TS8

Dylan Hayashi

MSDS 413, Fall 2022, Section 55

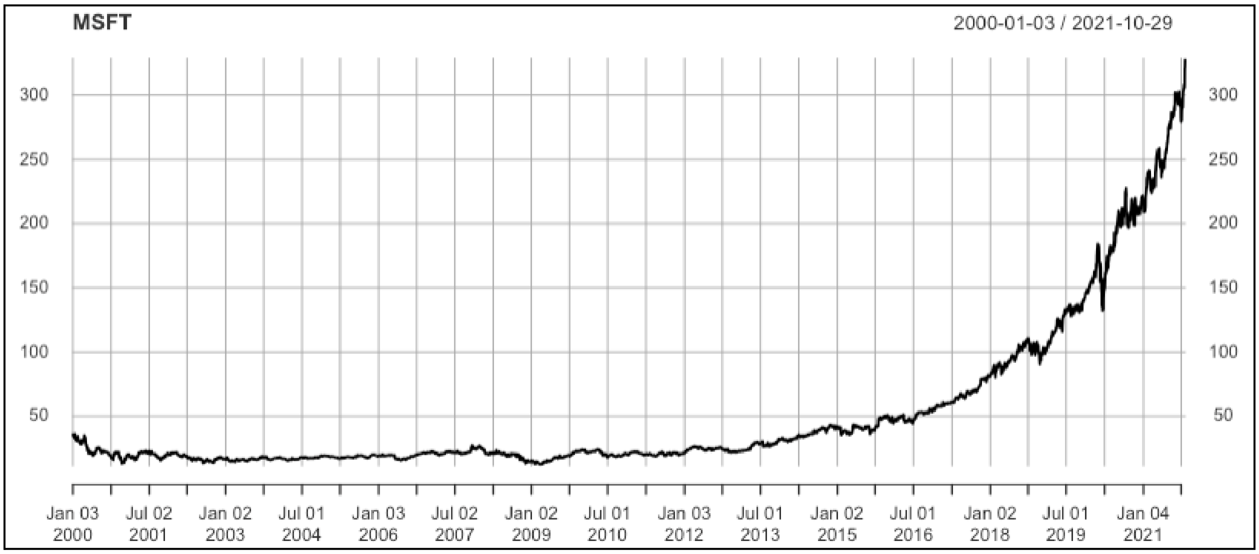
Northwestern University, Time-Series Analysis & Forecasting

November 14, 2022

1. Closing Prices

1.1 EDA

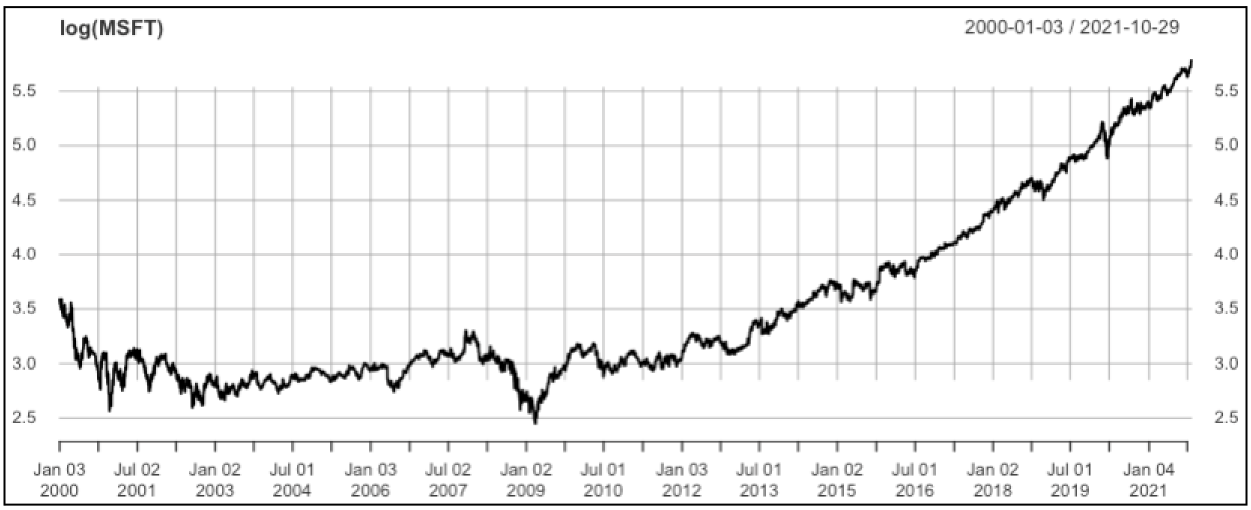
Below are plots of the closing prices of Microsoft and the S&P 500 from 1/3/2000 - 10/29/2021.

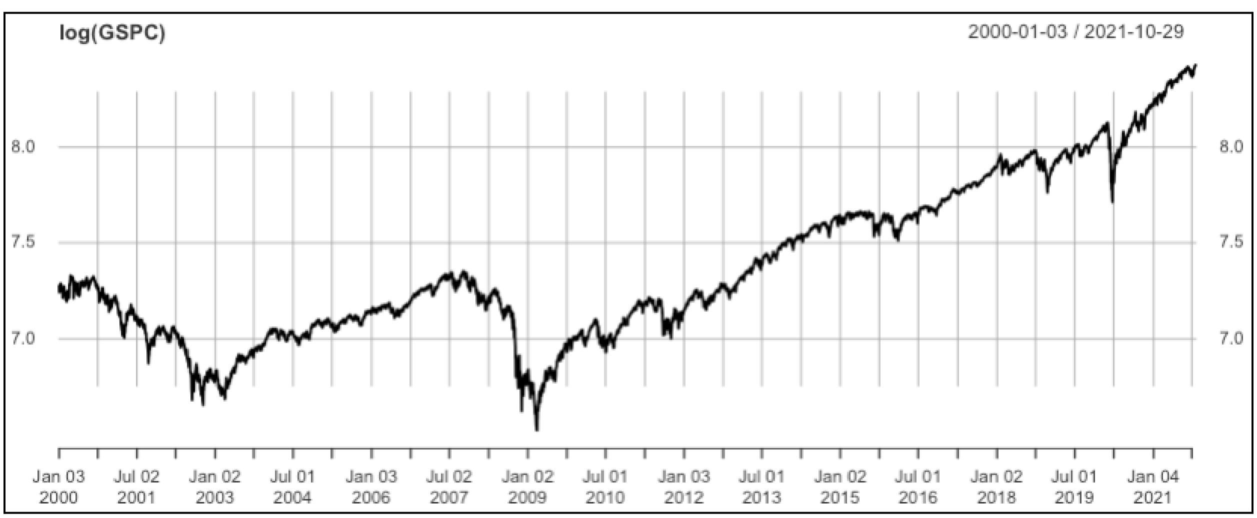


To be considered time series data, the data must be series of observations of stochastic variables over constant time intervals. The variables for prices of MSFT and GSPC have variances of 3454.519 and 675530.413, indicating they are stochastic. The variables are indexed by identical series of closing price timestamps, each of which are unique, as demonstrated by the fact that the vectors MSFT and GSPC closing prices, timestamps, and unique timestamps all have lengths of 5,493. These timestamps are the closing times of trading hours for 5,493 consecutive trading days, indicating they are over constant intervals of trading days. Thus, we have confirmed each part of the definition of time series data for both variables.

