Time Series Analysis | MS4218

Colm Kenny 18225012

21/04/2022

Introduction

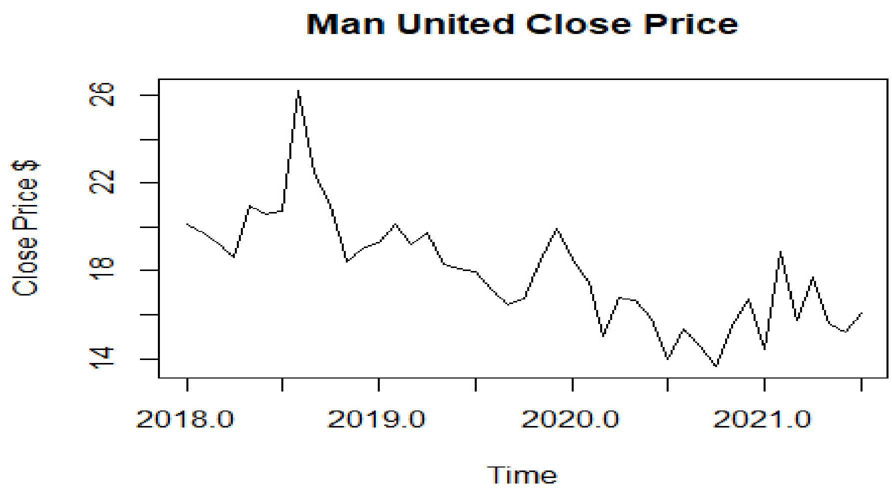
Manchester United plc (MANU) is not an average company listed on the NYSE. Operating as a professional sports team, they provide entertainment directly to consumers, sell licensed products and receive sponsorships (Yahoo finance, n.d.). The trajectory of MANU share price may differ to orthodox stocks due to the football calendar and transfer news etc. For example, no ticket sales in the summer could hinder earnings reports in the summer months whereas jersey sales and a congested fixture list could boost turnover in the Christmas period increasing demand for shares. As a football club, Manchester United is a shell of its former self. It is worth analysing the financial performance of the club to discover any links to the football performance of the club. Is the share price of the club also in decline? And if so, is there any signs of a recovery? In this project, I will use a MANU share price dataset to uncover trends in the share price. I will also attempt to fit an appropriate model to the data and create predictions.

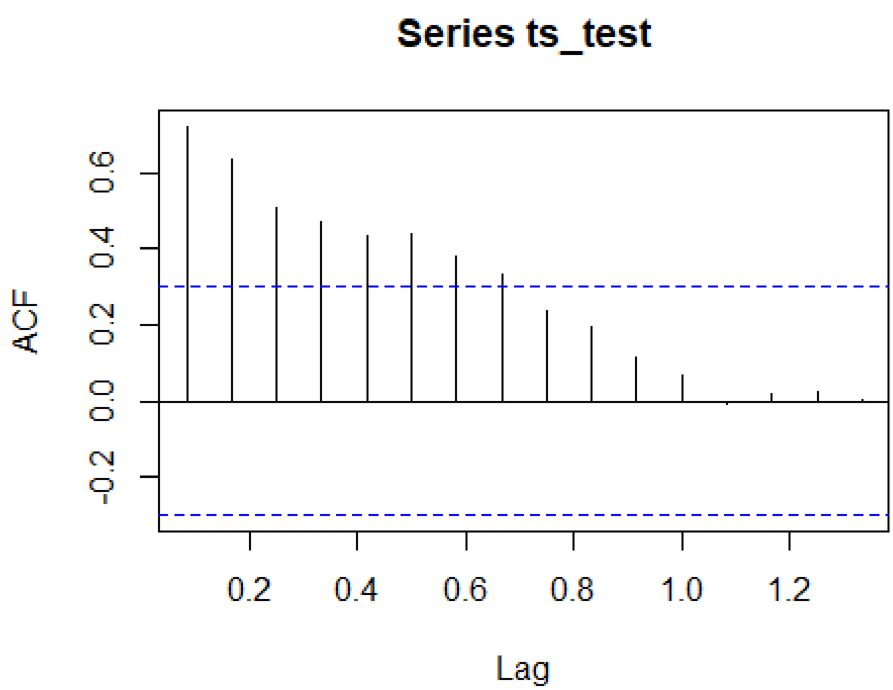
Data

From the monthly dataset in the range 01/01/2018 to 31/12/2021, I am concerned with the closing price variable which I will store as a time series object in R. Before any methods are applied to this series, the last 10% of observations will be put aside to be used later for comparisons to our predictions. The remaining 90% (43 months) will be called the test series and will be used to create our model.

Stationarity and Transformations

The first step is to identify whether our series is stationary or non-stationary. If a series is stationary the expected value of the series does not depend on time, or in other words the mean function must be constant with time. We can assess this by looking at a simple time series plot or by looking at a correlogram of the series. If an ACF value at the beginning of the correlogram exceeds the significance bands, then the series can de declared as non-stationary as it depends on the values nearest to it.

