

Time-Series Forecasting of Banana Prices

Kaisa Roggeveen, Scott Graham, Irwin Khuu, Johnson Tran

April 8th, 2018

Abstract

Introduction

(Data Collection) (Introduce where we got banana data, talk about how bananas are important) (Background info about bananas and possible reasons why the price is the price?)

Methodology

Original Time Series Plot

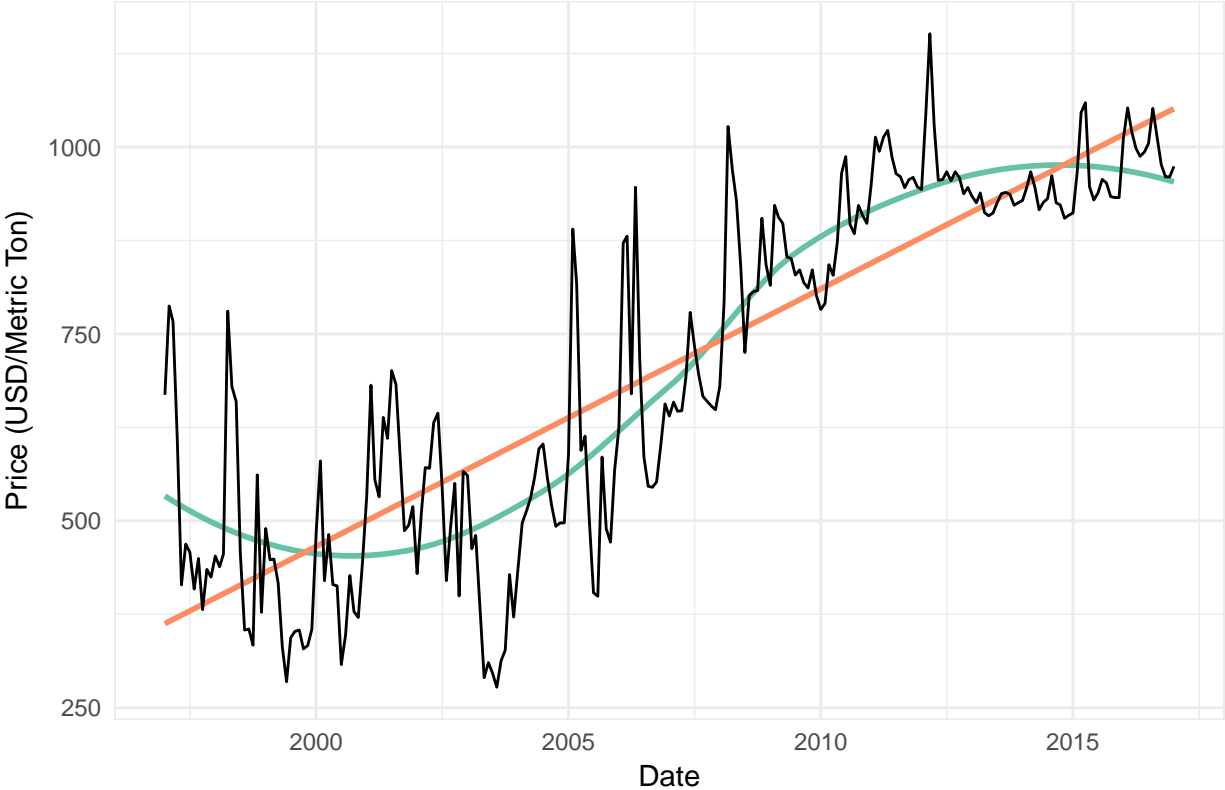
To begin analysis of the stock price of Bananas a plot of the price and the date was created, and is shown in Figure 01.

In Figure 01, 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 1998.

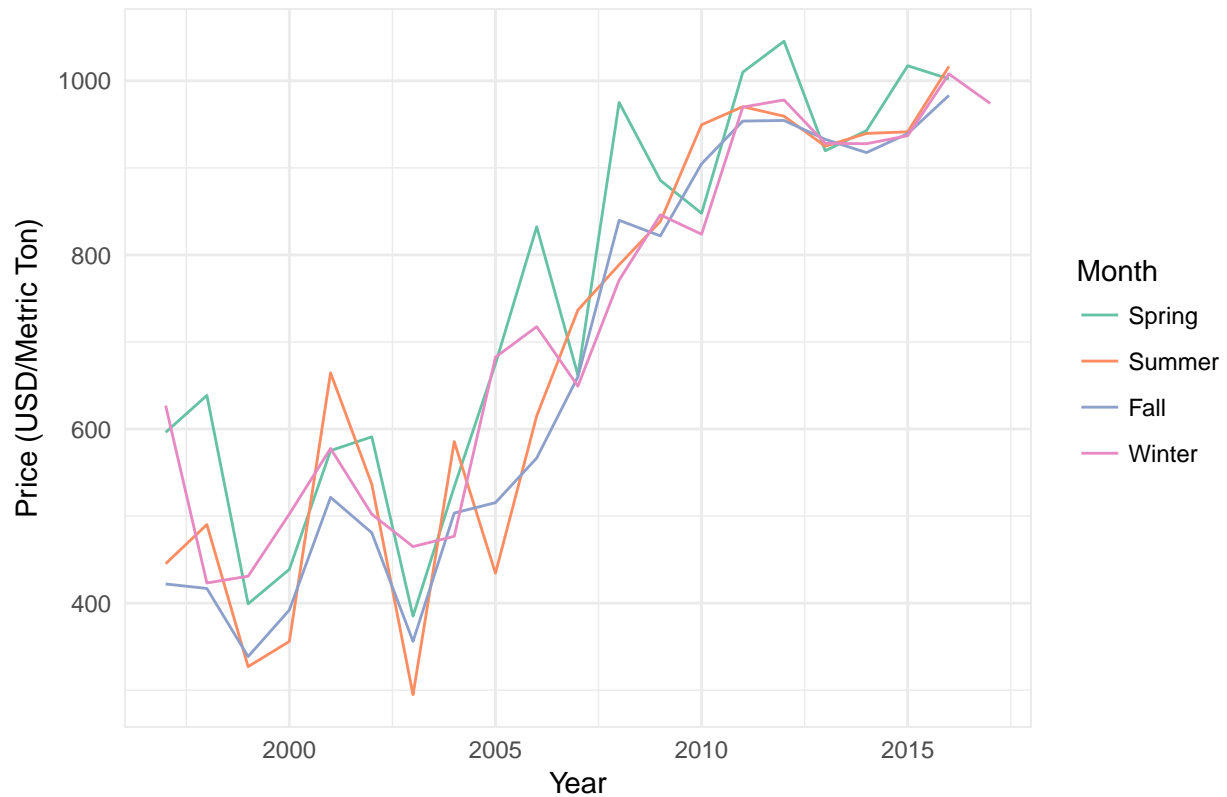
A local regression (loess) was fitted to the data. This adds a smooth curve to all of the data. The loess curve is used to test the fit of our linear regression model, where the best model would have a loess model similar to the linear regression model. Therefore, further analysis is required.