The MVN function applied to the data indicates that the variables are not univariate or multivariate normally distributed, as shown in its output on the right. Thus, I will apply a log transformation to the data. This will make models additive. Below are timplots of log price:

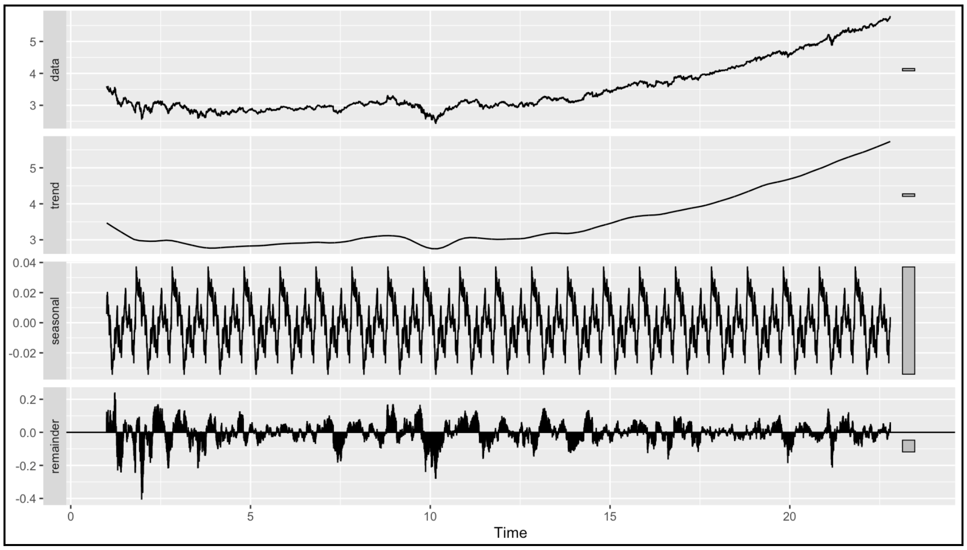
```
$multivariateNormality
             Test        Statistic p value Result
1 Mardia Skewness 6804.19309687317       0     NO
2 Mardia Kurtosis 41.8716823356405       0     NO
3             MVN             <NA>    <NA>     NO

$univariateNormality
              Test      Variable Statistic   p value Normality
1 Anderson-Darling MSFT.Adjusted    819.80  <0.001      NO
2 Anderson-Darling GSPC.Adjusted    287.16  <0.001      NO
```



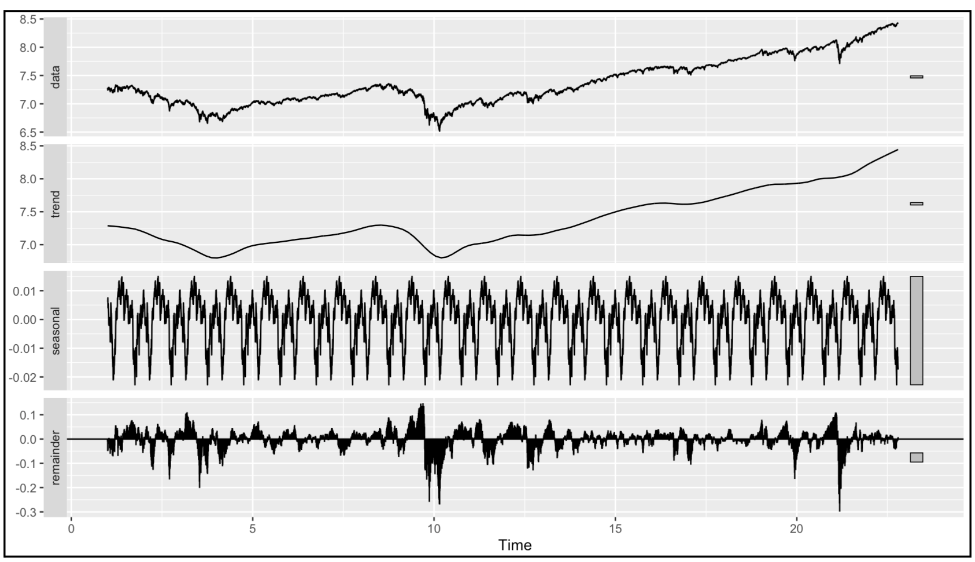


Below are decompositions of the log transformed price data. Both time series are set with frequencies of 252, for 252 trading days per year. The trend lines depict clear upward trends for both variables. The seasonal components of both variables do not show easily discernible, simple seasonal patterns. Rather, they show very complicated patterns that don't take a distinguishable general form when manipulating frequency. Overall, the most notable features of these decompositions are the discernibility of the trend lines.