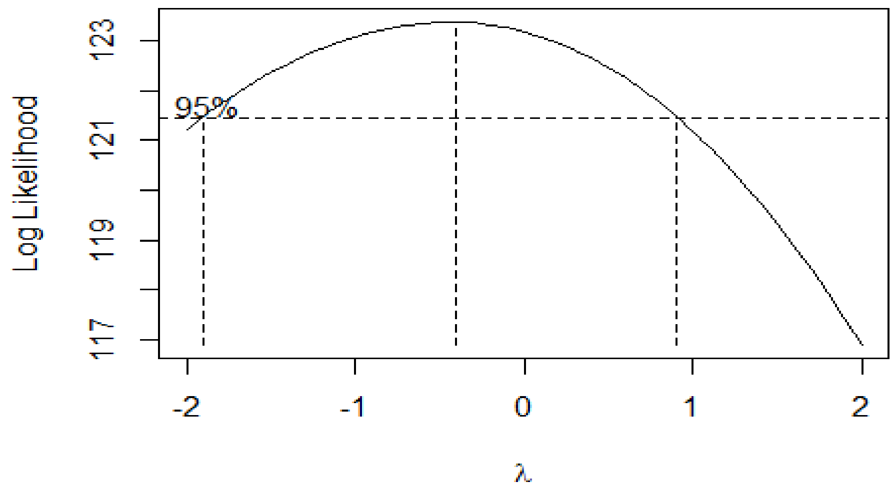




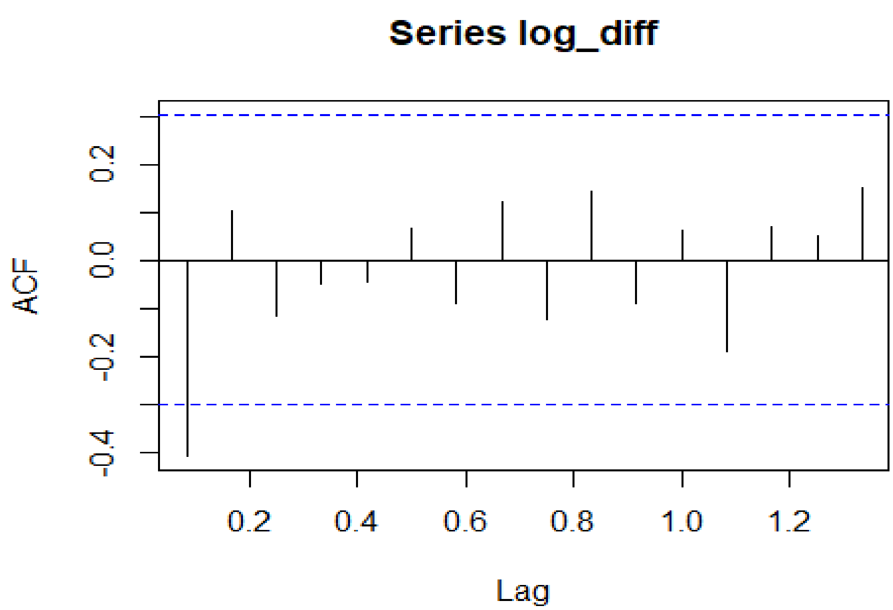
Above are the time series plot and correlogram for our test series. The first 8 ACF values are statistically significant. It is clear to see that this series is non-stationary and also showing downward trend. The close price of Man United stock is dependent on the values of the last 8 months. There is also no evidence of seasonal trend in the data. We can judge this by looking at the 12th lag on the ACF plot which is not significant. An Augmented Dickey Fuller Test will also show if our series is stationary. A p-value of 0.437 suggests that the series is non-stationary and that a transformation is needed.

```
## Augmented Dickey-Fuller Test
## alternative: stationary
##
## Type 1: no drift no trend
##      lag    ADF p.value
## [1,]   0 -0.659   0.437
## [2,]   1 -0.696   0.424
## [3,]   2 -0.646   0.442
## [4,]   3 -0.594   0.461
## Type 2: with drift no trend
##      lag   ADF p.value
## [1,]   0 -2.50   0.141
## [2,]   1 -1.69   0.446
## [3,]   2 -1.57   0.489
## [4,]   3 -1.33   0.576
## Type 3: with drift and trend
##      lag   ADF p.value
## [1,]   0 -4.06  0.0166
## [2,]   1 -3.03  0.1639
## [3,]   2 -3.24  0.0933
## [4,]   3 -3.38  0.0721
## ----
## Note: in fact, p.value = 0.01 means p.value <= 0.01
```

