Time-Series Forecasting of Banana Prices

Kaisa Roggeveen, Scott Graham, Irwin Khuu, Johnson Tran

April 8th, 2018

Abstract

This is a time-series analysis used for forecasting the price of the commodity, bananas. To conduct this analysis the original data set was examined to identify trend and seasonal components. Since this data set had both components, transformations were made, followed by using the minimum AIC and BIC for model selection. Once the model was selected, and diagnostics were completed, it was used to forecast the price of bananas to the year 2020. The expected price will increase to $1150 USD/Metric Ton by 2020.

Introduction

The intent of this paper is to forecast the price of Bananas. This paper uses time-series modeling techniques.The data was collected from the Federal Reserve Bank of St. Louis webpage, and ranges from January 1998 to December 2017 (FRED, 2017).

Bananas are a type of fruit that is consumed worldwide for its nutritional value and the availability of Bananas throughout the year (Britannica T.E, 2018). The main type banana that is exported is the Cavendish Banana (Britannica T.E, 2018). Though bananas can be frozen and used for baking, they are most commonly consumed fresh and therefore it is important for suppliers to always have a new supply of bananas for their customers (Britannica T.E, 2018).

There are a few important considerations when evaluating the price of bananas, the first is evaluating if there exists a seasonal component and a trend component. These components might occur because Banana plants only fruit once in their lifetime, and the banana crop is mostly exported (Britannica T.E, 2018). When banana plants fruit they only produce about 50 to 150 bananas each, once a banana plant has fruited, it's trunk is cut down to the ground and will be replaced by a rhizome where it will take between 9 to 12 months for the plant to fruit again (Britannica T.E, 2018),(Banana Link, n.d). Therefore, this paper will be evaluating the seasonal components and the trend components of the price of bananas overtime.

Methodology

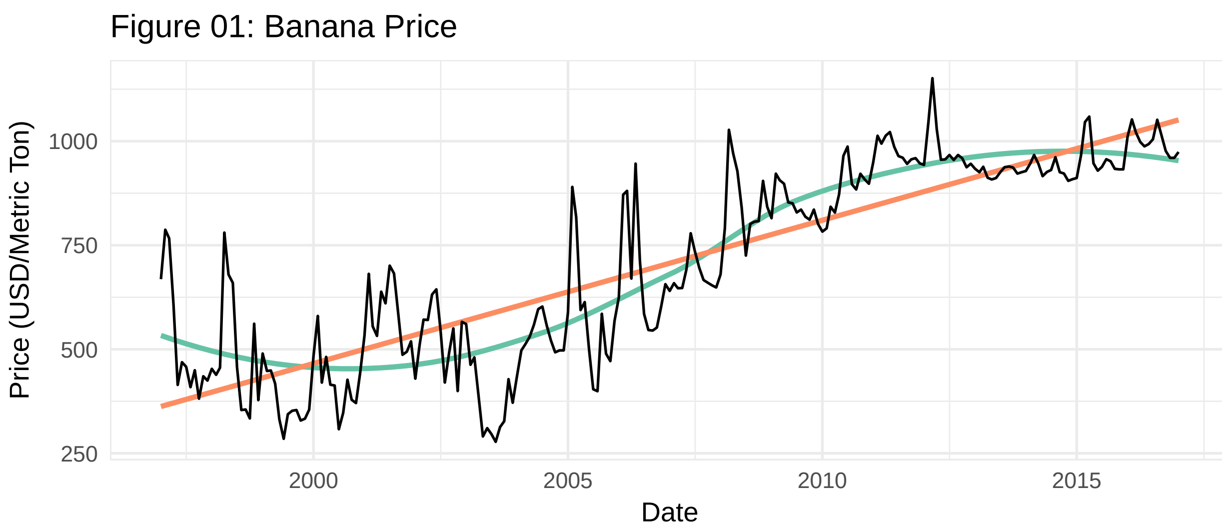
Original Time Series Plot

To begin analysis of the stock price of Bananas a plot of the price and the date was created, and is shown in Figure 1.

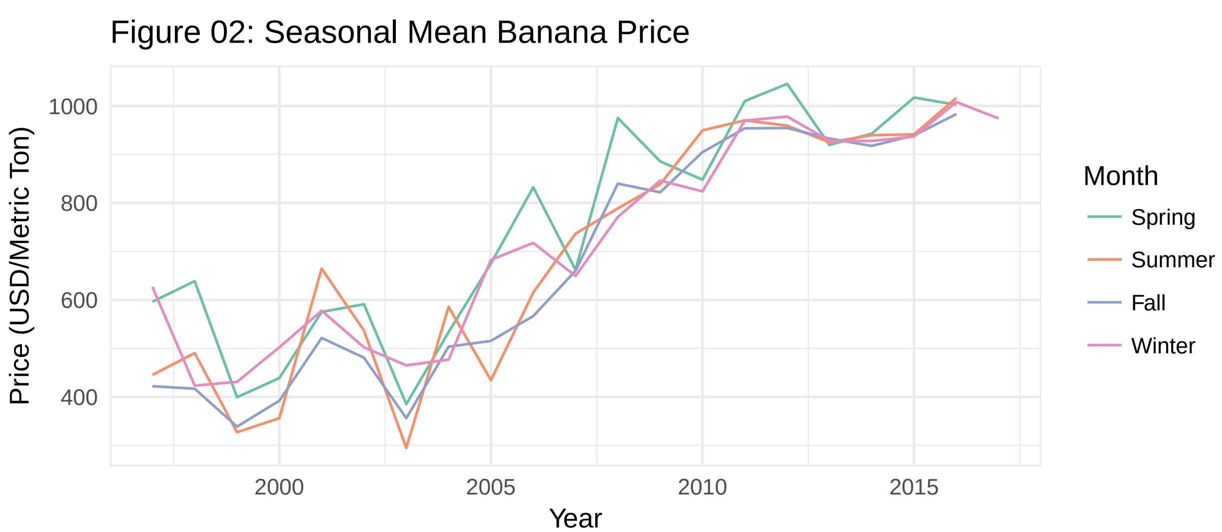
In Figure 1, there appears to be a seasonal effect on the price of bananas. This is noticed because of the rhythmic spikes and drops in the price that appears to follow a similar trend throughout multiple years.

A linear regression line was fitted to the data which has positive slope indicating that there is likely an increasing trend in the price of bananas since 1997.

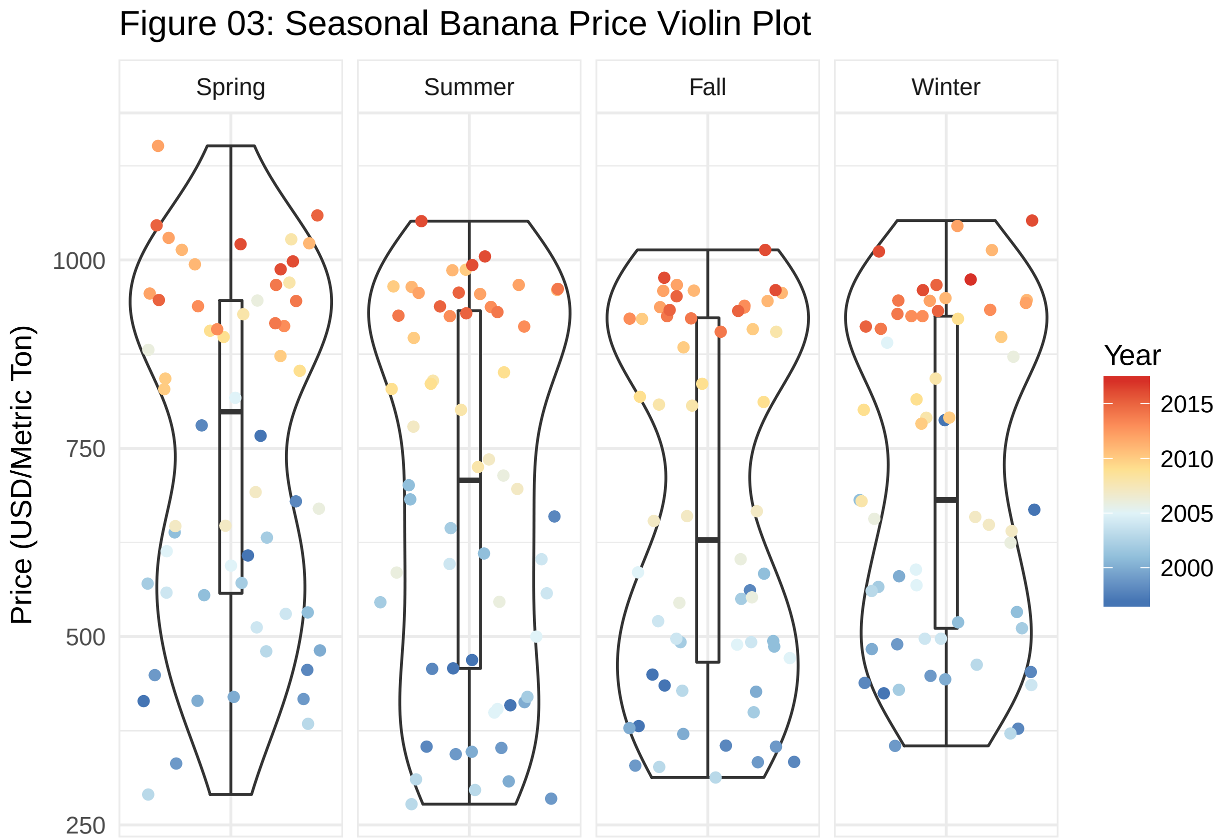
A local regression (loess) curve was fitted to the data. This adds a smooth curve to all of the data. The loess curve is used to gauge the fit of the linear regression model, where a good model would have a loess model similar to the linear regression model. Since the loess model and the linear regression model deviate significantly, it requires further analysis and adjustments to the data.



Orginial Seasonal Plot

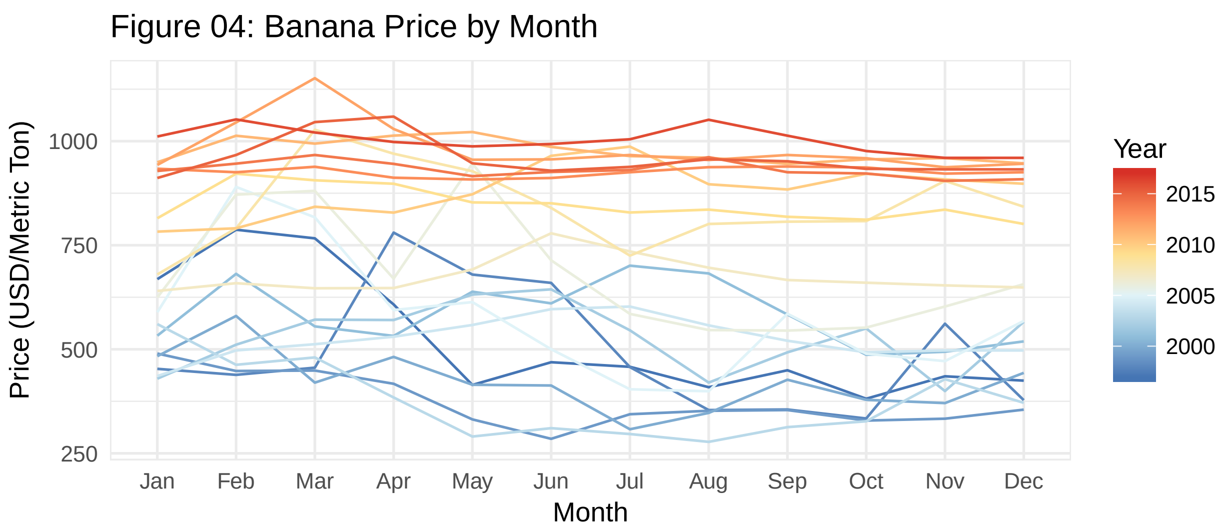


In Figure 02, the price of bananas was observed over each year, and stratified by the four seasons. From the seasonal plot, each month appears to follow a similar overall increasing trend in the price from 1997. In the graph, the spring months are the highest priced months over the time frame while fall months generally have the lowest prices over the time period. The winter and the summer months are generally in between. This indicates that over the time-series there is likely a seasonal component to this data, which is investigated further in Figure 03.



In Figure 03, violin plots of the price of bananas were separated by the season. Winter included December, January and February; spring included March, April and May; summer included June, July and August; and fall included September, October and November. Overall the trend shows that the mean and the median for the price of bananas in the spring is the highest, and the mean lowest price of bananas happens in the fall. From this, it can be concluded that there is some seasonal affect on the price of bananas.

Figure 03 also shows the price of bananas and the year as coloured points. These points show an overall trend that the cost of bananas has been increasing since 1998, as we can see the change in the colour from blue (indicating the earlier years) at the bottom and red (indicating the later years) at the top. Figure 04 shows this more clearly.



According to Figure 04, there has been an increase in the price of Bananas over the last twenty years. Banana prices have approximately doubled between 1998 and 2018, though the prices have been relatively steady in approximately the last 7 years. This may indicate that inflationary effects have been relatively muted since 2010.

