AMAZON STOCK PRICE PREDICTION

MADHURI DILLIKER

NISHTHA CHAUDHARY

INTRODUCTION

- Amazon ranks as one of the world's top companies by market value
- Amazon.com, Inc. (AMZN) shares closed the first half of 2019 at $1,893.63
- The weekly charts are positive from 21 June
- Amazon will continue to generate annual sales growth in the 15%-20% range
- Amazon's digital ad business is projected to grow 50% a year through 2020, expanding its U.S. market share from 4.1% in 2018 to 7%. That makes Amazon the only significant competitor to Google and Facebook, which together control 57.7% of the U.S. digital ad market.

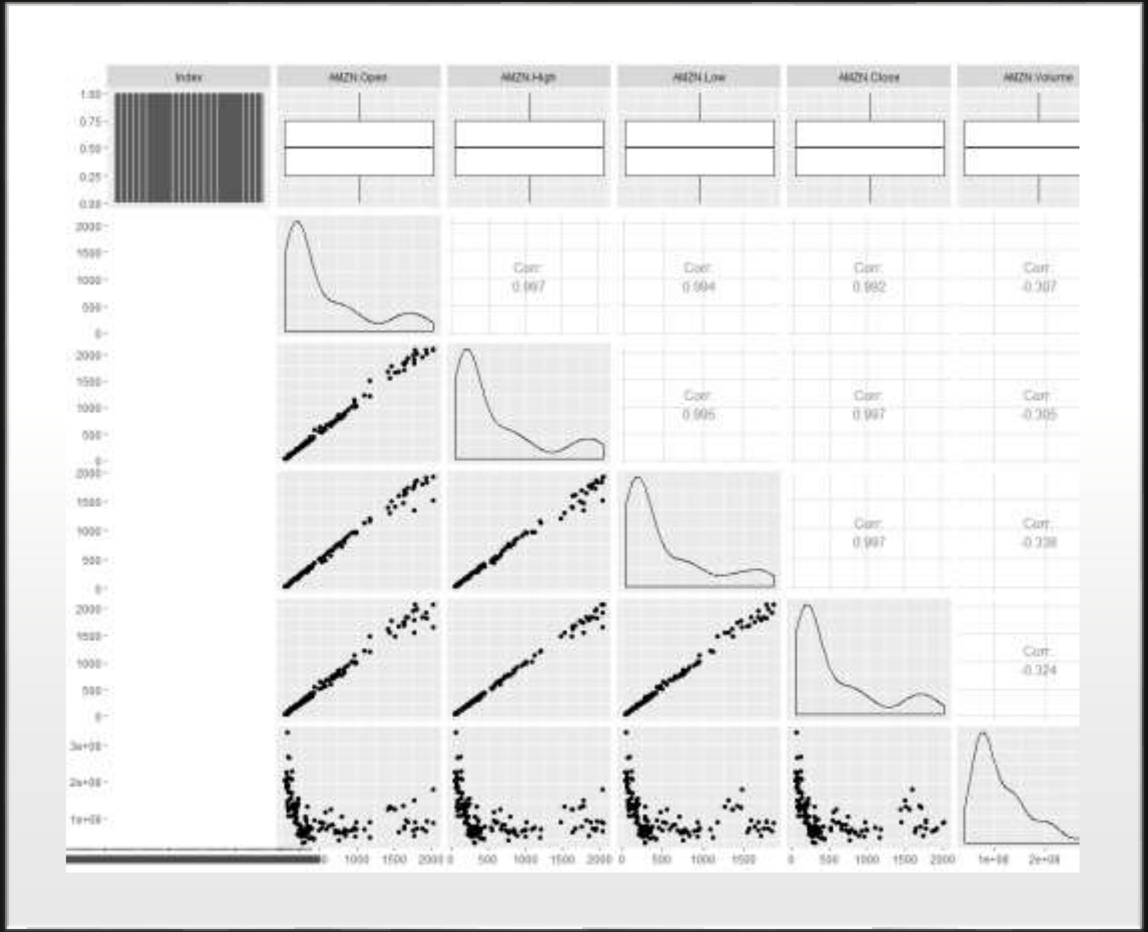
DATASET

- The dataset has 6 columns
- We are considering monthly stock values from Jan 2008 to Sep 2019

	AMZN.Open	AMZN.High	AMZN.Low	AMZN.Close	AMZN.Volume	AMZN.Adjusted
Jan 2008	93.26	93.40	68.84	77.70	265354200	77.70
Feb 2008	79.02	79.40	63.97	64.47	211217800	64.47
Mar 2008	63.59	76.93	61.20	71.30	192685600	71.30
Apr 2008	72.99	82.64	70.65	78.63	197849400	78.63
May 2008	78.40	84.75	71.56	81.62	152573200	81.62
Jun 2008	81.15	84.88	72.62	73.33	152905700	73.33
Jul 2008	72.24	82.38	62.99	76.34	215813300	76.34
Aug 2008	76.36	91.75	74.05	80.81	149044400	80.81
Sep 2008	83.16	86.77	61.32	72.76	192463400	72.76
Oct 2008	71.78	71.99	43.31	57.24	332074700	57.24
Nov 2008	56.35	58.73	34.68	42.70	228051500	42.70
Dec 2008	42.00	54.85	38.82	51.28	206194300	51.28
Jan 2009	51.35	59.74	47.63	58.82	204072300	58.82
Feb 2009	58.57	67.36	58.13	64.79	186780200	64.79
Mar 2009	63.94	75.61	59.82	73.44	227672300	73.44