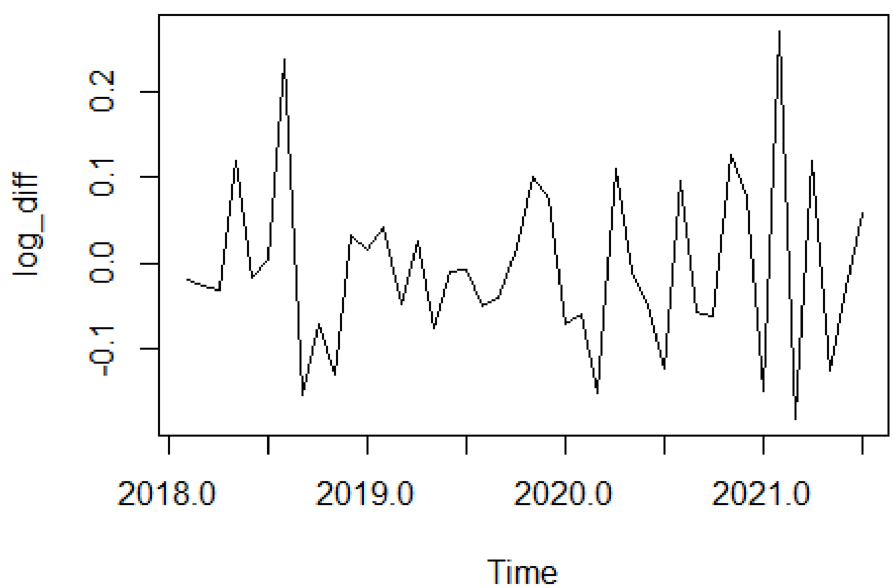
From first glance, there appears to be no issue of increasing variance. However, it is important to assess whether a log transformation is necessary. The Box Cox function can be used to assess whether a box cox transformation can be used, and which value of lambda will create a stationary series and remove increasing variance in the series.



Looking at the Log likelihood plot, $\lambda = 0$ lies within the 95% confidence interval. Therefore a log-transformation is supported. Since we are dealing with financial data, we will investigate whether a difference-log transformation will create a stationary series. This is helpful due to the difference-log series being approximately equal to the percentage change of the stock price which is easy to interpret. Once we derive the difference-log series, it is important to make sure that it is a stationary series, once again using the ADF test.



```
## Augmented Dickey-Fuller Test
## alternative: stationary
##
## Type 1: no drift no trend
##       lag   ADF p.value
## [1,]   0 -9.70    0.01
## [2,]   1 -5.55    0.01
## [3,]   2 -4.51    0.01
## [4,]   3 -4.10    0.01
## Type 2: with drift no trend
##       lag   ADF p.value
## [1,]   0 -9.63    0.01
## [2,]   1 -5.52    0.01
## [3,]   2 -4.49    0.01
## [4,]   3 -4.14    0.01
## Type 3: with drift and trend
##       lag   ADF p.value
## [1,]   0 -9.50  0.0100
## [2,]   1 -5.45  0.0100
## [3,]   2 -4.43  0.0100
## [4,]   3 -4.07  0.0165
## ----
## Note: in fact, p.value = 0.01 means p.value <= 0.01
```



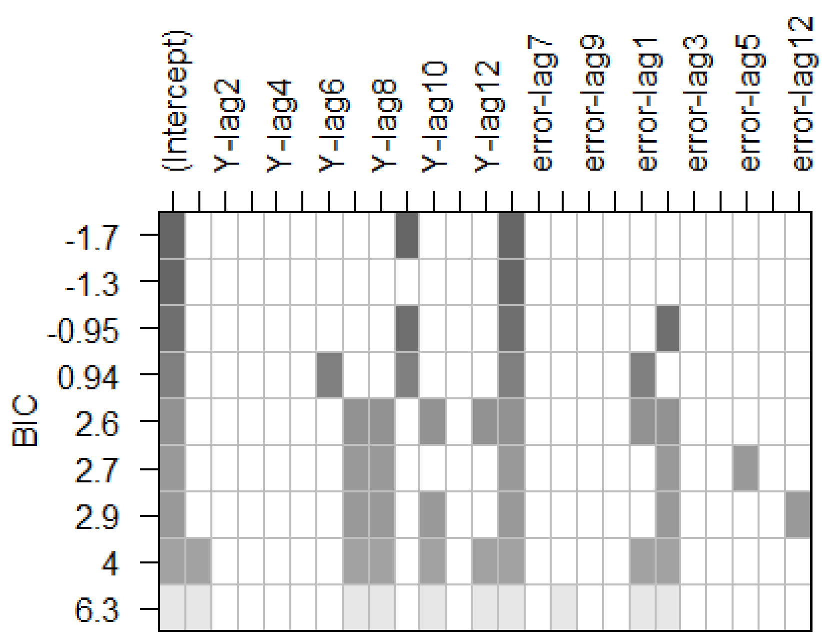
The results of the ADF test prove the difference-log transformation created a stationary series. This is an approximation of percentage change of the stock price. Looking at the plot of the series there is no evidence of trend or increasing variance.

Model Identification

Using this difference-log series, we will be able to model the data using an ARIMA model, where the value of d is 1. ARIMA models are useful for non-stationary series and are very powerful. Since the differencing is part of the model we will only need to input the log series. The next step requires us to choose parameters to use for both AR and MA components of the model. There are different ways we can choose the parameters. Choosing a few combinations of parameters will give us the ability to decide on the final model by running some tests and checks.

