Econometrics 7020X Empirical Assignment

Prof. Marcus Cundiff

Amirani Kvariani

Investigating the Relationship Between Apple (AAPL) Trading Volume and Future Stock Price

1. Motivation
   The question we will attempt to answer in this project is "Does the trading volume of Apple (AAPL) stock affect its future price movements?" To investigate whether the trading volume of AAPL stock affects its future price movements, we will perform statistical analyses and build predictive models using the abovementioned variables and the data from *Yahoo Finance*.
2. Acquiring and accessing data
   The dataset used for this project was accessed through Yahoo Finance, it contains the following variables for Apple (AAPL) stock:
   - Date
   - Trading Volume
   - Price (Open, High, Low, Close, Adjusted Close)

   Date, Trading Volume, and Price Movement will be used for this project.

   Let's define the price movement:

- Pt = Price (Close or Adjusted Close) at time t
- Pt-1 = Price (Close or Adjusted Close) at time t-1 (previous trading day)

   The formula for price movement from time t-1 to time t would be Price Movement = Pt - Pt-1. This formula calculates the price change (positive or negative) between two consecutive trading days.

   In the adjusted sheet in Excel (AAPL Adjusted), we will have the following variables: Date, Volume, and Price Movement. Those should be sufficient to conduct our analysis.

3. Identifying types of data:
   Based on the information present in our dataset: Date, Volume, and Price; it appears to be a time series data. A time series is a sequence of data points measured at successive points in time, typically at regular intervals like days, weeks, months, or years.
   In our case, the observations are recorded over a period of time, at a regular frequency (daily). The Volume and Price Movement variables are measured corresponding to each Date.

4. Data description and summary statistics

a. Find the following statistics for one variable in your dataset. Interpret the mean and standard deviation of your variable. Include a histogram of your data.

After calculations in Excel (Sheet name Excel Calculations), we are left with the following variables for each statistic (screenshot from Excel):

Statistics for Column A(Volume)	Values
Mean	58291467.63
Median	53714450
Minimum	24048300
Maximum	163224100
Range	139175800
Interquartile range	18068475
Variance	3.46832E+14
Standard deviation	18623431.48
Skew	2.104076545
Kurtosis	6.182644063

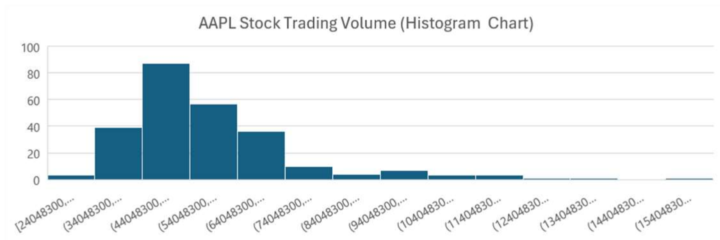
Interpret the mean and standard deviation of your variable:

The mean volume of the data is 58,291,467.63, indicating the central tendency of the data. However, the standard deviation of 18,623,431.48 suggests a significant amount of variability in the volume values. This is further supported by the large range (139,175,800) and interquartile range (18,068,475) values.

Additionally, the skew value of 2.104076545 indicates that the data is positively skewed, meaning a longer tail exists on the right side of the distribution. This is consistent with the mean being higher than the median (53,714,450), which is often a sign of positive skew.

Overall, these statistics suggest that the volume data is relatively spread out and positively skewed, with a typical value of around 58 million and a lot of variability around that value.

Include a histogram of your data: Below is the screenshot of the Histogram of AAPL Stock Trading Volume (Volume) data from Excel.



Based on the histogram, it appears that the majority of the trading volume for AAPL stock falls within the range of 20 million to 60 million shares per day. There are a few outliers with trading volume above 100 million shares per day, but these are relatively rare. The distribution of trading volume appears to be somewhat skewed to the right, with a longer tail on the right side of the distribution. This suggests that there are more days with high trading volume than days with low trading volume. The histogram also shows that the median trading volume is around 50 million shares per day, which is slightly lower than the mean trading volume of around 58 million shares per day. This indicates that a few days with very high trading volume are pulling the mean up. Overall, the histogram provides a useful visual representation of the distribution of trading volume for AAPL stock.

