TS6

Dylan Hayashi

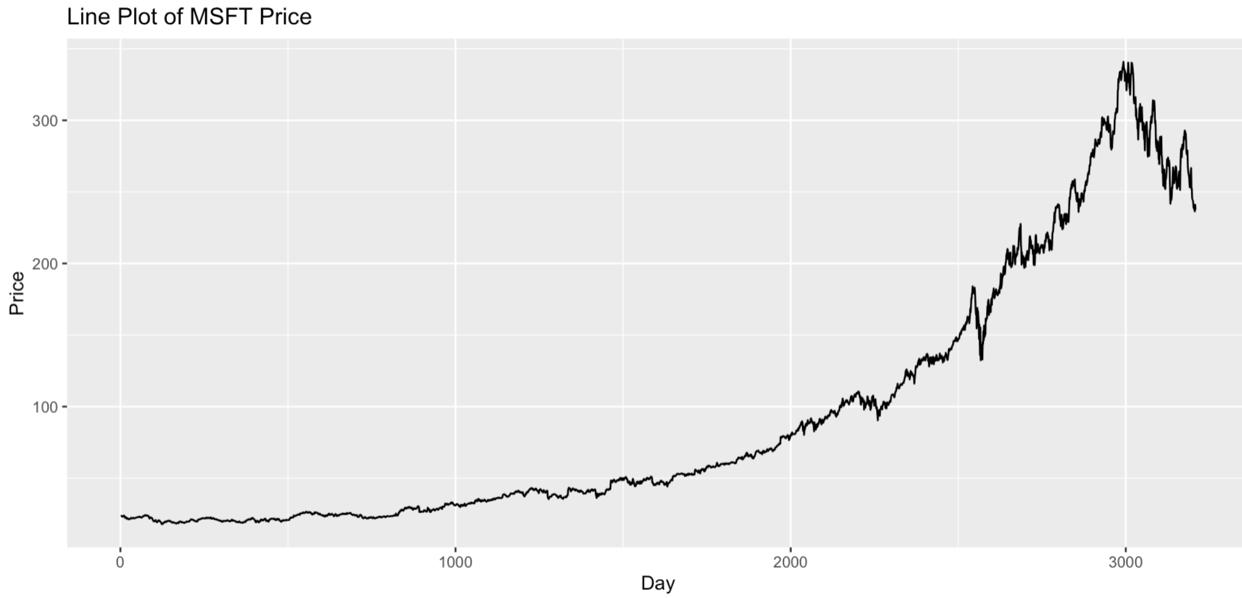
MSDS 413, Fall 2022, Section 55

Northwestern University, Time-Series Analysis & Forecasting

October 31, 2022

1. MSFT

1.1 EDA



Above is a line plot of MSFT price data. Before analysis, confirmation that the data is, in fact, time series data is required. In order to fit the definition, the data must be a time ordered sequence of observations of a stochastic variable over constant time intervals.

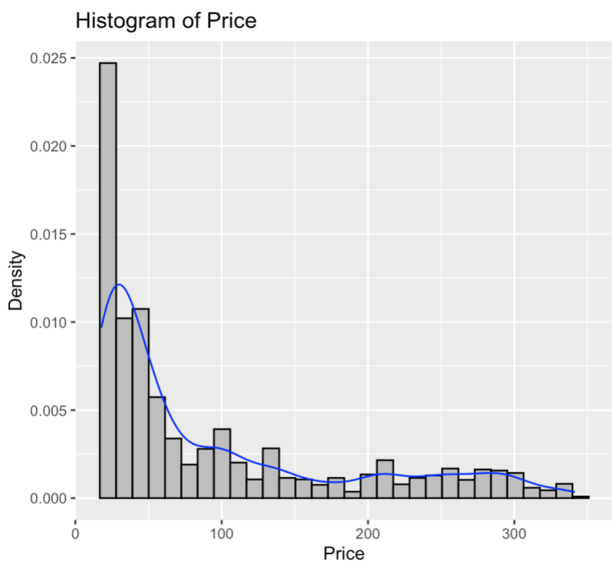
nobs	3208.000000
NAs	0.000000
Minimum	17.819582
Maximum	340.882782
1. Quartile	25.821381
3. Quartile	132.263676
Mean	92.713259
Median	49.183710
Sum	297424.133608
SE Mean	1.554055
LCL Mean	89.666216
UCL Mean	95.760301
Variance	7747.603384
Stdev	88.020471
Skewness	1.254206
Kurtosis	0.294634

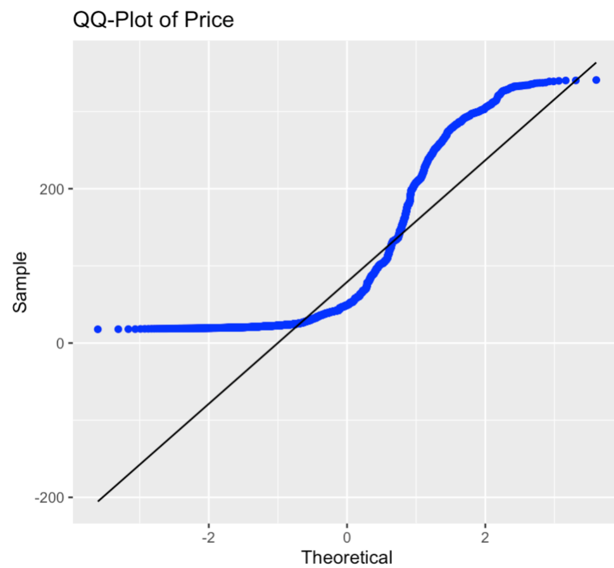
To test that this is a time ordered sequence, we can confirm that the data is indexed by unique time periods and that there is an observation of temperature for each time period. The vectors Day, unique(Day), and Price all have 3208 observations, confirming that this is the case.

To test for constant intervals between time periods, we can sum the difference between each pair of successive periods, and this should equal one less than the total number of observations. The sum of these differences was 3207, one less than the index length of 3208, confirming that the dataset has constant time intervals.

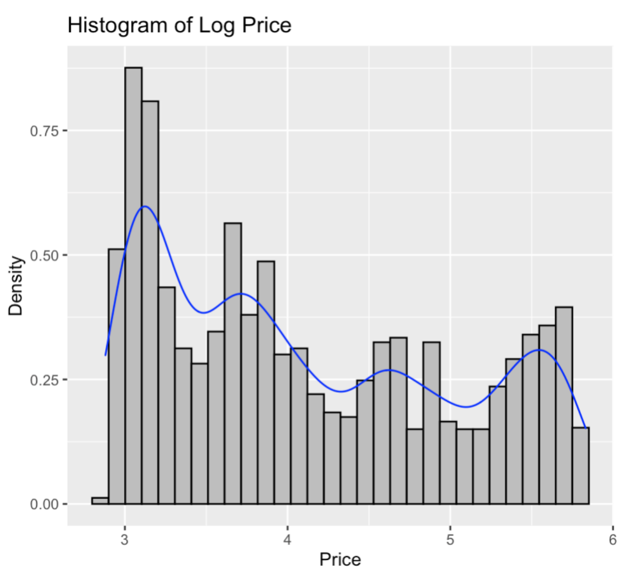
The data summary above on the right, which is of the Price variable, shows that the variable has variance. Having variance is indicative of stochasticity. This confirms that the data match the definition of time-series in its entirety.

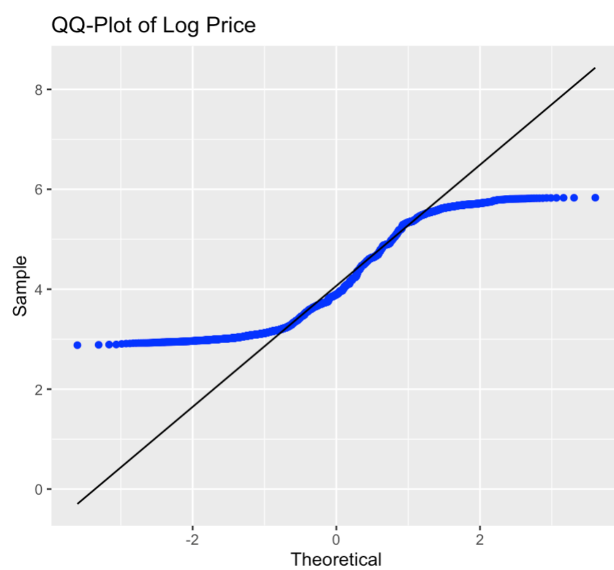
The data are not normally distributed. As the histogram and QQ plot on the next page show, the distribution is highly right skewed and the tails deviate from the normal line.



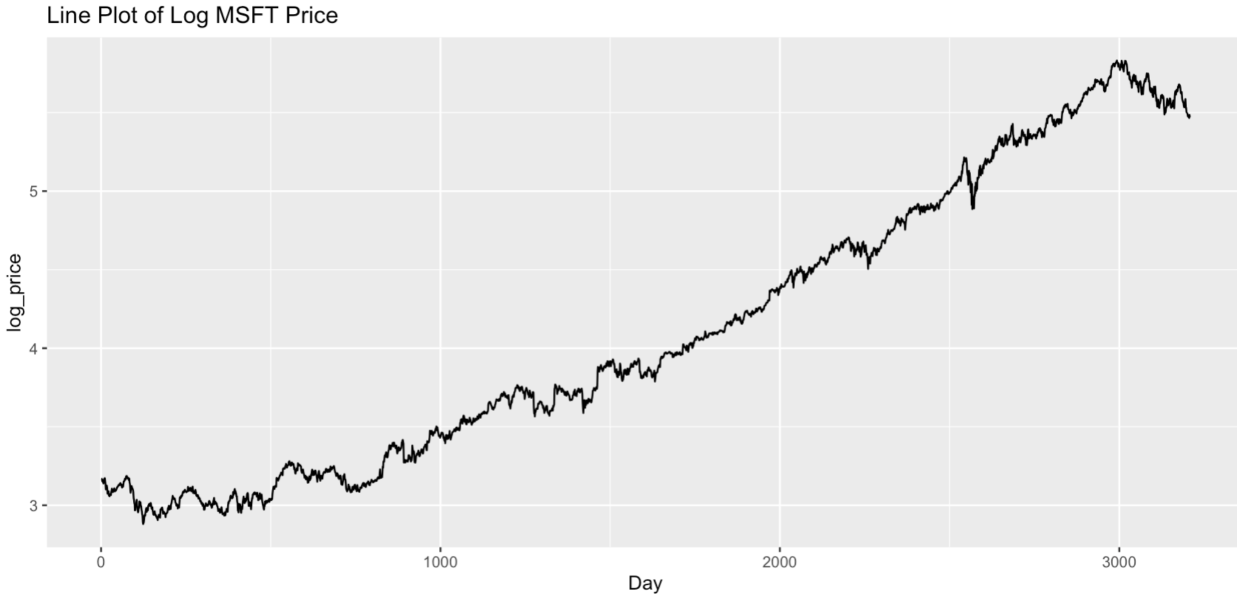


Applying log transformation to the data improves the distribution. The histogram and QQ plot of the new data are shown below. Skew has been improved, demonstrated by the tails in the QQ plot being more identical. The previous 95% confidence interval for skew was (1.251, 1.256), the new is (0.406, 0.41). Kurtosis has worsened, from a 95% confidence interval of (0.29, 0.305 to (-1.177, -1.173).