SCATTER PLOT MATRIX

- Variables are highly correlated and the relationship is also linear
- Volume is not highly correlated

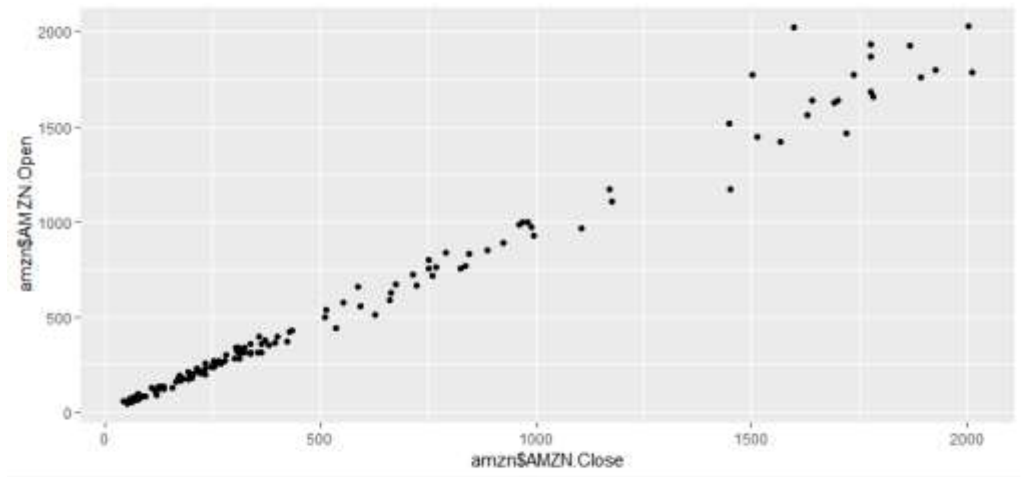


AMAZON CLOSING AND VOLUME

- There is inverse relationship between closing price and volume

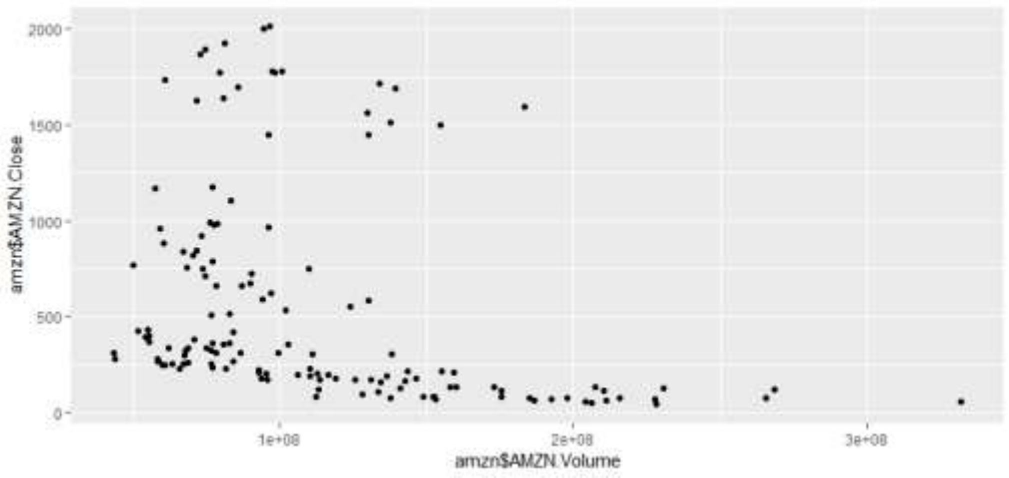