Figure 01: Banana Price



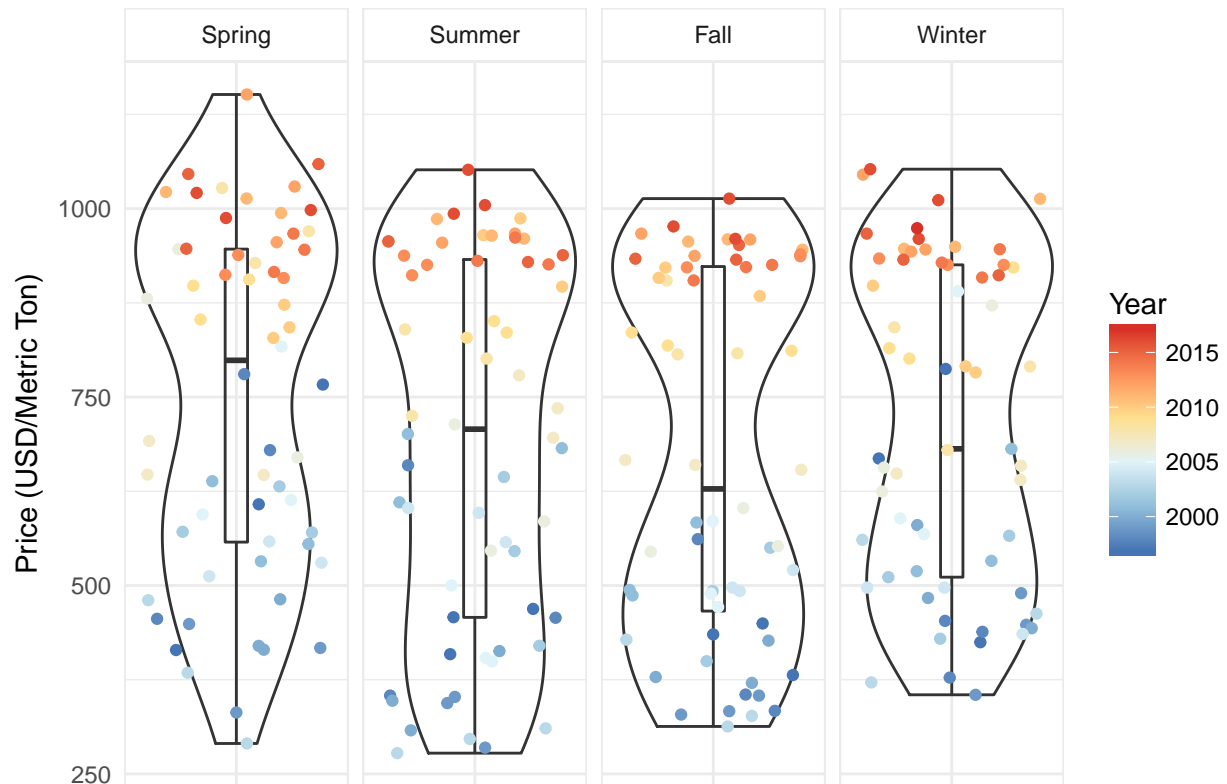
Orginial Seasonal Plot

Figure 02: Seasonal Mean Banana Price



In Figure 02, the price of bananas was over each year, and seperated by season. In Figure 02, each month appears to follow a similar overall increasing trend in the price from 1998. In the graph, the spring and winter months are the highest priced months over the time frame and the summer and fall months follow a trend of being the lowest priced over the time frame. This indicates that over the time-series there is likely a seasonal component to this data and this is investigated further in Figure 03.