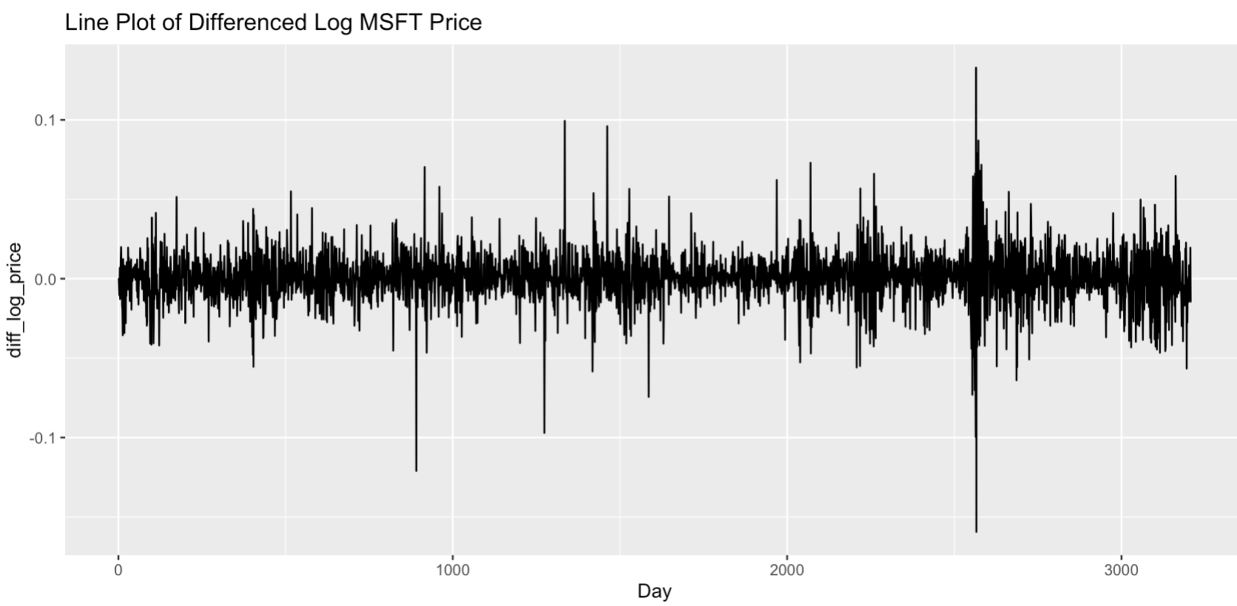




The line plot of the log data, shown on the next page, displays an obvious linear trend. This indicates that the expected return of the log data is not zero, as a zero return would be indicated by a flat line. This also indicates that the data are not stationary, as they are not mean zero. The mean is 4.113, with a t-test 95% confidence interval of (4.082, 4.1444).

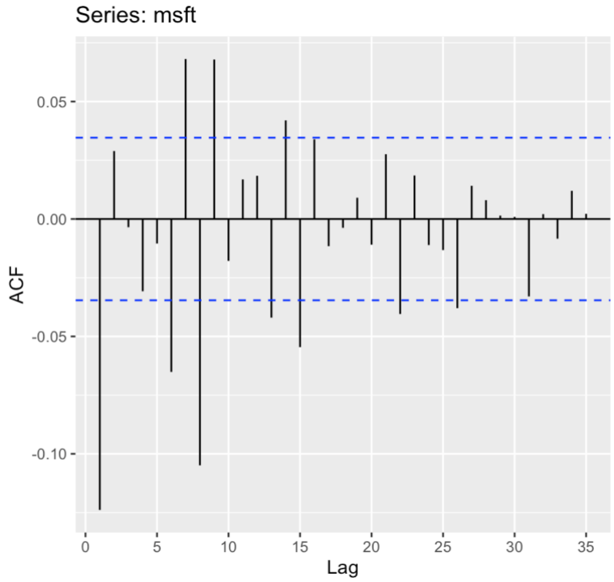


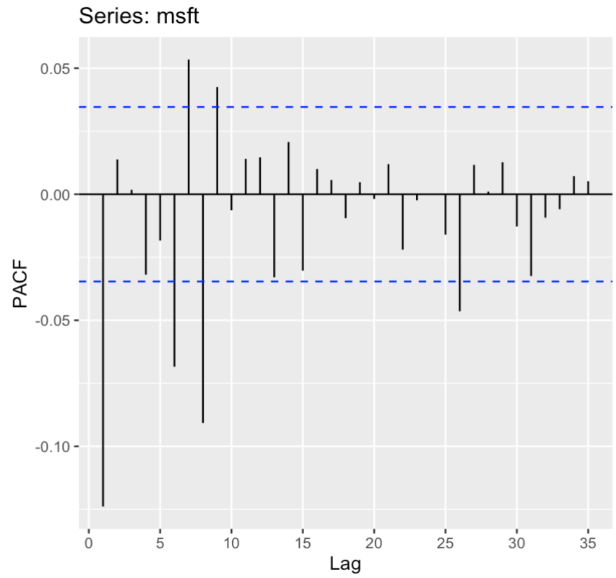
Applying first differencing produces the line plot below. The data now have a mean of 0.00072 with a t-test 95% confidence of (0.00016, 0.0013) and a p-value of 0.01. The p-value indicates a rejection of the null hypothesis that the data are mean zero, but the differenced log data are closer to mean zero than the undifferenced data.



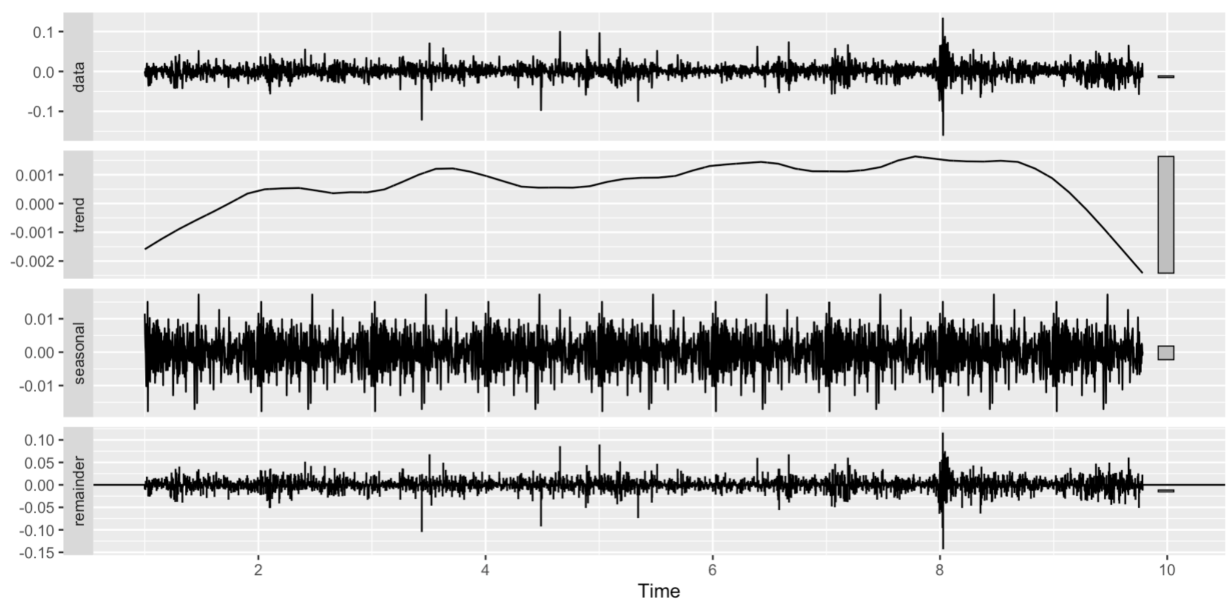
Linear trend stationarity is suggested by ADF and KPSS tests, which each returned a p-value/test-statistic beneath the critical value required to assert stationarity (ADF p-value: 0.01 < 0.05, KPSS test statistic: 0.0996 < 0.146.)

The ACF and PACF plots on the top of the next page indicate that the first lag has significant auto-correlation. This indicates that the log returns do have serial correlation, as differenced log price data is equivalent to log returns.





Below is a decomposition of the data. Overall, it does not provide very much information. The trend line is smooth, but with a very small overall range. The seasonal component does not appear to have an easily identifiable pattern. Additionally, the residuals appear similar to the line plot, indicating that the trend and seasonal components don't explain that much of the variance.



1.2 Mean Model

The ACF and PACF both show one lag with significant auto-correlation. So, for my mean model, I will use an ARIMA(1,0,1) model. This model has two degrees of freedom. The output is below:

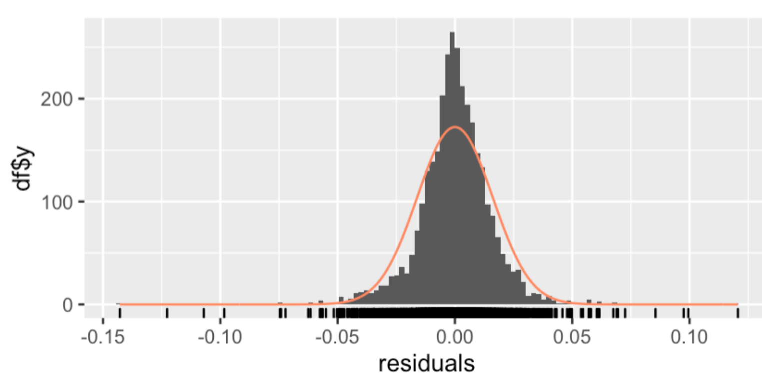
```
Series: msft
ARIMA(1,0,1) with non-zero mean

Coefficients:
          ar1     ma1    mean
      -0.2260  0.1037  0.0007
s.e.   0.1256  0.1281  0.0003

sigma^2 = 0.0002568:  log likelihood = 8707.2
AIC=-17406.39   AICc=-17406.38   BIC=-17382.1
```

The residuals of this model indicate it is a decent fit of the data. The residuals are linear-trend stationary, approximately normally distributed, and do not feature business cycles. However, they are not strictly stationary and they feature auto-correlation.

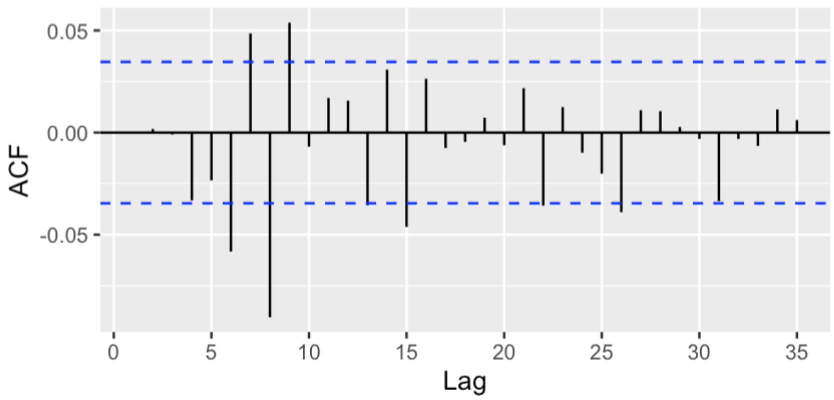
As the summary table on the right shows, their mean was -0.00003, which resulted in a t-test of 95% confidence interval of (-0.0006, 0.0006) with a p-value of 0.9925. Thus, the mean is statistically indistinguishable from zero. Linear-trend stationarity is indicated by ADF and KPSS tests, as each had a p-value/test statistic beneath the critical value required to assert linear-trend stationarity (ADF p-value: $0.01 < 0.05$, KPSS test statistic: $0.1015 < 0.146$.). However, lack of strict stationarity is indicated by a McLeod-Li test, which returned a set with every tested lag, indicating non-constant variance.

```
nobs         3207.000000
NAs             0.000000
Minimum        -0.142803
Maximum         0.120625
1. Quartile    -0.007615
3. Quartile     0.008375
Mean           -0.000003
Median          0.000045
Sum            -0.008541
SE Mean         0.000283
LCL Mean       -0.000557
UCL Mean        0.000552
Variance        0.000257
Stdev           0.016021
Skewness       -0.285908
Kurtosis        7.548099
```

The histogram of the residuals, shown below to the right, shows that the residuals' distribution appears approximately normal; they are unimodal and have good symmetry. The 95% confidence for skew (-0.931, 0.323) included zero, however the confidence interval for excess kurtosis (4.259, 11.258) did not. Overall, however, the residuals are approximately normal.



Auto-correlation in the residuals is demonstrated in their ACF plot and a Box-Ljung test. The ACF plot, shown below, shows that there are several lags with significant auto-correlation. A Box-Ljung of the null hypothesis of independence of observations returned a p-value of less than 0.01, indicating a rejection of the null hypothesis that the residuals are independently distributed.



1.3 ARMA-GARCH Model with Gaussian Distribution

For this section, I fit an ARMA(1,1)-GARCH(1,1) model to the data. The general form of this equation is:

$$dr_t = \omega + \phi_1 dr_{t-1} + \theta_1 z_t + \theta_2 z_{t-1} + \beta_1 \sigma^2_{t-1|t-2} + \alpha_1 a^2_{t-1}$$

The precise coefficients of this model, as well as their relevant summary statistics, are shown in the model output below. This model has 5 degrees of freedom.