EXPLORATORY ANALYSIS

- Amazon close and amazon open has a linear relationship

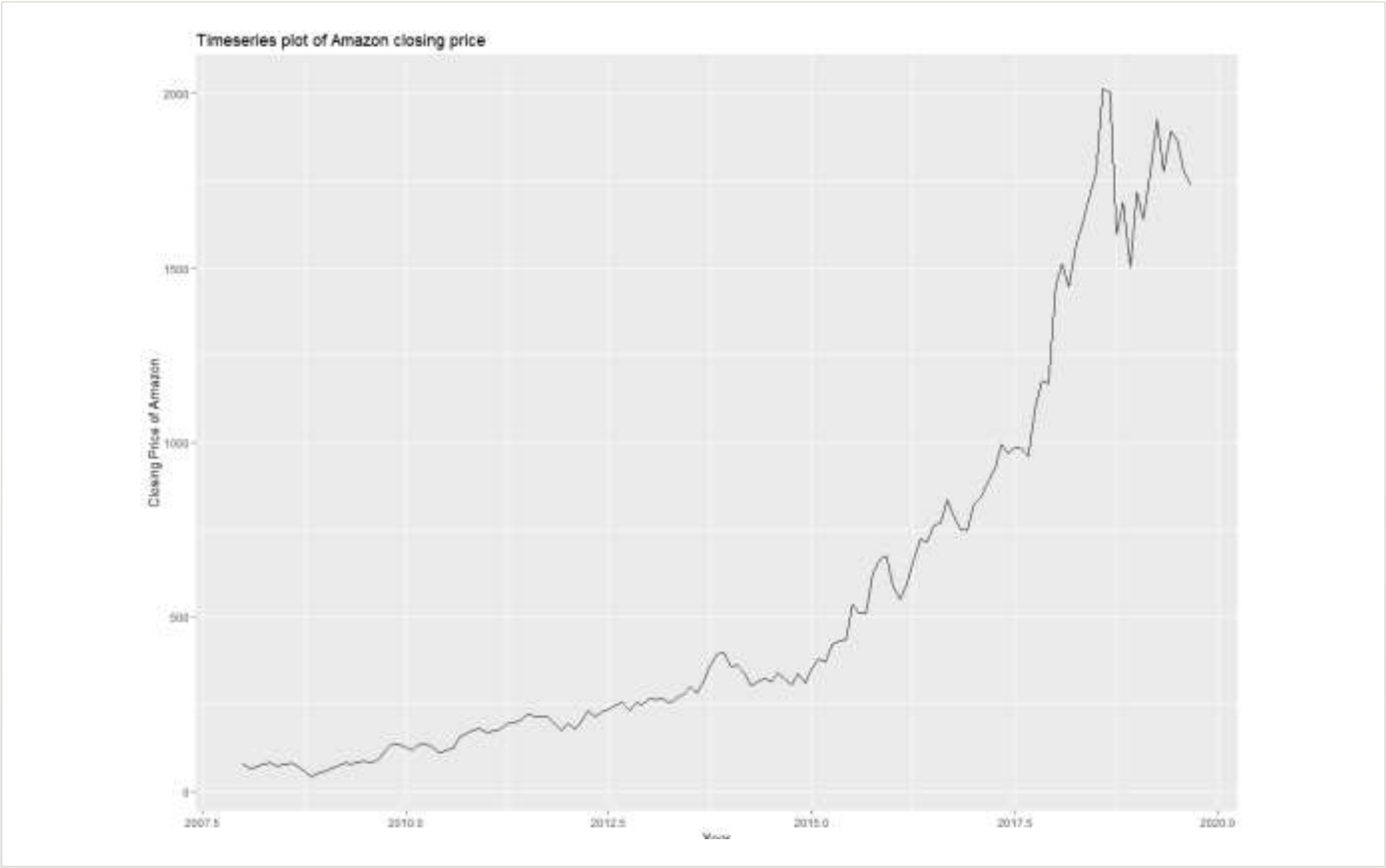


EXPLORATORY ANALYSIS

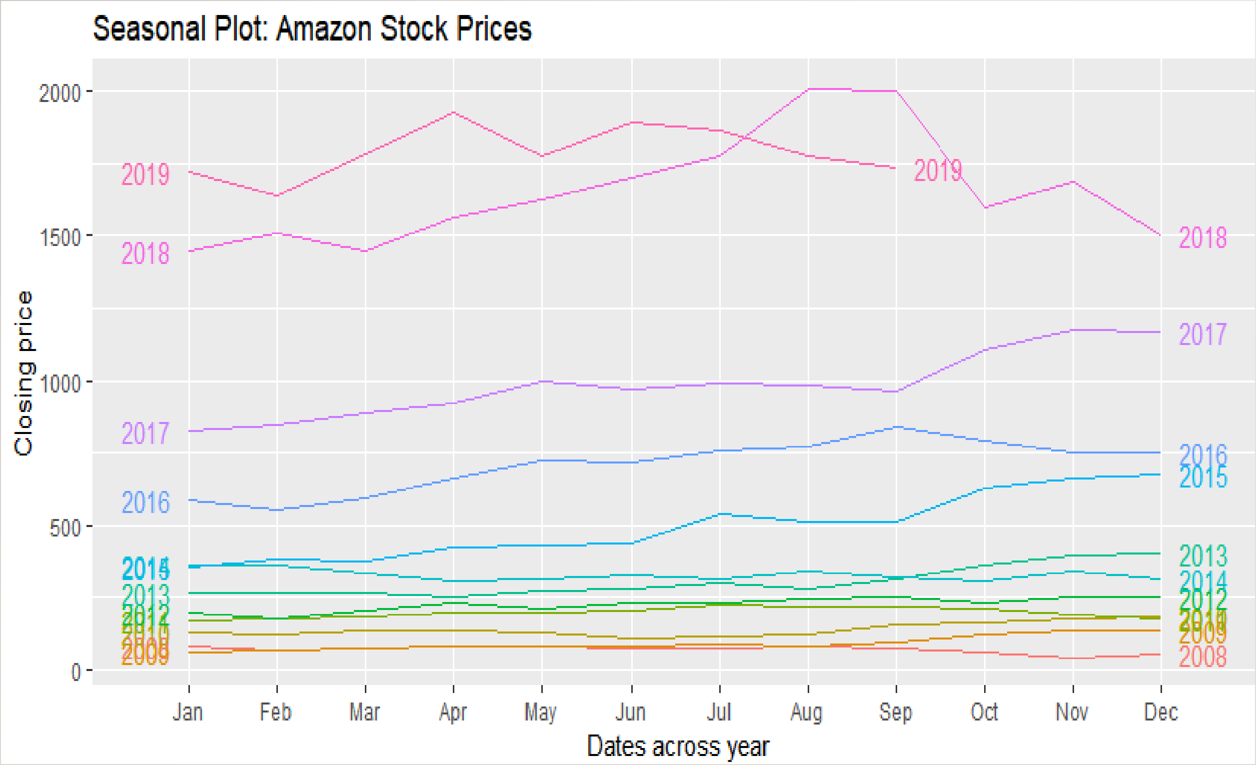
- For amazon volume and amazon close plot, the volume is high only when the closing price is low