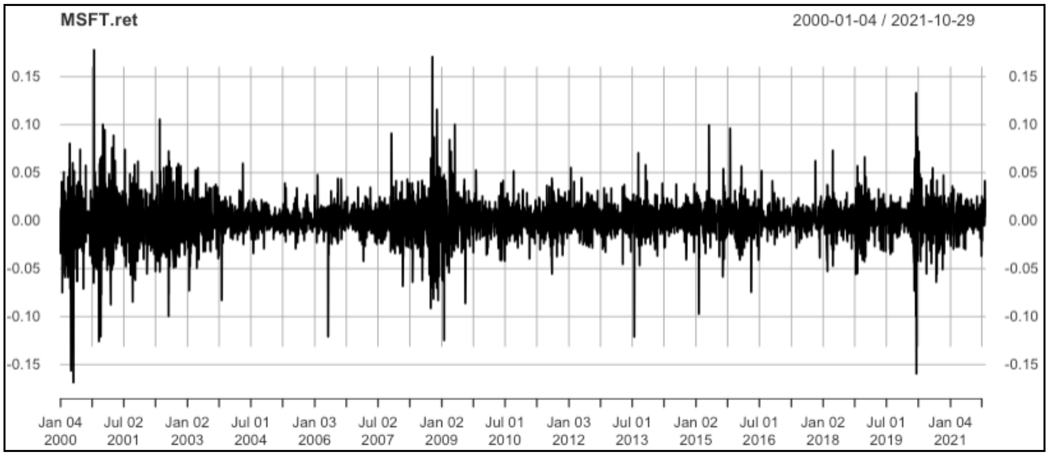
MSFT

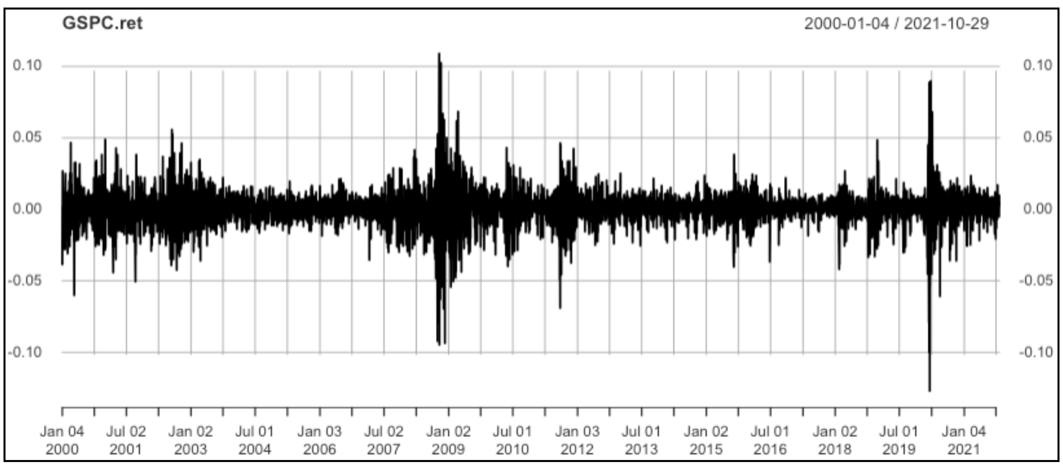


GSPC



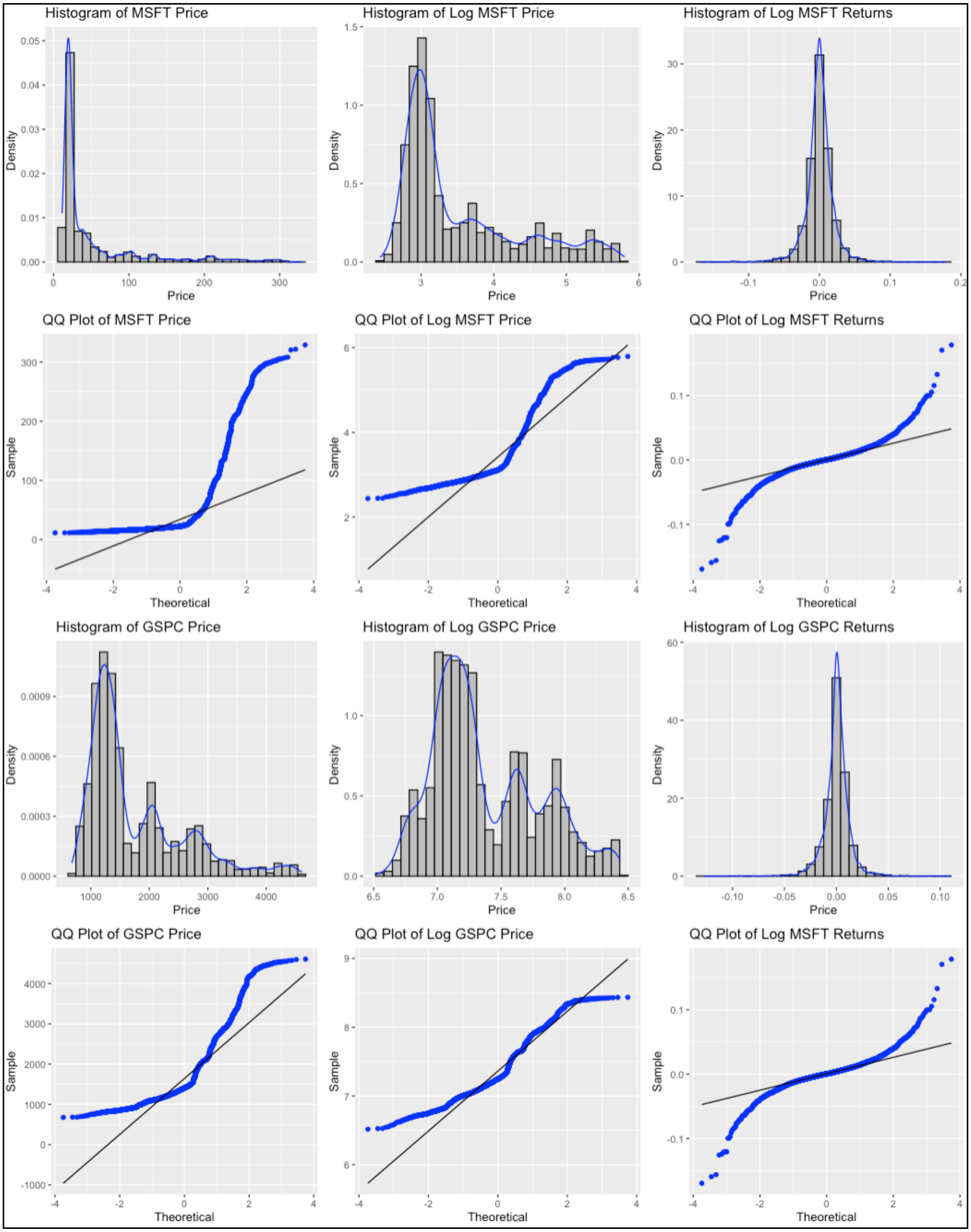
We are interested in the daily return of stocks, which is the difference between price at a time t and t minus one. So, I will difference the data as well. Below are time plots of log returns.





According to the MVN function, the log return variables are not univariate or multivariate normally distributed, as shown in the output on the right. However, as the histograms and QQ on the top of the next page show, the variables' distributions have improved throughout the transformations. Raw and log price data feature high right skew and various degrees of multi-modality, whereas log returns are unimodal with good symmetry. Kurtosis has stayed high throughout.

```
$multivariateNormality
             Test        Statistic p value Result
1 Mardia Skewness 6804.19303964932       0     NO
2 Mardia Kurtosis 41.8716812510657       0     NO
3             MVN             <NA>    <NA>     NO

$univariateNormality
              Test      Variable Statistic   p value Normality
1 Anderson-Darling MSFT.Adjusted    819.80  <0.001      NO
2 Anderson-Darling GSPC.Adjusted    287.16  <0.001      NO
```