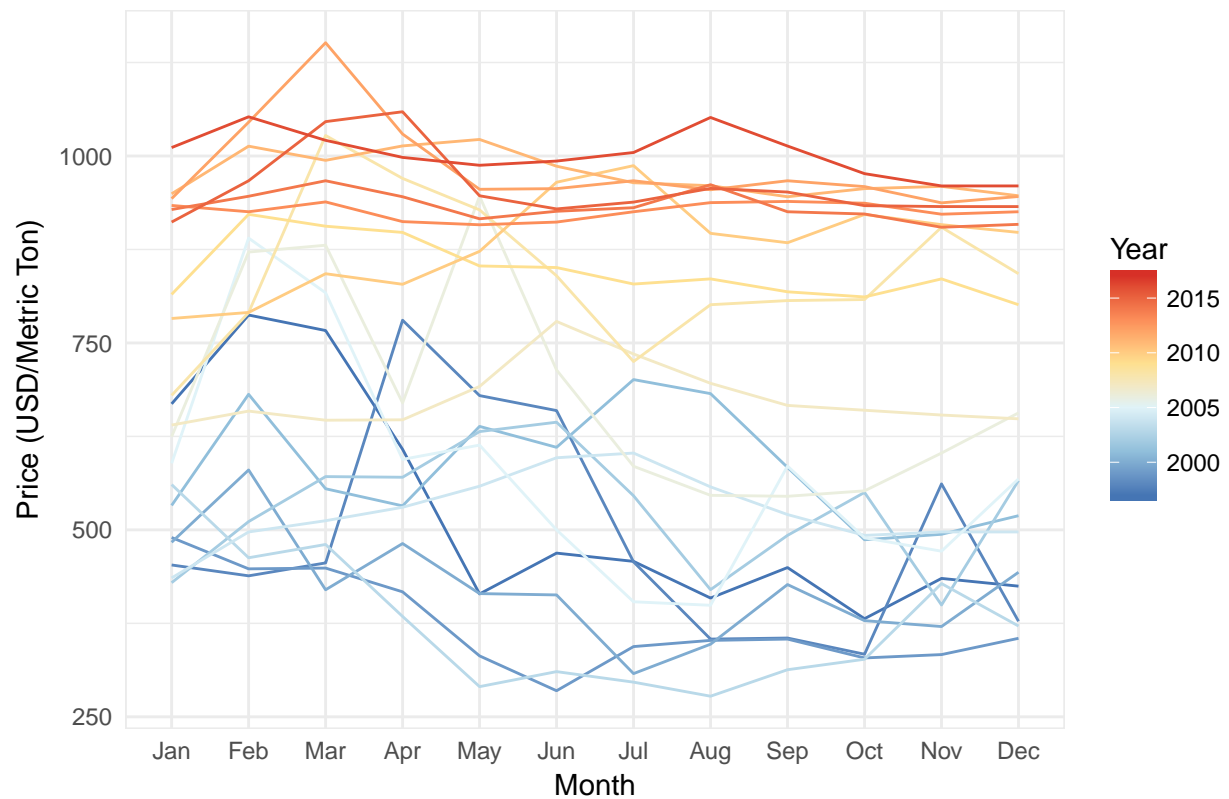
Figure 03: Seasonal Banana Price Violin Plot



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 in 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 (being the earlier years) at the bottom and red (being the later years) at the top. Figure 04 shows this more clearly.

Figure 04: Banana Price by Month



As described in Figure 03, there has been an increase in the price of Bananas over the last twenty years.
 HOW HAS THE PRICE OF BANANAS INCREASED WITH INFLATION?

Lagged Scatter Plot

Figure 05: Lag Scatter Plot

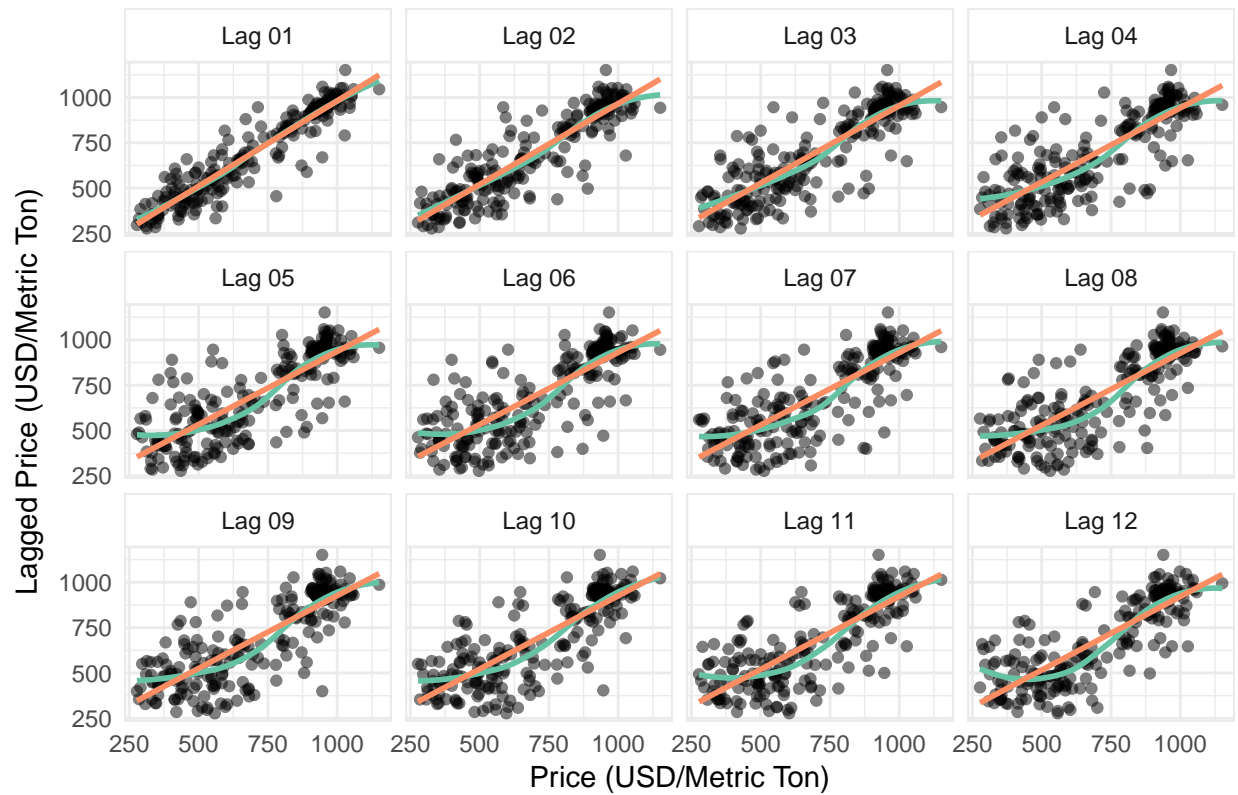


Figure 05 show the lag plots at each month in our dataset. The lag plots are showing the potential correlation between one lag and the previous lag. In these plots we are comparing the first month to all the other months. There appears to be significant correlation in lags 1 through 4, and in lags 4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

