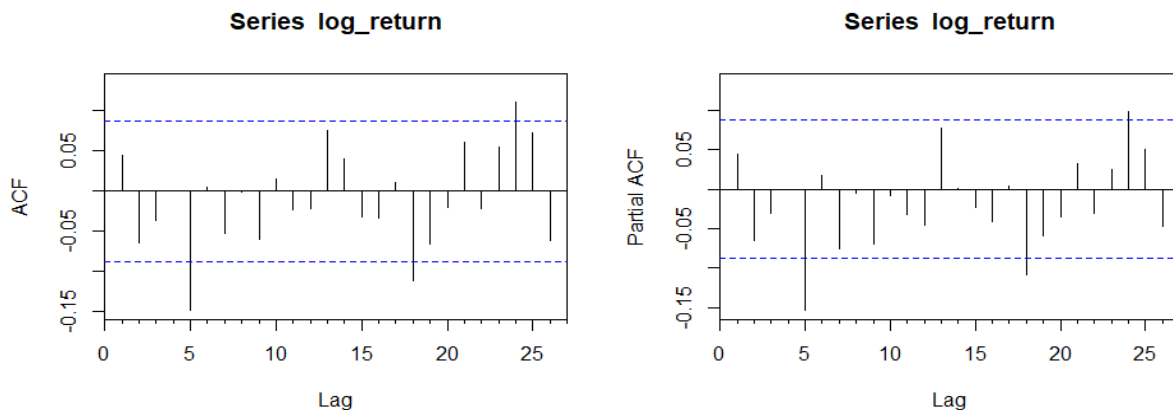
3.1. Kiểm định tính dừng của chuỗi lợi suất.

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 5. 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 ĐTDV	0.1652
(5,0,18)	2177.96	2.039023	Có nghiệm nằm trên ĐTDV	0.1046
(5,0,5)	2171.48	2.089456	Nằm trong ĐTDV	0.4208
(5,0,0)	2168.33	2.104128	Nằm trong ĐTDV	0.2987

Bảng 8. So sánh và kiểm tra nhiễu 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.

PHỤ LỤC

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.

```
## 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|)  
## (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
```

```
## 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
## 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
## ---
```

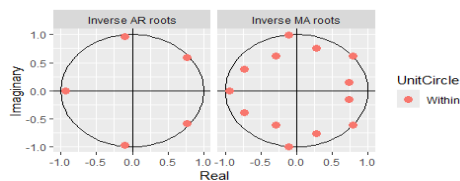
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2465 on 493 degrees of freedom
## Multiple R-squared:  0.4789, Adjusted R-squared:  0.4768
## F-statistic: 226.5 on 2 and 493 DF,  p-value: < 2.2e-16
##
##
## Value of test-statistic is: -16.0717 129.1513
##
## 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).

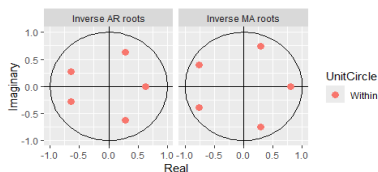
```
## arima(x = price, order = c(5, 1, 13))
##
## Coefficients:
##      ar1      ar2      ar3      ar4      ar5      ma1      ma2      ma3
##      0.4001 -0.3227 -0.1816  0.2864 -0.8311 -0.3567  0.2578  0.1894
## s.e.  0.1341  0.1760  0.2064  0.1690  0.1206  0.1405  0.1742  0.1963
##      ma4      ma5      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
```



```
## Ljung-Box test
##
## data: Residuals from ARIMA(5,1,13)
## Q* = 8.0318, df = 3, p-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
```

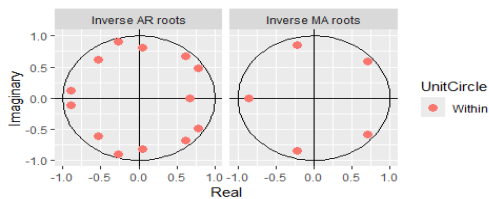


```
## Ljung-Box test
##
## data: Residuals from ARIMA(5,1,5)
## Q* = 4.0782, df = 3, p-value = 0.2531
## Model df: 10. Total lags used: 13
```

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-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
## 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

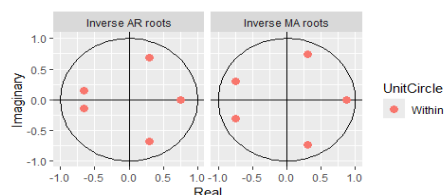
## Test regression none
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.3253 -0.8714  0.0410  1.0683  6.8047
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## z.lag.1      -1.01360    0.06155 -16.469 <2e-16 ***
## z.diff.lag    0.06055    0.04444   1.362   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
```

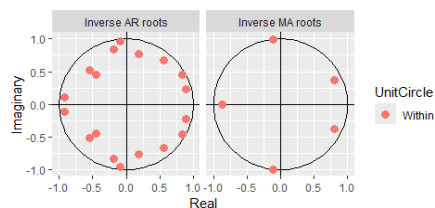
```
##          ma5 intercept
##          -0.3636      0.1628
## s.e.      0.3854      0.0595
##
## sigma^2 estimated as 4.366: log likelihood = -1073.74, aic = 2171.48
##
## Training set error measures:
##              ME      RMSE      MAE MPE MAPE      MASE      ACF1
## Training set 0.003570057 2.089456 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).

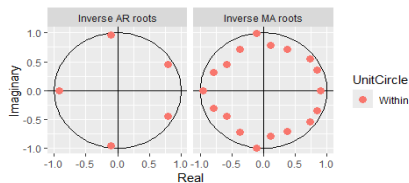
```
## arima(x = log_return, order = c(18, 0, 5))
##
## Coefficients:
##          ar1      ar2      ar3      ar4      ar5      ar6      ar7      ar8
##          0.5553 -0.3288 0.1875 0.4344 -0.8178 0.1805 -0.1834 0.0672
## s.e.      0.1725 0.1881 0.1911 0.1947 0.1525 0.0752 0.0759 0.0840
##          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
##          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).

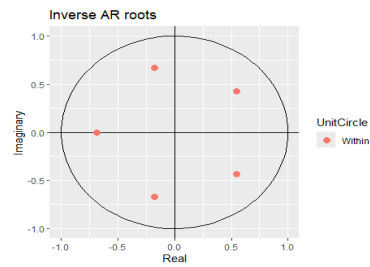
```
## arima(x = log_return, order = c(5, 0, 18))
##
## Coefficients:
##      ar1      ar2      ar3      ar4      ar5      ma1      ma2      ma3
##      0.4585 -0.1509 -0.0217  0.4310 -0.7173 -0.4125  0.0653  0.0077
## s.e.  0.1700  0.1478  0.1446  0.1343  0.1417  0.1760  0.1536  0.1414
##      ma4      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:
##              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
```