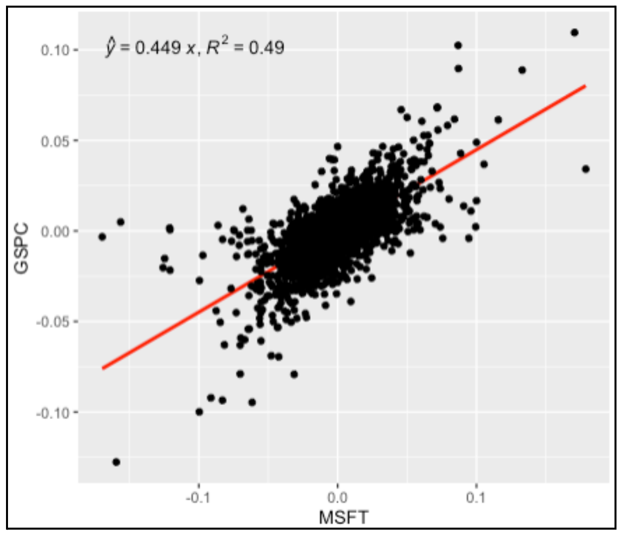


Histogram of MSFT Price
Histogram of Log MSFT Price
Histogram of Log MSFT Returns
Density
Price
QQ Plot of MSFT Price
QQ Plot of Log MSFT Price
QQ Plot of Log MSFT Returns
Sample
Theoretical
Histogram of GSPC Price
Histogram of Log GSPC Price
Histogram of Log GSPC Returns
Density
Price
QQ Plot of GSPC Price
QQ Plot of Log GSPC Price
QQ Plot of Log MSFT Returns
Sample
Theoretical

The time plots of the log returns data appear to be mean zero. T-tests produce 95% confidence intervals that include zero for both variables, as shown in the table below. Linear-trend stationarity is indicated for both variables by ADF and KPSS tests, which return a p-value and test statistic beneath that required to assert linear-trend stationarity in both tests. The values of these tests are also shown in the table. McLeod-Lit tests of both variables returned sets with all tested lags, indicating that the variables feature non-constant variance. Thus, the variables are linear-trend stationary, but are not strictly stationary. MARCH tests return a p-value of 0 for each test, indicating rejections of their null hypotheses, showing that the variables feature multivariate ARCH effects.

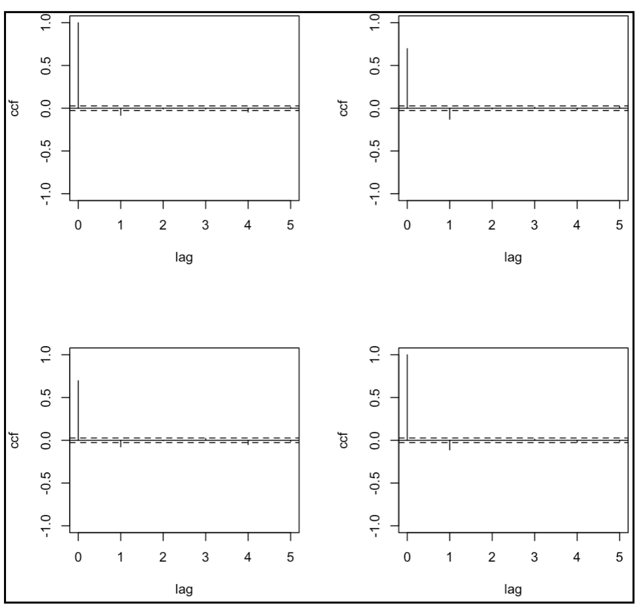
Variable	Mean LL	Mean	Mean UL	ADF p-value	ADF crit val	KPSS t-stat	KPSS crit val
MSFT	-0.0001	0.0004	0.0009	0.0100	0.0500	0.0235	0.1460
GSPC	-0.0001	0.0002	0.0005	0.0100	0.0500	0.0321	0.1460

To the right is a scatter plot of the variables, as well as a regression line of one on another. They show a positive relationship between their returns with a correlation coefficient of 0.697 and R^2 of 0.49.



Below are simplified cross correlation matrices and ACF plots of the univariate and cross correlations ACFS. They indicate significant auto and cross correlation of all variables at lag 1, as well some statistically significant but relatively small correlation at lag 4.

```
CCM at lag:  1
- -
- -
CCM at lag:  2
. .
. .
CCM at lag:  3
. .
. .
CCM at lag:  4
- .
- .
CCM at lag:  5
. .
. .
```



1.2. ARCH(5) Models

ARCH tests against the first five lags of the log return variables produce p-values of 0, indicating that both variables have statistically significant auto-correlation to at least five lags. In this section, I fit ARCH(5) models to each variable. Below are the coefficients of each model. Both models have five degrees of freedom and all of their variables are statistically significant.

MSFT

```
Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 0.00010798  0.00000171     63.2   <2e-16 ***
a1 0.18625448  0.01323140     14.1   <2e-16 ***
a2 0.17190471  0.01293356     13.3   <2e-16 ***
a3 0.12699725  0.00765684     16.6   <2e-16 ***
a4 0.15022732  0.01146464     13.1   <2e-16 ***
a5 0.14315663  0.01227620     11.7   <2e-16 ***
```

GSPC

```
Coefficient(s):
    Estimate  Std. Error  t value Pr(>|t|)
a0 0.00002814  0.00000101     27.9   <2e-16 ***
a1 0.10923519  0.01018547     10.7   <2e-16 ***
a2 0.20595863  0.01576188     13.1   <2e-16 ***
a3 0.18613247  0.01510665     12.3   <2e-16 ***
a4 0.19166281  0.01468747     13.1   <2e-16 ***
a5 0.14270135  0.01320723     10.8   <2e-16 ***
```

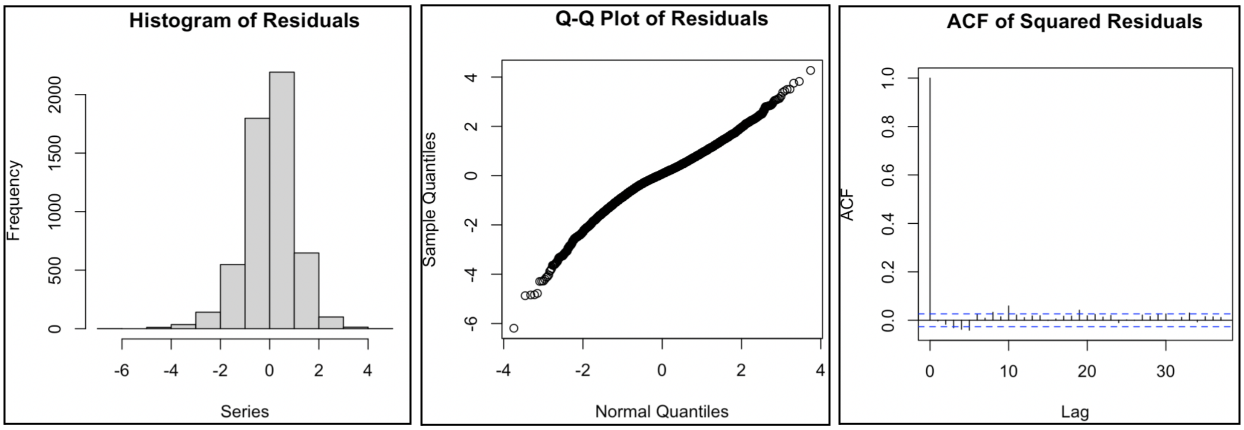
The sum of the models' coefficients of the models are 0.77865 (MSFT) and 0.83572 (GSPC). These values indicate that the models feature persistence. The sum is higher in the GSPC model, indicating that returns volatility is higher in the S&P 500 than MSFT.