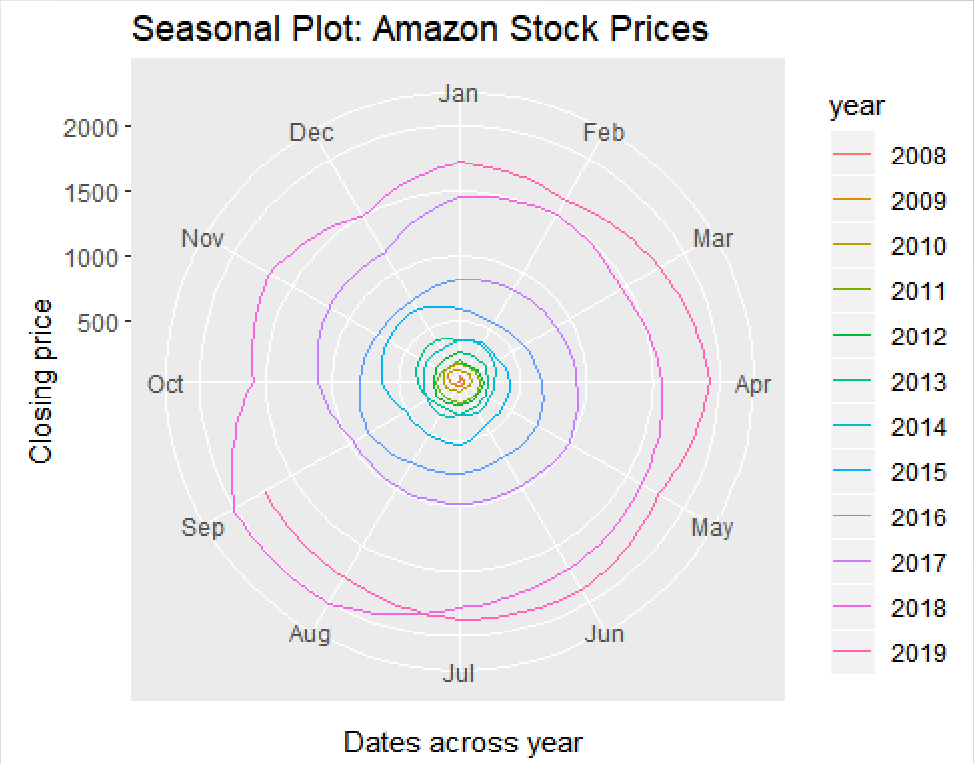


TIMESERIES PLOT OF AMAZON CLOSING PRICE



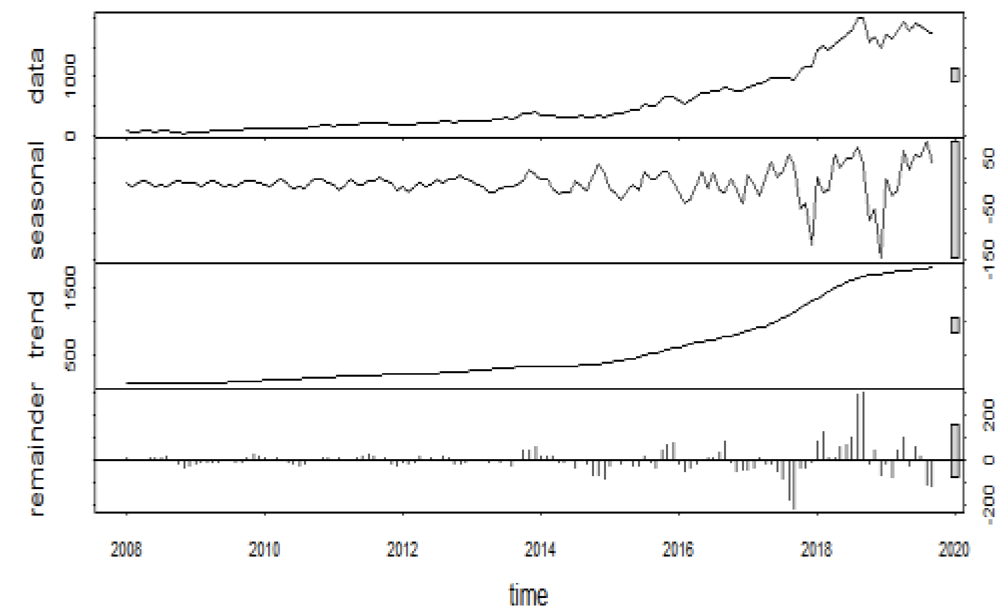
SEASONAL PLOT



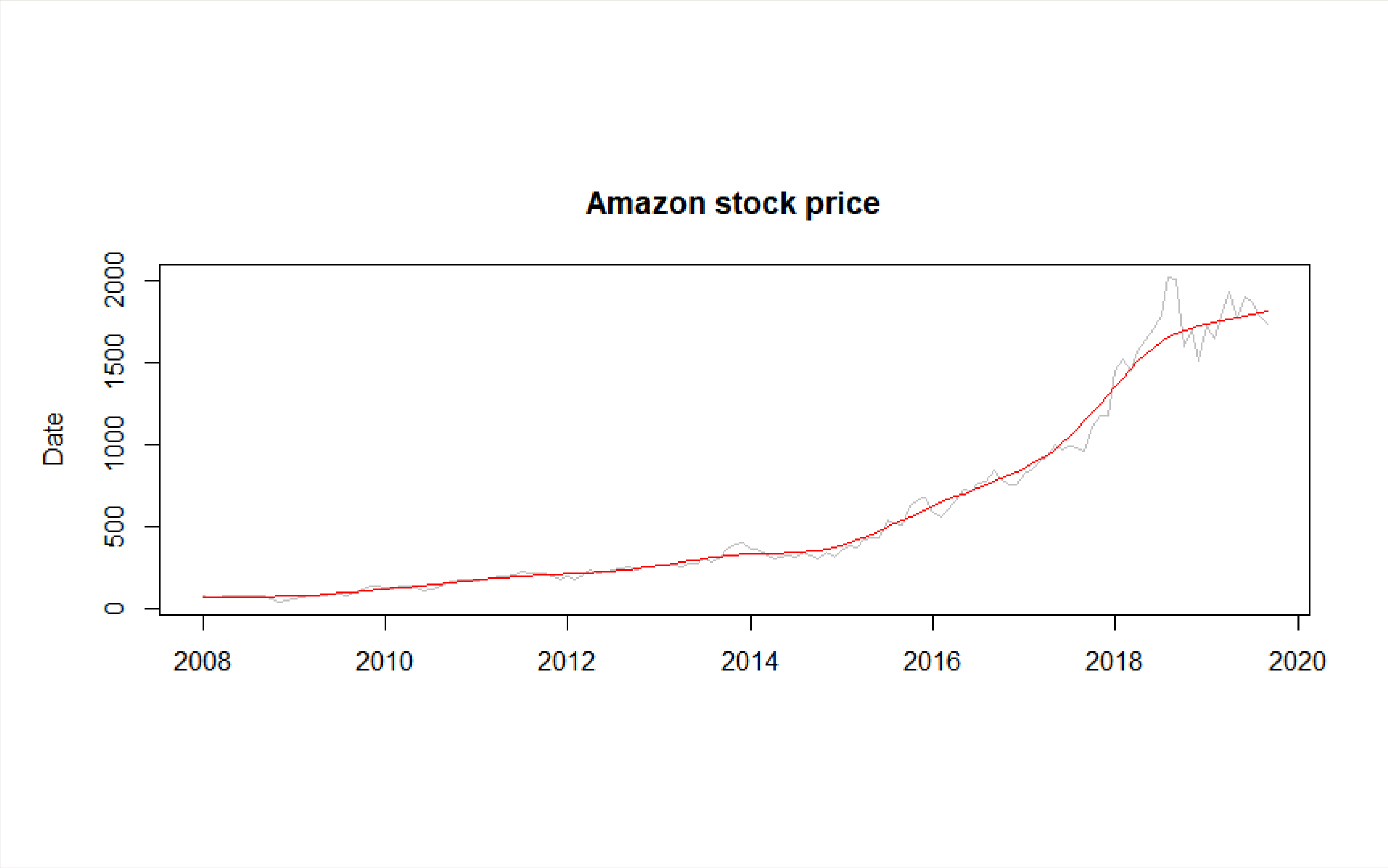


DECOMPOSITION

- The large grey bar in the bottom panel shows that the variation in the remainder component is small compared to the variation in the data. If we shrunk the bottom three panels until their bars became the same size as that in the data panel, then all the panels would be on the same scale.



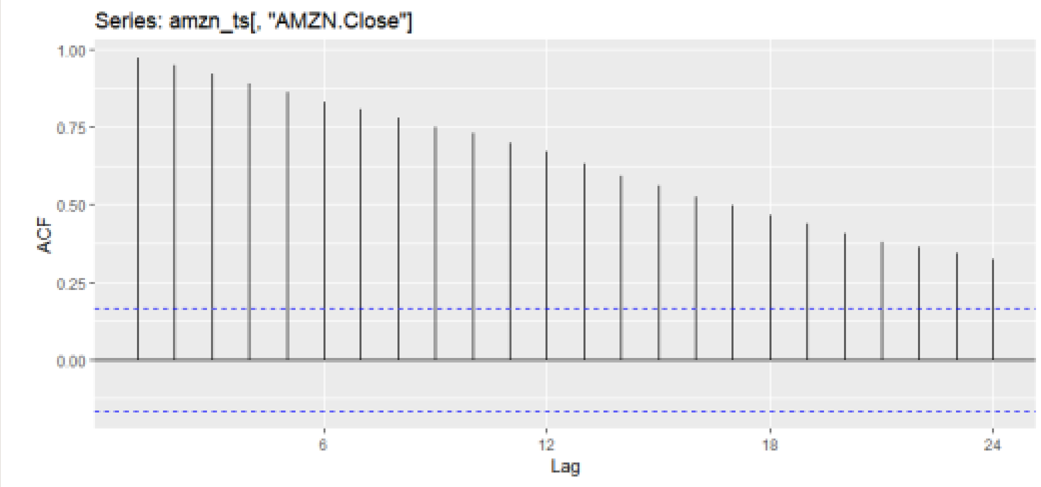
TREND CYCLE



Amazon stock price
Date
0
500
1000
1500
2000
2008
2010
2012
2014
2016
2018
2020

CORRELOGRAM

- They are highly correlated at initial lags but tend to decrease gradually. This shows data has trend
- We see no seasonality/cyclic behaviour in the correlogram



STATIONARY OR NON-STATIONARY?

- When there is large autocorrelation within our lagged values, we see geometric decay in our plots, indicating trend and hence the data properties depend upon time making it non-stationary data.
- For a stationary time series, the ACF will drop to zero relatively quickly, while the ACF of non-stationary data decreases slowly. Also, for non-stationary data, the value of r1 is often large and positive.
- Huge indicator that we will have to take the difference of our time series object.
- In our dataset, we only observe trend.. no seasonality or cyclic pattern..