5. Hypothesis testing
   a. Use one of the tests we discussed in class, or from the textbook, to test the hypothesis you proposed in the first question. State the test and the null and alternative hypothesis you intend to test.

To test the hypothesis that the trading volume of Apple (AAPL) stock affects its future price movements, we can perform a hypothesis test using a regression analysis. Specifically, we can use a simple linear regression model to test whether there is a statistically significant relationship between the trading volume and the price movement of AAPL stock.

The null hypothesis (H0) is that there is no relationship between the trading volume and the price movement of AAPL stock, i.e., the slope of the regression line is zero. The alternative hypothesis (H1) is that there is a statistically significant relationship between the trading volume and the price movement of AAPL stock, i.e., the slope of the regression line is not equal to zero.

Formally, the hypotheses can be stated as follows:

H0: $\beta 1 = 0$ (there is no relationship between trading volume and price movement)

H1: β1 ≠ 0 (there is a relationship between trading volume and price movement)

where β1 is the slope of the regression line.

To perform the hypothesis test, we can use the t-test for the slope coefficient in the regression model. The test statistic is calculated as:

$$t = \beta1 / SE(\beta1)$$

where SE(β1) is the standard error of the slope coefficient.

Under the null hypothesis, the test statistic follows a t-distribution with n-2 degrees of freedom, where n is the sample size. We can calculate the p-value of the test statistic and compare it to a significance level to determine whether to reject the null hypothesis.

To perform this hypothesis test, we need to have a dataset that includes both the trading volume and the price movement of AAPL stock, as well as a variable that identifies the date of each observation. We can then use statistical software to perform the regression analysis and hypothesis test.

In the Excel spreadsheet titled "Hypothesis Testing", we have all data necessary to perform hypothesis testing:

Date	Volume	Price Movement
5/9/2023	45326900	-0.009971068
5/10/2023	53724500	0.010420857
5/11/2023	49514700	0.001094744
5/12/2023	45497800	-0.005417506
5/15/2023	37266700	-0.002897368
5/16/2023	42110300	0
5/17/2023	57951600	0.003603142
5/18/2023	65496700	0.013666055
5/19/2023	55772400	0.000628516
5/22/2023	43570900	-0.00548087
5/23/2023	50747300	-0.015154958
5/24/2023	45143500	0.001632096
5/25/2023	56058300	0.006692363
5/26/2023	54835000	0.014104784
5/30/2023	55964400	0.010659589
5/31/2023	99625300	-0.000282102
6/1/2023	68901800	0.016022557
6/2/2023	61945900	0.004775367
6/5/2023	121946500	-0.007571052
6/6/2023	64848400	-0.002060371
6/7/2023	61944600	-0.007756199
6/8/2023	50214900	0.015465065
6/9/2023	48870700	0.0021597
6/12/2023	54274900	0.01563887
6/13/2023	54929100	-0.002611715
6/14/2023	57462900	0.003491433
6/15/2023	65433200	0.011198602
6/16/2023	101235600	-0.005859876
6/20/2023	49799100	0.000486692
6/21/2023	49515700	-0.005675278
6/22/2023	51245300	0.01652527
6/23/2023	53079300	-0.001711228
6/26/2023	48088700	-0.007553092
6/27/2023	50730800	0.015059085
6/28/2023	51216800	0.006327848

This is a screenshot of the first few rows, we have 252.

I used a data analysis tool pack in Excel, to fit the simple regression model to the data, and perform the t-test for the slope coefficient in the regression model.