On the top of the next page are visualizations of these models' residuals, with the MSFT model on top of the GSPC model. The residuals are quite similar and have generally desirable features. Their histograms depict unimodal distributions with good symmetry and slight left skew. The QQ plots show that the tails do deviate from the normal line (and slightly more so in the GSPC model), but overall, their distributions are approximately normal.

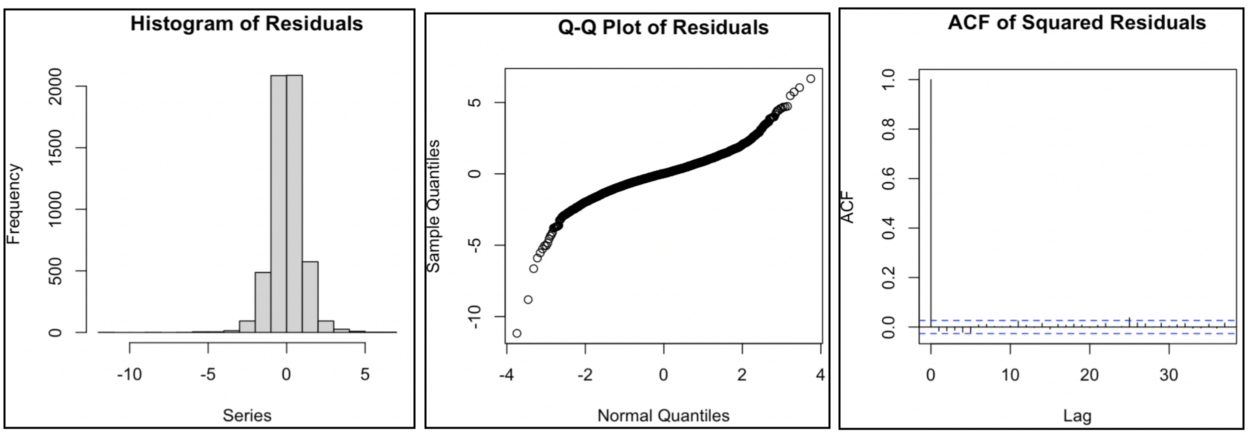
The ACF plots depict that the residuals do not feature significant auto-correlation after the first lag, showing a desirable, quick drop. This is indicative of stationarity, and linear-trend stationarity is suggested for both sets of residuals by ADF and KPSS tests, which all produce p-values/tests-statistics beneath those required to assert stationarity. The values are shown below.

Residuals	ADF p-value	ADF crit val	KPSS test stat	KPSS crit val
MSFT	0.01	0.05	0.0194	0.146
GSPC	0.01	0.05	0.0604	0.146

MSFT



GSPC



1.3. GARCH(1,1) Models

Below are the model coefficients of GARCH(1,1) models fitted to each of the log returns data. The outputs indicate that the variables of both models are all statistically significant. These models have three degrees of freedom. The sum of each models' coefficients are quite close to one, they are 0.98 (MSFT) and 0.984 (GSPC). These sums indicate that the models feature persistence.

	Estimate	Std. Error	t value	Pr(>|t|)	
mu	0.000777026	0.000193789	4.010	6.08e-05	***
omega	0.000007817	0.000001335	5.855	4.76e-09	***
alpha1	0.090663825	0.011232899	8.071	6.66e-16	***
beta1	0.888678163	0.013462270	66.013	< 2e-16	***

	Estimate	Std. Error	t value	Pr(>|t|)	
mu	0.000613434	0.000106124	5.780	7.45e-09	***
omega	0.000002372	0.000000301	7.881	3.33e-15	***
alpha1	0.127042618	0.009845072	12.904	< 2e-16	***
beta1	0.856099943	0.009996907	85.636	< 2e-16	***

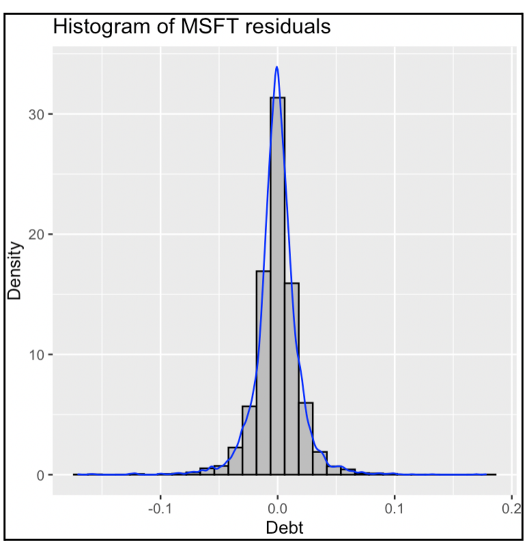
Below are the results of standard residuals tests. The tests indicate that the residuals of these models are similar. Jarque-Bera tests for both sets of residuals return p-values of 0, indicating their alternative hypotheses of non-normality are accepted. Ljung-Box tests show p-values of 0 for the first 10 lags, indicating that the residuals feature auto-correlation. Neither of the model's had ARCH effects present in their residuals, indicated by the p-values of ARCH tests being greater than 0.05. Linear trend-stationarity is suggested for both sets of residuals by ADF and KPSS tests, which all return p-values/test-statistics beneath those required to assert stationarity (these values are shown in the table beneath the tests.)