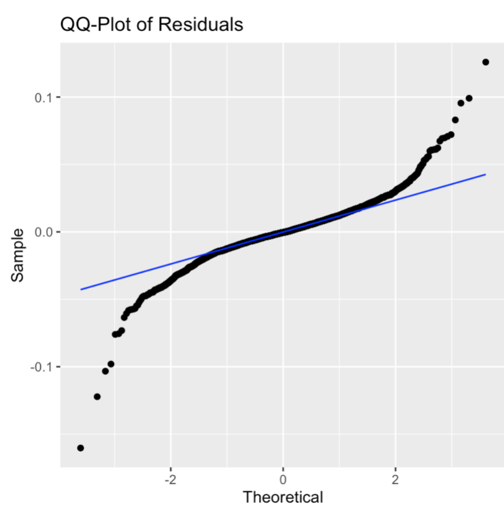
```
Error Analysis:
          Estimate  Std. Error  t value              Pr(>|t|)
mu      0.00024541  0.00012718    1.930                0.0537 .
ar1     0.76994090  0.11233256    6.854  0.00000000000717537 ***
ma1    -0.80790472  0.10312586   -7.834  0.00000000000000466 ***
omega   0.00002404  0.00000420    5.724  0.00000001040712649 ***
alpha1  0.14678870  0.02122236    6.917  0.00000000000462275 ***
beta1   0.76164987  0.03162132   24.087 < 0.0000000000000002 ***
```

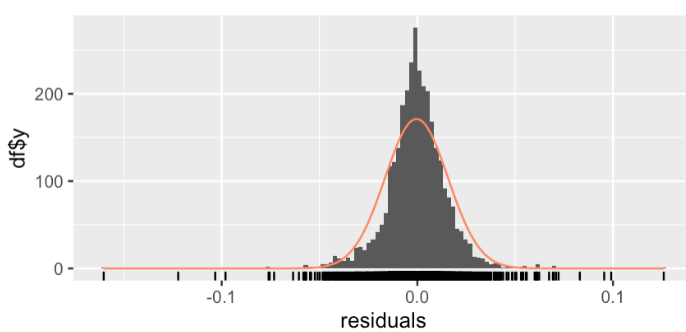
The model output indicates that all coefficients but the constant are statistically significant at all regularly tested significant levels. The sum of the alpha and beta model coefficients are 0.907. While close to one, they are still beneath, and so this model does not feature persistence and is instead mean reverting. The sum of alpha and beta being less than one also indicates a lack of unit root.

The residuals of this model indicate it is a decent fit of the data. The residuals are linear-trend stationary and arguably approximately normally distributed. However, they are not strictly stationary, they feature auto-correlation, and have business cycles.

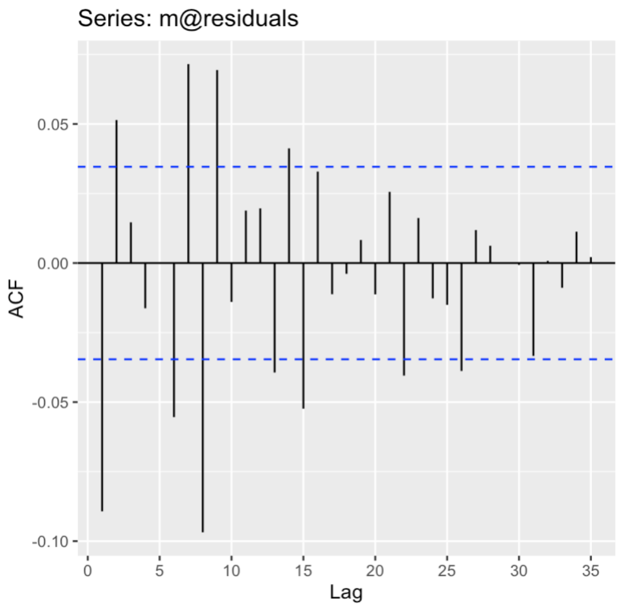
The mean of the residuals was -0.0004 with a t-test of 95% confidence interval of (-0.001, 0.0001) with a p-value of 0.1442. The interval and p-value demonstrate the mean is statistically indistinguishable from zero. Linear-trend stationarity is indicated by ADF and KPSS tests, as each had a p-value/test statistic beneath the critical value required to assert linear-trend stationarity (ADF p-value: 0.01 < 0.05, KPSS test statistic: 0.1153 < 0.146.). A McLeod-Li test of the residuals returned a set with every tested lag, indicating non-constant variance. Thus, the residuals are linear-trend stationary, but not strictly stationary.

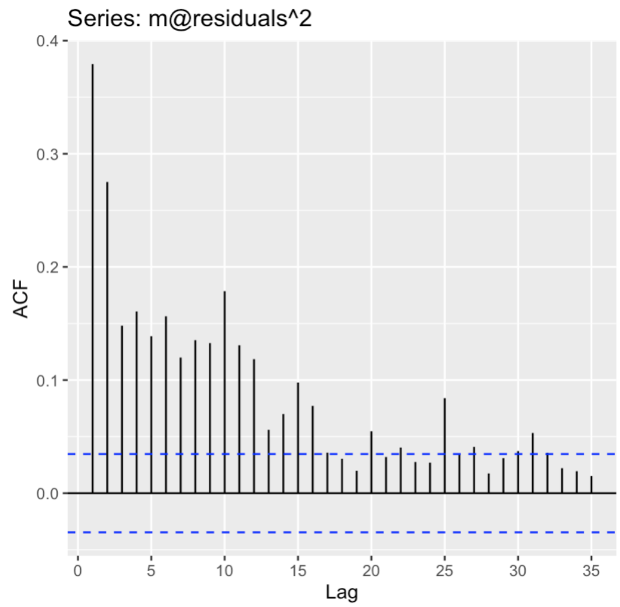
The histogram of the residuals, shown below on the right, depict a distribution that is unimodal and approximately normally distributed. However, there is visibly high kurtosis and a slight left skew. The 95% confidence for skew (-1.079, 0.399) included zero, however the confidence interval for excess kurtosis (3.933, 13.78) did not. The QQ plot, shown below on the left, shows how the tails deviate from the normal line significantly. Despite this, the unimodal shape of the residuals and the inclusion of 0 skew in the confidence interval lead me to conclude that the residuals are approximately normal. Given the size of the tails and their deviation, it is worth investigating the use of an alternative distribution.





Auto-correlation in the residuals is demonstrated by both ACF plots of the residuals and squared residuals, as well as by a Box-Ljung test. The ACF plots shown on the next page indicate multiple lags with significant auto-correlation. The Box-Ljung of the ARCH test test produced a p-value of less than 0.01, indicating a rejection of the null hypothesis that the residuals do not feature time-dependent heteroskedasticity. Thus, the residuals feature auto-correlation, indicating that the model did not explain all auto-correlation.





As previously stated, the residuals do have business cycles. Their frequencies are 12.431, 3.582, and 2.635. These indicate that the model failed to explain all of the variance in the data.

1.4 ARMGA-GARCH Model with Student's t-distribution

As stated in the analysis of the previous model, the QQ plot of its residuals indicate that using an alternative distribution may be worth investigating. In this section, I fit an ARMA(0,1)-GARCH(1,1) model that assumes the data follow Student's t-distribution. This model has four degrees of freedom. Its general form is:

$$dr_t = \omega + \theta_1 z_t + \theta_2 z_{t-1} + \beta_1 \sigma^2_{t-1|t-2} + \alpha_1 a^2_{t-1}$$

The precise coefficients of this model, as well as their relevant summary statistics, are shown in the model output below. This model has four degrees of freedom.

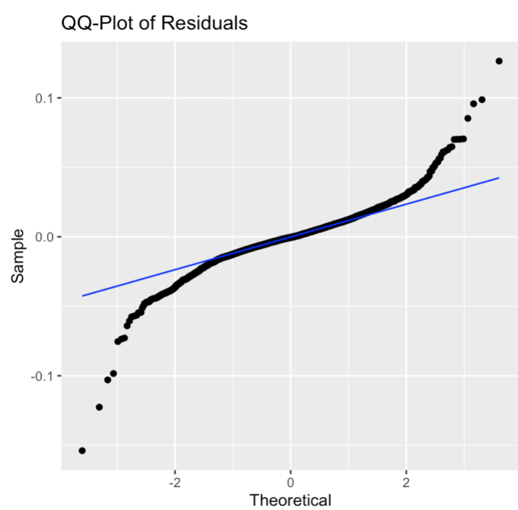
```
Error Analysis:
           Estimate   Std. Error  t value            Pr(>|t|)
mu      0.001055308  0.000196351    5.375        0.00000007675 ***
ma1    -0.051837496  0.017849209   -2.904             0.003682 **
omega   0.000008772  0.000002471    3.550             0.000385 ***
alpha1  0.112127783  0.018963348    5.913        0.00000000336 ***
beta1   0.860658941  0.023101371   37.256 < 0.0000000000000002 ***
shape   4.579712707  0.372809971   12.284 < 0.0000000000000002 ***
```

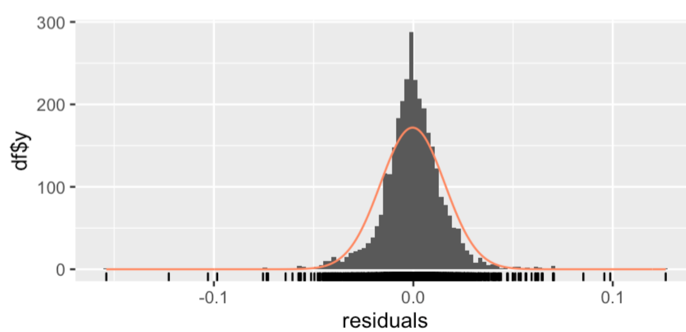
The model output indicates that all coefficients are statistically significant. The sum of the alpha and beta model coefficients are 0.97. While quite close to one, they are still beneath, and so this model does not feature persistence and is instead mean reverting. The sum of alpha and beta being less than one also indicates a lack of unit root, which does not give an indication of business cycles.

The residuals of this model are quite similar to those of the previous. They are linear-trend stationary and approximately normally distributed, but are not strictly stationary and do feature auto-correlation.

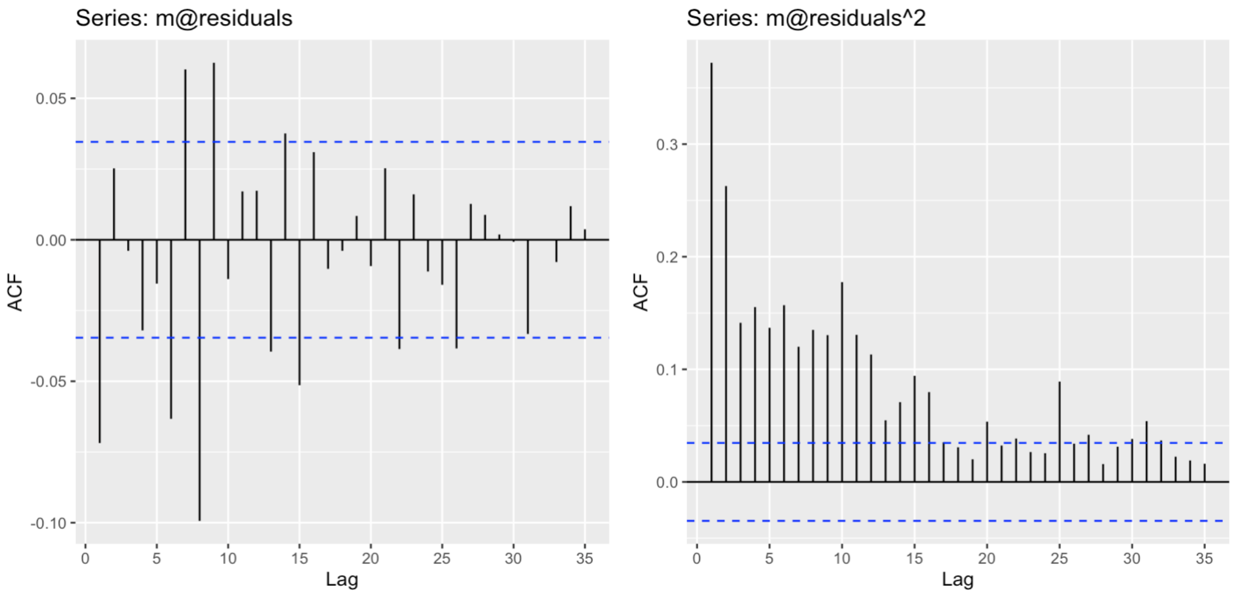
The mean of the residuals was -0.0004 with a t-test of 95% confidence interval of (-0.001, 0.0002) with an accompanying p-value of 0.2106, demonstrating that the mean is statistically indistinguishable from zero. ADF and KPSS tests both indicate linear-trend stationarity, as each had a p-value/test statistic beneath the critical value (ADF p-value: 0.01 < 0.05, KPSS test statistic: 0.1153 < 0.146.). However, A McLeod-Li test of the residuals returned a set with every tested lag, indicating non-constant variance. Therefore, the residuals are linear-trend stationary but not strictly stationary.