Lagged Scatter Plot

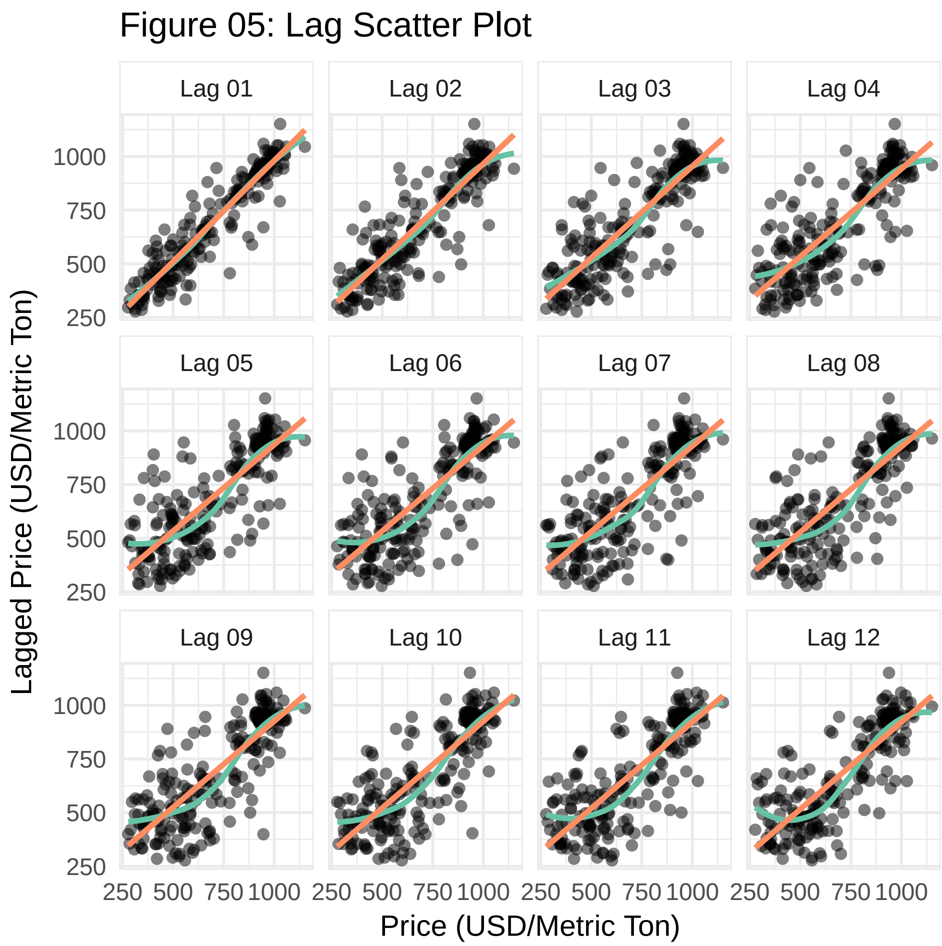


Figure 05 show the lag plots at each month in our data set. The lag plots are showing the potential correlation between one lag and the previous lag. In these plots we are comparing the first month to all the other months. There appears to be significant correlation in lags 1 through 4, and in lags 4 through 12 the distribution appears to be more random.

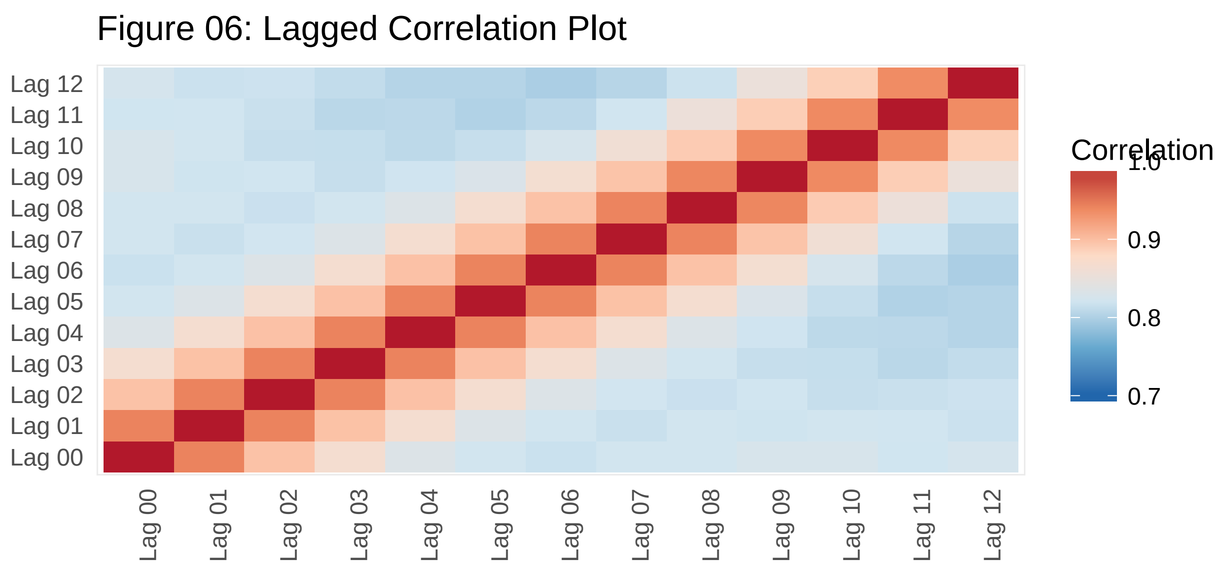


Figure 06 is a correlation plot between the lags. In general the correlations range from around 0.75 to perfect correlation. It can be seen that the correlation remains very high (between 0.85 and 1) within the first 4 lags, then starts to decrease below 0.8.

Autocorrelation Function and Partial Autocorrelation Function

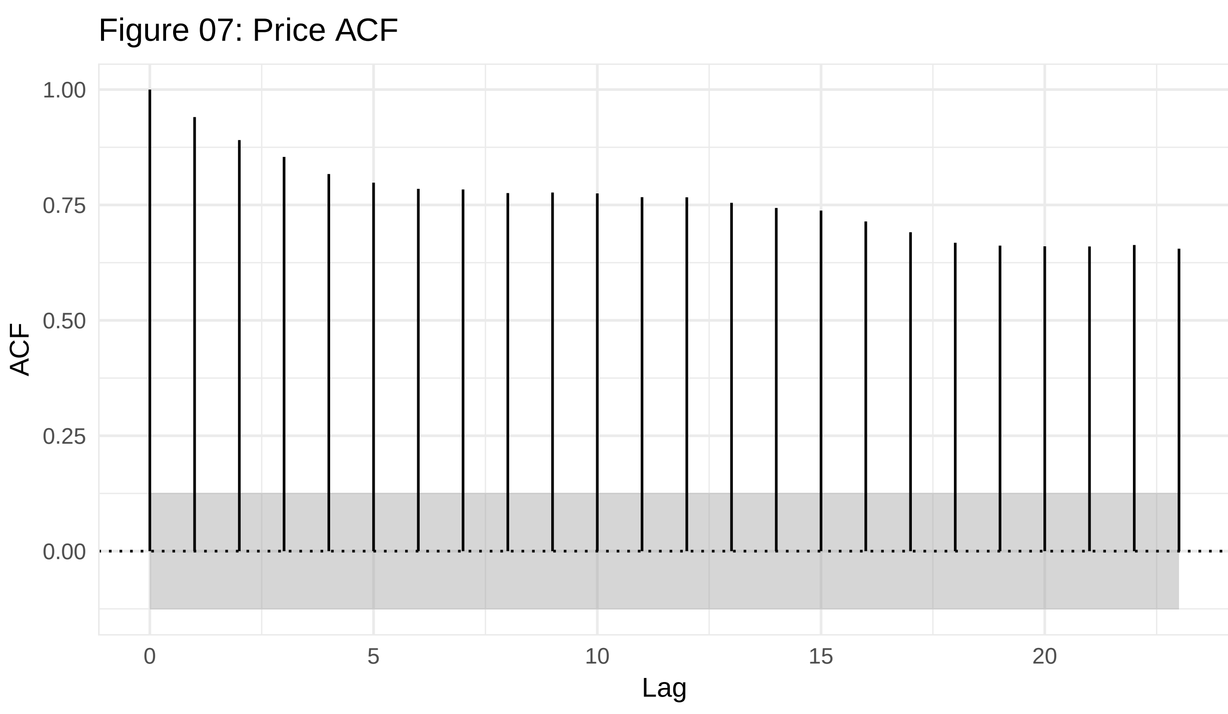
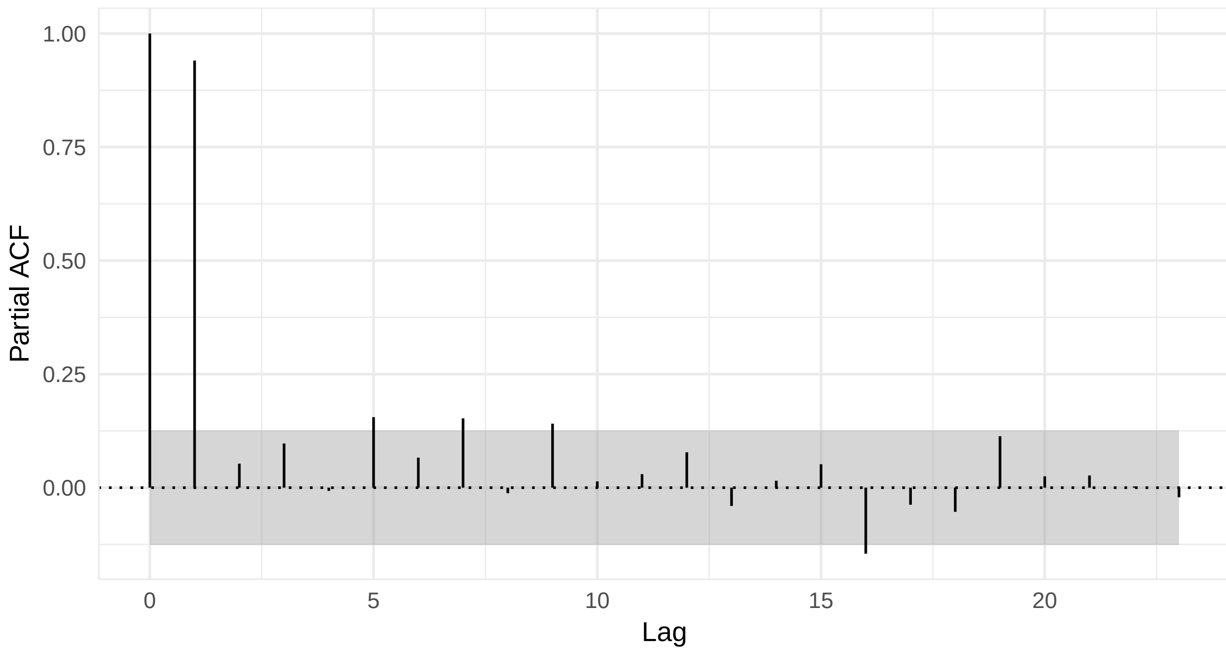


Figure 07 is a plot of the Autocorrelation function, which is checking the correlation between a lag and the previous lag. In this case, there is a gradual decrease in the spikes along the plot, which might indicate that there is some autoregressive component required in the model.

Figure 08: Price PACF



In the Partial Autocorrelation plot in Figure 08, there is a significant spike at lag 1, followed by correlations that are mostly not significant, therefore it is suspected that this has an autoregressive component. Furthermore, there are noticeable spikes approximately every 4 lags, indicating a seasonality component.

Transformations

Since there are both seasonal and trend components in the Banana Price data set, transformations are required. Multiple different transformations were tested on the Banana Price data set and can be seen in Figure 09. Note that once again the orange line represents the results from fitting a linear regression model while the green line represents the results from fitting a loess regression model.

The first transformation was to take the log of the Banana Price. This transformation did not have a significant impact on removing the trend or the seasonality from the data set.

The second transformation was using the Box-Cox method in R. Again, there was not a significant reduction in the trend or the seasonality of the data set.

The third transformation performed was to change the data set from the stock price of bananas to the return of the bananas. The new data set, Banana Return, resulted in a significant reduction to the trend of the Banana Price.

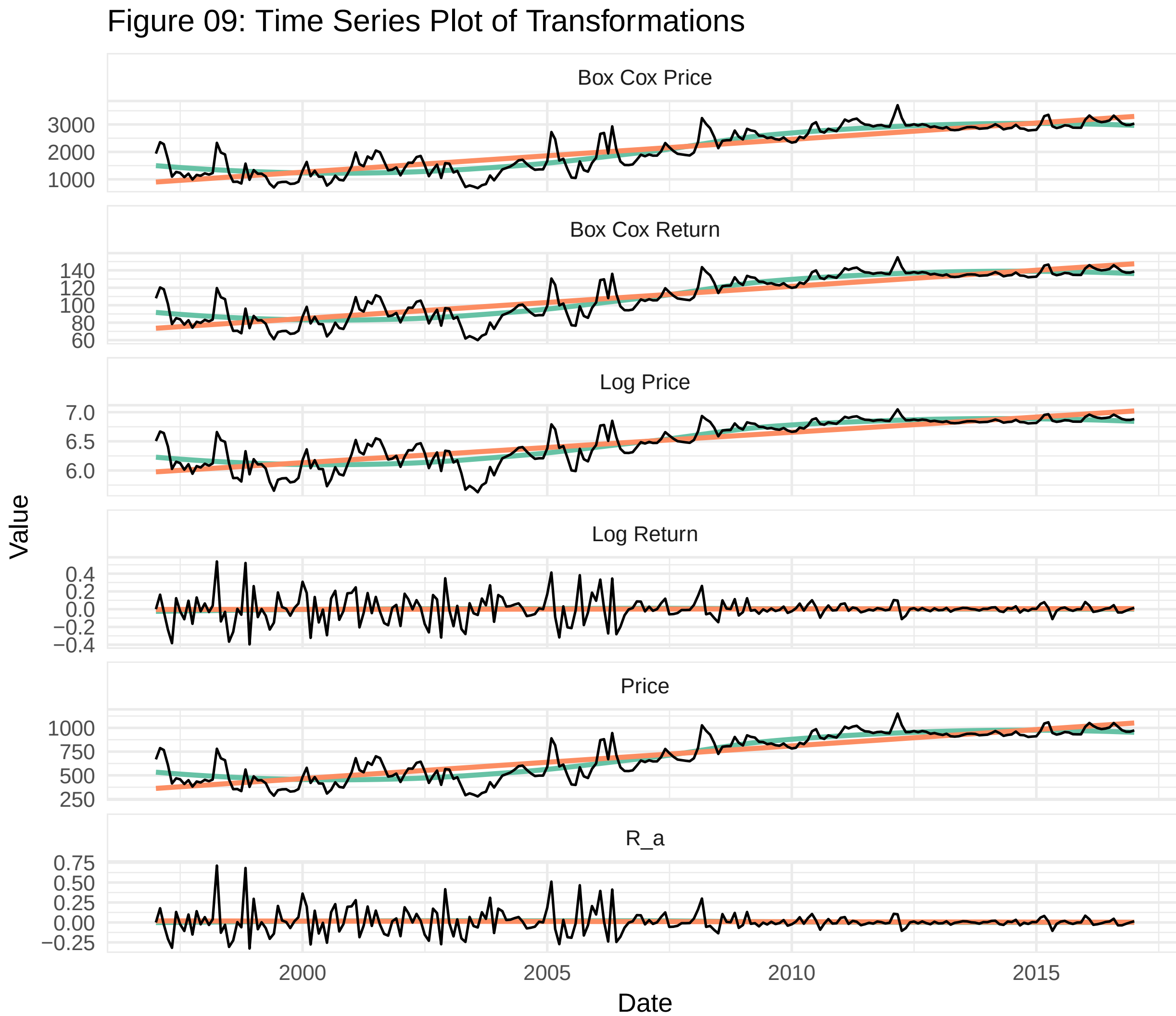
Following the initial transformation of the Banana Price to Banana Return, additional transformations were conducted to further reduce the trend.