Figure 06: Lagged Correlation Plot

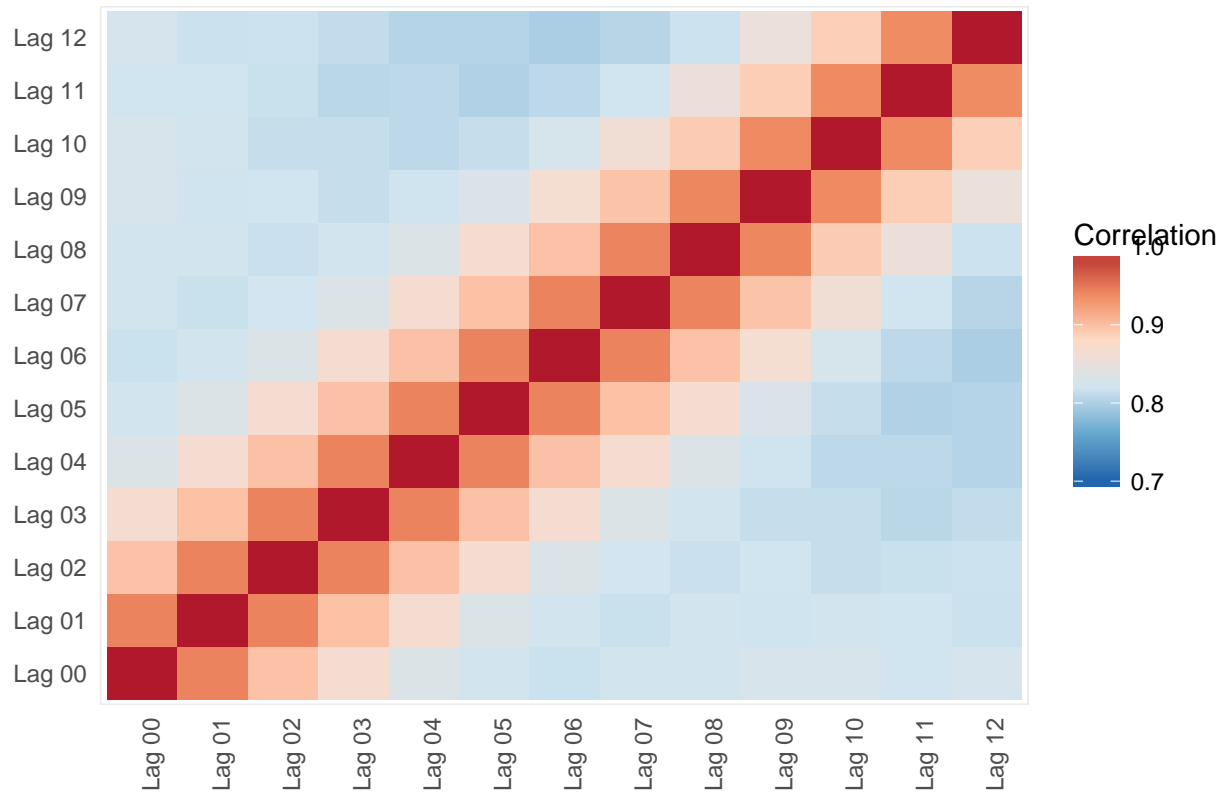


Figure 06 is a correlation plot between the lags. There is a strong correlation between each lag and itself. SOMEONE REWRITE THIS