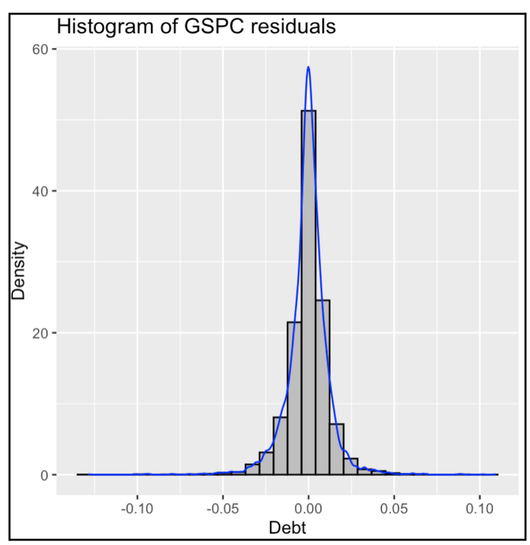
```
Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  18008     0
 Shapiro-Wilk Test  R    W      NA        NA
 Ljung-Box Test     R    Q(10)  20.336    0.026227
 Ljung-Box Test     R    Q(15)  25.799    0.040187
 Ljung-Box Test     R    Q(20)  30.787    0.058059
 Ljung-Box Test     R^2  Q(10)  2.2455    0.99409
 Ljung-Box Test     R^2  Q(15)  4.871     0.99316
 Ljung-Box Test     R^2  Q(20)  5.9731    0.99893
 LM Arch Test       R    TR^2   2.6184    0.99769
```

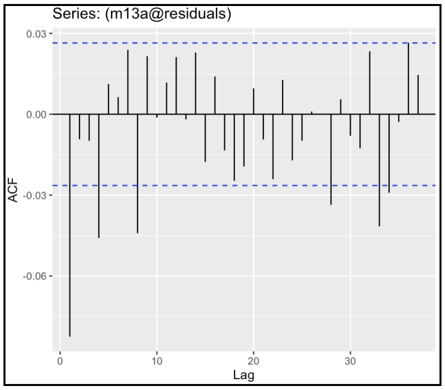
```
Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  964.85    0
 Shapiro-Wilk Test  R    W      NA        NA
 Ljung-Box Test     R    Q(10)  21.88     0.015726
 Ljung-Box Test     R    Q(15)  29.601    0.013446
 Ljung-Box Test     R    Q(20)  37.036    0.011585
 Ljung-Box Test     R^2  Q(10)  15.487    0.1153
 Ljung-Box Test     R^2  Q(15)  18.238    0.25038
 Ljung-Box Test     R^2  Q(20)  19.415    0.49505
 LM Arch Test       R    TR^2   16.339    0.17621
```

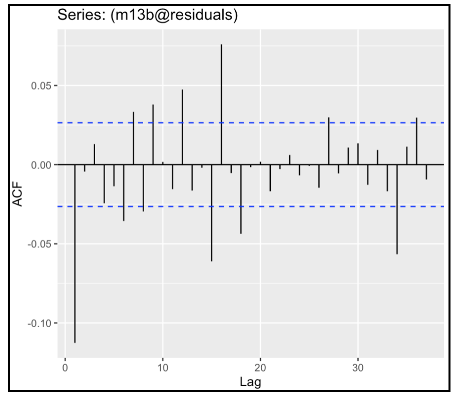
Residuals	ADF p-value	ADF crit val	KPSS test stat	KPSS crit val
MSFT	0.01	0.05	0.0321	0.146
GSPC	0.01	0.05	0.0321	0.146

While the residuals failed the Jarque-Bera test for normality, their histograms (shown below) indicate their distributions are unimodal with good symmetry and low skew. Thus, they are approximately normal. Their ACF plots, shown on the top of the next page, demonstrate the auto-correlation described by the Ljung-Box tests.



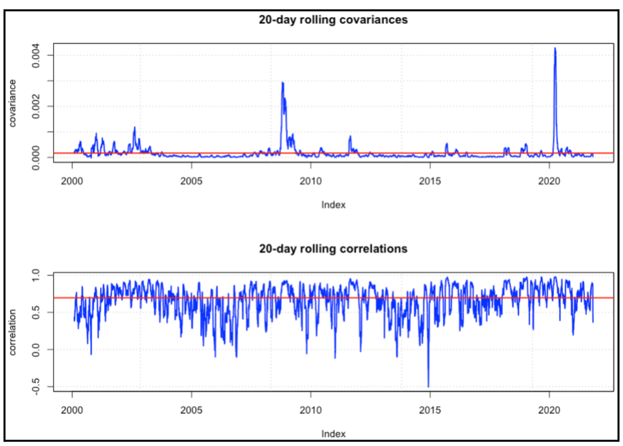


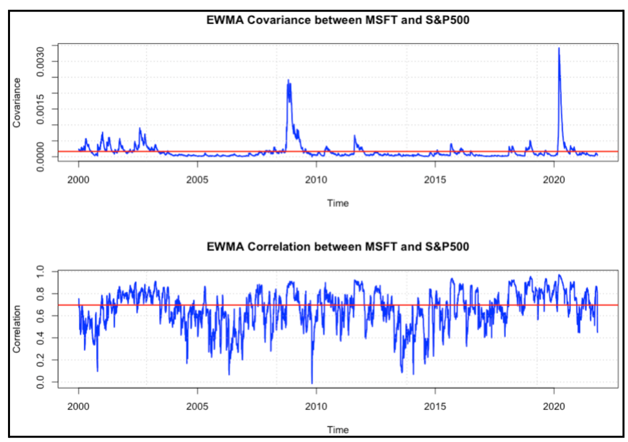


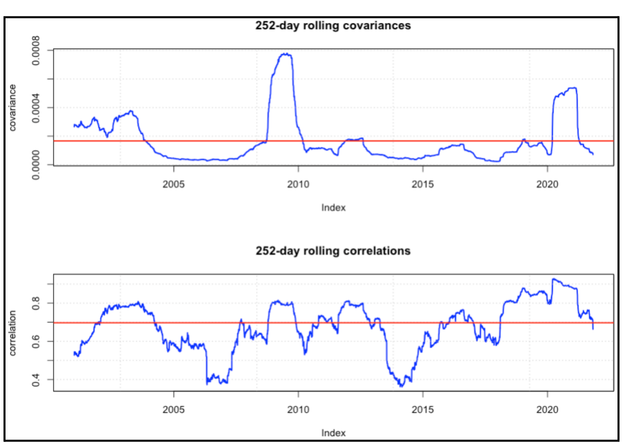


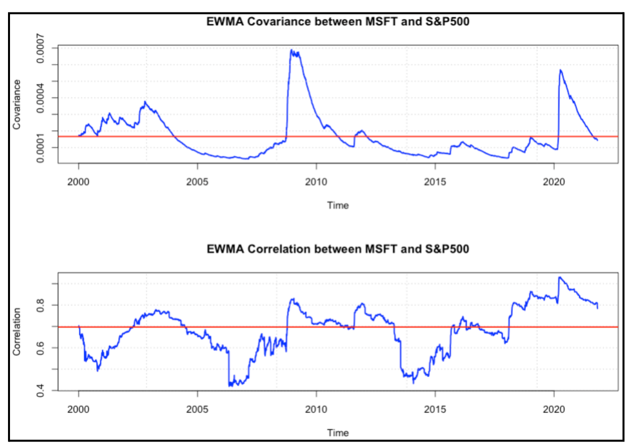
1.4. Rolling Covariances and Correlations

Below are plots of the rolling regular and EWMA covariances and correlations of the variables using two different sliding windows. In all of the plots, the correlation lines fluctuate around a red correlation line at about + 0.7. These demonstrate that MSFT generally leads GSPC. Expanding the sliding window makes it easier to see how the magnitude of the relationship has trended.