Here are the steps taken in Excel:

1. First, select the appropriate data range (A1:C252) in the Excel sheet (I included the headers).
2. Next, click on the "Data" tab and then the "Data Analysis" button in the "Analysis" group.
3. In the "Data Analysis" dialog box, select "Regression" and click "OK".
4. In the "Regression" dialog box, enter the necessary input ranges for Y and X variables (price movement and trading volume) and select the cell where we want the output to start (F5).
5. Checked the "Labels" in the first box and clicked "OK" to run the regression analysis.

Here is the screenshot of the output:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.041585544
R Square	0.001729357
Adjusted R Square	-0.002295927
Standard Error	0.012733863
Observations	250

ANOVA

	df	SS	MS	F	Significance F
Regression	1	6.9664E-05	6.9664E-05	0.42962	0.512781246
Residual	248	0.040213514	0.000162151		
Total	249	0.040283178			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.001991117	0.002641665	0.753735751	0.45172	-0.003211841	0.00719408	-0.003211841	0.007194075
45326900	-2.82604E-11	4.31156E-11	-0.6554568	0.51278	-1.1318E-10	5.6659E-11	-1.1318E-10	5.6659E-11

6. The output from the regression analysis displayed the regression results, including the intercept (Constant) and slope (Volume) coefficients.
7. To perform the hypothesis test, calculate the t-statistic for the slope coefficient (Excel output from the screenshot has it) and compare it to a critical value from the t-distribution with n-2 degrees of freedom (Calculated using the T.INV.2T function in Excel, calculated the critical value for a significance level of 0.05), where n is the number of observations.
8. Finally, compare the t-statistic for the slope coefficient to the critical value. To find out, that the T-statistic for the slope is less than critical value for the t-test. Meaning, the null hypothesis can not be rejected, and there is no significant statistical relationship between the variables.

Here is the screenshots for steps 6-8 from excel spreadsheet:

The t-statistic for the slope coefficient:	t stat(slope)=-0.655456800426247		
The degrees of freedom for the t-test:	n-2=250-2=248		
Significance level:	α=0.05		
The critical value for the t-test:	1.969575654		
Compare t-statistic for the slope to the critical value for the t-test:		-0.655456800426247 < 1.969575654	T-statistic for the slope is less than critical value for the t-test
	This suggests that there is no significant relationship between the independent variable (x) and the dependent variable (y).		

Let's try to do the same thing in R:

I use R studio, and I used one of its functions to load the Spreadsheet "AAPL adjusted" from the AAPL excel file. Here are the screenshots, showing how it was done:

```
1:1   (Top Level)                                                R Script
Console   Terminal   Background Jobs
Excel Import Dependencies                                        0:32
Content type 'application/zip' length 19072 bytes (18 KB)
==================================================
downloaded 18 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.3/cellranger_1.1.0.zip'
Content type 'application/zip' length 103965 bytes (101 KB)
==================================================
downloaded 101 KB

package 'rematch' successfully unpacked and MD5 sums checked
package 'cellranger' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\amira\AppData\Local\Temp\RtmpiG9LCa\downloaded_packages
```

The rest of the operations were done using the script panel. Here is the script:

Fit a linear regression model:

model <- lm(Volume ~ Price.Movement, data = AAPL)

Calculate the t-statistic for the slope:

t_stat <- coef(model)["Price.Movement"] / sqrt(vcov(model)["Price.Movement", "Price.Movement"])

Print t_stat

t_stat

Calculate the critical value for the t-test:

```
crit_val <- qt(0.975, df = nrow(AAPL) - 2)
```