A transformation was applied by taking the log of the Banana Return. This transformation also had a significant effect on the removal of trend in the Banana Return data set.

A Box-Cox transformation was also performed on the Banana Return data set, this transformation did not result in any significant removal of the trend in the Banana Return data set.

Therefore, due to the ease of interpretation and removal of trend, the Banana Return and the log of Banana Return will be used for further modeling of the Price of Bananas.

Transformation Time-Series Plot



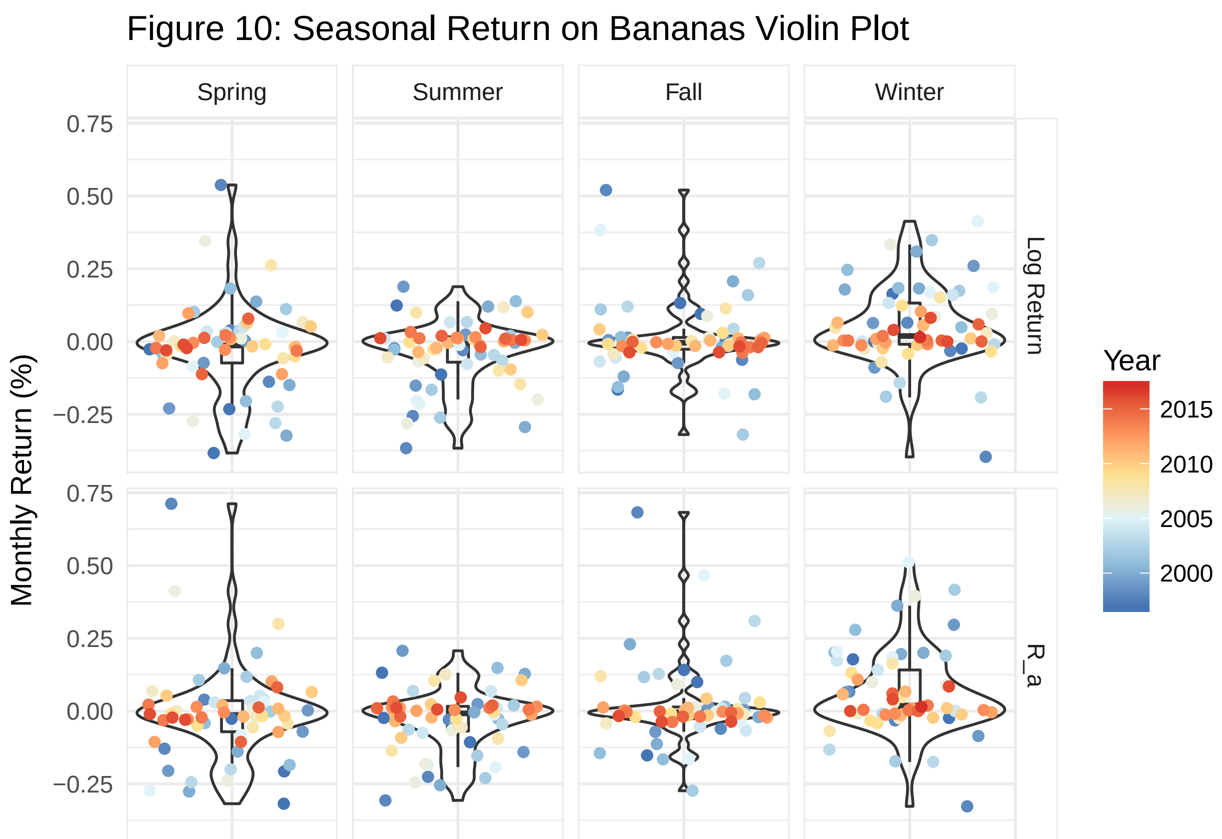
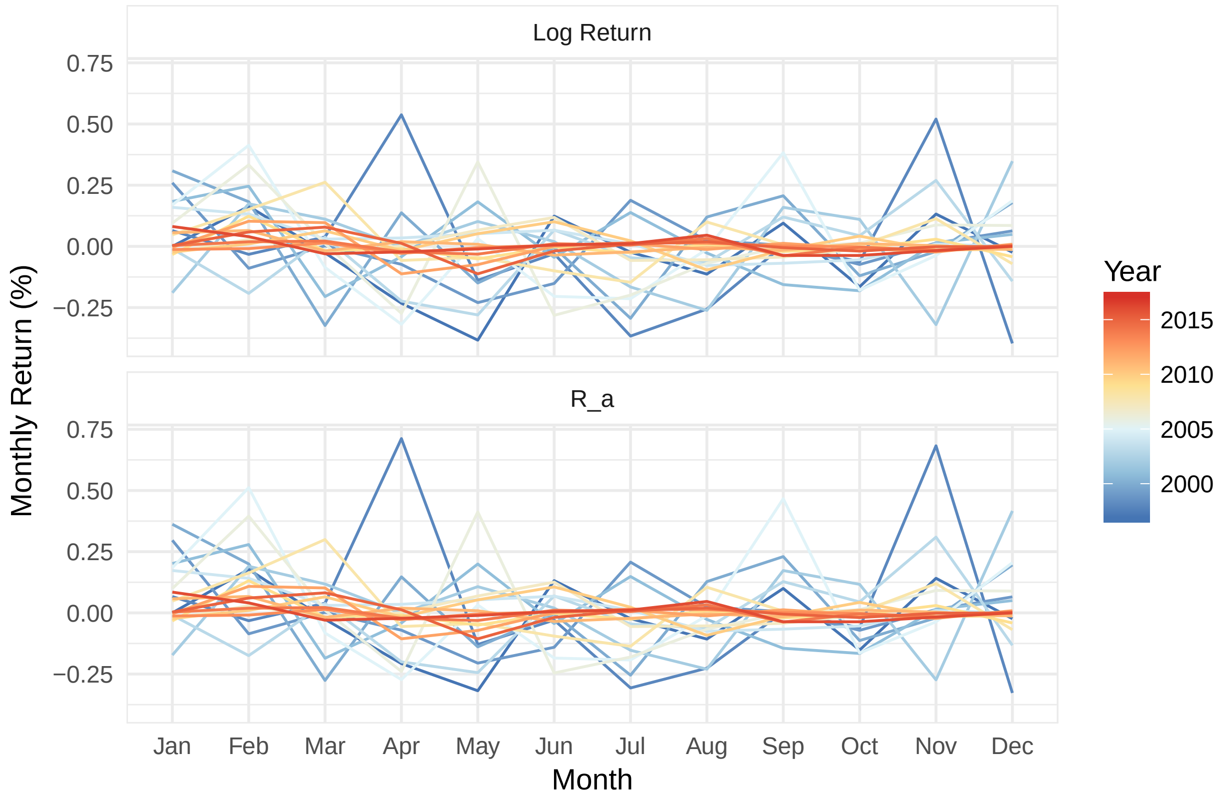


Figure 10 displays violin plots separating the banana data by season. After the transformation the effects of the increasing trend have been removed. This is observed as the data points are not showing the same trend where the earlier years, seen in blue, appeared near the bottom and the later years, seen in red, appeared near the top and the observed pattern is more random. In addition, the median of the log Banana Return, is more consistent, indicating a reduction in seasonal effects.

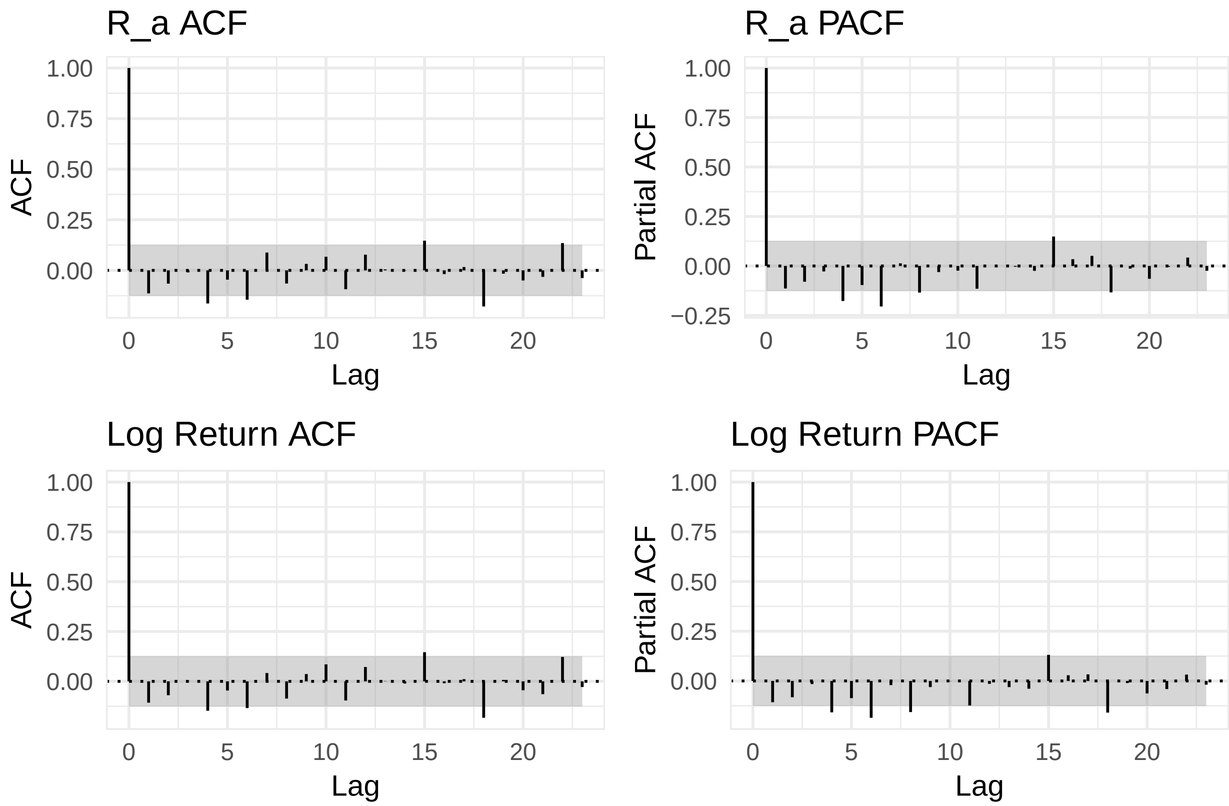
Figure 11: Return on Bananas by Month



From Figure 11, the elimination of the trend in the data set is apparent since the gradual increasing trend from the earlier years, in blue, and the later years, in red, is no longer present. Additionally, the plot also shows that there is higher variation on return rate during the earlier years compared to the later years.

Transformation Autocovariance Function

Figure 12: ACF & PACF Plots for Return on Bananas



The plots in Figure 12 all show a significant correlation at the first lag, followed by correlations that are not significant.

Decomposition of Transformation

To confirm if there exists a trend or seasonal component in the Banana Price, the Banana Return and the Log of the Banana Return data sets, the decomposition method was used.

Table 1: Price Decomposition Summary Statistics

Seasonal	Trend	Remainder
Min. :-49.45736	Min. : 365.6	Min. :-167.4500
1st Qu.:-34.29768	1st Qu.: 513.6	1st Qu.: -41.3469
Median : -6.21290	Median : 673.0	Median : -7.3479
Mean : -0.02578	Mean : 707.4	Mean : -0.6704
3rd Qu.: 26.06852	3rd Qu.: 934.7	3rd Qu.: 26.9518
Max. : 71.54570	Max. :1017.6	Max. : 241.0151

Table 2: Return Decomposition Summary Statistics

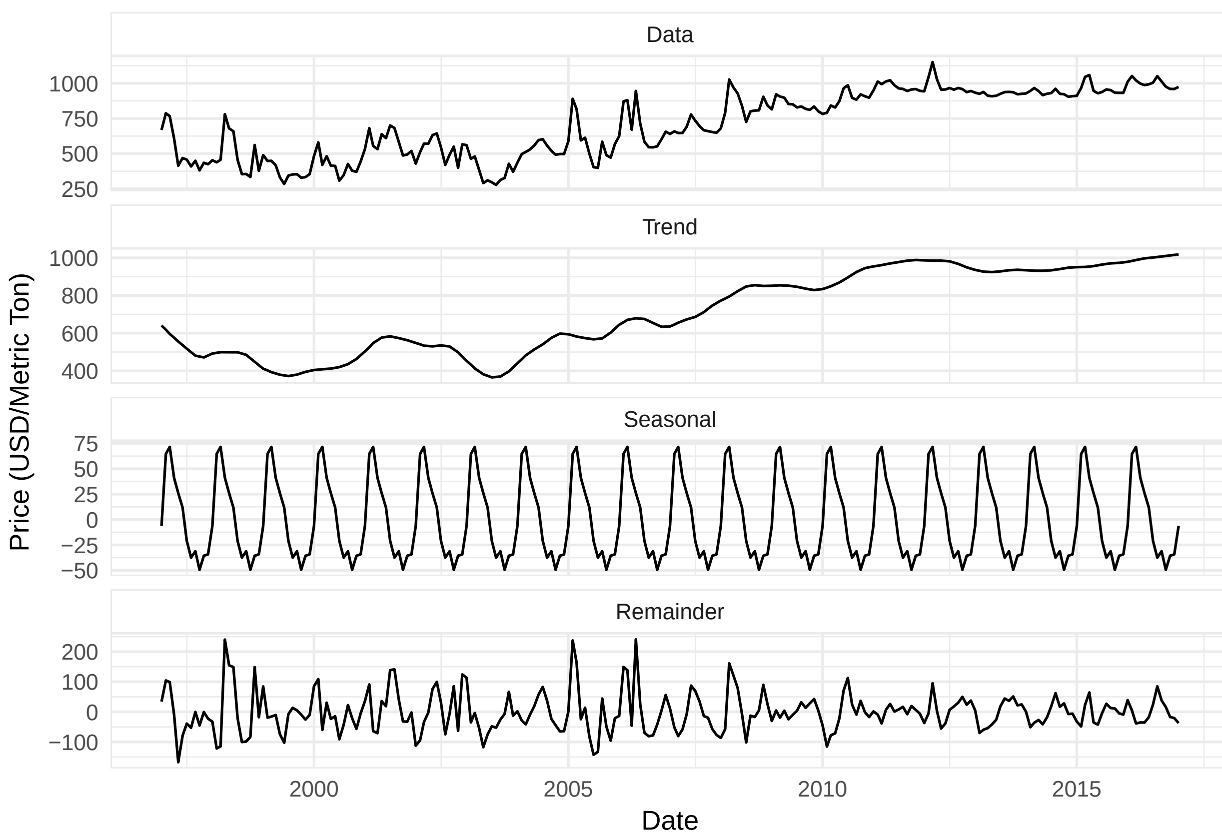
Seasonal	Trend	Remainder
Min. :-0.0587901	Min. :-0.0641957	Min. :-0.3340027