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x o o o o o o o o o o  o  o  o
## 1 o o o o o o o o o o o  o  o  o
## 2 x o o o o o o o o o o  o  o  o
## 3 x o o o o o o o o o o  o  o  o
## 4 x o o o o o o o o o o  o  o  o
## 5 o o o o o o o o o o o  o  o  o
## 6 o x o o o o o o o o o  o  o  o
## 7 o x o o o o o o o o o  o  o  o
```

The above extended ACF suggests an ARIMA(0,1,1) model. This will be our first model to test.



Looking at the arma subsets figure, we can see that the BIC is minimised with an ARIMA model of parameters (9,1,6). The two methods have given us separate models. I can now create these models and run some tests to decide on the final model to apply predictions.

Model Checking

Call: arima(x = log_test, order = c(0, 1, 1))

Coefficients

	ma1
	-0.4893
s.e.	0.1628

sigma^2 estimated as 0.007974: log likelihood = 41.73, aic = -81.46
Call: arima(x = log_test, order = c(9, 1, 6))

Coefficients (continued below)

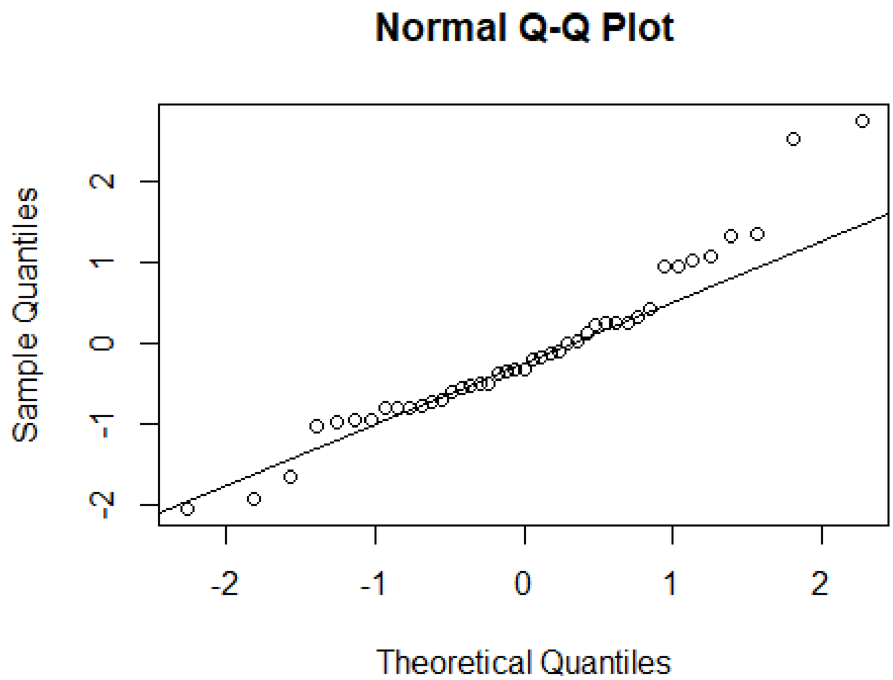
	ar1	ar2	ar3	ar4	ar5	ar6	ar7
	0.05694	0.2626	-0.2939	0.3509	-0.007465	-0.4134	-0.2042
s.e.	0.3953	0.3964	0.2966	0.2328	0.4126	0.2481	0.2109

	ar8	ar9	ma1	ma2	ma3	ma4	ma5	ma6
	0.0272	-0.2824	-0.585	-0.2238	0.3579	-0.84	0.3406	0.5592