1.5. DCC Model

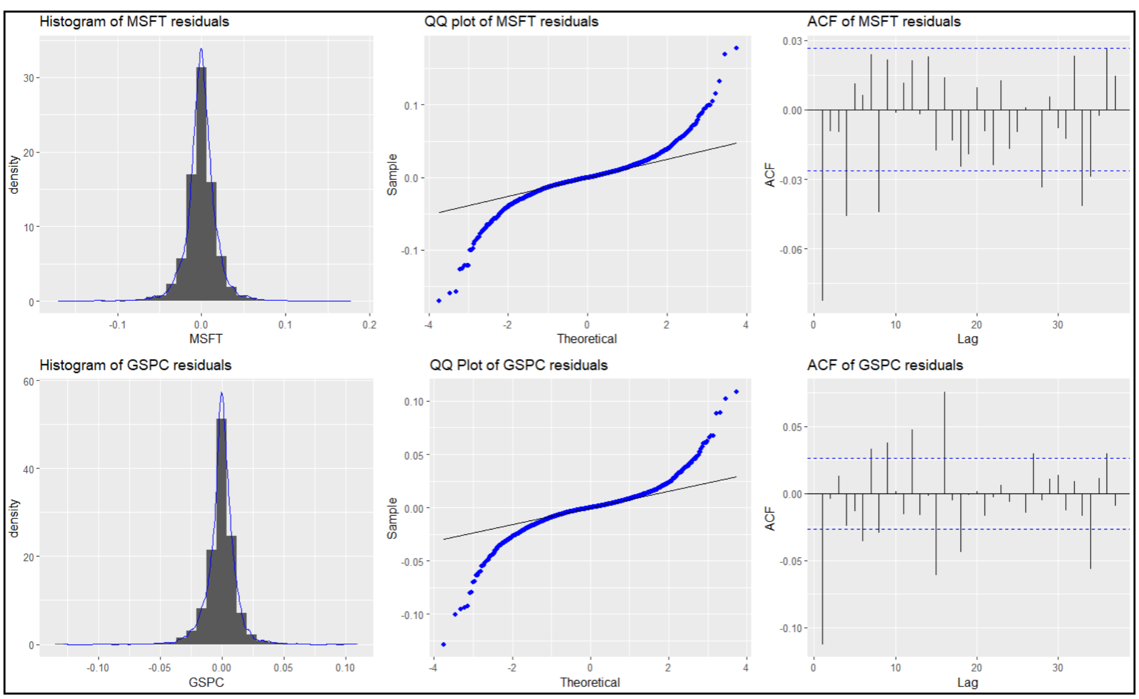
To the right are the coefficients of a DCC(1,1) model fitted to the data. The sums of alpha and beta coefficients for the variables and their joint are all 0.98, quite close to one, indicating the model features persistence and the effect of the variables on one another is very strong.

As shown in its output to the right, the MVN function indicates that this model's residuals are not univariate or multivariate normally distributed. However, their histograms and QQ plots below show that the distributions are approximately normal, with unimodality and good symmetry. The ACF plots show that the residuals do feature auto-correlation, as both have multiple lags with spikes above the significance line. Linear-trend stationarity is suggested for both sets of residuals by ADF and KPSS tests, showing p-values and test statistics sufficiently low to assert stationarity. The values are shown in the table below.

```
$multivariateNormality
             Test         Statistic               p value Result
1 Mardia Skewness 768.585106447845 4.89390250908838e-165     NO
2 Mardia Kurtosis 287.915415859858                     0     NO
3             MVN             <NA>                  <NA>     NO

$univariateNormality
              Test  Variable Statistic   p value Normality
1 Anderson-Darling  Column1      101.6  <0.001      NO
2 Anderson-Darling  Column2      118.5  <0.001      NO
```

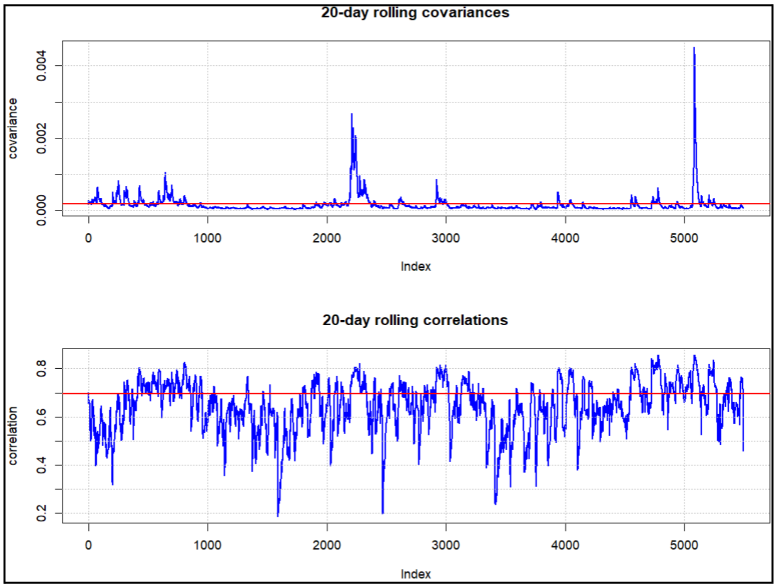
```
Optimal Parameters
------------------------------------
                         Estimate  Std. Error  t value Pr(>|t|)
[MSFT.Adjusted].mu       0.000777    0.000194  4.01377 0.000060
[MSFT.Adjusted].omega    0.000008    0.000007  1.12790 0.259361
[MSFT.Adjusted].alpha1   0.090238    0.016933  5.32923 0.000000
[MSFT.Adjusted].beta1    0.889240    0.029174 30.48070 0.000000
[GSPC.Adjusted].mu       0.000614    0.000108  5.66541 0.000000
[GSPC.Adjusted].omega    0.000002    0.000001  1.81364 0.069733
[GSPC.Adjusted].alpha1   0.126858    0.014539  8.72558 0.000000
[GSPC.Adjusted].beta1    0.856194    0.015650 54.71062 0.000000
[Joint]dcca1             0.038113    0.048117  0.79209 0.428310
[Joint]dccb1             0.937079    0.109313  8.57242 0.000000
```

Residuals	ADF p-value	ADF crit val	KPSS t-stat	KPSS crit val
MSFT	0.010	0.050	0.024	0.146
GSPC	0.010	0.050	0.032	0.146