Seasonal	Trend	Remainder
1st Qu.:-0.0316258	1st Qu.:-0.0002795	1st Qu.:-0.0588012
Median :-0.0018949	Median : 0.0053194	Median :-0.0055948
Mean : 0.0002256	Mean : 0.0104364	Mean : 0.0002157
3rd Qu.: 0.0378455	3rd Qu.: 0.0235565	3rd Qu.: 0.0524438
Max. : 0.1012764	Max. : 0.0630801	Max. : 0.7124953

Table 3: Log Return Decomposition Summary Statistics

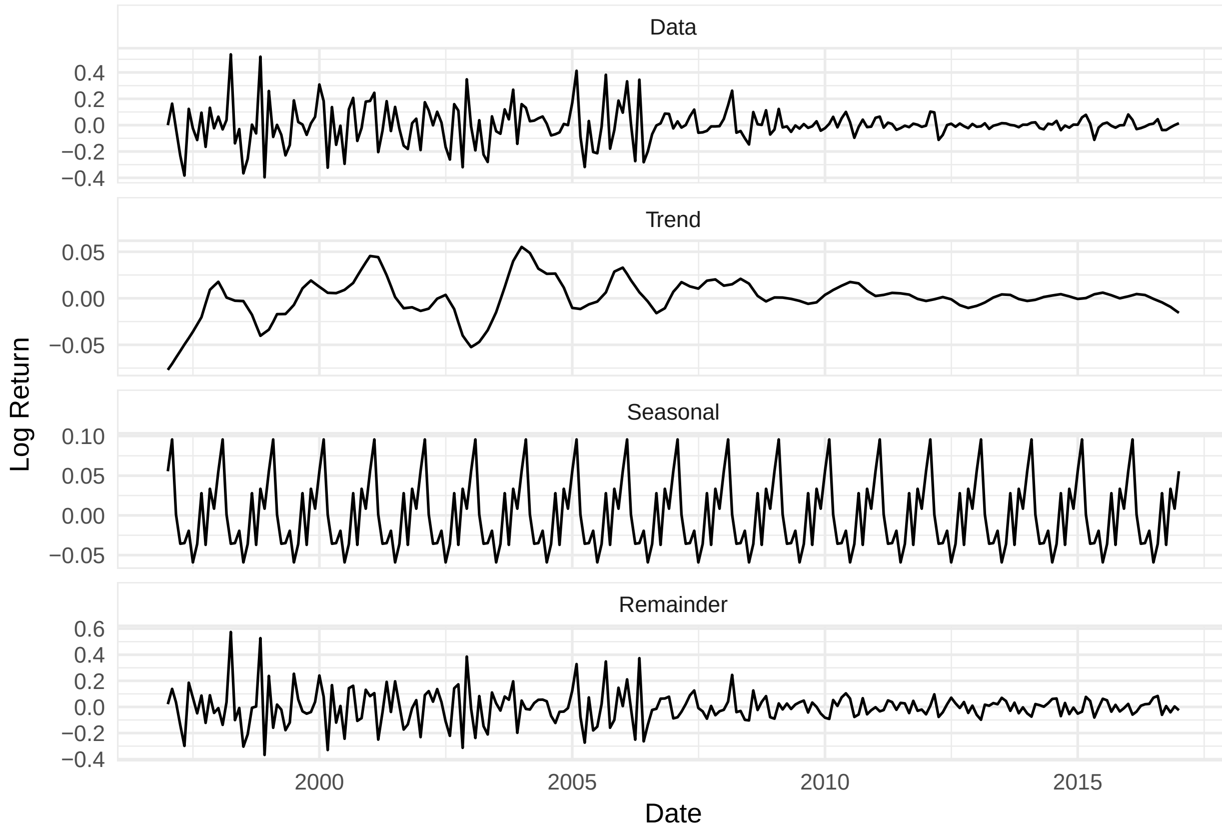
Seasonal	Trend	Remainder
Min. :-0.0591749	Min. :-0.076904	Min. :-0.3676951
1st Qu.:-0.0357003	1st Qu.:-0.005612	1st Qu.:-0.0549211
Median : 0.0011470	Median : 0.001408	Median : 0.0002603
Mean : 0.0002305	Mean : 0.001110	Mean : 0.0002209
3rd Qu.: 0.0334869	3rd Qu.: 0.011565	3rd Qu.: 0.0548785
Max. : 0.0956195	Max. : 0.055215	Max. : 0.5743404

Figure 13: Banana Prices Decomposition



From figure 13, we see that the mean seasonal effect is 0.0002305%, the mean trend effect is 0.001110%, and the mean remainder effect is 0.0002209.

Figure 14: Banana Log Returns Decomposition



In Figure 14: The Banana Log Returns and the Remainder show a similar pattern. This is evidence that there is no significant seasonal nor trend component in the model.

ADF Test (Dickey's Test)

The Augmented Dickey-Fuller test is used to evaluate whether the time-series of Banana Price, Banana Return and the Banana Log Return are stationary or integrated. The Price model is not stationary at lag 12, but is stationary at lag 6. The Return and Log Return models are both stationary at lags 6 and 12.

Table 4: Augmented Dickey-Fuller Test: `.$Price`

Test statistic	Lag order	P value	Alternative hypothesis
-3.767	6	0.02131 *	stationary

Table 5: Augmented Dickey-Fuller Test: `.$Price`

Test statistic	Lag order	P value	Alternative hypothesis
-2.487	12	0.3712	stationary

Table 6: Augmented Dickey-Fuller Test: `.$R_a`

Test statistic	Lag order	P value	Alternative hypothesis
-8.172	6	0.01 * *	stationary

Table 7: Augmented Dickey-Fuller Test: `.$R_a`

Test statistic	Lag order	P value	Alternative hypothesis
-6.563	12	0.01 * *	stationary

Table 8: Augmented Dickey-Fuller Test: `.$`Log Return“

Test statistic	Lag order	P value	Alternative hypothesis
-8.1	6	0.01 * *	stationary

Table 9: Augmented Dickey-Fuller Test: `.$`Log Return“

Test statistic	Lag order	P value	Alternative hypothesis
-6.296	12	0.01 * *	stationary

Model Selection

Auto ARIMA

The auto-ARIMA function recommends an ARIMA(1,0,1) model with non-zero mean for the Banana Return model. For the Banana Log Return model, an ARMIA(1,0,1) model with zero mean is recommended instead.

```
## Series: banana_price$`Log Return`
## ARIMA(1,0,1) with zero mean
##
## Coefficients:
##          ar1      ma1
##       0.7300  -0.9256
## s.e.  0.0649   0.0340
##
## sigma^2 estimated as 0.01693:  log likelihood=150.27
## AIC=-294.53   AICc=-294.43   BIC=-284.08
##
## Training set error measures:
##                       ME      RMSE        MAE MPE MAPE      MASE
## Training set 0.006712473 0.1295891 0.08346482 NaN  Inf 0.6522896
##                    ACF1
## Training set 0.01384834
```

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t$$

where

$$\phi(z) = 1 - \phi_1(z) - ... - \phi_p z^p, \Phi(z) = 1 - \Phi_1(z) - ... - \Phi_p z^p$$

$$\theta(z) = 1 + \theta_1 z + ... + \theta_q z^q, \Theta(z) = 1 + \Theta_1 z + ... + \Theta_q z^q$$

Auto-ARIMA(1,0,1) where

$$\phi(z) = 1 - 0.7300z$$

$$\theta(z) = 1 - 0.9256z$$

Minimum AIC and BIC

Table 10: Minimum AIC

p	d	q	P	D	Q	AIC	BIC
1	0	1	1	0	0	-294.3741	661.6259
1	0	1	1	0	1	-294.3741	661.6259
1	0	1	1	1	1	-294.3741	661.6259
1	0	1	0	0	1	-294.3741	661.6259
1	0	1	0	1	1	-294.3741	661.6259

Table 11: Minimum BIC

p	d	q	P	D	Q	AIC	BIC
0	1	0	0	1	1	-87.16703	151.833
0	1	0	0	1	0	-87.16703	151.833
0	1	0	0	0	0	-87.16703	151.833
0	1	0	0	0	1	-87.16703	151.833
0	1	0	1	1	0	-87.16703	151.833
0	1	0	1	1	1	-87.16703	151.833
0	1	0	1	0	0	-87.16703	151.833
0	1	0	1	0	1	-87.16703	151.833

```
## Series: banana_price$`Log Return`
## ARIMA(1,0,1) with non-zero mean
##
## Coefficients:
##          ar1      ma1    mean
##       0.7573  -0.9568  0.0031
## s.e.  0.0672   0.0377  0.0018
##
## sigma^2 estimated as 0.01686:  log likelihood=151.19
## AIC=-294.37   AICc=-294.2   BIC=-280.43
##
## Training set error measures:
##                        ME      RMSE        MAE  MPE MAPE      MASE
## Training set -0.004216673 0.1290165 0.08398509 -Inf  Inf 0.6563556
##                    ACF1
## Training set 0.01066458
```

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	-0.004217	0.129	0.08399	-Inf	Inf	0.6564	0.01066

```
## Series: banana_price$`Log Return`
## ARIMA(0,1,0)
##
## sigma^2 estimated as 0.04038:  log likelihood=44.58
## AIC=-87.17   AICc=-87.15   BIC=-83.69
##
## Training set error measures:
##                        ME      RMSE       MAE MPE MAPE      MASE      ACF1
## Training set 6.045479e-05 0.2005319 0.1274257 NaN  Inf 0.9958506 -0.516853
```

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	6.045e-05	0.2005	0.1274	NA	Inf	0.9959	-0.5169

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t$$

where

$$\phi(z) = 1 - \phi_1(z) - ... - \phi_p z^p, \Phi(z) = 1 - \Phi_1(z) - ... - \Phi_p z^p$$
$$\theta(z) = 1 + \theta_1 z + ... + \theta_q z^q, \Theta(z) = 1 + \Theta_1 z + ... + \Theta_q z^q$$

AIC (1,0,1) where

$$\phi(z) = 1 - 0.7573z$$
$$\theta(z) = 1 - 0.9568z$$

BIC (0,1,0)(0,1,1) where

$$\phi(z) = 1 - 0.7573z$$
$$\theta(z) = 1 - 0.9568z$$

Forecasting Banana Price

Auto-ARIMA

Table 14: Auto-ARIMA Forecasted Log Return

Date	Mean	Lower 95%	Upper 95%
2017-02-01	0.0004338	-0.2546168	0.2554843
2017-03-01	0.0003166	-0.2595688	0.2602021
2017-04-01	0.0002311	-0.2621942	0.2626565
2017-05-01	0.0001687	-0.2636001	0.2639375
2017-06-01	0.0001232	-0.2643587	0.2646050
2017-07-01	0.0000899	-0.2647711	0.2649509
2017-08-01	0.0000656	-0.2649972	0.2651284
2017-09-01	0.0000479	-0.2651224	0.2652182

Date	Mean	Lower 95%	Upper 95%
2017-10-01	0.0000350	-0.2651926	0.2652625
2017-11-01	0.0000255	-0.2652325	0.2652835
2017-12-01	0.0000186	-0.2652556	0.2652929
2018-01-01	0.0000136	-0.2652693	0.2652965
2018-02-01	0.0000099	-0.2652776	0.2652975
2018-03-01	0.0000072	-0.2652828	0.2652972
2018-04-01	0.0000053	-0.2652860	0.2652966
2018-05-01	0.0000039	-0.2652881	0.2652959
2018-06-01	0.0000028	-0.2652896	0.2652952
2018-07-01	0.0000021	-0.2652905	0.2652946
2018-08-01	0.0000015	-0.2652912	0.2652942
2018-09-01	0.0000011	-0.2652916	0.2652938
2018-10-01	0.0000008	-0.2652920	0.2652936
2018-11-01	0.0000006	-0.2652922	0.2652934
2018-12-01	0.0000004	-0.2652924	0.2652932
2019-01-01	0.0000003	-0.2652925	0.2652931
2019-02-01	0.0000002	-0.2652926	0.2652930
2019-03-01	0.0000002	-0.2652926	0.2652930
2019-04-01	0.0000001	-0.2652927	0.2652929
2019-05-01	0.0000001	-0.2652927	0.2652929
2019-06-01	0.0000001	-0.2652927	0.2652929
2019-07-01	0.0000000	-0.2652928	0.2652929
2019-08-01	0.0000000	-0.2652928	0.2652928
2019-09-01	0.0000000	-0.2652928	0.2652928
2019-10-01	0.0000000	-0.2652928	0.2652928
2019-11-01	0.0000000	-0.2652928	0.2652928
2019-12-01	0.0000000	-0.2652928	0.2652928
2020-01-01	0.0000000	-0.2652928	0.2652928

Figure 15: Forecasted Banana Log Return

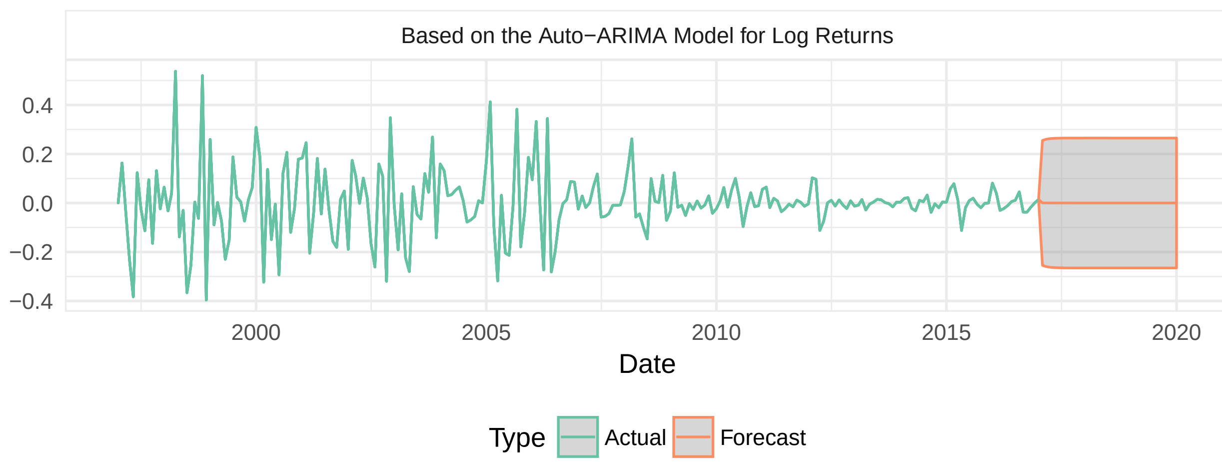


Table 15: Auto-ARIMA Forecasted Price

Date	Mean	Lower 95%	Upper 95%
2017-02-01	974.4557	755.0837	1257.561
2017-03-01	974.7643	751.6797	1264.056