KPSS TEST

Transforming our data to adjust for non-stationary

- In this test, the null hypothesis is that the data are stationary, and we look for evidence that the null hypothesis is false.
- Consequently, small p-values (e.g., less than 0.05) suggest that differencing is required.
- The test can be computed using the ur.kpss() function from the urca package

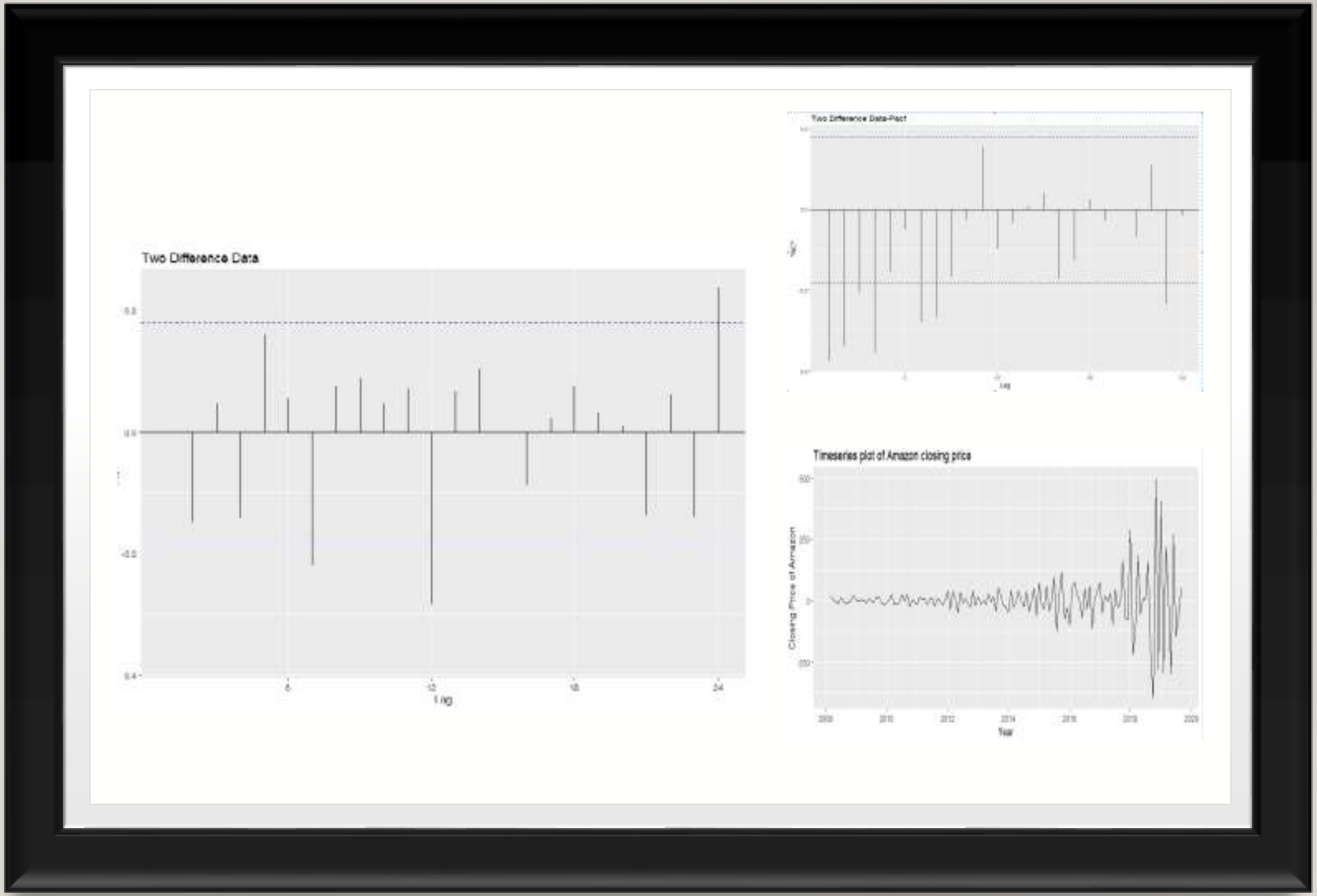
```
#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 4 lags.

Value of test-statistic is: 2.1342

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

```
#######################
# KPSS Unit Root Test #
#######################

Test is of type: mu with 4 lags.

Value of test-statistic is: 0.8417

Critical value for a significance level of:
                10pct  5pct 2.5pct  1pct
critical values 0.347 0.463  0.574 0.739
```

```
> Box.test(train_diff2, lag=12, type="Ljung-Box")

        Box-Ljung test

data:  train_diff2
X-squared = 45.776, df = 12, p-value = 7.581e-06
```

TEST FOR STATIONARY DATA

STATIONARY TIME SERIES DATA



DIVIDING DATA INTO TRAINING AND TESTING DATA