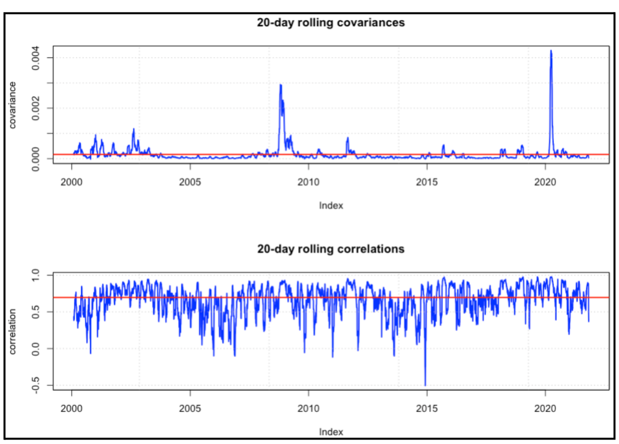


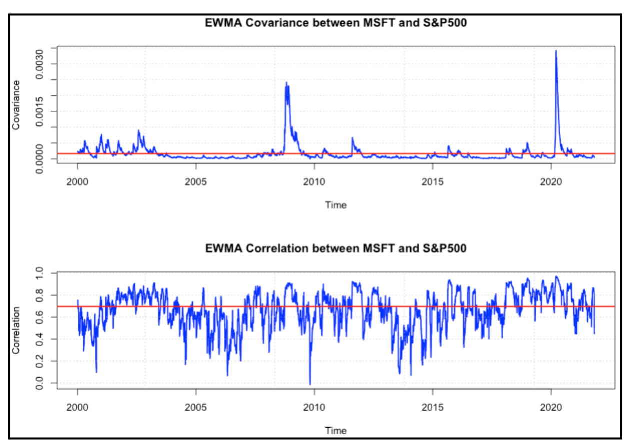
1.6. Conditional Covariances and Correlations

Below are plots of the DCC model's rolling covariances and correlations.



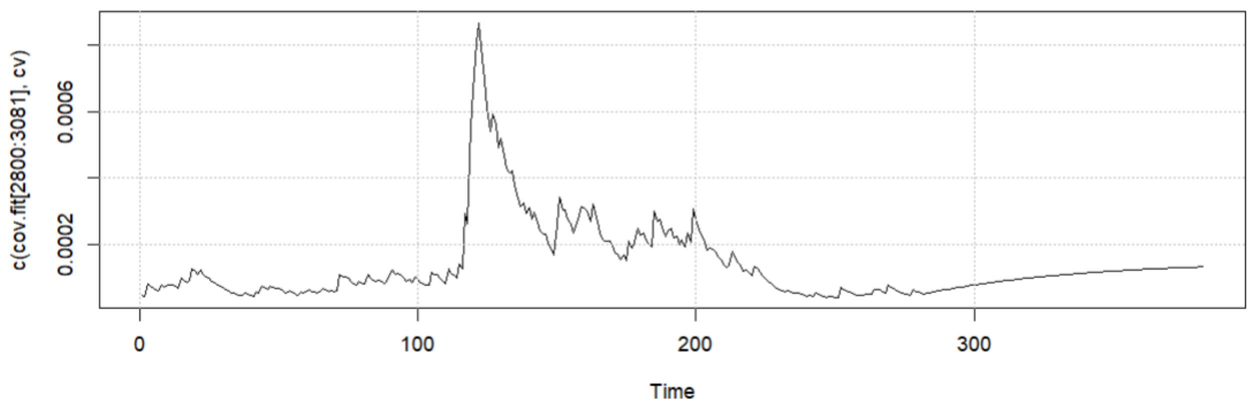
The covariance plot produced by the DCC model are similar to those of the previous regular and EWMA rolling covariance plots, shown below. However, the rolling correlation plot of the DCC model looks significantly more similar to the EWMA correlation plot than the regular. The plots of the DCC and EWMA correlations have extreme values at identical points, notably not where they are in the original plot. It also appears that the correlations in the original plot stay closer to the red line than the others.

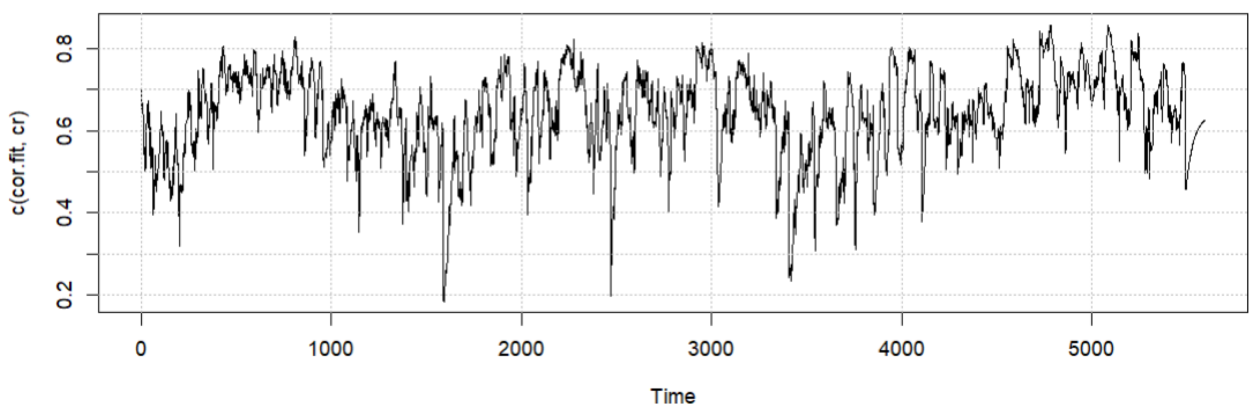




1.7. Forecasts

Below are the DCC(1,1) model's 100-day forecasts for covariance and correlation of MST and GSPC log returns.





1.8. Model Comparison

I prefer the DCC model to the combinations of univariate ARCH and GARCH models. The main reason for this is that DCC allows for the modeling of the variable's effects on one another, which is not possible with univariate ARCH and GARCH models. The sum of the DCC model's coefficients for the variables' joint was very close to 1, indicating that the magnitude of these effects is large. The rolling covariance and correlation plots show that there is an unstable lead / lag relationship between the variables, which is advantageous to be able to model.

2. Report

I have built several models that analyze the relationship between the prices and returns of Microsoft and the S&P 500. Microsoft is itself part of the S&P 500 and is closely economically connected to the index's other constituents. Thus, their prices have strong, positive correlation. It follows that their returns are positively correlated as well.

My models were particularly interested in modeling volatility in returns, as well the relationship between the returns' volatilities. Analysis showed that the volatilities are highly correlated. Volatility in price and returns of Microsoft lead those of the S&P 500, most strongly around times of economic crisis, such as in 2008 and 2020.

Below are my best model's predictions for the covariance and correlation of the returns over the next 100 days. They both suggest an increase in these values over this time. The convexity of the prediction lines indicate possible leveling off or reversal of the slope in the near future. This frequency would seem to fit with the plots' recent behavior. These features and this interpretation of the plots is more easily discernible in the correlation plot due to its rescaling of the covariance values.