Autocorrelation Function and Partial Autocorrelation Function

Figure 07: Price ACF

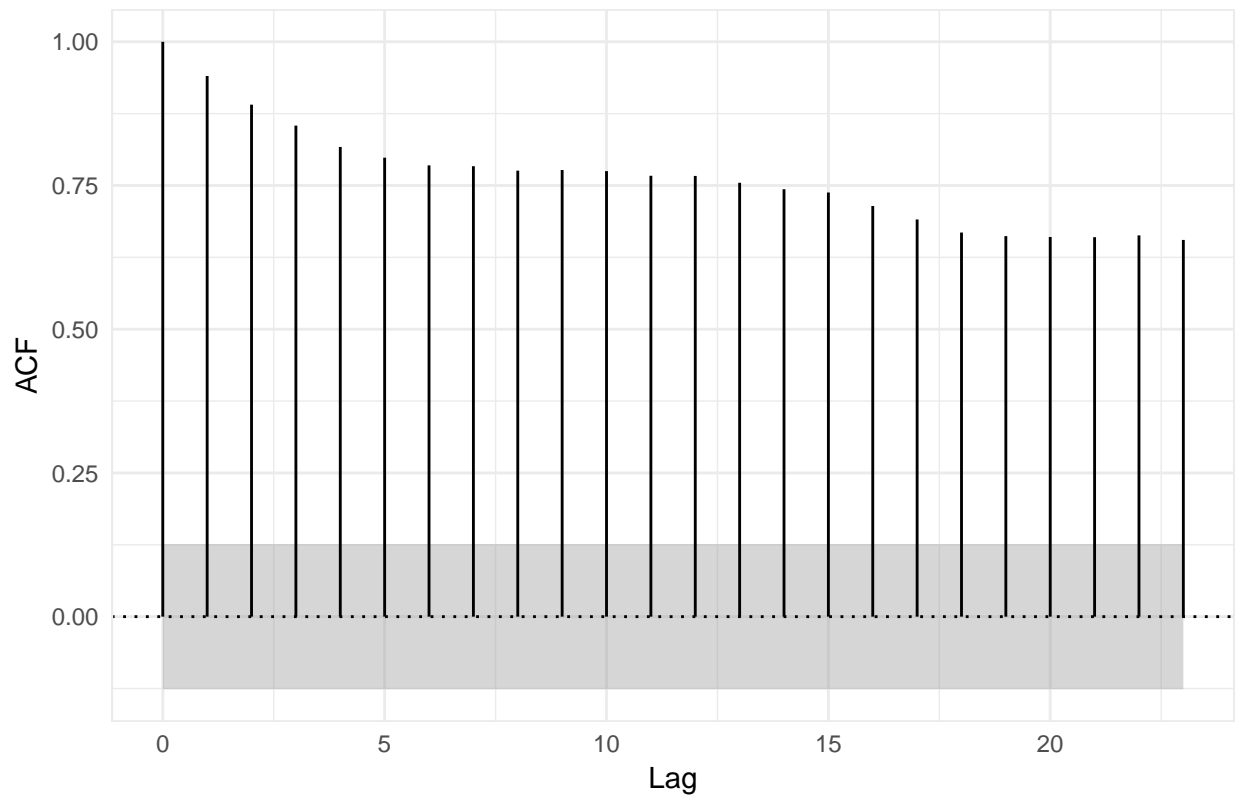
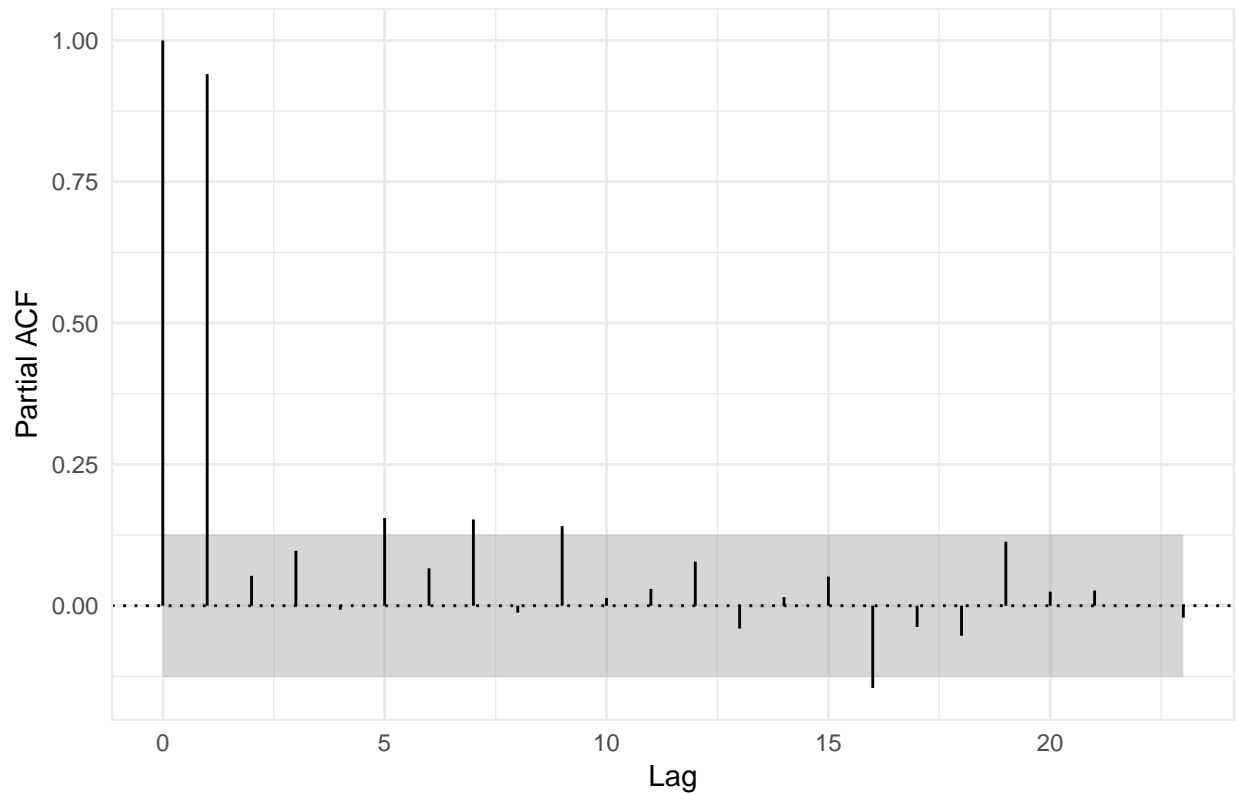


Figure 07 is a plot of the Autocorrelation function, which is checking the correlation between a lag and the previous lag. In this case we see a gradual decrease in the spikes along the plot, which might indicate that there is some autoregressive component required in the model.

Figure 08: Price PACF



In 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.

Transformations

Since there is a seasonal and trend component in the Banana Price dataset, transformations are required.

There were multiple different transformations made on the Banana Price dataset. 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 dataset.