The distribution of this model's residuals are closer to normal than those of the previous. The histogram shown below on the right shows that the residuals are slightly less left-skewed, and the QQ plot on the left shows that the tails lie closer to the normal line than before. Similar to the previous model, the residuals' 95% confidence interval for skew (0.976, 0.441) include zero, however, the confidence interval for excess kurtosis (4.095, 12.884) did not. Overall, these results indicate that the t-distribution is a better fit than the normal.





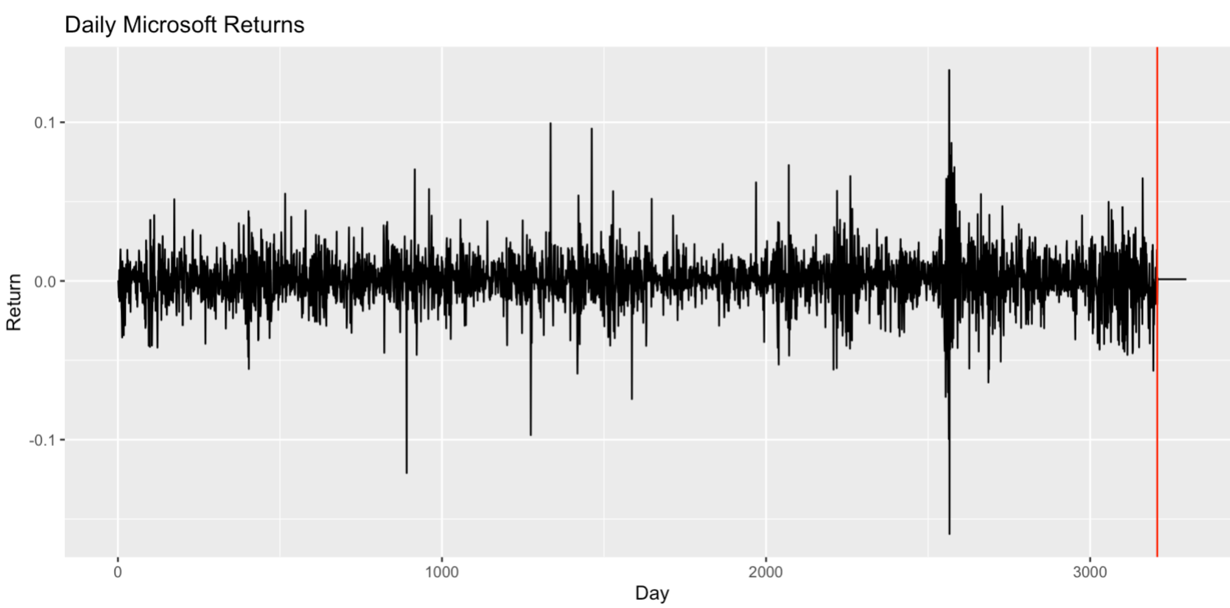
The auto-correlation features of these residuals are also similar to those of the previous model. As shown below, the ACF for the residuals and squared residuals are similar to the previous model. Notably, the ACF of the residuals has fewer lags with statistically significant auto-correlation, indicating that this model is a better fit of the data. A Box-Ljung test in the ARCH function corroborates the notion of auto-correlation, returning a p-value below 0.01, rejecting the null hypothesis of independence.



The residuals have three business cycles, and their respective frequencies are 11.593, 2.381, and 3.898. This indicates that there are patterns of variance in the data that the model failed to explain.

1.5 Forecast

Below is a 90-day forecast of returns based on the model using Student's t-distribution:



The prediction line appears to essentially be the mean of the data. The line also appears constant, indicating that the model did not sufficiently capture the patterns of variance in the data. This is not a positive indication of the model's explanatory power.

1.6 IGARCH Model

In this section, I fit an IGARCH model to the data. The general form of this model is:

$$\sigma^2_{t|t-1} = \omega + \beta_1 \sigma^2_{t-1|t-2} + (1 - \beta_1)\ a^2_{t-1}$$

The coefficients of the fitted model are shown below in the model output. This model has one degree of freedom. As it is an IGARCH model, it features persistence. The beta coefficient is indicated as statistically significant.

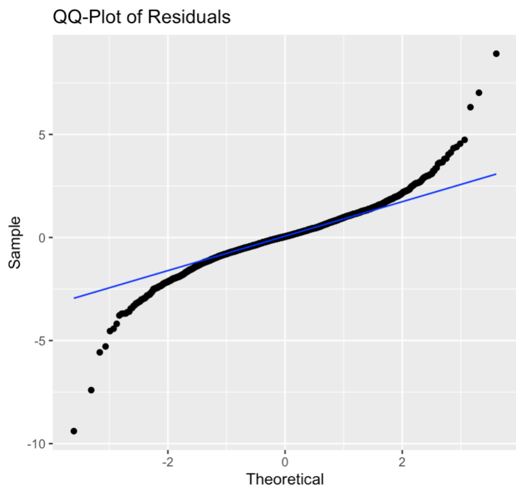
```
Coefficient(s):
       Estimate  Std. Error  t value              Pr(>|t|)
beta 0.96337148  0.00321593  299.562 < 0.000000000000000222 ***
```

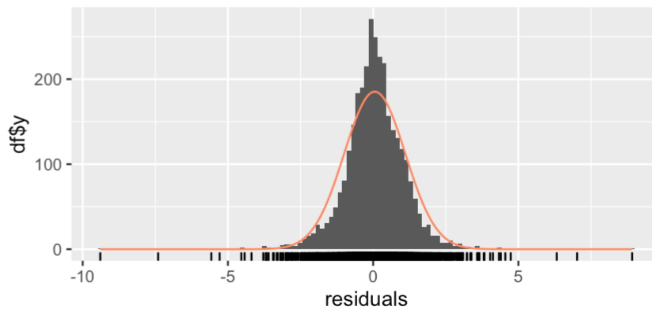
The residuals of this model indicate that it is a better fit of the data than the previous models. The residuals are linear-trend stationary, arguably strictly stationary, approximately normally distributed, and do not feature auto-correlation in the squared residuals. However, they still have business cycles.

The mean of the residuals is 0.055 with a t-test 95% confidence interval of (0.018, 0.091) and a p-value of 0.003. This indicates that the mean is not statistically indistinguishable from zero. However, linear-trend stationarity is suggested by ADF and KPSS tests, which each produced a p-value/test statistic beneath the critical value (ADF p-value: 0.01 < 0.05, KPSS test statistic: 0.1272 < 0.146.). A McLeod-Li test produced an empty set, indicating that the residuals feature constant variance. Linear-trend stationarity and constant variance indicates strict stationarity. However, because the t-test indicated that the residuals mean is not statistically indistinguishable from zero, I have hesitancy to assert strict stationarity. However, linear-trend stationarity and arguable strict stationarity are an improvement relative to the previous models.

The residuals of this model feature a distribution more closely approximating a normal distribution than the previous models. The histogram, shown below on the right, depicts a further improvement of skew relative to the previous models, and the QQ Plot on the left shows more observations falling on the normal line. Similar to the previous models, the 95% confidence interval for skew (-0.736, 0.542) included zero but the confidence interval for excess kurtosis (2.929, 10.716) did not.

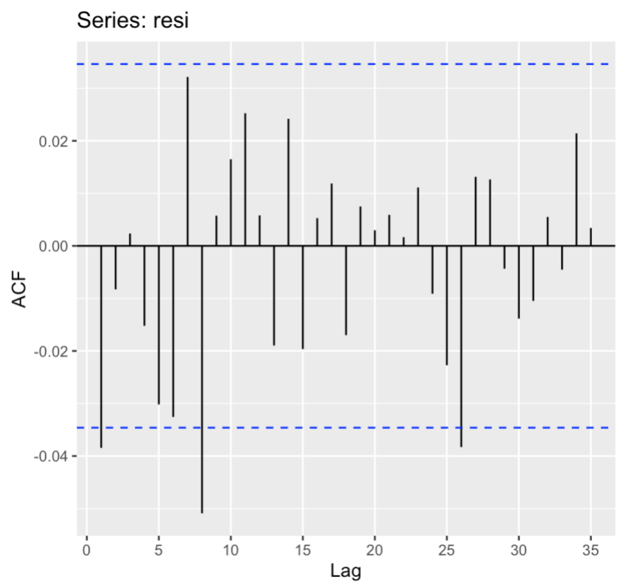
The residuals do have business cycles, and their respective frequencies are 11.593, 2.381, and 3.898. The previous two models also had business cycles of similar frequencies. The cycles indicate that the model did not explain all patterns of variance, however, it is not an indication that this model is better or worse than the previous two.





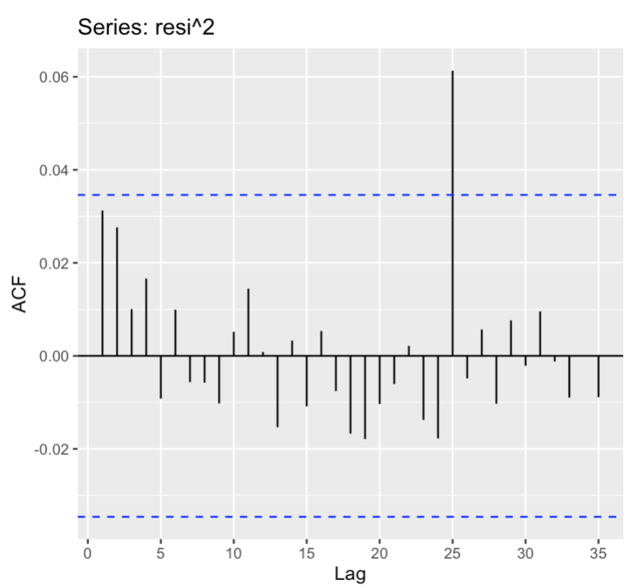
1.7 Serial Correlation in Residuals

To the right is an ACF plot of the residuals. It shows a number of lags with significant auto-correlation. A Box-Ljung test of the residuals agrees with this notion, returning a p-value of 0.006, indicating a rejection of the null hypothesis that the residuals do not feature serial correlation.



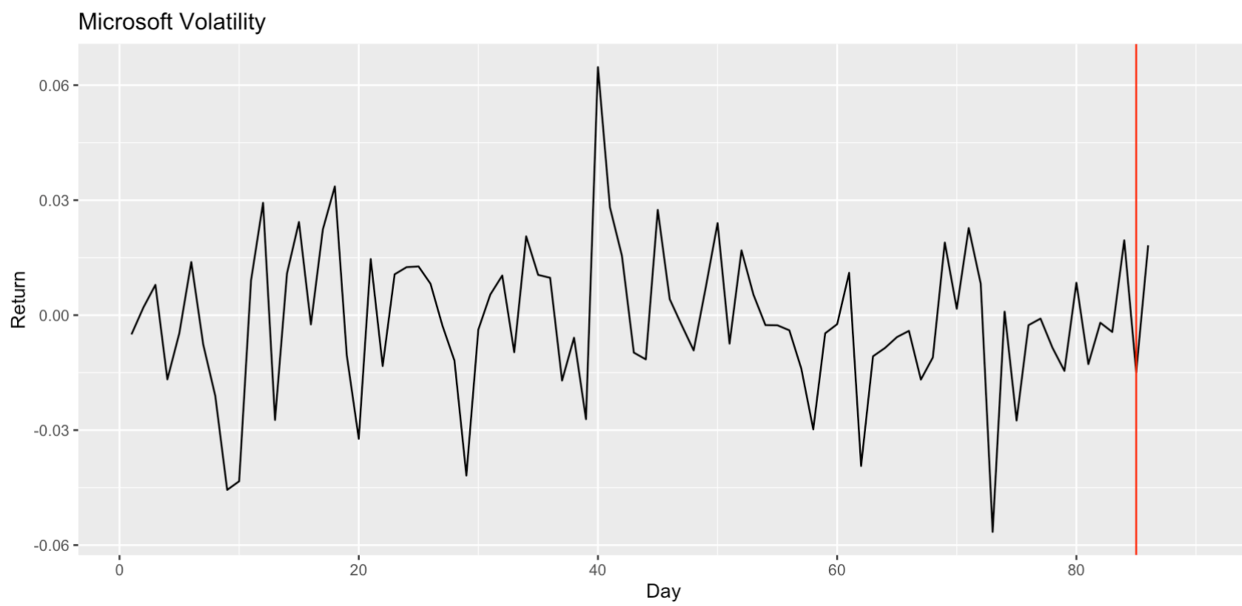
1.8 Serial Correlation in Squared Residuals

Below is an ACF plot of the squared residuals, which features only one lag with significant auto-correlation. This is the 25th lag, and there is a possibility that this was by chance and not indicative of auto-correlation. Lack of serial correlation in the squared residuals is corroborated by a Box-Ljung test, which produced a p-value of 0.6268, supporting the null hypothesis that the residuals are independent. Relative to the previous two models, this is an improvement.