Date	Mean	Lower 95%	Upper 95%
2017-04-01	974.9897	749.9462	1267.564
2017-05-01	975.1542	749.0658	1269.482
2017-06-01	975.2743	748.6241	1270.544
2017-07-01	975.3620	748.4075	1271.140
2017-08-01	975.4260	748.3056	1271.480
2017-09-01	975.4727	748.2610	1271.678
2017-10-01	975.5068	748.2444	1271.795
2017-11-01	975.5317	748.2406	1271.866
2017-12-01	975.5499	748.2424	1271.911
2018-01-01	975.5632	748.2461	1271.939
2018-02-01	975.5728	748.2501	1271.958
2018-03-01	975.5799	748.2537	1271.970
2018-04-01	975.5851	748.2567	1271.978
2018-05-01	975.5888	748.2590	1271.984
2018-06-01	975.5916	748.2609	1271.988
2018-07-01	975.5936	748.2623	1271.991
2018-08-01	975.5951	748.2633	1271.993
2018-09-01	975.5961	748.2641	1271.995
2018-10-01	975.5969	748.2647	1271.996
2018-11-01	975.5975	748.2651	1271.996
2018-12-01	975.5979	748.2654	1271.997
2019-01-01	975.5982	748.2656	1271.997
2019-02-01	975.5984	748.2658	1271.998
2019-03-01	975.5986	748.2659	1271.998
2019-04-01	975.5987	748.2660	1271.998
2019-05-01	975.5988	748.2661	1271.998
2019-06-01	975.5989	748.2661	1271.998
2019-07-01	975.5989	748.2662	1271.998
2019-08-01	975.5989	748.2662	1271.998
2019-09-01	975.5990	748.2662	1271.998
2019-10-01	975.5990	748.2662	1271.998
2019-11-01	975.5990	748.2662	1271.998
2019-12-01	975.5990	748.2662	1271.998
2020-01-01	975.5990	748.2662	1271.998

Figure 16: Forecasted Banana Price

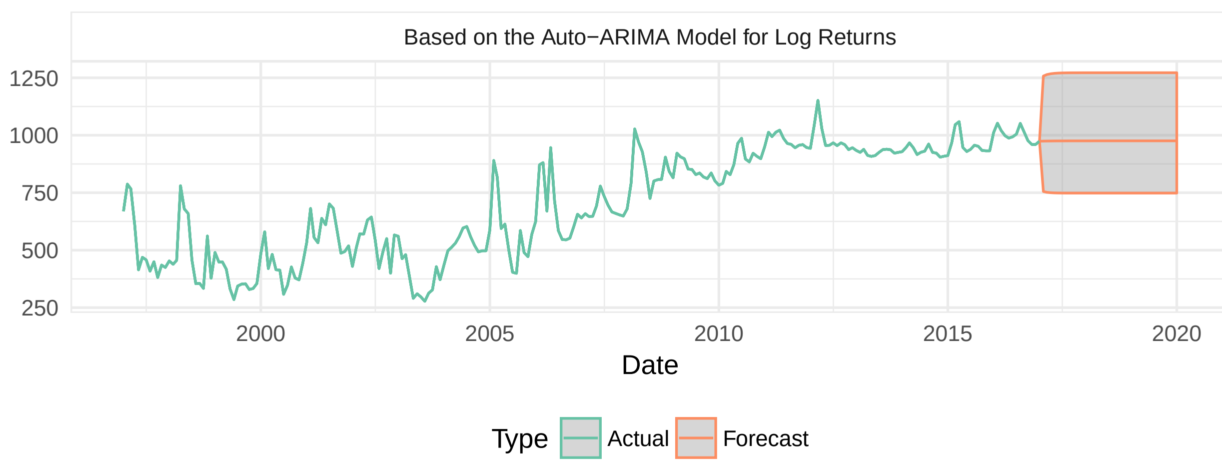


Table 16: Auto-ARIMA Accuracy

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
0.0067125	0.1295891	0.0834648	NaN	Inf	0.6522896	0.0138483

Minimum AIC

From the lowest AIC model, the RMSE of the model is 0.129. The RSME is the lowest of the 3 models, with the Auto-ARIMAs RSME of 0.130 being almost identical. Looking at the forecasted banana log return model, the expected forecast steadies around zero. In addition, over the past 10 years, the variance of the log return model is decreasing and approaching zero.

From the minimum AIC log returns model with respect to price, it is expected that the price will increase to $1150 USD/Metric Ton by 2020.

Table 17: Minimum AIC Forecasted Log Return

Date	Mean	Lower 95%	Upper 95%
2017-02-01	0.0150559	-0.2394006	0.2695124
2017-03-01	0.0121627	-0.2473083	0.2716336
2017-04-01	0.0099715	-0.2523322	0.2722753
2017-05-01	0.0083122	-0.2556026	0.2722269
2017-06-01	0.0070555	-0.2577788	0.2718898
2017-07-01	0.0061037	-0.2592565	0.2714640
2017-08-01	0.0053830	-0.2602785	0.2710444
2017-09-01	0.0048371	-0.2609969	0.2706712
2017-10-01	0.0044237	-0.2615093	0.2703567
2017-11-01	0.0041106	-0.2618791	0.2701004
2017-12-01	0.0038735	-0.2621487	0.2698958
2018-01-01	0.0036940	-0.2623469	0.2697349
2018-02-01	0.0035580	-0.2624936	0.2696096
2018-03-01	0.0034550	-0.2626027	0.2695128
2018-04-01	0.0033770	-0.2626842	0.2694383
2018-05-01	0.0033180	-0.2627453	0.2693812
2018-06-01	0.0032732	-0.2627912	0.2693377
2018-07-01	0.0032394	-0.2628258	0.2693045
2018-08-01	0.0032137	-0.2628518	0.2692792
2018-09-01	0.0031943	-0.2628714	0.2692600
2018-10-01	0.0031795	-0.2628863	0.2692454
2018-11-01	0.0031684	-0.2628975	0.2692343
2018-12-01	0.0031600	-0.2629060	0.2692259
2019-01-01	0.0031536	-0.2629124	0.2692195
2019-02-01	0.0031487	-0.2629172	0.2692147
2019-03-01	0.0031451	-0.2629209	0.2692111
2019-04-01	0.0031423	-0.2629237	0.2692083
2019-05-01	0.0031402	-0.2629258	0.2692062
2019-06-01	0.0031386	-0.2629274	0.2692046
2019-07-01	0.0031374	-0.2629286	0.2692034
2019-08-01	0.0031365	-0.2629295	0.2692025
2019-09-01	0.0031358	-0.2629302	0.2692018
2019-10-01	0.0031353	-0.2629307	0.2692013
2019-11-01	0.0031349	-0.2629311	0.2692009

Date	Mean	Lower 95%	Upper 95%
2019-12-01	0.0031346	-0.2629314	0.2692006
2020-01-01	0.0031343	-0.2629317	0.2692003

Figure 17: Forecasted Banana Log Return

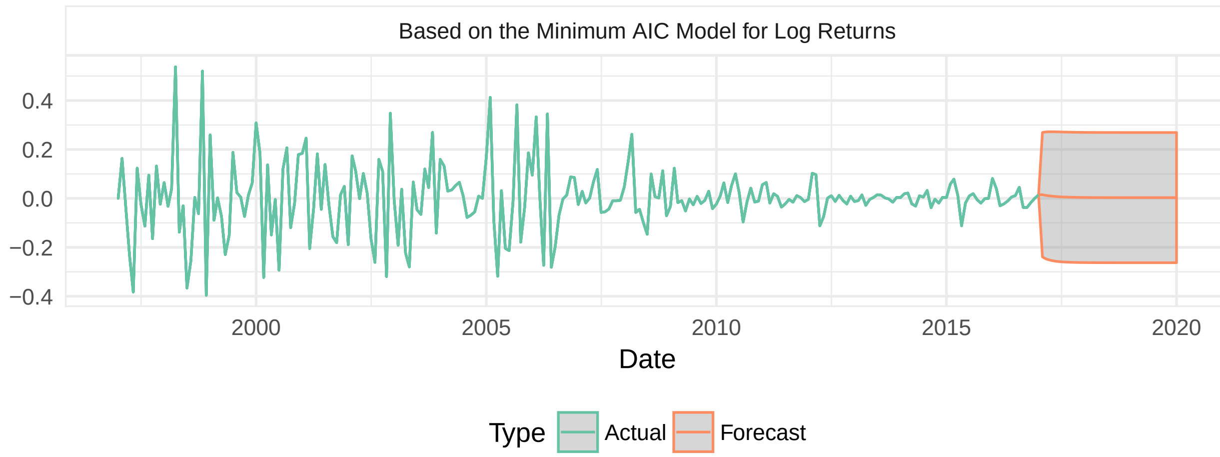


Table 18: Minimum AIC Forecasted Price

Date	Mean	Lower 95%	Upper 95%
2017-02-01	988.809	766.6610	1275.327
2017-03-01	1000.909	772.1609	1297.422
2017-04-01	1010.940	777.6929	1314.142
2017-05-01	1019.378	782.9218	1327.247
2017-06-01	1026.595	787.7405	1337.874
2017-07-01	1032.881	792.1466	1346.774
2017-08-01	1038.455	796.1824	1354.451
2017-09-01	1043.491	799.9049	1361.253
2017-10-01	1048.117	803.3718	1367.423
2017-11-01	1052.434	806.6352	1373.134
2017-12-01	1056.519	809.7395	1378.508
2018-01-01	1060.429	812.7210	1383.635
2018-02-01	1064.209	815.6091	1388.582
2018-03-01	1067.892	818.4269	1393.396
2018-04-01	1071.504	821.1925	1398.115
2018-05-01	1075.065	823.9201	1402.764
2018-06-01	1078.590	826.6204	1407.365
2018-07-01	1082.090	829.3019	1411.932
2018-08-01	1085.573	831.9710	1416.478
2018-09-01	1089.046	834.6326	1421.010
2018-10-01	1092.514	837.2905	1425.535
2018-11-01	1095.981	839.9475	1430.059
2018-12-01	1099.450	842.6059	1434.585
2019-01-01	1102.923	845.2673	1439.117
2019-02-01	1106.401	847.9330	1443.655
2019-03-01	1109.886	850.6040	1448.203
2019-04-01	1113.379	853.2810	1452.761
2019-05-01	1116.881	855.9647	1457.330

Date	Mean	Lower 95%	Upper 95%
2019-06-01	1120.392	858.6554	1461.911
2019-07-01	1123.912	861.3536	1466.505
2019-08-01	1127.443	864.0595	1471.112
2019-09-01	1130.984	866.7732	1475.732
2019-10-01	1134.536	869.4950	1480.366
2019-11-01	1138.098	872.2251	1485.014
2019-12-01	1141.671	874.9634	1489.676
2020-01-01	1145.255	877.7101	1494.353

Figure 18: Forecasted Banana Price

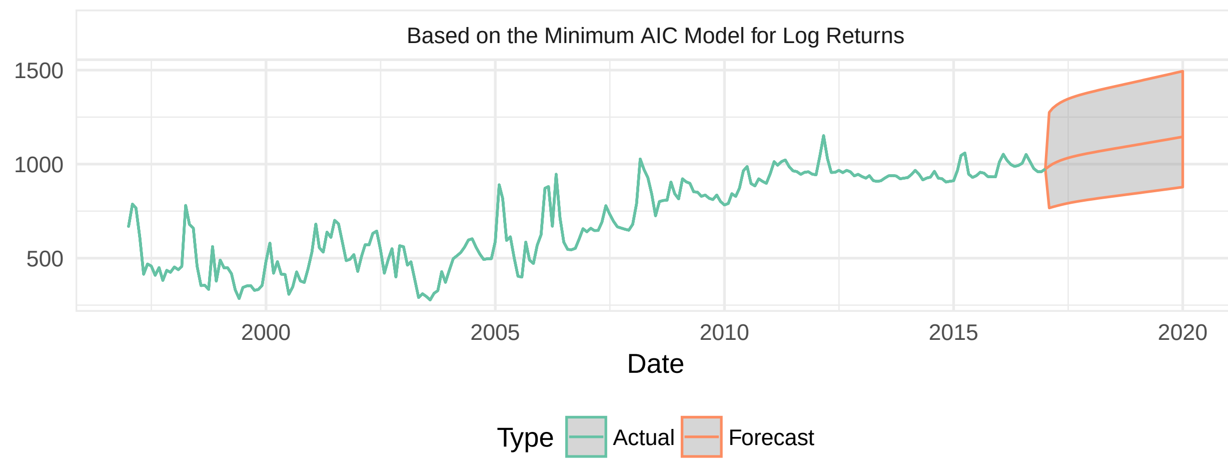


Table 19: Minimum AIC Accuracy

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
-0.0042167	0.1290165	0.0839851	-Inf	Inf	0.6563556	0.0106646

Minimum BIC

From the lowest BIC model the RMSE of the model is 0.201, the largest RMSE of the models. The forecasted banana log returns suggests that the expected log return to be approximately 0, since the past 10 years have been close to 0 as well. Looking at the minimum BIC models plot with respect to price, it can be seen that the expected price is increasing slightly. However, the 95% confidence interval, the range of the interval is significantly larger than the other models, with the upper bound of the minimum BIC model being approximately 16000 which suggests that this may not be the best model to forecast.