```
> train <- ts(amzn[,"AMZN.Close"],start=c(2008,1),end=c(2017,12),freq = 12)
> test<-ts(amzn[,"AMZN.Close"],start=c(2018,1),end=c(2019,9),freq = 12)
> dim(train)
[1] 120   1
> dim(test)
[1] 21  1
>
```

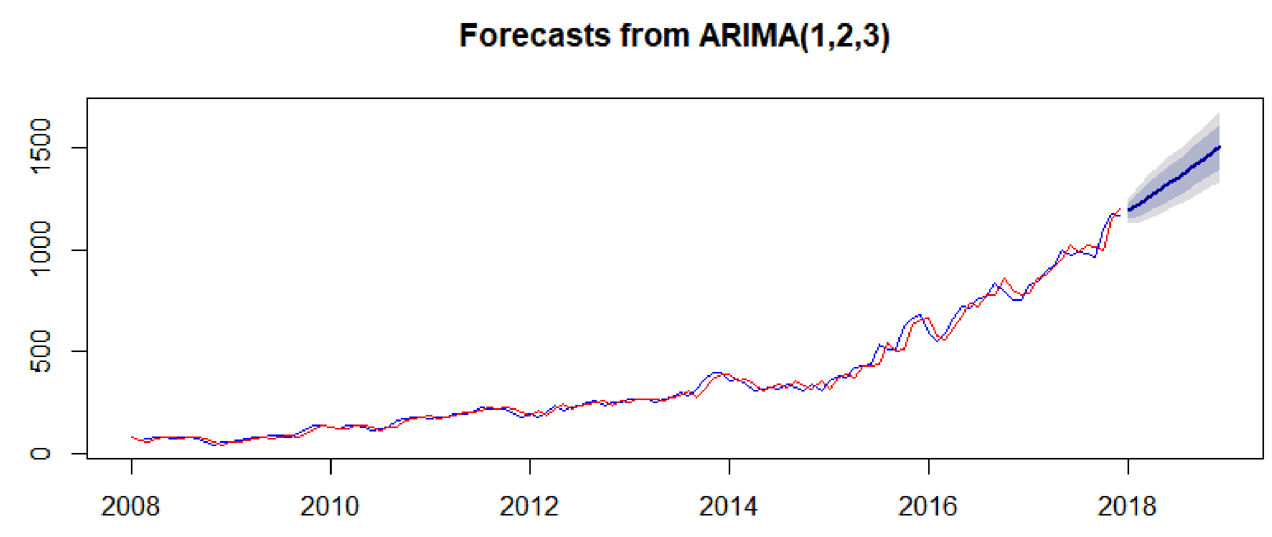
- Building model on the data from year 2008 to 2017.
- Testing model on the data from year 2018 to 2019

ARIMA MODEL - RESULT

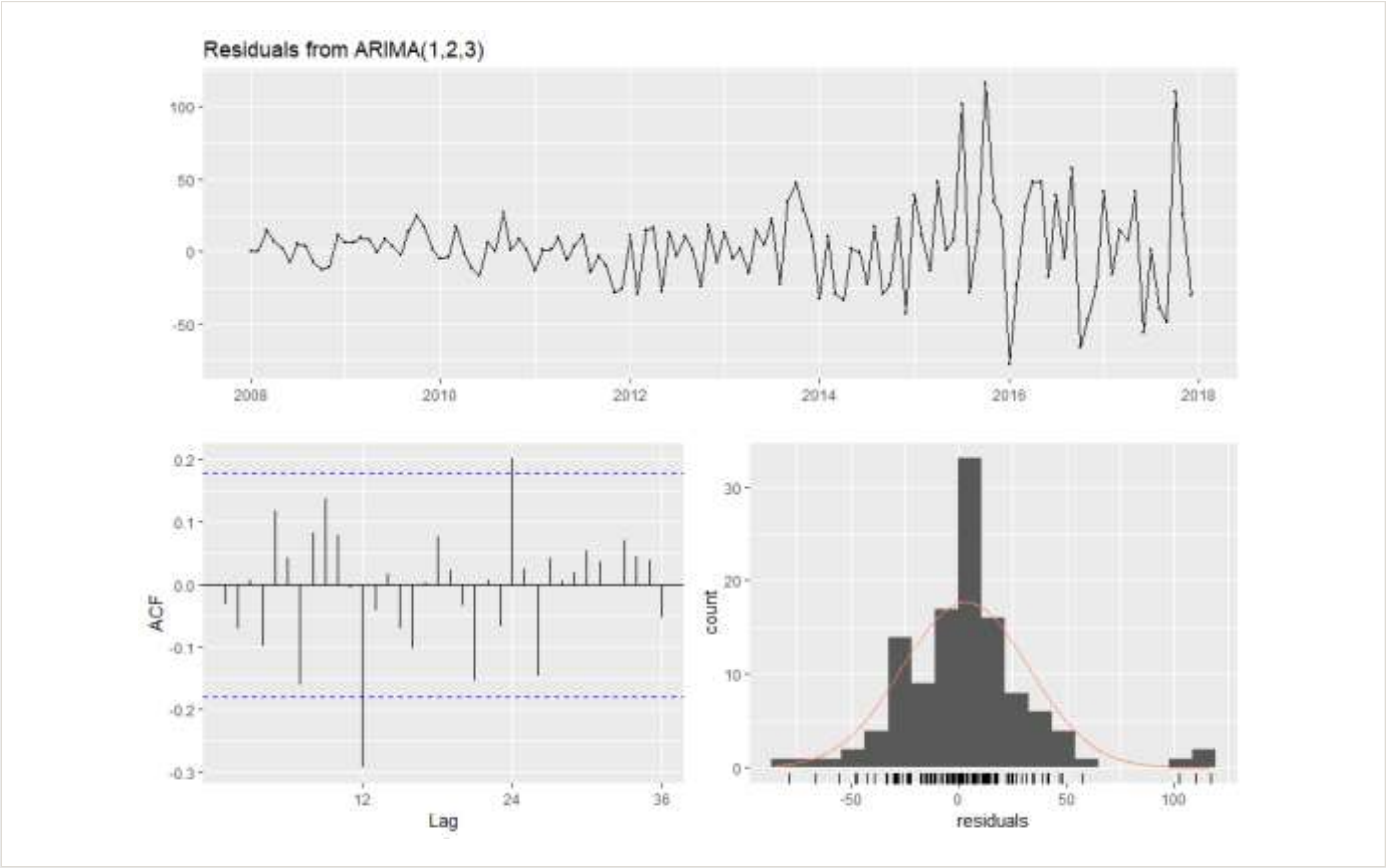
```
> fit<-auto.arima(train,seasonal = FALSE)
> summary(fit)
Series: train
ARIMA(1,2,3)