s.e.	0.2416	0.2926	0.3987	0.5696	0.4351	0.2591	0.5499	0.4479

sigma^2 estimated as 0.005339: log likelihood = 46.5, aic = -63

The first brief check involves us analysing the AIC values of both models. The aim is to produce a model which minimises the AIC value. By this rationale, we would choose the ARIMA(0,1,1) as the AIC is considerably lower than the (9,1,6) model. However, we have plenty of checks to perform before choosing a final model.

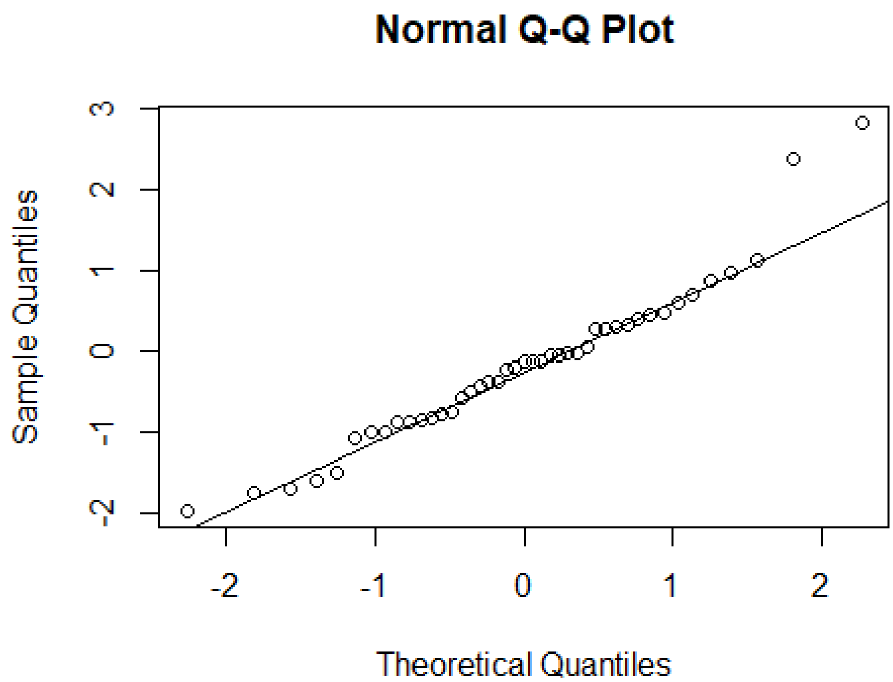
Normality of Residuals



Shapiro-Wilk normality test: resid_01

Test statistic	P value
0.9411	0.02839 *

Looking at the Q-Q plot for the (0,1,1) model, the residuals do not appear to be relatively normal. There are a lot of outliers. This is supported by Shapiro-Wilk test with a p-value of 0.02839. We reject the null hypothesis and conclude that the residuals of this model do not follow a normal distribution.

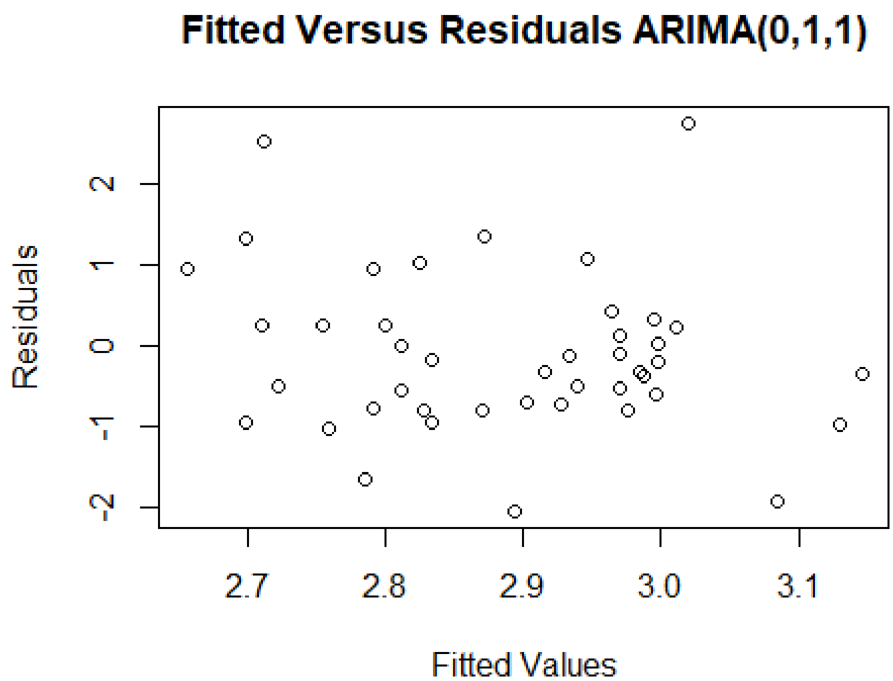


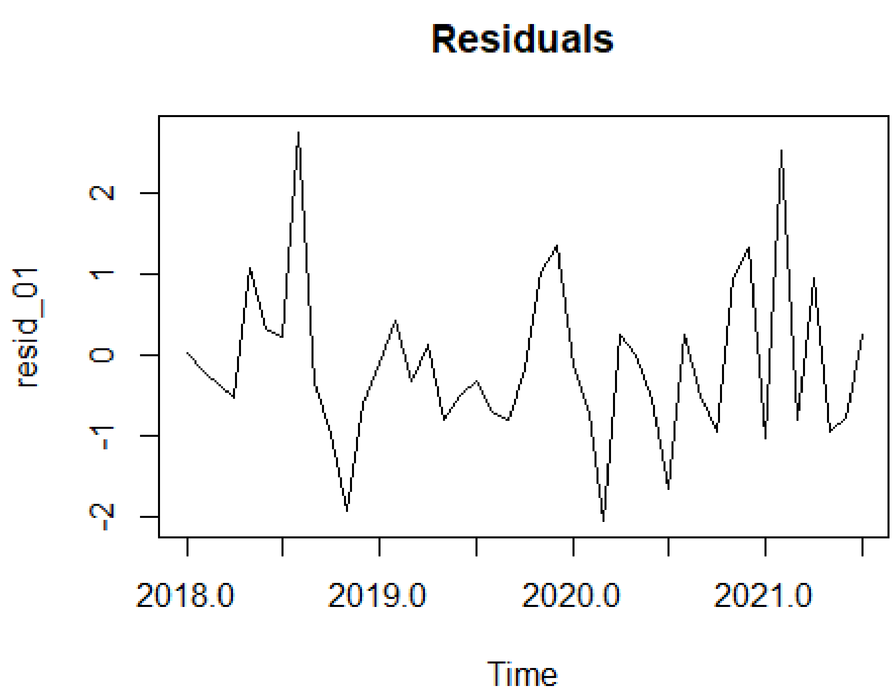
Shapiro-Wilk normality test: resid_96

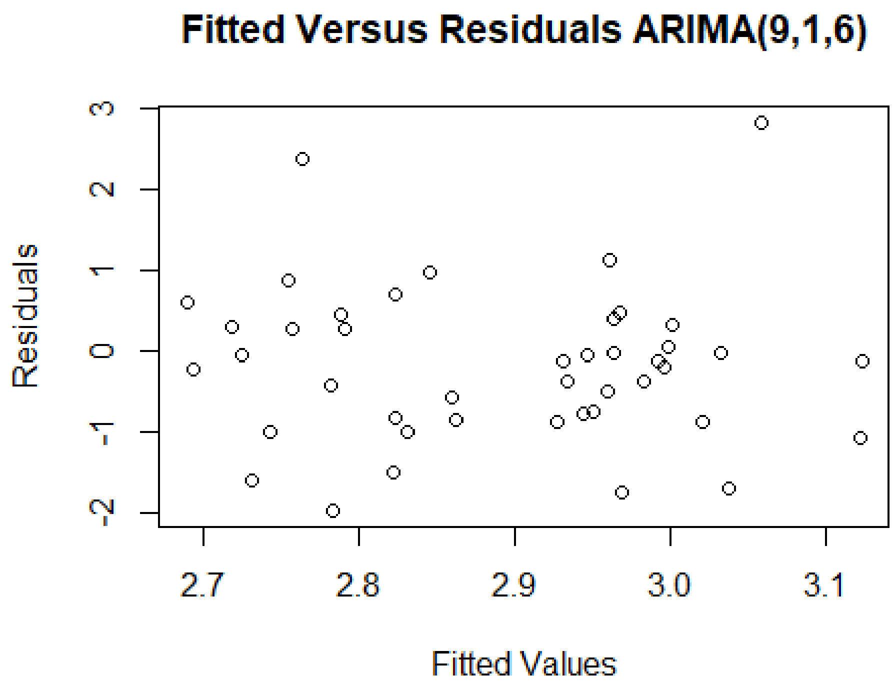
Test statistic	P value
0.9508	0.06388

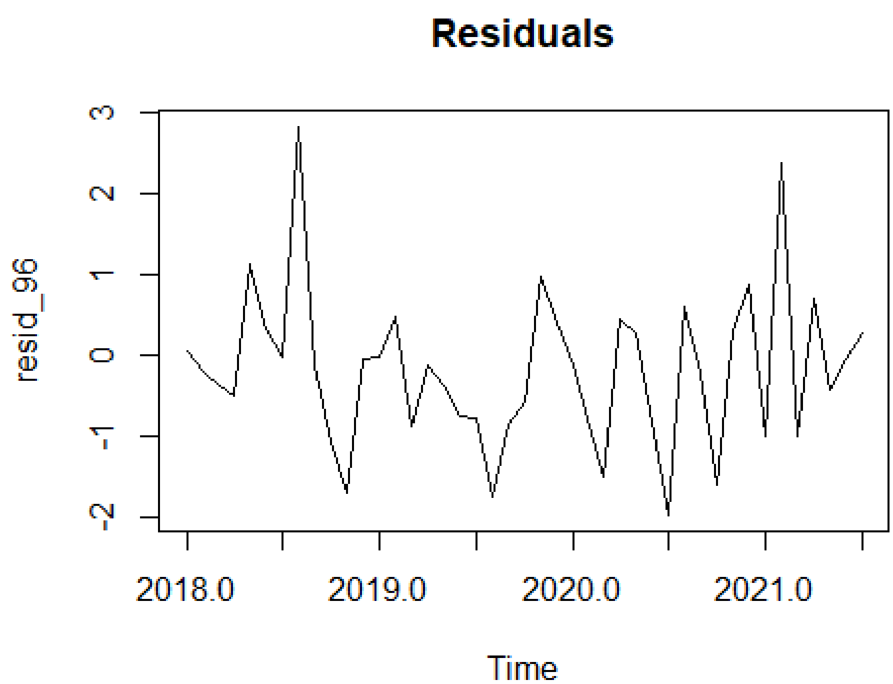