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

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

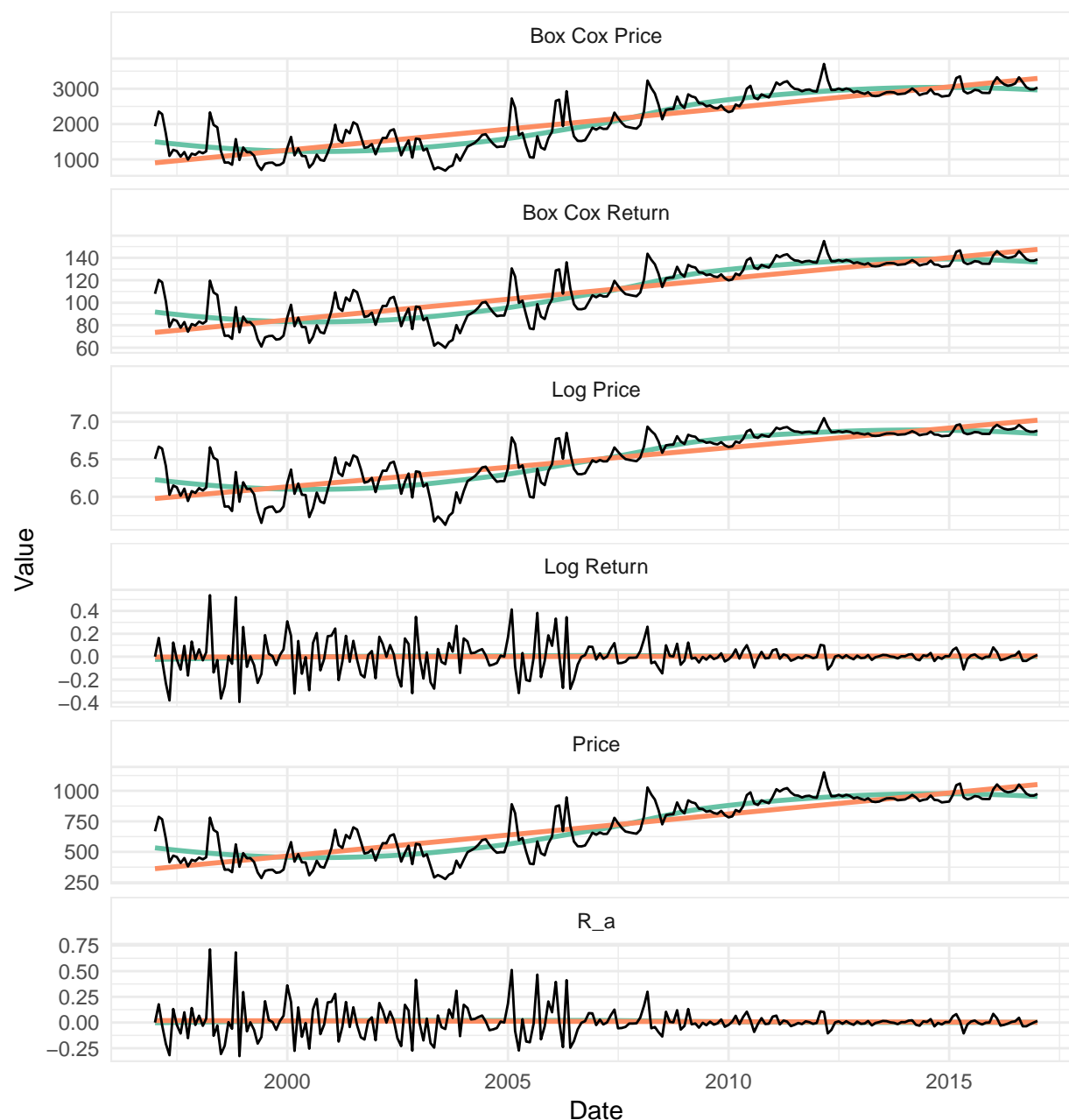
Following the initial transformation of the Banana Price to Banana Return a second was applied by taking the log of the Banana Return. This transformation also had a significant effect on the removal of trend in the Banana Return dataset.

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

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 09: Time Series Plot of Transformations



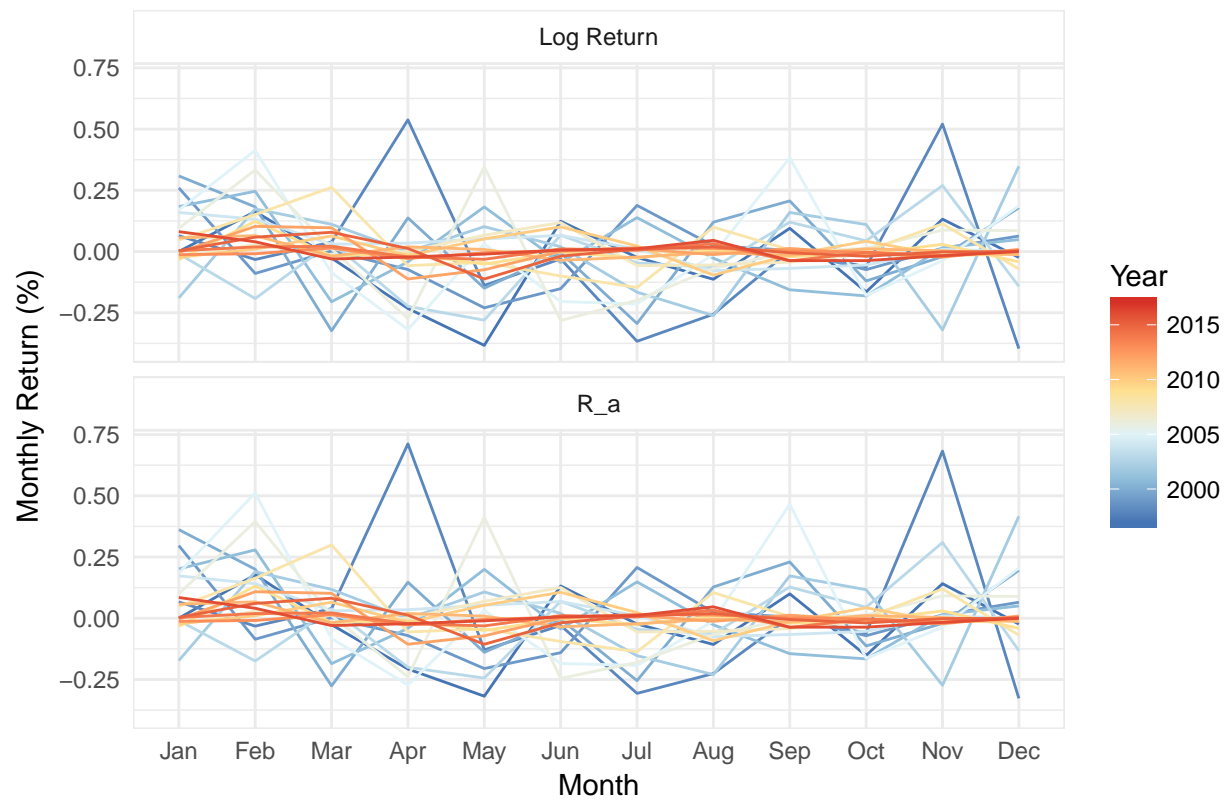
From Figure XX: Violin plots separating the banana data by season were created. After the transformation the effects of the increasing trend have been removed. This is observed as the data points are not showing the same trend where the earlier years, seen in blue, appeared near the bottom and the later 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.