1.9 IGARCH Model Fit and Forecasts

Based on the residuals, I think that the IGARCH model is a good fit of the data. The residuals are linear-trend stationary, arguably strictly stationary, and feature squared residuals without serial correlation. Below is a line plot of the model's predictions of the next 5 days. Overall, it looks like a reasonable prediction given the recent behavior of the graph. This model predicts more than just the mean of the data, and the level of variance that the model predicts is not outsized relative to the previous observations.



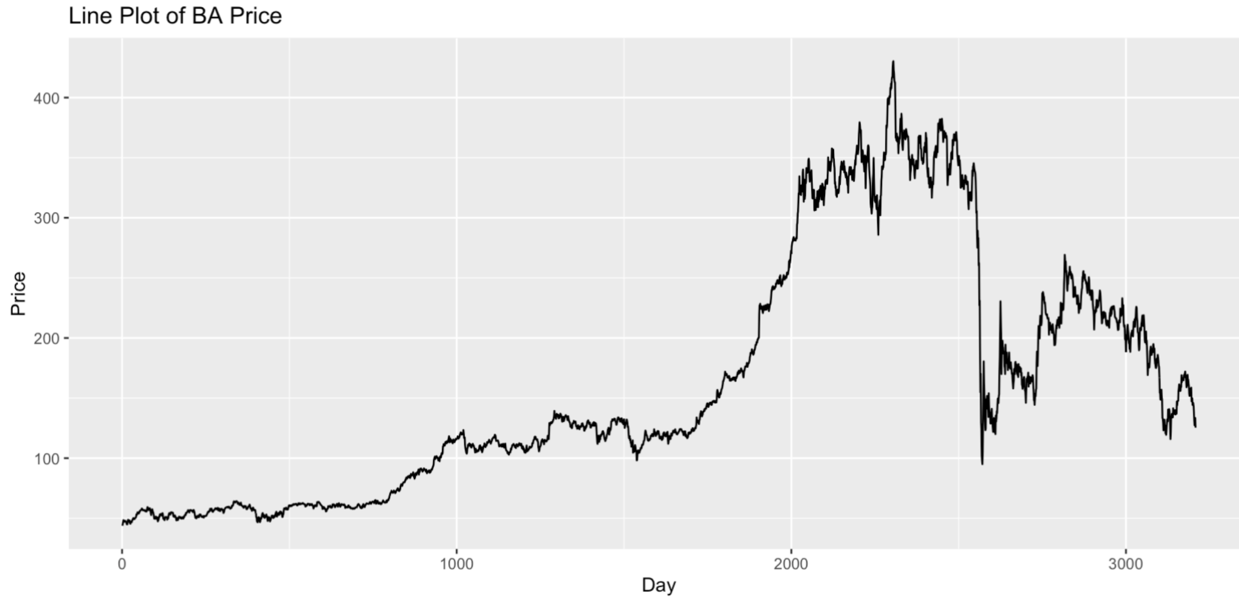
1.10 Model Comparison

I prefer the IGARCH model (model 3) relative to the two ARMA-GARCH models (models 1 and 2.) As I noted in my analysis of the residuals of model 3, they demonstrate a better fit of the data. The residuals of model 3 are linear-trend stationary and arguably strictly stationary, while the residuals of models 1 and 2 had McLeod-Li tests indicating significant auto-correlation at every lag, excluding the possibility of strict stationarity. The distribution of the model 3 residuals also more closely approximate a normal distribution compared to models 1 and 2, with far less skew. Lastly, the squared residuals of model 3 did not feature auto-correlation, and the squared residuals of the first two models did.

2. Boeing Returns

2.1 EDA

Below is a time plot of Boeing's price data:



Prior to analysis, confirmation that the data is, in fact, time series data is required. In order to fit the definition, the data must be a time ordered sequence of observations of a stochastic variable over constant time intervals.

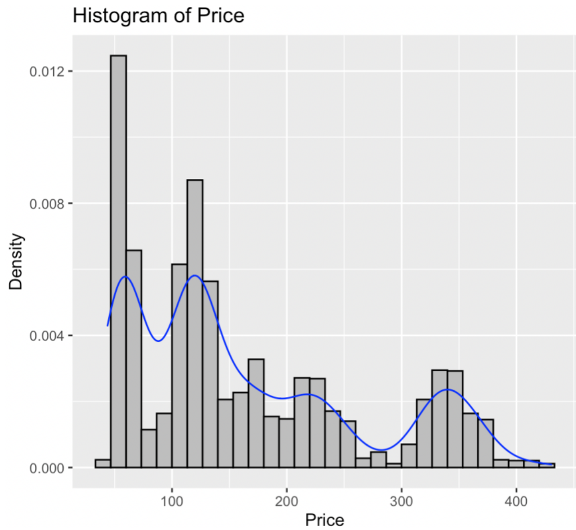
To test that this is a time ordered sequence, we can confirm that the data is indexed by unique time periods and that there is an observation of temperature for each time period. The vectors Day, unique(Day), and Price all have 3208 observations, confirming that this is the case.

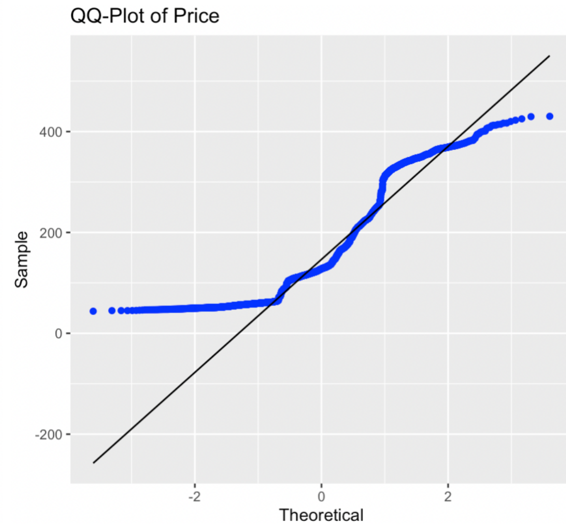
To test for constant intervals between time periods, we can sum the difference between each pair of successive periods, and this should equal one less than the total number of observations. The sum of these differences was 3207, one less than the index length of 3208, confirming that the dataset has constant time intervals.

The data summary on the right, which is of the Price variable, shows that the variable has variance. Having variance is indicative of stochasticity. This confirms that the data match the definition of time-series in its entirety.

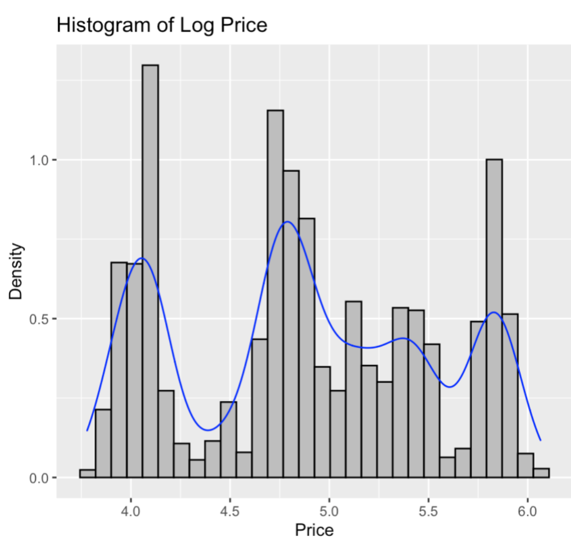
The histogram and QQ plot on the top of the next page show the data are not normally distributed. The distribution appears to be tri-modal and the observations deviate from the normal line in both the tails and the center of the distribution. To remedy this, I will apply log transformation to the data.

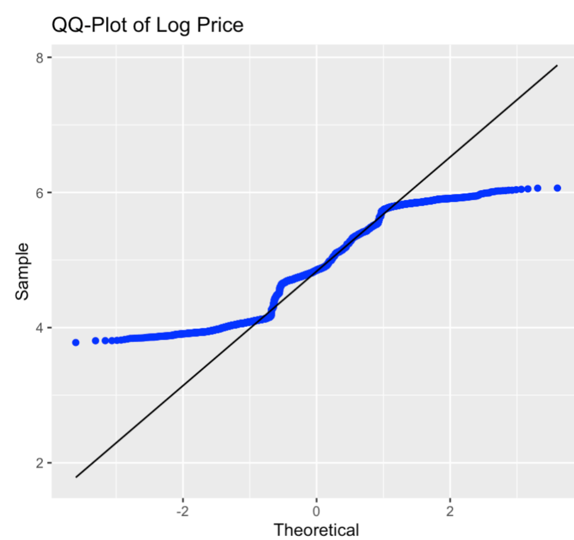
nobs	3208.000000
NAs	0.000000
Minimum	43.777557
Maximum	430.299957
1. Quartile	70.963442
3. Quartile	222.131356
Mean	161.348721
Median	127.782306
Sum	517606.698309
SE Mean	1.761120
LCL Mean	157.895687
UCL Mean	164.801756
Variance	9949.751357
Stdev	99.748440
Skewness	0.815219
Kurtosis	-0.508319



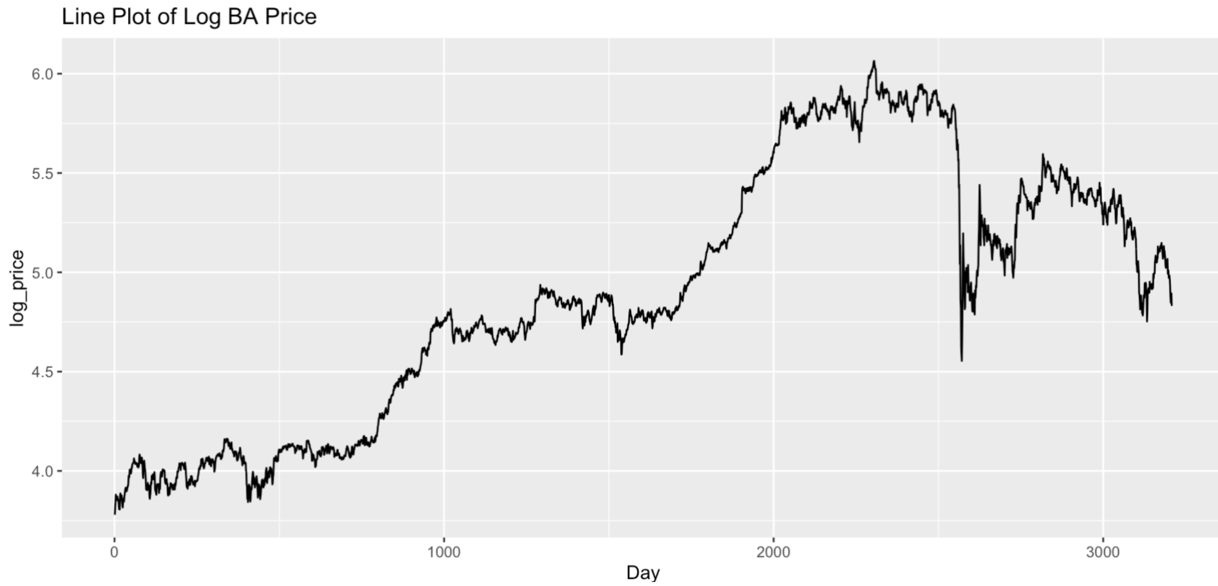


As the histogram and QQ plots below show, log transformation did not make the data normally distributed. The data still appears tri-modal and the observations still deviate from the normal line in the QQ plot. However, the distribution did improve. Overall, the distribution appears to have less skew and fewer observations deviate from the normal line. The previous 95% confidence interval for skew was (0.758, 0.874) and the new is (-0.021, 0.069), which contains zero. Unfortunately, excess kurtosis did not improve, changing from (-0.645, -0.373) to (-1.177, -1.08). However, the log transformation made an overall improvement to the distribution of the data.