The residuals of the (9,1,6) model follow a more normal distribution, although there are two large outliers. This is evident in the Shapiro-Wilk test where we cannot reject the hypothesis that the residuals are normally distributed. This would appear to be a more appropriate model if the residuals are normally distributed.

Constant Variance



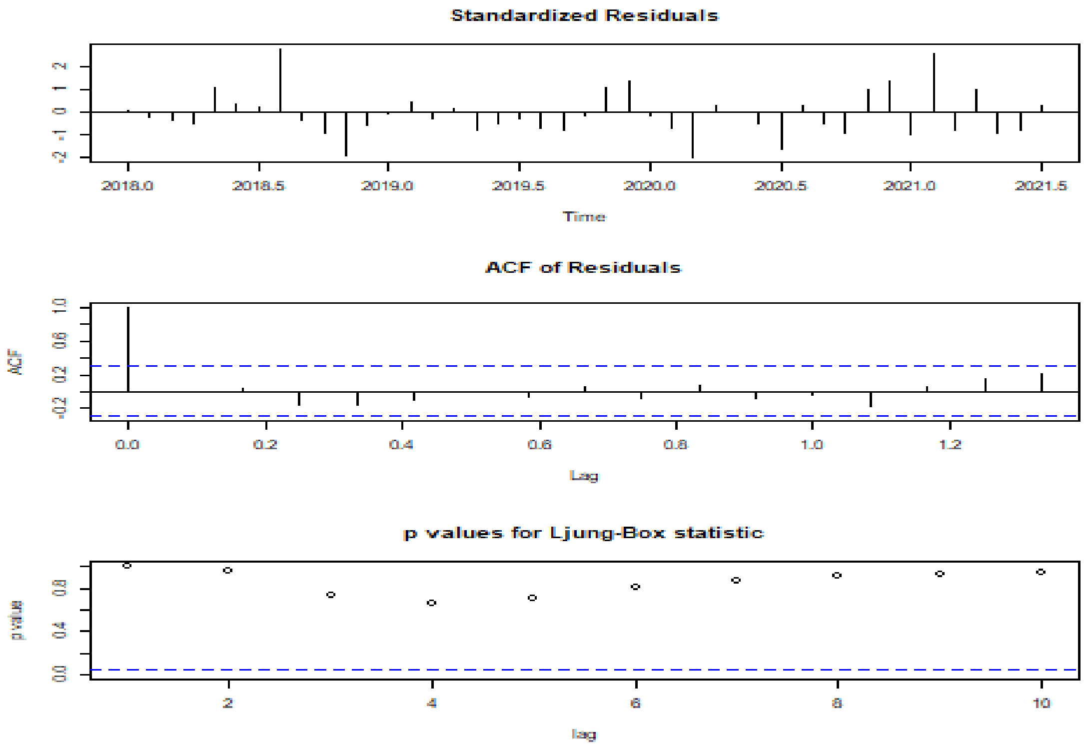




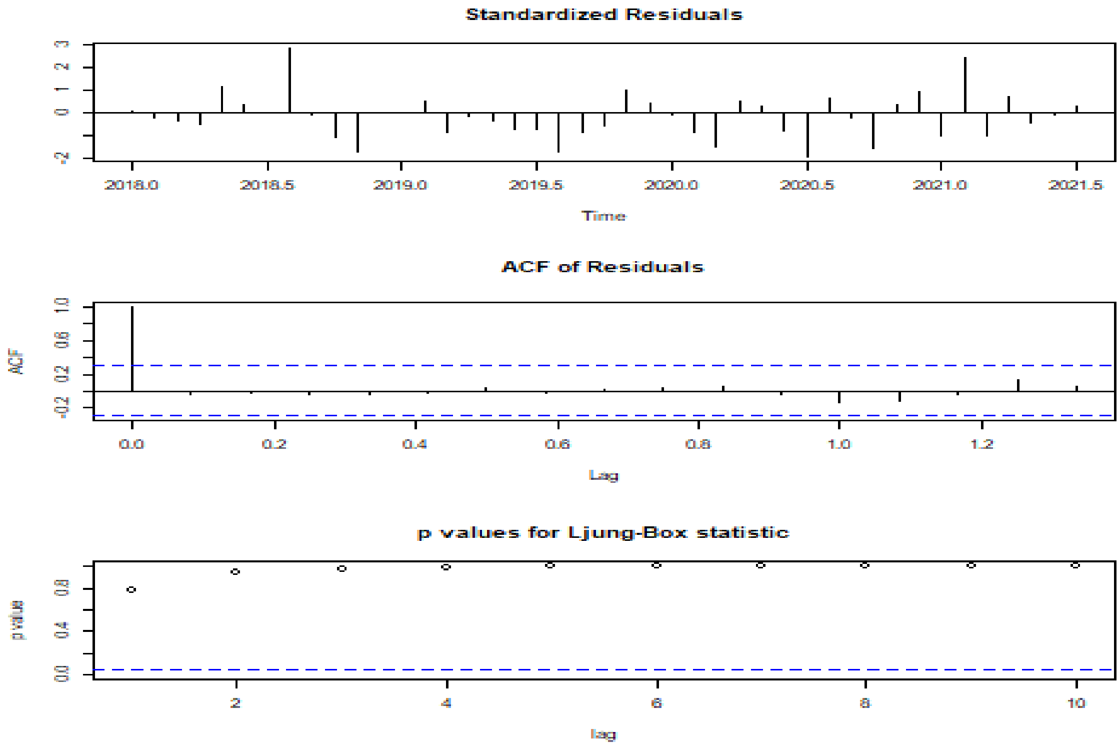


Fitted v Residual plots and time series plots show no evidence of non-constant variance or trend for both models. This check did not suggest any favourable model.

Autocorrelation



- **pvalue:** *0.893*
- **observed.runs:** *21*
- **expected.runs:** *21.09*
- **n1:** *27*
- **n2:** *16*
- **k:** *0*



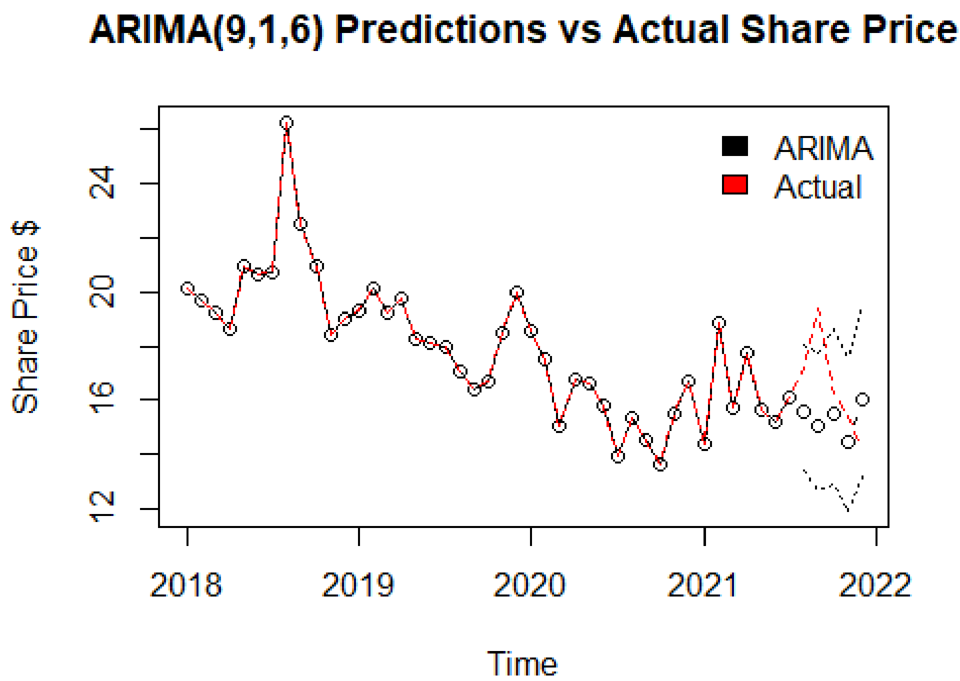
- **pvalue:** *0.974*
- **observed.runs:** *21*
- **expected.runs:** *20.53*
- **n1:** *28*
- **n2:** *15*
- **k:** *0*

Looking at the Ljung-Box plots and runs tests for each model, we can conclude that the residuals in both models are white noise. Large p-values suggest we cannot reject the null hypothesis. Residuals in neither model suffer from autocorrelation.

From all tests and checks I believe it is a good idea to use the (9,1,6) model due to the residuals being normally distributed. Although the AIC is higher on this model, I believe the normality of the residuals provides a more appropriate model for our final predictions.