DOUBLE CHECK THIS, MIGHT NEED TO BE MOVED BELOW THE NEXT FIG

Figure 10: Seasonal Return on Bananas Violin Plot



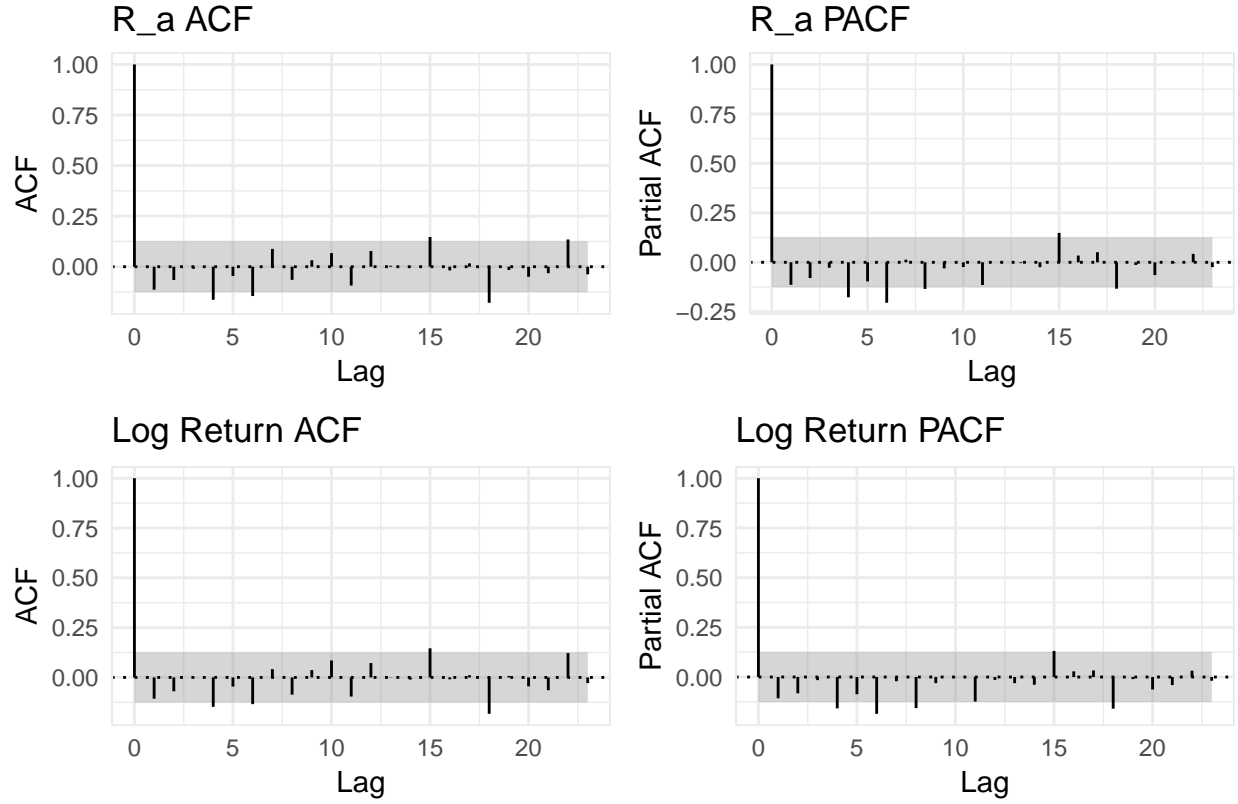
Figure 11: Return on Bananas by Month



From Figures 11, the elimination of the trend in the dataset since the gradual increasing trend from the earlier years, in blue, and the later years, in red, is no longer present.

Transformation Autocovariance Function

Figure 12: ACF & PACF Plots for Return on Bananas



Figures XX through to YY allk show a significant correlation at the first lag, followed by correlations that are not significant. This indicates that there is an autoregressive term in the data.

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 datasets, 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

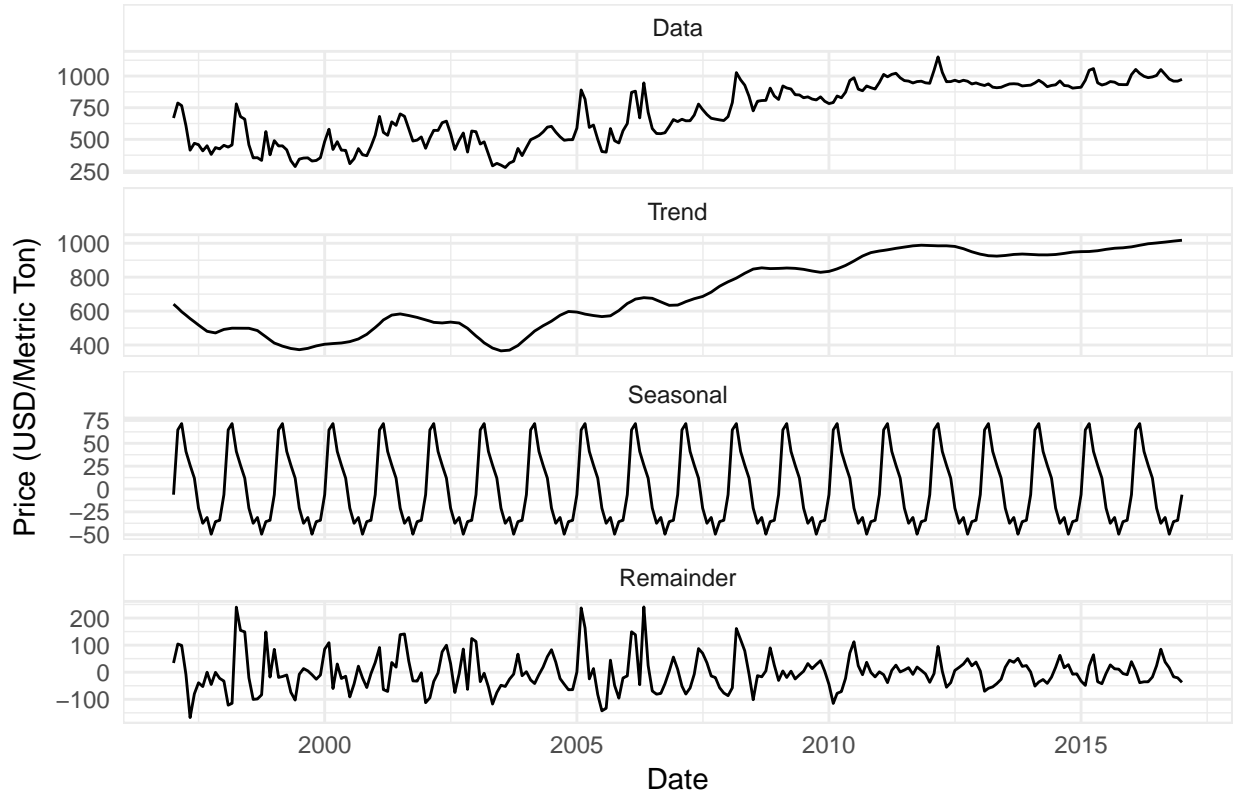
Seasonal	Trend	Remainder
Min. :-0.0587901	Min. :-0.0641957	Min. :-0.3340027