Table 20: Minimum BIC Forecasted Log Return

Date	Mean	Lower 95%	Upper 95%
2017-02-01	0.0145696	-0.3792836	0.4084228
2017-03-01	0.0145696	-0.5424229	0.5715621
2017-04-01	0.0145696	-0.6676041	0.6967433
2017-05-01	0.0145696	-0.7731368	0.8022760
2017-06-01	0.0145696	-0.8661129	0.8952521
2017-07-01	0.0145696	-0.9501697	0.9793089

Date	Mean	Lower 95%	Upper 95%
2017-08-01	0.0145696	-1.0274680	1.0566072
2017-09-01	0.0145696	-1.0994154	1.1285546
2017-10-01	0.0145696	-1.1669899	1.1961291
2017-11-01	0.0145696	-1.2309035	1.2600427
2017-12-01	0.0145696	-1.2916936	1.3208328
2018-01-01	0.0145696	-1.3497778	1.3789170
2018-02-01	0.0145696	-1.4054882	1.4346274
2018-03-01	0.0145696	-1.4590941	1.4882333
2018-04-01	0.0145696	-1.5108172	1.5399564
2018-05-01	0.0145696	-1.5608431	1.5899823
2018-06-01	0.0145696	-1.6093287	1.6384679
2018-07-01	0.0145696	-1.6564079	1.6855471
2018-08-01	0.0145696	-1.7021966	1.7313358
2018-09-01	0.0145696	-1.7467954	1.7759346
2018-10-01	0.0145696	-1.7902924	1.8194316
2018-11-01	0.0145696	-1.8327656	1.8619048
2018-12-01	0.0145696	-1.8742839	1.9034231
2019-01-01	0.0145696	-1.9149090	1.9440483
2019-02-01	0.0145696	-1.9546963	1.9838355
2019-03-01	0.0145696	-1.9936955	2.0228347
2019-04-01	0.0145696	-2.0319516	2.0610908
2019-05-01	0.0145696	-2.0695055	2.0986447
2019-06-01	0.0145696	-2.1063947	2.1355339
2019-07-01	0.0145696	-2.1426531	2.1717923
2019-08-01	0.0145696	-2.1783121	2.2074513
2019-09-01	0.0145696	-2.2134004	2.2425396
2019-10-01	0.0145696	-2.2479447	2.2770839
2019-11-01	0.0145696	-2.2819693	2.3111086
2019-12-01	0.0145696	-2.3154972	2.3446364
2020-01-01	0.0145696	-2.3485495	2.3776887

Figure 19: Forecasted Banana Log Return

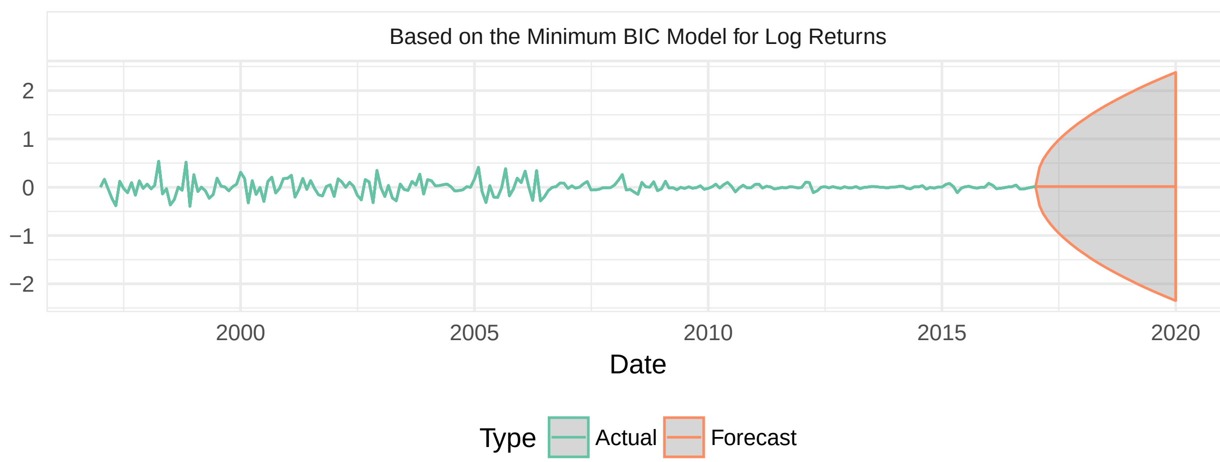


Table 21: Minimum BIC Forecasted Price

Date	Mean	Lower 95%	Upper 95%
2017-02-01	988.3283	666.5811	1465.377
2017-03-01	1002.8333	574.5528	1750.361
2017-04-01	1017.5511	514.3893	2012.892
2017-05-01	1032.4850	469.6639	2269.762
2017-06-01	1047.6380	434.2459	2527.474
2017-07-01	1063.0134	405.0958	2789.457
2017-08-01	1078.6144	380.4653	3057.859
2017-09-01	1094.4445	359.2495	3334.197
2017-10-01	1110.5068	340.7034	3619.645
2017-11-01	1126.8049	324.2998	3915.172
2017-12-01	1143.3422	309.6516	4221.620
2018-01-01	1160.1222	296.4662	4539.754
2018-02-01	1177.1484	284.5168	4870.286
2018-03-01	1194.4246	273.6243	5213.900
2018-04-01	1211.9542	263.6448	5571.258
2018-05-01	1229.7412	254.4607	5943.014
2018-06-01	1247.7892	245.9751	6329.819
2018-07-01	1266.1021	238.1071	6732.325
2018-08-01	1284.6837	230.7885	7151.190
2018-09-01	1303.5381	223.9611	7587.082
2018-10-01	1322.6691	217.5753	8040.681
2018-11-01	1342.0810	211.5881	8512.678
2018-12-01	1361.7777	205.9622	9003.780
2019-01-01	1381.7635	200.6651	9514.712
2019-02-01	1402.0426	195.6680	10046.216
2019-03-01	1422.6194	190.9459	10599.055
2019-04-01	1443.4981	186.4762	11174.010
2019-05-01	1464.6832	182.2390	11771.886
2019-06-01	1486.1793	178.2166	12393.509
2019-07-01	1507.9909	174.3929	13039.731
2019-08-01	1530.1225	170.7536	13711.426
2019-09-01	1552.5790	167.2856	14409.496
2019-10-01	1575.3650	163.9773	15134.869
2019-11-01	1598.4855	160.8179	15888.500
2019-12-01	1621.9453	157.7978	16671.373
2020-01-01	1645.7494	154.9081	17484.503

Figure 20: Forecasted Banana Price

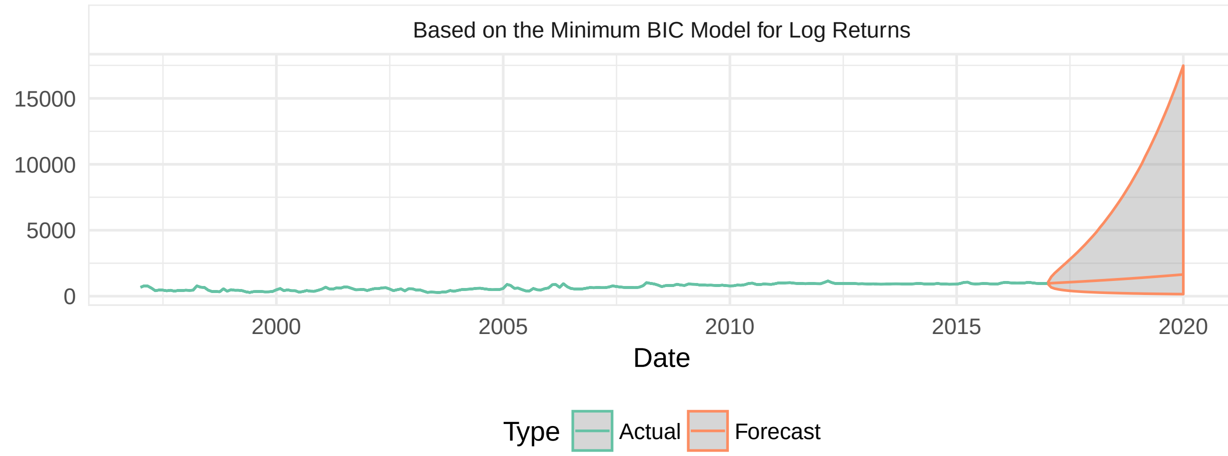


Table 22: Minimum BIC Accuracy

ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
6.05e-05	0.2005319	0.1274257	NaN	Inf	0.9958506	-0.516853

Comparison

Figure 21: Forecasted Banana Log Return

Figure 22: Forecasted Banana Price

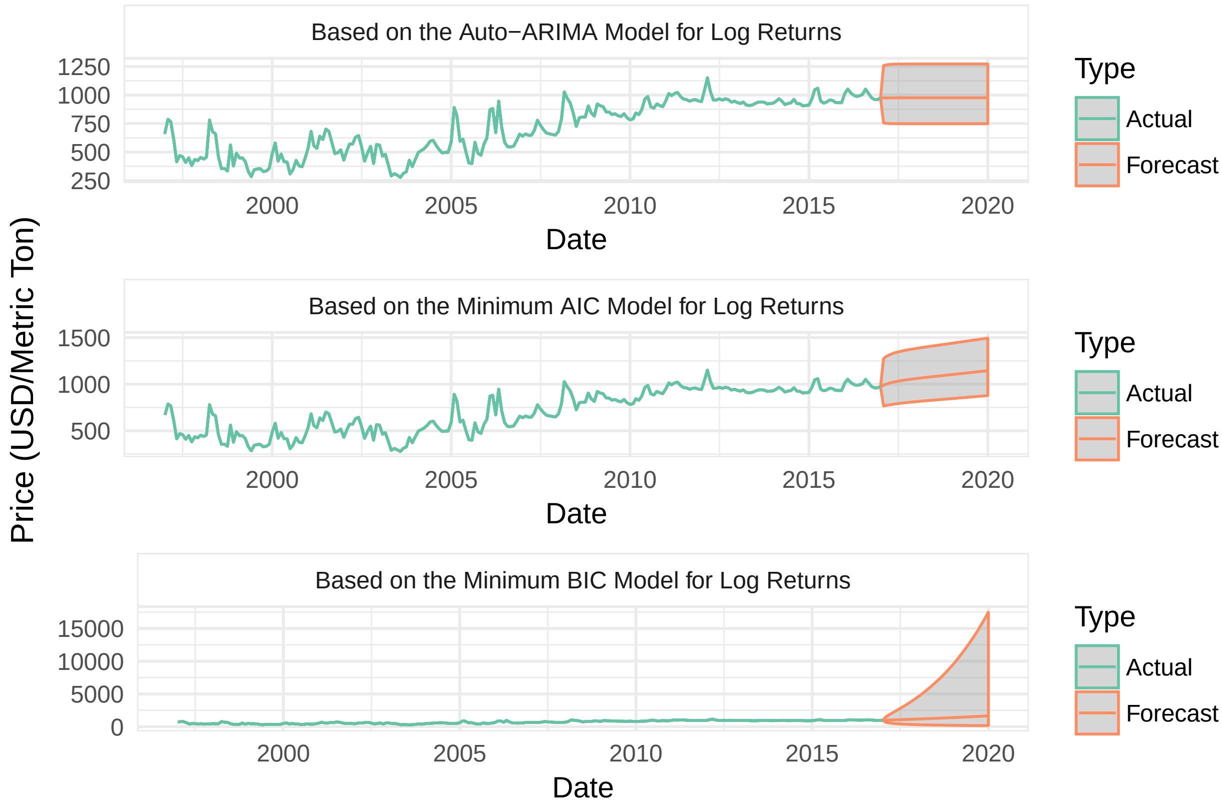


Table 23: Comparison of Model Accuracy

Model	ME	RMSE	MAE	MASE	ACF1
Auto-ARIMA	0.0067125	0.1295891	0.0834648	0.6522896	0.0138483
Minimum AIC	-0.0042167	0.1290165	0.0839851	0.6563556	0.0106646
Minimum BIC	0.0000605	0.2005319	0.1274257	0.9958506	-0.5168530