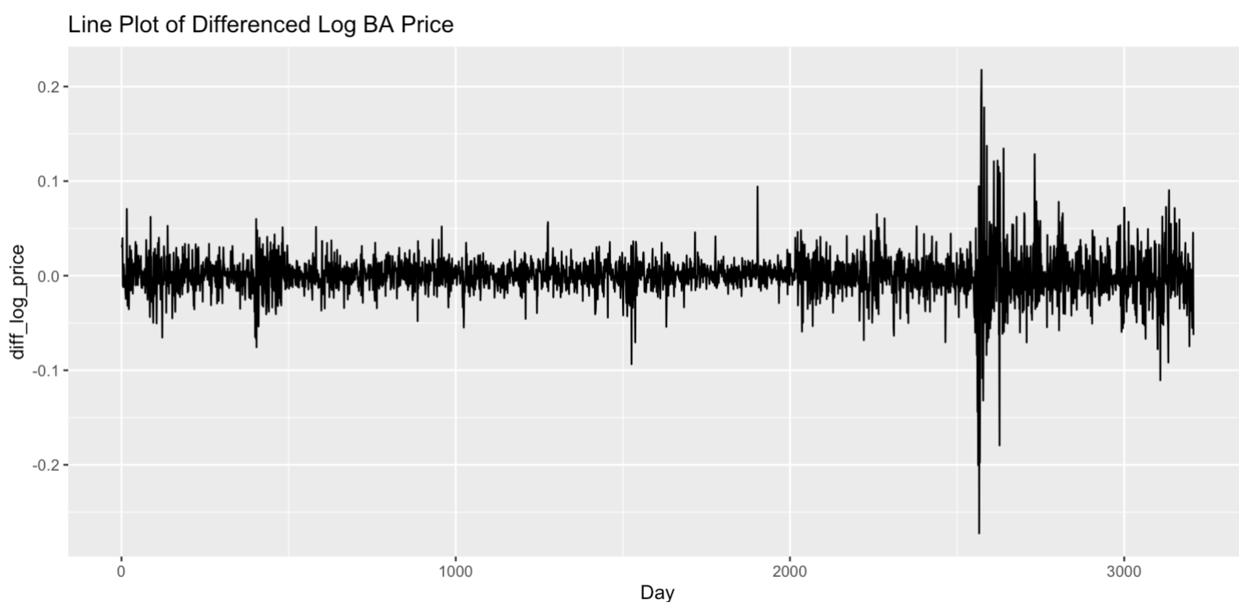




The line plot of the log data, shown on the next page, displays an obvious linear trend. This indicates that the expected return of the log data is not zero, as a zero return would be indicated by a flat line. This also indicates that the data are not stationary, as they are not mean zero. The mean is 4.89, with a t-test 95% confidence interval of (4.868, 4.9117).

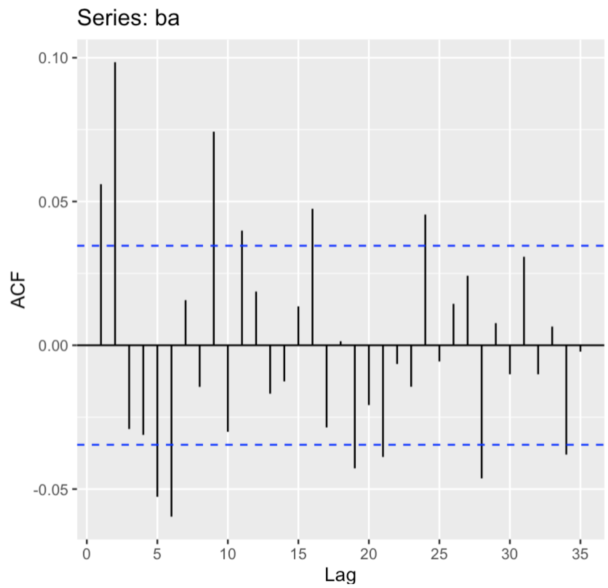


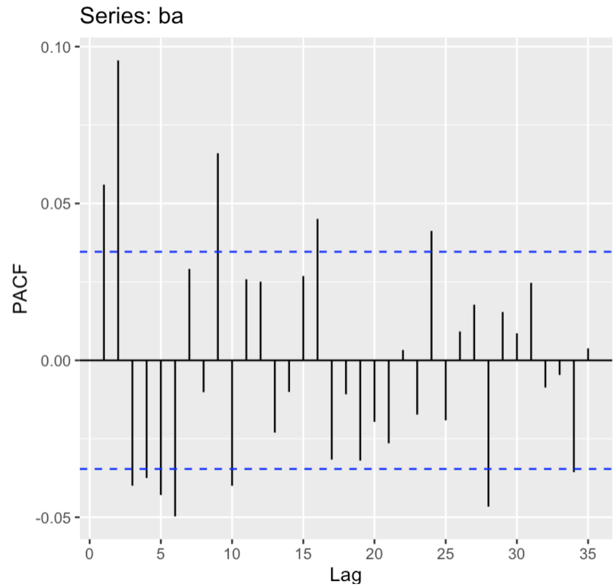
Applying first differencing produces the line plot below. The data now have a mean of 0.0003 with a t-test 95% confidence of (-0.0005, 0.001) and a p-value of 0.4186. Thus, we can conclude that the differenced log data are mean zero.



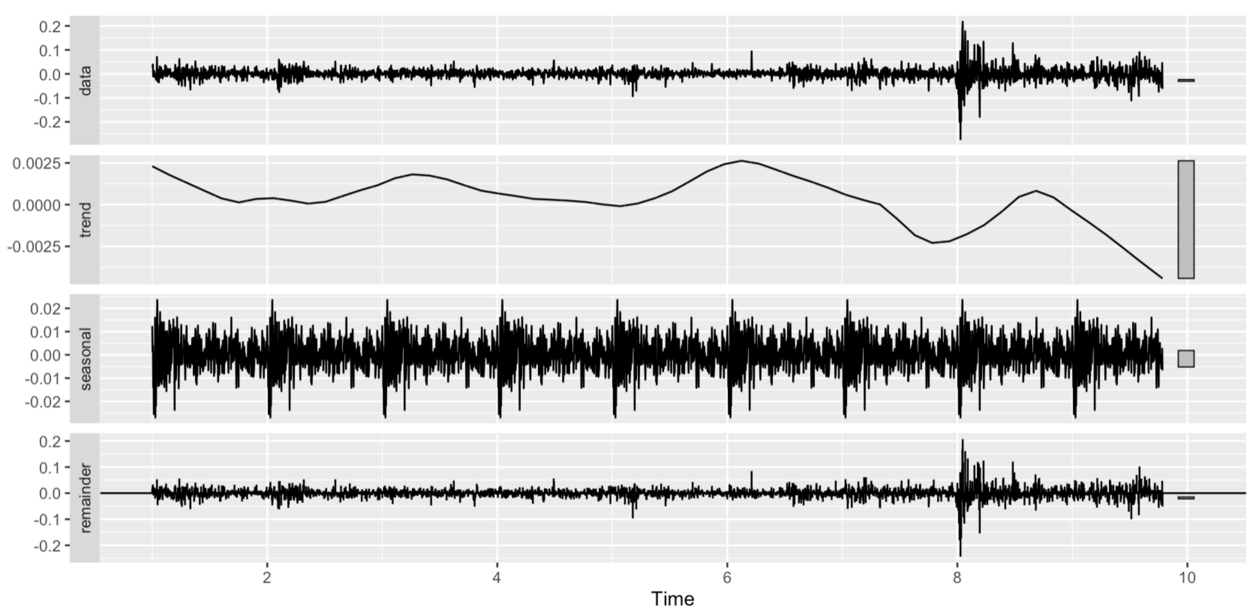
After differencing, ADF and KPSS tests each returned a p-value/test-statistic beneath the critical value required to assert linear-trend stationarity (ADF p-value: 0.01 < 0.05, KPSS test statistic: 0.0996 < 0.146.)

The ACF and PACF plots on the top of the next page indicate the time series has multiple lags with significant auto-correlation. This indicates that the log returns do have serial correlation, as differenced log price data is equivalent to log returns. Thus, ARCH effects are present.





Below is a decomposition of the data. Overall, it does not provide very much information. The trend line is smooth, but with a very small overall range. The seasonal component does not appear to have a discernable pattern. The residuals appear similar to the line plot, indicating that the trend and seasonal components don't explain that much of the variance.



2.2 GARCH Model with Gaussian Distribution

In this section, I use a GARCH(1,1) model that assumes a Gaussian distribution. The general form of this model is:

$$\sigma^2_{t|t-1} = \omega + \beta_1 \sigma^2_{t-1|t-2} + \alpha_1 a^2_{t-1}$$

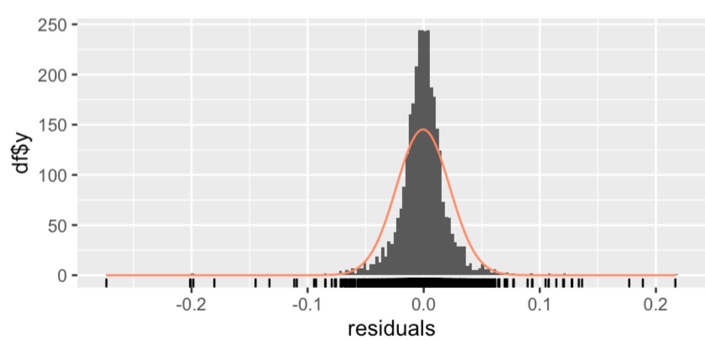
The precise coefficients of the model are shown below in the model output. The sum of the alpha and beta coefficients is slightly below one, indicating that this model is mean-reverting. All variables are indicated as statistically significant. This model has two degrees of freedom.

```
Error Analysis:
          Estimate  Std. Error  t value              Pr(>|t|)
mu      0.000913829 0.000271442    3.367             0.000761 ***
omega   0.000008151 0.000001392    5.857         0.0000000047 ***
alpha1  0.089476407 0.009772840    9.156 < 0.0000000000000002 ***
beta1   0.889891810 0.011303171   78.729 < 0.0000000000000002 ***
```

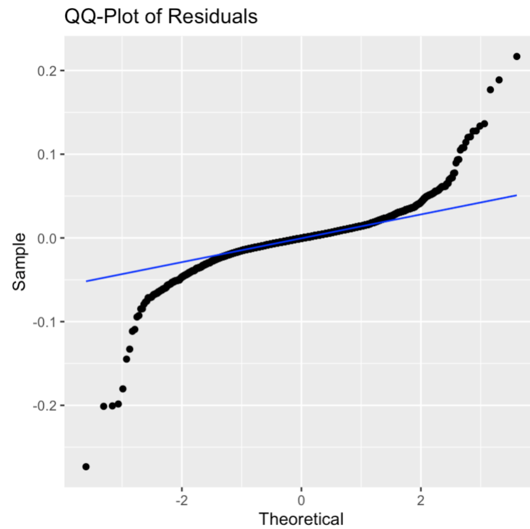
The residuals indicate that this model is not a great fit of the data. While they are linear-trend stationary and approximately normally distributed, they are not strictly stationary, they do feature auto-correlation, and do have detectable business cycles.

The mean of the residuals was -0.0006 with a t-test 95% confidence interval of (-0.0014, 0.0002) with an accompanying p-value of 0.1485, demonstrating that the mean of the residuals is statistically indistinguishable from zero. Linear-trend stationarity is suggested by both ADF and KPSS, which each had a p-value/test statistic beneath the critical value (ADF p-value: 0.01 < 0.05, KPSS test statistic: 0.0737 < 0.146.). However, A McLeod-Li test of the residuals returned a set with every tested lag, indicating non-constant variance. Therefore, the residuals are linear-trend stationary but not strictly stationary.