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

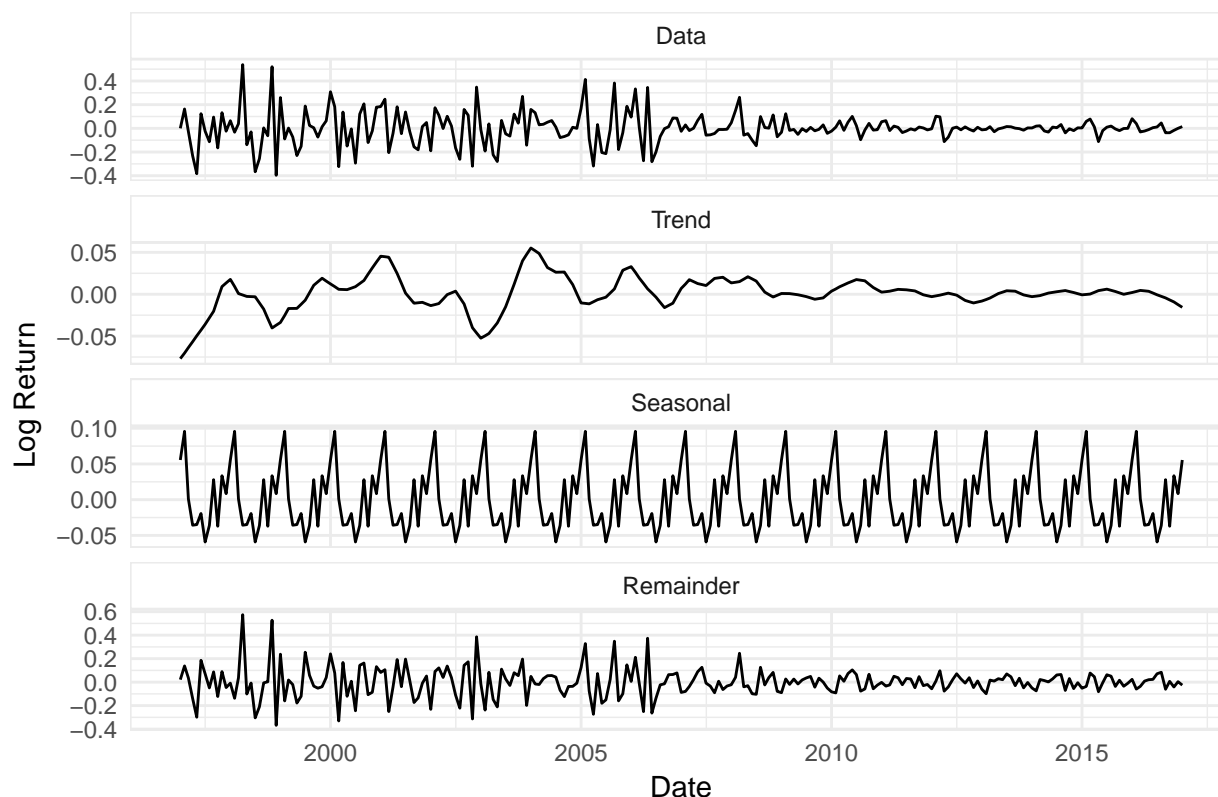
Seasonal	Trend	Remainder
Min. : -0.0591749	Min. : -0.076904	Min. : -0.3676951
1st Qu.: -0.0357003	1st Qu.: -0.005612	1st Qu.: -0.0549211
Median : 0.0011470	Median : 0.001408	Median : 0.0002603
Mean : 0.0002305	Mean : 0.001110	Mean : 0.0002209
3rd Qu.: 0.0334869	3rd Qu.: 0.011565	3rd Qu.: 0.0548785
Max.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

Figure 13: Banana Prices Decomposition



In Figure XX: The remainder or random data and the Banana Price data do not have a similar trend. This indicates that there is some seasonal or trend component in the dataset. The smoothed trendline shows an increasing pattern and the seasonality also has a regular pattern.

Figure 14: Banana Log Returns Decomposition



In Figure YY: The Banana log Returns and the Remainder show a similar trend. This provides evidence towards not including a seasonal or trend component in the model.

ADF Test (Dickey's Test)

Dickey's test is used to evaluate whether the time-series of Banana Price, Banana Return and the Banana Log Return are stationary or integrated.

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

Test statistic	Lag order	P value	Alternative hypothesis
----------------	-----------	---------	------------------------

Table 7: Augmented Dickey-Fuller Test: $\ln R_a$

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

Table 8: Augmented Dickey-Fuller Test: $\ln \text{Log Return}$

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

Table 9: Augmented Dickey-Fuller Test: $\ln \text{Log Return}$

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

SARIMA

AIC and BIC Selection

Model Diagnostics

Model Selection

Forecasting of Banana Price for Next 3 Years

Results

Conclusions

Appendix

References