Model Diagnostics

White Noise

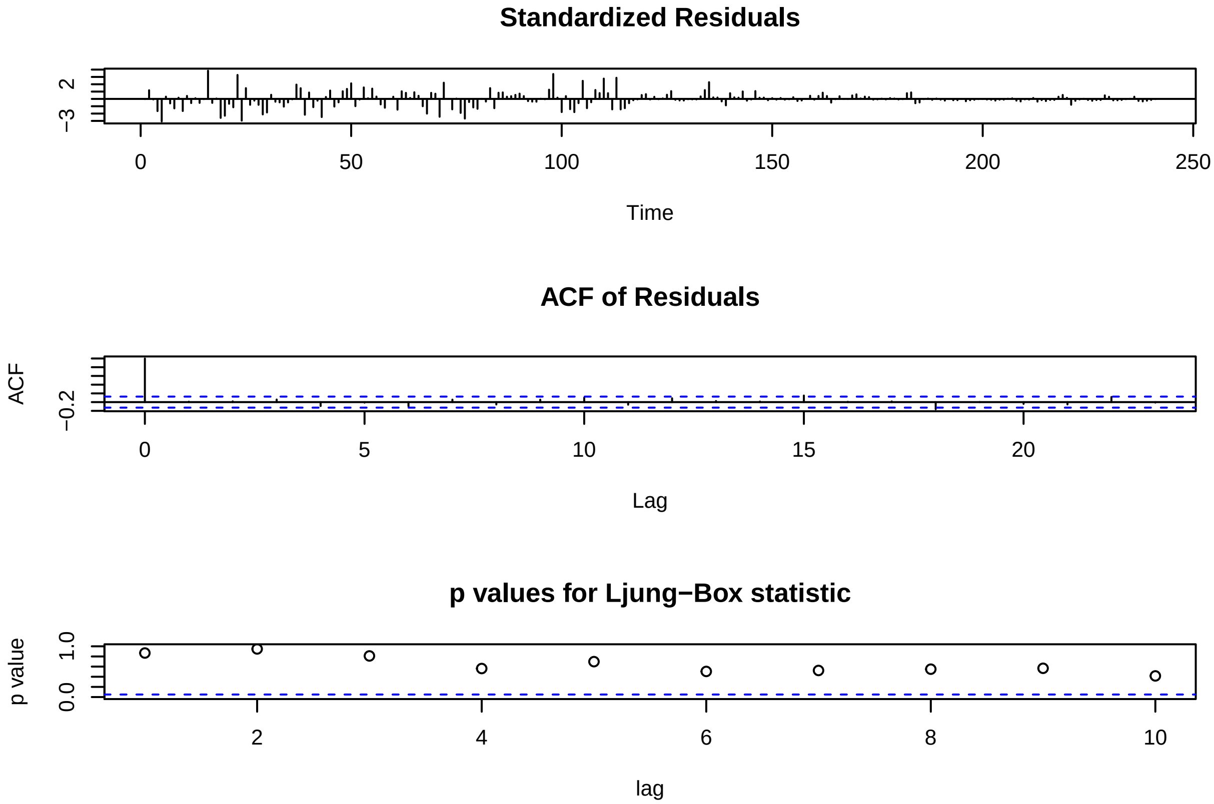


Figure 23: ARIMA(1,0,1) Model Residuals

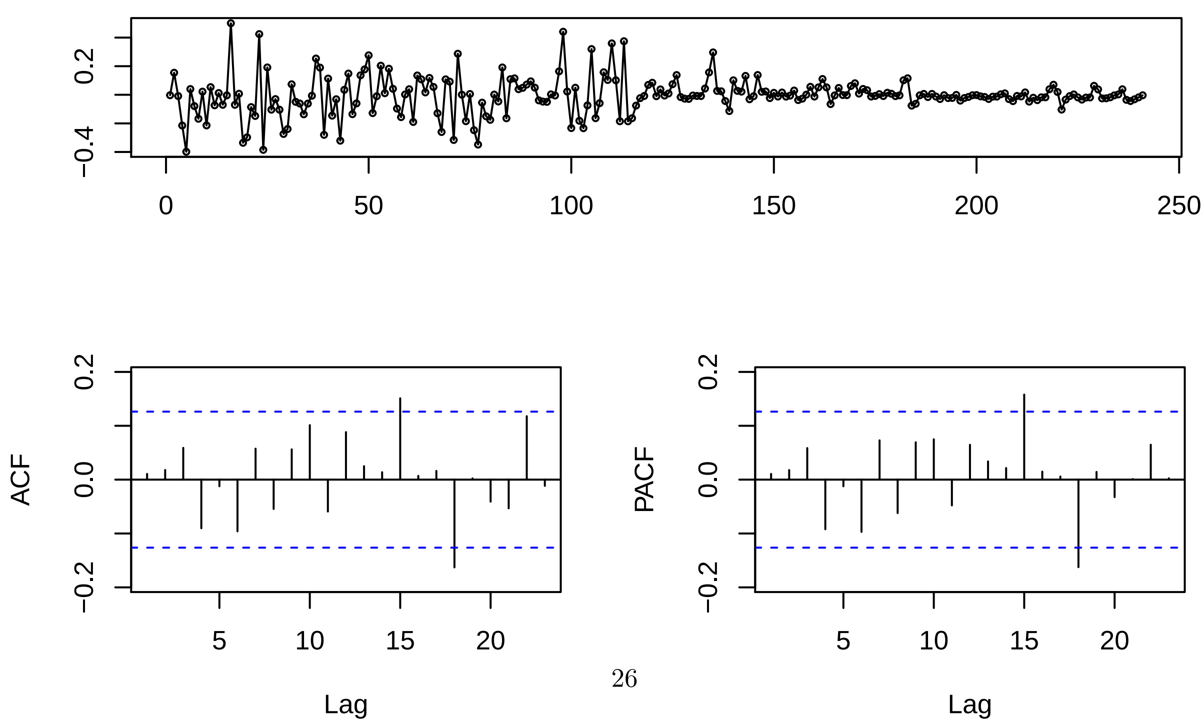
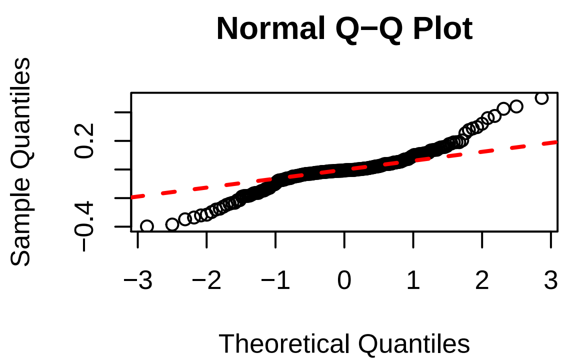
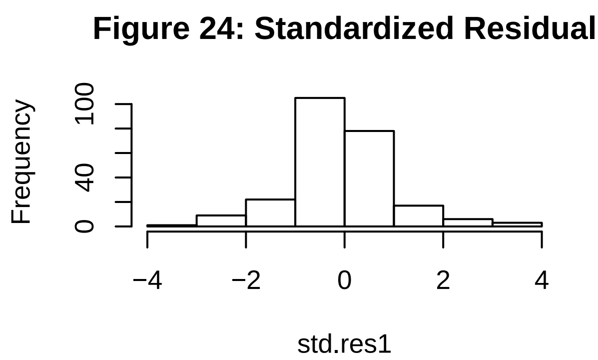


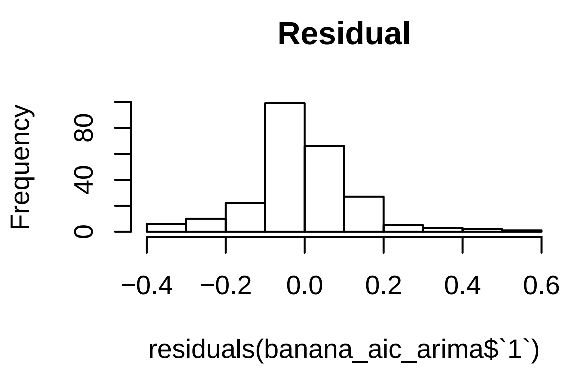
Figure 23: ACF show slight jumps spikes at certain lags, but not enough to be significant. In addition, the Ljung-Box statistics are all significantly greater than 0.05, which indicates that the Residuals are independently distributed.

Normality of Residuals

```
##
##  Jarque Bera Test
##
## data:  residuals(banana_aic_arima$`1`)
## X-squared = 69.562, df = 2, p-value = 7.772e-16
```

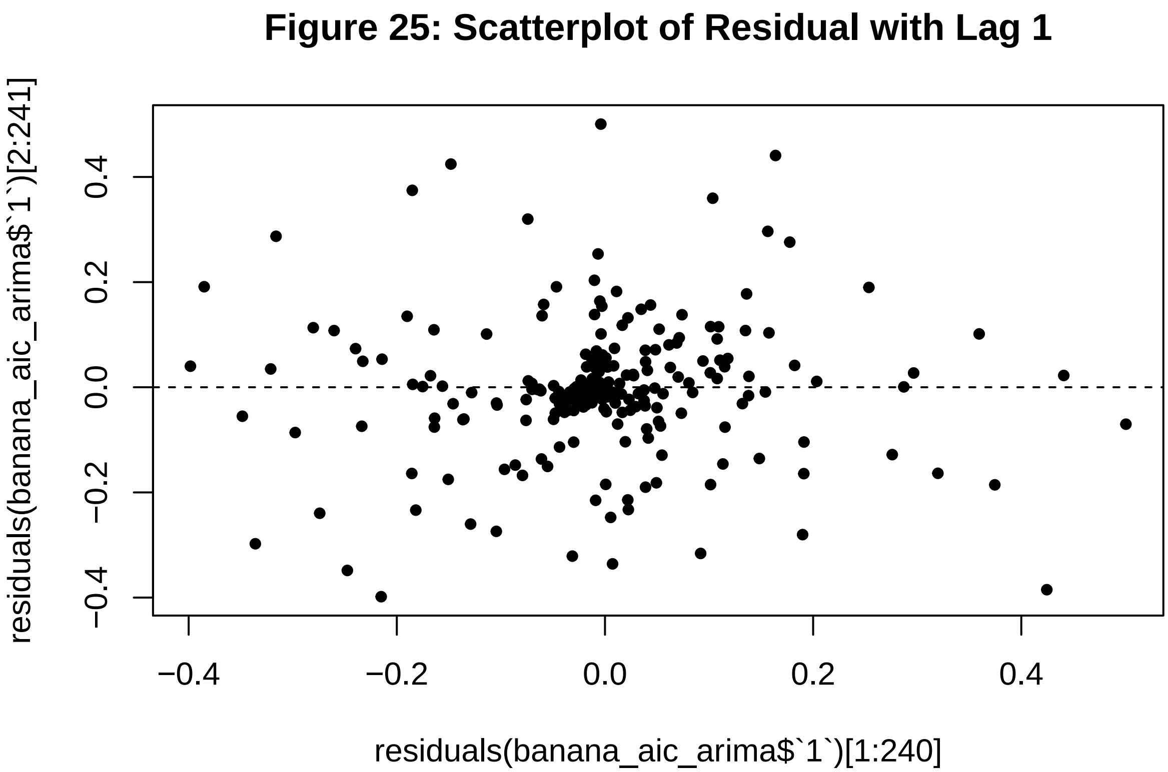






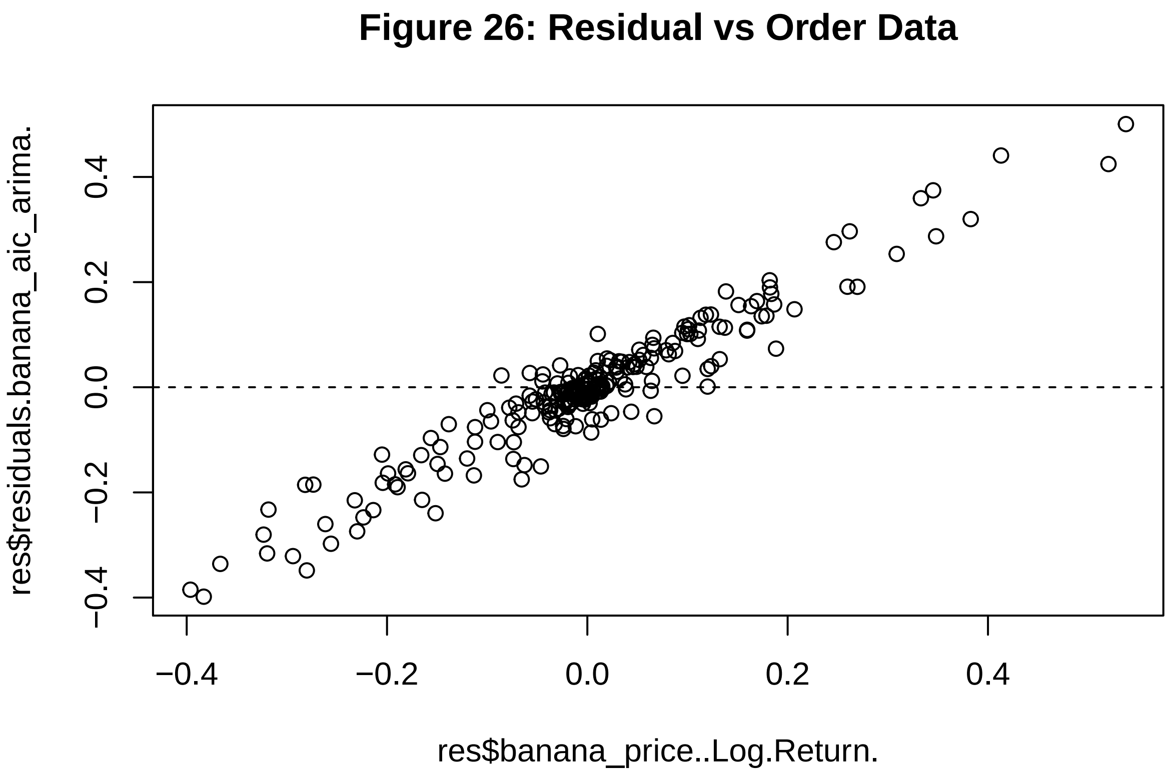
From the Figure 24, the histograms are relatively symmetric at 0, with a mean of 0, this resembles white noise. In addition, the Jarque Bera test statistic is less than 0.05, thus indicating the the residuals are not normally distributed

Constant Variance



In Figure 25, the points are randomly distributed around the 0 line, indicating that variances of residuals are random.

Drift



From Figure 26, the residuals are not randomly scattered, thus indicating signs of drift.

Homoscedasticity

The p-values indicate that the residuals of a ARIMA model is heteroscedastic.