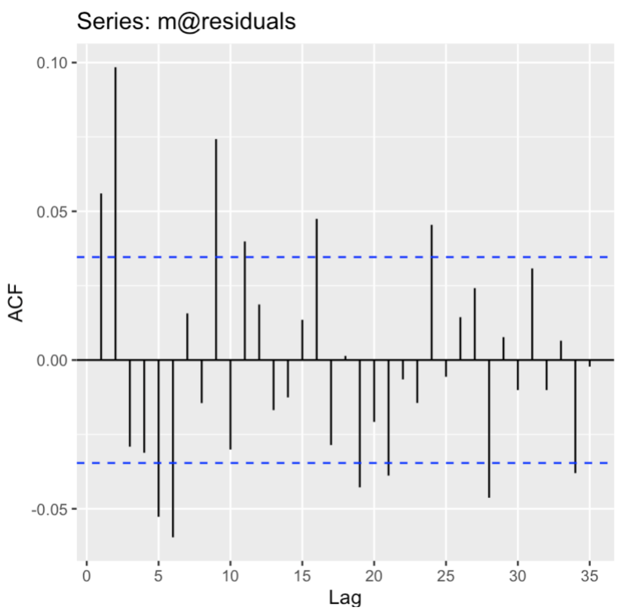
A histogram of the residuals is shown below. The distribution is approximately normal with slight left skew (not supported by confidence interval) and high kurtosis. The 95% confidence interval for skew was (-1.819, 0.752), which does contain zero. The confidence interval for kurtosis was (12.91, 30.89), which is quite high. The residuals' QQ plot, shown on the top of the next page,

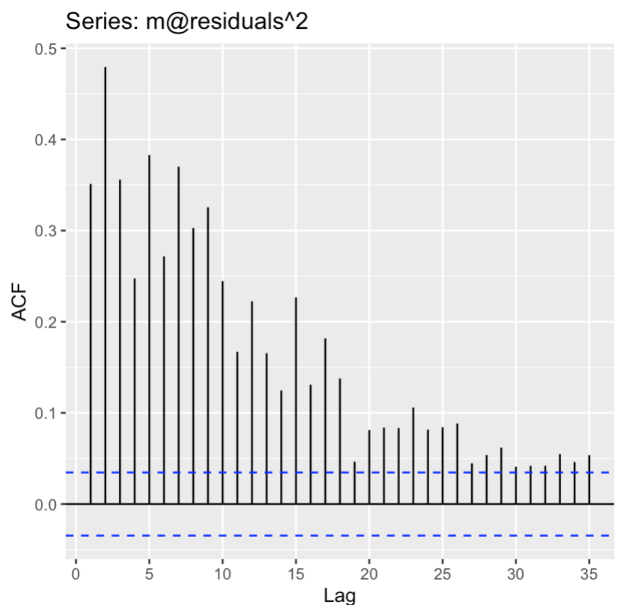


indicates that the tails deviate from the normal line. This indicates that using a different distribution may provide better results. Despite this, I would argue that the residuals are still approximately normal given their shape and lack of skew.



Auto-correlation in the residuals is demonstrated both visually and statistically. Below are the ACF and PACF plots of the residuals, which both display multiple lags with significant auto-correlation. A Box-Ljung test of the null hypothesis that the residuals are independently distributed produced a p-value of less than 0.01, indicating a rejection of the null hypothesis, therefore asserting auto-correlation.





Applying the business cycle functions to the residuals returned two cycles, their respective frequencies were 7.832 and 2.646. These indicate that the model failed to explain some pattern of variance present in the data.

2.3 GARCH Model with Student's t-Distribution

As I described in my analysis of the previous model's residuals, the QQ plot showed that the tails deviated from the normal line and that an alternative distribution may be useful. In this section, I use a GARCH(1,1) model that assumes Student's t-distribution. The general form of this model is:

$$\sigma^2_{t|t-1} = \omega + \beta_1 \sigma^2_{t-1|t-2} + \alpha_1 a^2_{t-1}$$

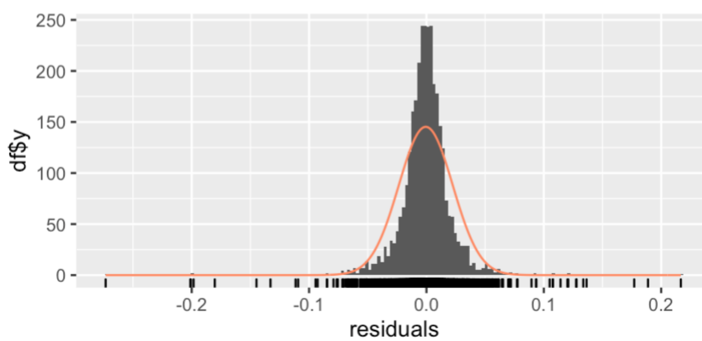
The precise coefficients of this model, as well as their relevant summary statistics, are shown in the model output below. All variables are indicated as statistically significant. The model has two degrees of freedom. The sum of the alpha and beta coefficients is just less than 1, indicating that the model is mean-reverting. The mean of the data is 0.009, indicating that the returns are positively skewed. The standard error of the mean is 0.002, which is sufficiently small relative to 0.009 to indicate significance.

```
        Estimate    Std. Error   t value            Pr(>|t|)
mu      0.000918036 0.000259078    3.543            0.000395 ***
omega   0.000006110 0.000001648    3.708            0.000209 ***
alpha1  0.102350430 0.015032358    6.809     0.00000000000985 ***
beta1   0.888421357 0.015067729   58.962 < 0.0000000000000002 ***
skew    0.996033264 0.024284342   41.015 < 0.0000000000000002 ***
shape   4.866019941 0.417173214   11.664 < 0.0000000000000002 ***
```

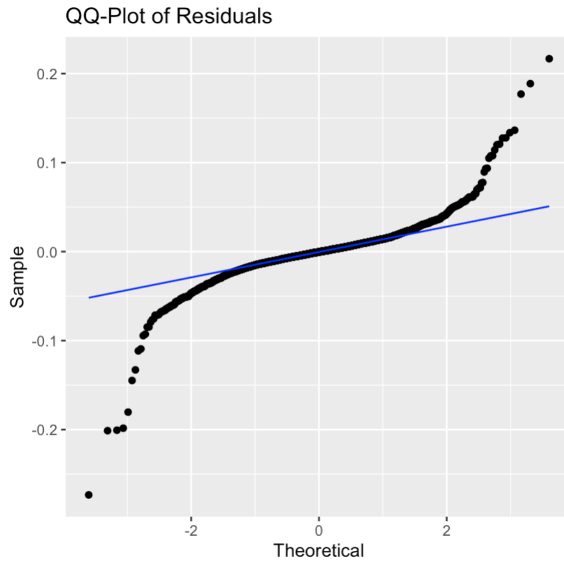
Unsurprisingly, the residuals of this model are quite close to those of the previous model. They are linear-trend stationary and approximately normally distributed, they are not strictly stationary, they do feature auto-correlation, and do have detectable business cycles.

The mean of the residuals was -0.000586, and it had a t-test 95% confidence interval of (-0.0014, 0.0002) and a p-value of 0.1485. Thus we accept the null hypothesis that the residuals' mean is statistically indistinguishable from zero. ADF and KPSS both indicate linear-trend stationarity, as each had a p-value/test statistic beneath the critical value (ADF p-value: 0.01 < 0.05, KPSS test statistic: 0.0737 < 0.146.). However, a McLeod-Li test returned a set with every tested lag, indicating non-constant variance. Therefore, the residuals are linear-trend stationary but not strictly stationary.

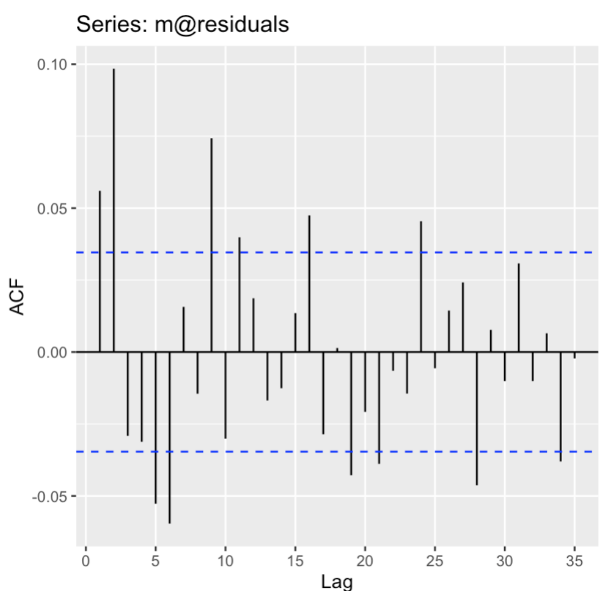
The distribution of this models' residuals appear similar to those of the previous model. Their histogram, shown on the right, is unimodal and with decent skew, although it appears slightly left skewed. The distribution also features high kurtosis. Their respective 95% confidence intervals

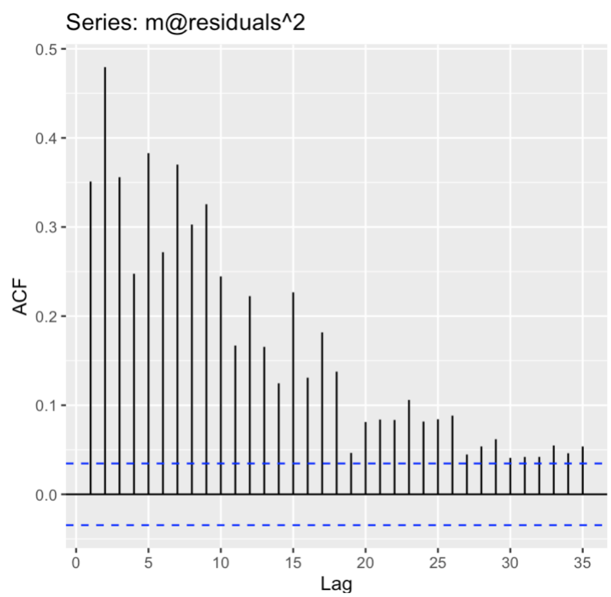


are (-1.91, 0.788) and (12.577, 30.86). These are quite similar to those of the previous model. The QQ plot, shown to the right, indicates that the tails deviate from the normal line. I cannot discern major differences between this and the previous models' plots, indicating that the change from Gaussian to Student's t-distribution may not have made that much of a difference.



Auto-correlation in the residuals is demonstrated both visually and statistically. The ACF and PACF plots of the residuals, shown below, both indicate multiple lags with significant auto-correlation. A Box-Ljung test produced a p-value of less than 0.01, indicating a rejection of the null hypothesis that the residuals are independently distributed.





The residuals of this model feature three business cycles with durations of 10.720, 2.414, and 3.893, indicating variance patterns the model did not account for. This is one more cycle than in the residuals of the previous model, indicating that this model may have failed to explain something that the previous model did.

2.4. GARCH-M Model

In this section, I fit a GARCH-M model to the data. The general form of this model is:

$$\sigma^2_{t|t-1} = \omega + \mu + \gamma(\beta_1 \sigma^2_{t-1|t-2} + \alpha_1 a^2_{t-1})$$

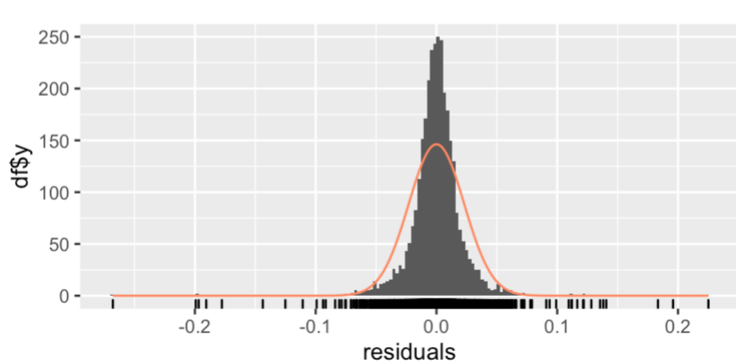
The precise coefficients of the fitted model are shown in the output below. This model has two degrees of freedom. Notably, the mean and gamma constants are not statistically significant. Gamma represents the risk premium, indicating that it is not significant. The sum of the alpha and beta coefficients are under 1, indicating that the model is mean-reverting.