Coefficients:
         ar1      ma1     ma2     ma3
      0.8016  -1.7524  0.5746  0.1943
s.e.  0.1272   0.1835  0.2770  0.1150

sigma^2 estimated as 907:  log likelihood=-569.5
AIC=1148.99   AICc=1149.53   BIC=1162.85

Training set error measures:
                   ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
Training set 3.255827 29.35461 20.19418 0.9430978 6.763287 0.1901981 -0.03183544
>
```

```
> accuracy(fit2,test)
                      ME       RMSE        MAE          MPE        MAPE       MASE        ACF1 Theil's U
Training set    3.255827   29.35461   20.19418    0.9430978    6.763287  0.1901981 -0.03183544        NA
Test set     -1403.129084 1414.10119 1403.12908 -1989.8841510 1989.884151 13.2153185  0.86108129  165.7647
>
```

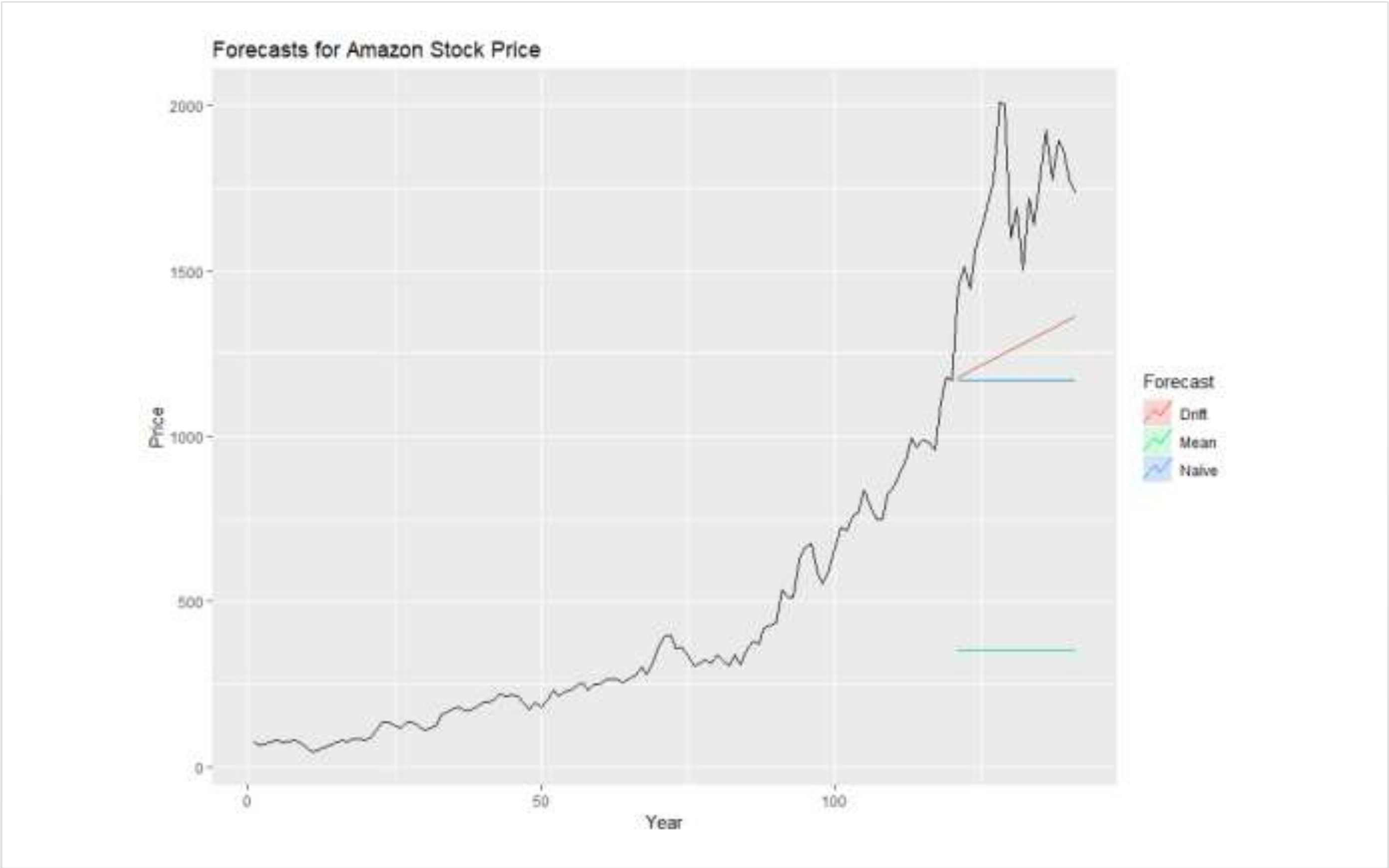
ARIMA MODEL


Forecasts from ARIMA(1,2,3)
0
500
1000
1500
2008
2010
2012
2014
2016
2018

CHECKING RESIDUALS: NORMAL DISTRIBUTION?



OTHER FORECASTING METHODS



CONCLUSION

Model	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
ARIMA	3.58	28.43	20.15	0.90	6.81	0.18	0.03
Mean Forecast Method	1.87e-14	286.35	226.61	-93.96	121.94	2.13	0.95
Naive Forecast Method	9.17	32.37	21.70	1.84	7.28	0.20	0.09
Drift Method	-1.19e+03	31.04	21.22	-3.12	7.57	0.19	0.09