Model with Drift Diagnostics

White Noise on Model with Drift

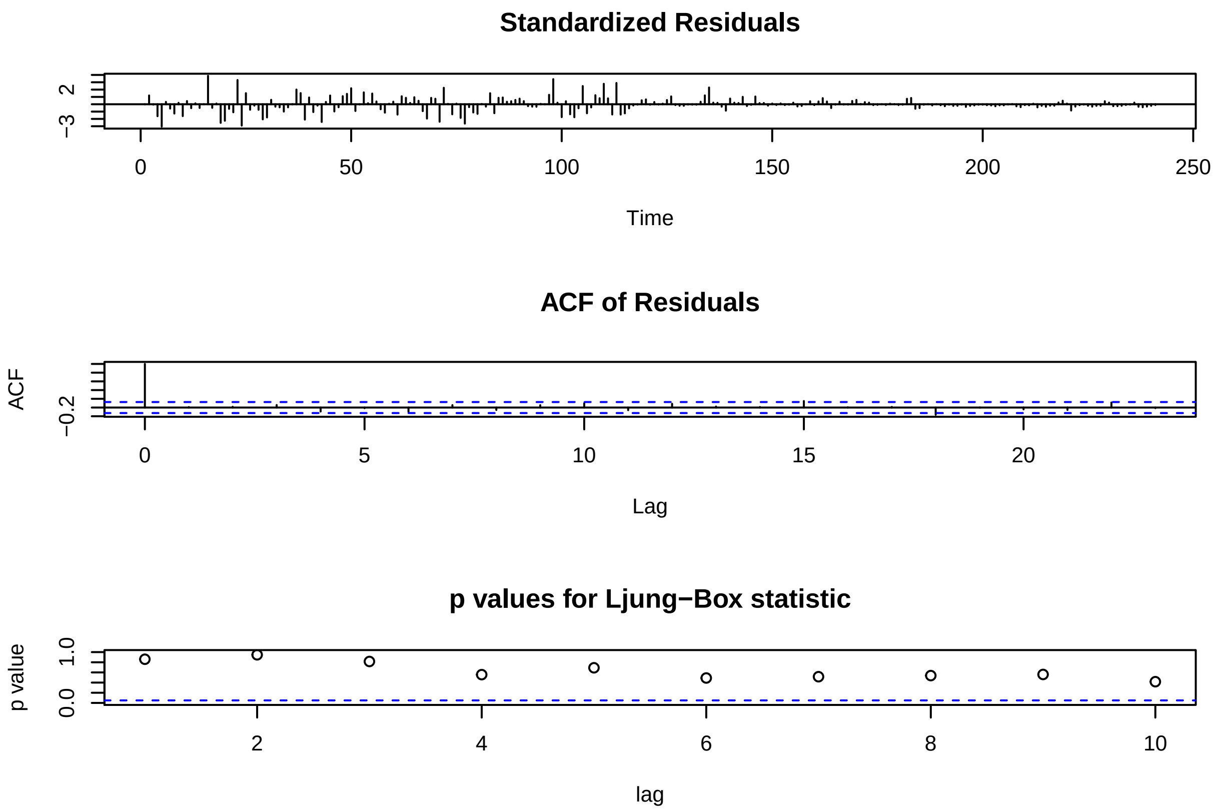
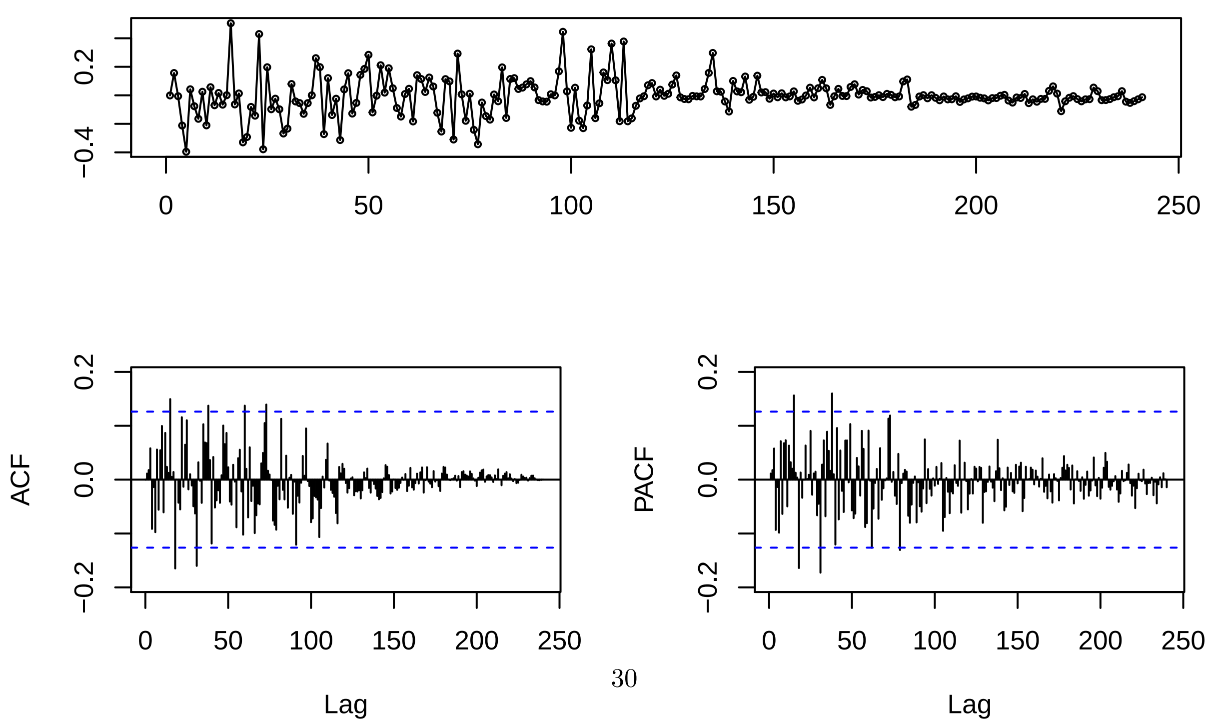
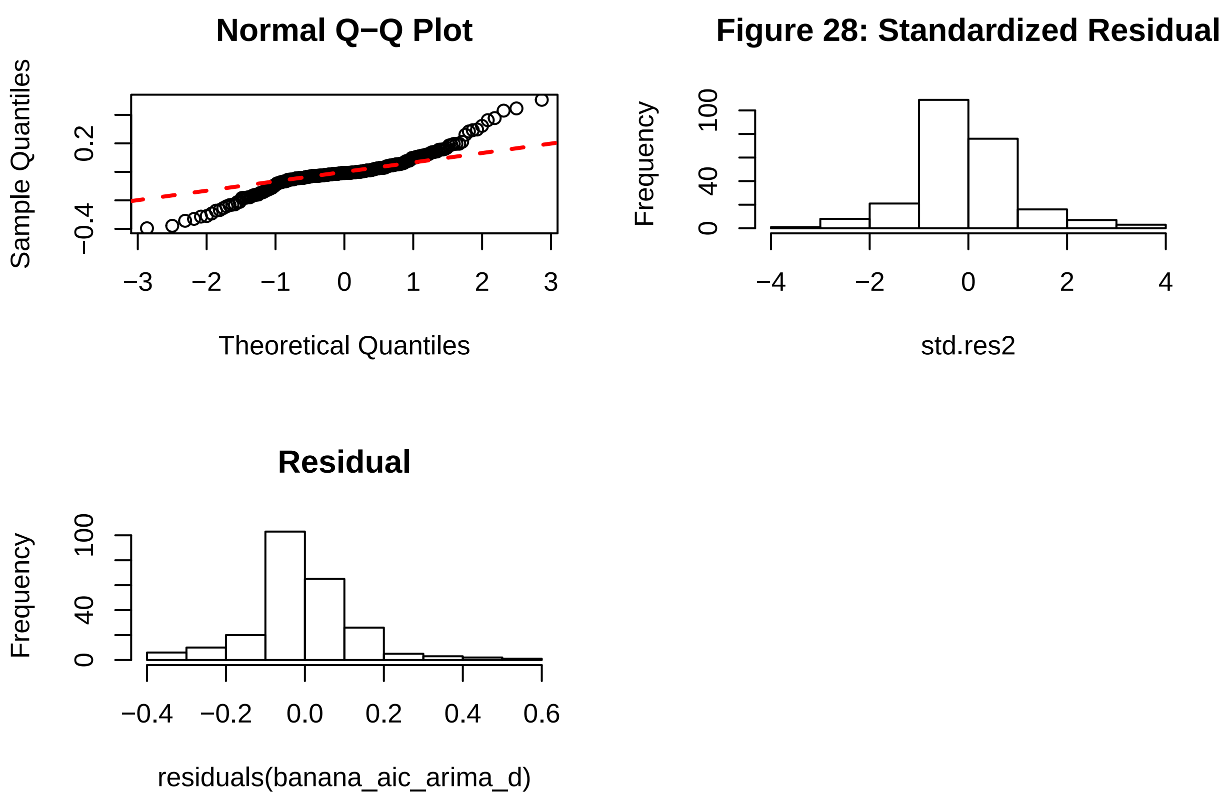


Figure 27: ARIMA(1,1,1)with drift Model Residuals



From Figure 27 the autocorrelation function there are some small jumps spikes at certain lags, however they are not enough to be significant. The Ljung-Box statistics are all significantly greater than 0.05, indicating that the residuals are independently distributed.

Normality of Residuals



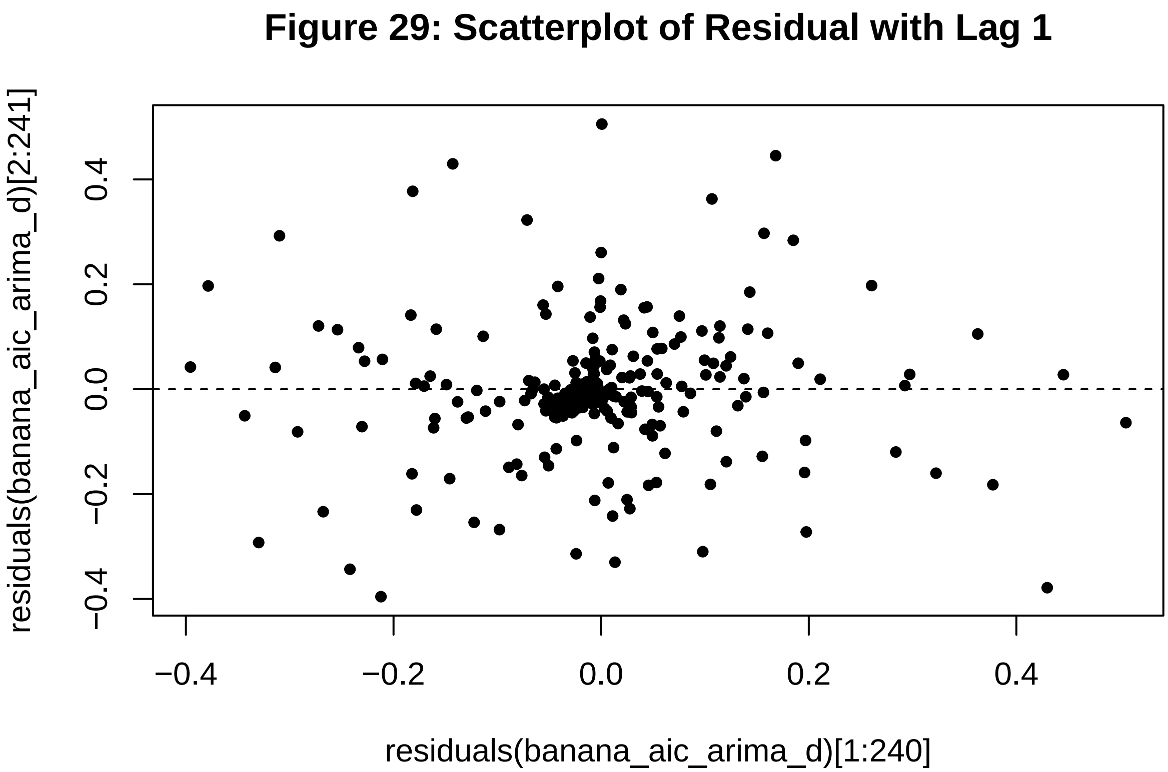
From Figure 28, the histograms are relatively symmetric at 0, with mean = 0, resembling white noise.

Homoscedasticity

```
##
##  Jarque Bera Test
##
## data:  residuals(banana_aic_arima_d)
## X-squared = 73.473, df = 2, p-value < 2.2e-16
```

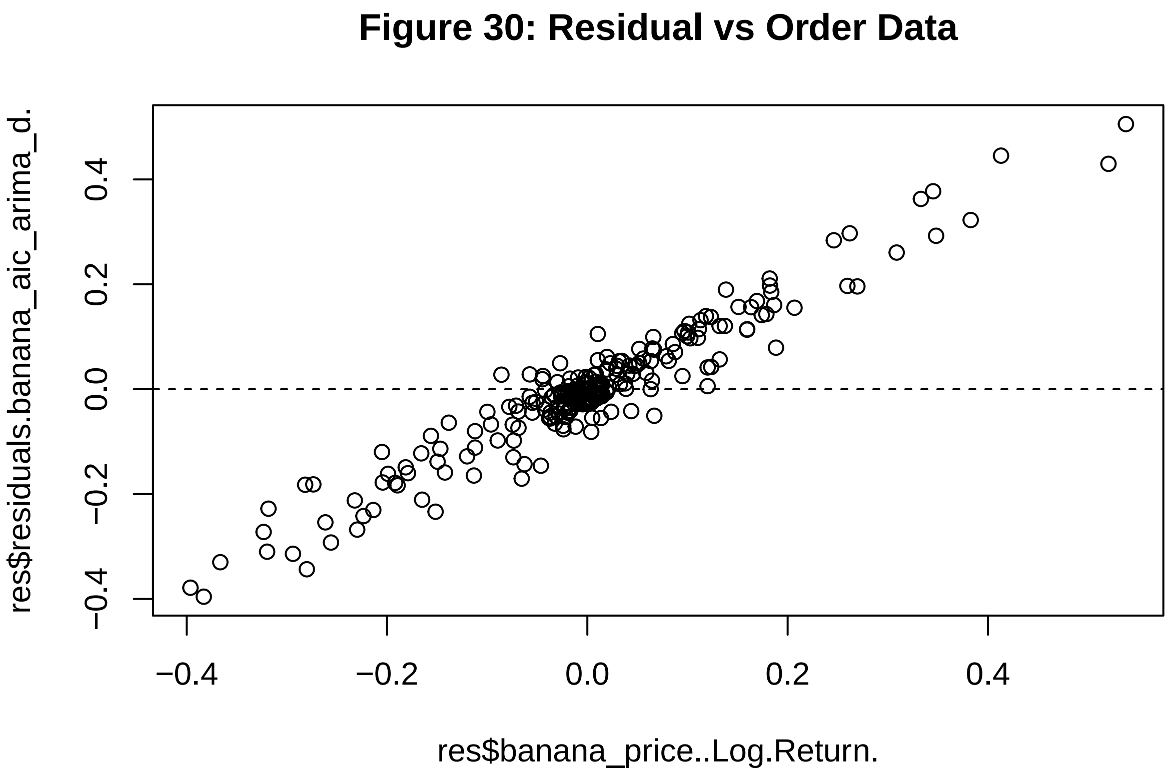
In addition, the jarque bera test statistic is less than 0.05, thus indicating the the residuals are not normally distributed.

Constant Variance



From Figure 29, the scatter plot shows random spread around 0 line, indicating variances of residuals are random.

Drift



Results

Based on the results of the model selection the minimum AIC model was chosen as the best model to forecast the price of Bananas.

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t$$

where

$$\phi(z) = 1 - \phi_1(z) - ... - \phi_p z^p, \Phi(z) = 1 - \Phi_1(z) - ... - \Phi_p z^p$$

AIC (1,0,1) where

$$\phi(z) = 1 - 0.7573z$$

$$\theta(z) = 1 - 0.9568z$$

Using this model to forecast the price of bananas to April 1st 2018 the expected price of bananas is $1071 USD per Metric Ton and the actual price of bananas was $1160 USD per Metric Ton. There is some error in the forecasted value, this error is likely due to predicting the price based on data set that ended a year and half ago.

Using the minimum AIC model the price of bananas was also forecasted to the year 2020, where the expected price will increase to $1150 USD/Metric Ton. This price is lower than the known price from April 1st 2018 and therefore a more accurate prediction for the price of bananas could be made more accurate by shortening the time between the end of the data set and the forecast date.

Conclusions

The analysis of the price of bananas shows that the price of bananas has been increasing since 1997, and in 2017 the price of bananas had doubled. The price of bananas was affected by the seasons and is usually highest during the northern hemispheres spring and the lowest during the northern hemispheres winter. The banana price data had seasonal components and trend and therefore transformations, model selection and diagnostics where used to determine the best model for forecasting the price of bananas.

The recommended model for forecasting the price of bananas is the minimum AIC model. The minimum AIC model suggests that the price will continue to increase, which follows the trend over the last 20 years.

In conclusion the forecasted price for bananas for April 1st 2018 is expected to be $1071 USD per Metric Ton and the log return is expected to be 0.003377 and the expected price will increase to $1150 USD/Metric Ton by 2020.

Appendix

References

[1] Federal Reserve Bank of St.Louis (FRED). (2017, July 12). Global price of Bananas. Retrieved April 09, 2018, from https://fred.stlouisfed.org/series/PBANSOPUSDM

[2] Britannica, T. E. (2018, March 15). Banana. Retrieved April 09, 2018, from http://www.britannica.com/plant/banana-plant

[3] Banana Link. (n.d.). Retrieved April 09, 2018, from http://www.bananalink.org.uk/