Prediction

Now that a final model has been selected we can predict into the future using the ARIMA(9,1,6) model. Predicting the share price of Man United could be advantageous to shareholders and potential investors to analyse the potential trend of the stock and when to buy or sell.



Looking at our predictions, the model shows signs that the share price of Man United should stall over the next five months. When comparing to the actual trajectory of the price, it is not significantly accurate. The price initially rises for two months until showing downward trend until December 2021. The actual share price in September 2021 was outside our 95% confidence prediction limit. This could suggest that an extraordinary or unexpected event caused the share price to rise to an unprecedented level. In the context of Manchester United this could be explained by the return of Cristiano Ronaldo to the club. Which could not be predicted and would definitely force demand for the stock to increase. More time might be needed to assess the accuracy of the ARIMA model.

Conclusion

Being able to model time series data and produce predictions is very advantageous to potential investors or current shareholders. In my analysis of the MANU share price I found that the price is related to prices of the past 8 months, with no significant seasonal trend. This makes sense when put into a football context as regulations and sponsorships may change by league season, which typically lasts 7/8 months. By creating and fitting an ARIMA model, the predictions suggest that the share price will continue to exhibit negative trend until the beginning of 2022. When compared to the actual figures, we can see there are some challenges with predicting the share price of MANU. It appears that football performance or major team news play a large role in changes to the share price. The season-by-season performance of Man United has been declining for the past 5 years, which could explain the negative trend seen in the share price time series. Monitoring of these events should also be considered to accurately gauge the trajectory of the share price.

Bibliography

Manchester United Plc (Manu) Stock Price, News, Quote & History (n.d.) *Yahoo! Finance*, Yahoo!, Available at: https://finance.yahoo.com/quote/MANU [Accessed 21 April 2022]