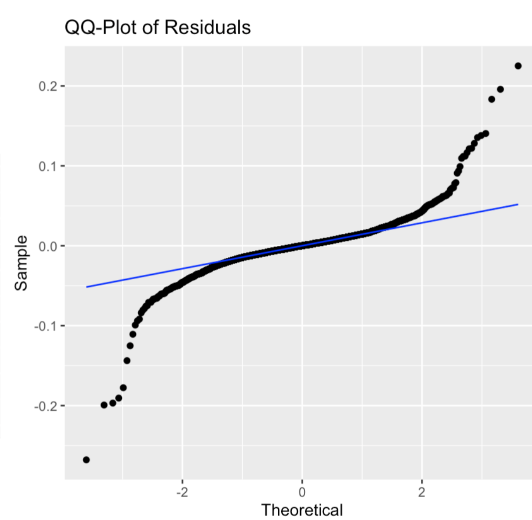
```
Coefficient(s):
            Estimate      Std. Error  t value               Pr(>|t|)
mu     0.00062363074  0.00037797591  1.64992              0.098959 .
gamma -0.59041803338  1.05877532726 -0.55764              0.577089
omega  0.00000814068  0.00000142145  5.72704          0.00000001022 ***
alpha  0.08947643736  0.00987769698  9.05843 < 0.0000000000000000222 ***
beta   0.88989193871  0.01147927281 77.52163 < 0.0000000000000000222 ***
```

The residuals of this model are quite similar to those of the previous two, indicating the model is not a great fit of the data. The residuals are linear-trend stationary and approximately normally distributed, they are not strictly stationary, and they do feature auto-correlation.

The mean of the residuals was 0 with a t-test of 95% confidence interval of (-0.0008, 0.0008) with a p-value of 0.9991. The actual mean, as well as the interval and p-value, demonstrate the residuals are mean zero. Linear-trend stationarity is indicated by ADF and KPSS tests, as each had a p-value/test statistic beneath the critical value required to assert linear-trend stationarity (ADF p-value: 0.01 < 0.05, KPSS test statistic: 0.0558 < 0.146.). A McLeod-Li test of the residuals returned a set with every tested lag, indicating non-constant variance. Thus, the residuals are linear-trend stationary, but not strictly stationary.

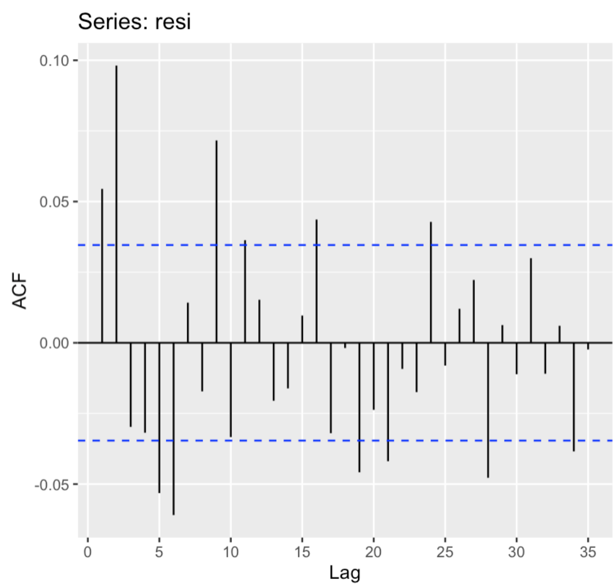
The histogram of the residuals, shown below, depict a distribution that is approximately normally distributed with features similar to those of the previous

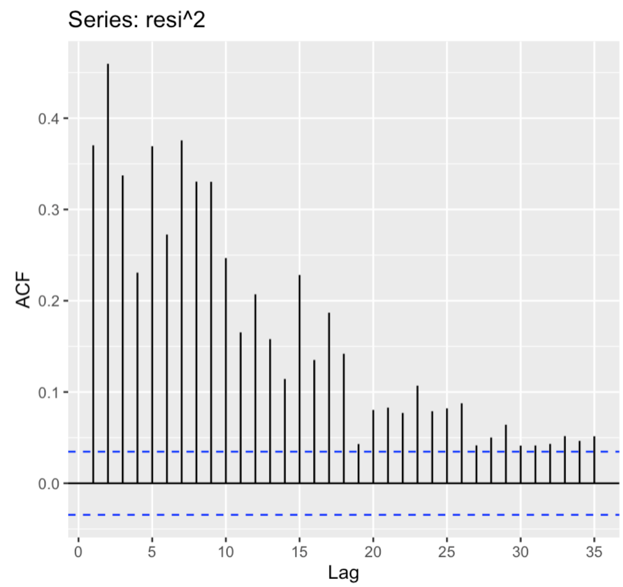




models. The distribution is unimodal with high kurtosis. The 95% confidence for skew (-1.715, 1.089) included zero, however the confidence interval for excess kurtosis (13.086, 30.556) did not. The QQ plot, shown on the previous page, shows how the tails deviate from normal. Again, there is not a clear difference between this QQ plot and those of the previous two models.

Auto-correlation in the residuals is demonstrated both visually and statistically. Below are the ACF plots of the residuals and squared residuals, which both display multiple lags with significant auto-correlation. A Box-Ljung test of the null hypothesis that the residuals are independently distributed produced a p-value of less than 0.01, indicating a rejection of the null hypothesis, therefore asserting auto-correlation.





2.5 TGARCH Model

The time series displays a larger degree of variance in the right side of the time plot. For this reason, it may be worth investigating whether the use of a TGARCH(1,1) model explains this variance well by considering it step wise. The general form of the model I use in this section is:

$$\sigma^2_{t|t-1} = \omega + \beta_1 \sigma^2_{t-1|t-2} + \alpha_1 a^2_{t-1|\gamma t-2}$$

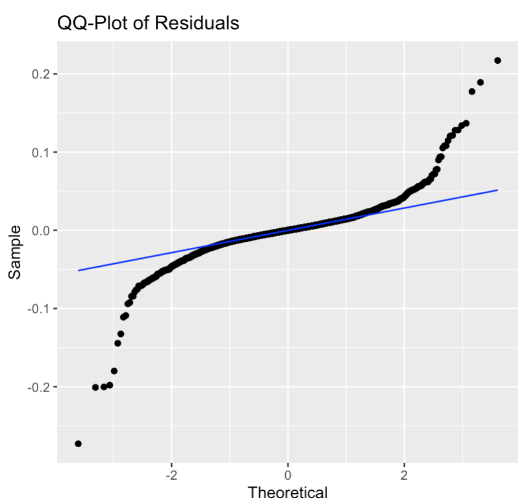
The precise coefficients of the model are shown in the model output on the top of the next page. All coefficients are indicated as statistically significant. The gam1 coefficient, which represents the leverage effect, is statistically significant as shown by its stars as well as its magnitude relative to its standard error. The sum of the alpha and beta coefficients is less than one, indicating that the model is mean-reverting.

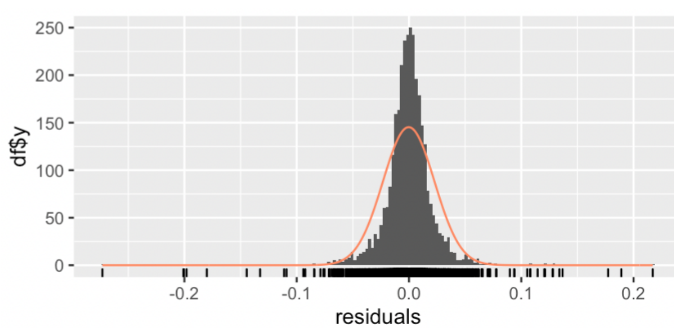
```
Coefficient(s):
           Estimate    Std. Error  t value               Pr(>|t|)
mu     0.00059076123 0.00027607200  2.13988               0.032364 *
omega  0.00000813490 0.00000140287  5.79876        0.0000000066805 ***
alpha  0.04002555449 0.00928341983  4.31151        0.0000162143673 ***
gam1   0.07739340346 0.01356774813  5.70422        0.0000000116878 ***
beta   0.89903105581 0.01217027488 73.87106 < 0.000000000000000222 ***
```

The residuals of this model are similar to those of the previous models. They are linear-trend stationary, approximately normally distributed, are not strictly stationary, and feature auto-correlation.

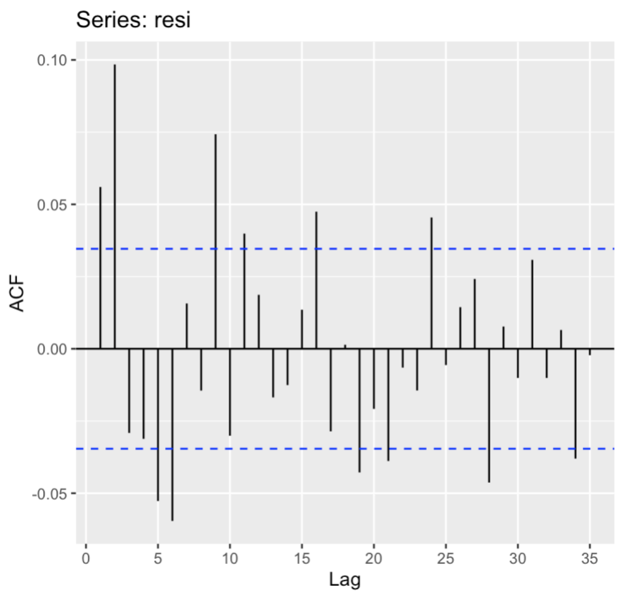
The mean of the residuals was -0.0003 and it had a t-test 95% confidence interval of (-1.91, 0.82) and a p-value of 0.5169. Thus, the null hypothesis that the residuals' mean is statistically indistinguishable from zero is accepted. ADF and KPSS both indicate linear-trend stationarity, as each had a p-value/test statistic beneath the critical value (ADF p-value: $0.01 < 0.05$, KPSS test statistic: $0.0737 < 0.146$.). However, a McLeod-Li test returned a set with every tested lag, indicating non-constant variance. Therefore, the residuals are linear-trend stationary but not strictly stationary.

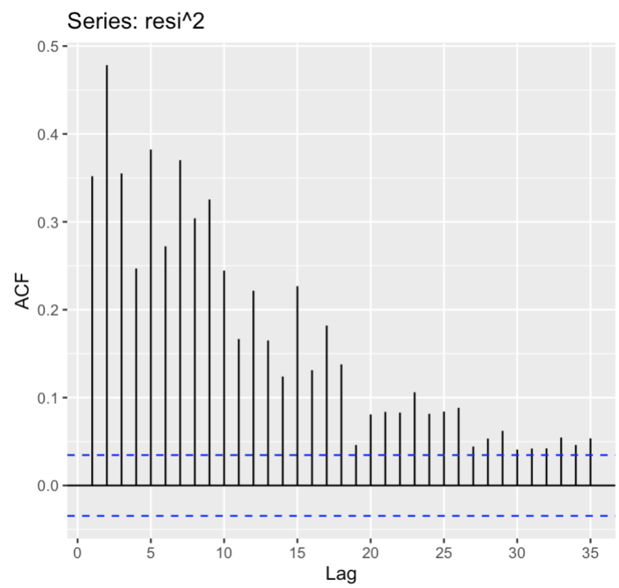
As with the previous models, the residuals of this model are approximately normally distributed with high kurtosis, as shown in the histogram below to the right. The QQ plot on the left also looks quite similar to the previous with similar tail deviation patterns.. The 95% confidence interval for skew (-1.91, 0.82) contains zero, however the confidence interval for excess kurtosis does not (13.141, 30.5). I would argue that given the shape and skew of the distribution, they are approximately normally distributed.





ACF plots for the residuals and squared residuals are shown on the top of the next page. They both show numerous lags with statistically significant auto-correlation. A Box-Ljung test supports this notion, returning a p-value of less than 0.01, indicating a rejection of the null hypothesis that the residuals are independently distributed.





3. Report

Understanding how the price of a stock behaves as a result of its volatility can provide valuable insight. It can enable one to predict future behavior of the stock based on real price data as well as in a simulated environment. I have created several models that analyze this feature of BA's stock by identifying patterns in sequential price movements.

All of the models indicate that BA's price features auto-correlation in its volatility. This means that the level of volatility in price is, in part, dependent on how much volatility there has been recently. My models only tested for the existence of a relationship between the variance of one day and the next, but they show that the relationship exists statistically.

The models also agree that BA's stock is mean-reverting in its volatility, meaning that volatility caused by price shocks decreases after the initial shock. However, the statistical measure that is used to determine how fast this happens indicates that it happens slowly relative to the entire range of possible paces. It is notable that it is also possible for volatility to increase after a shock, but this is not suggested to be the case for BA based on my models.