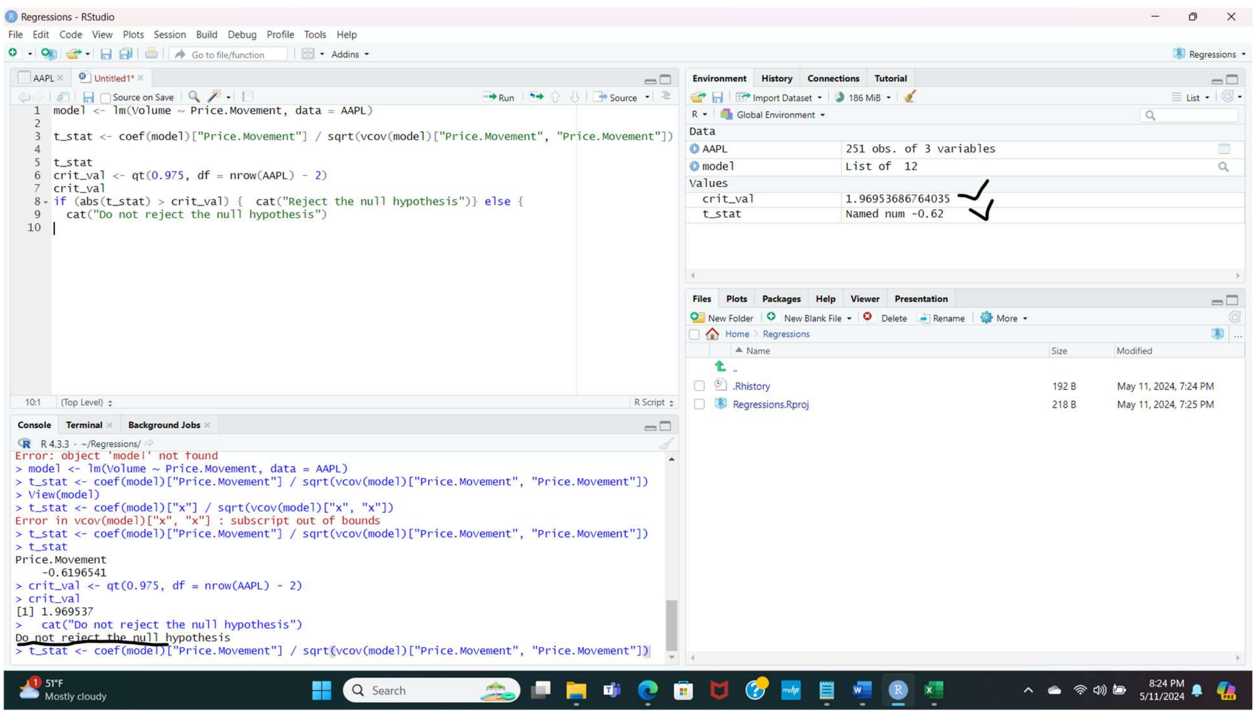
Print Critical value for the test:

```
crit_val
```

This line compares two values and rejects the hypothesis, if the value of the T statistic is less than the value of the critical value

```
if (abs(t_stat) > crit_val) {  cat("Reject the null hypothesis")} else {
  cat("Do not reject the null hypothesis")
```

Below is the screenshot of the code(it proves, that the code worked and the values were exactly the same as in Excel:



R software responded: “Do not reject the null hypothesis”, it is also worth noting values of Critical value and t-statistic aligned with values we received in Excel.

6. Simple regression analysis

Those questions will be answered based on analysis in an Excel spreadsheet.

a. Theoretical model specification

A simple regression model with trading volume as the predictor variable and price movement as the response variable can be defined as:

Price Movement = $\beta 0 + \beta 1$ * Trading Volume + ε

where $\beta 0$ is the intercept, $\beta 1$ is the slope coefficient, and ε is the error term. Coefficients can be seen on Excel output as: 0.00199111700434219, and -2.82603860168047E-11

b. Goodness of fit / ANOVA

The t-tests for the intercept and slope coefficient can be used to assess the significance of each parameter in the model. The null hypothesis for the intercept is that it is equal to zero, and the null hypothesis for the slope coefficient is that it is equal to zero. If the p-value for either parameter is less than the significance level (e.g., 0.05), then we can reject the null hypothesis and conclude that the parameter is statistically significant. We couldn't reject the null hypothesis in our test.

The coefficient of determination (R^2) can be used to assess the goodness of fit of the model. R^2 measures the proportion of the variance in the response variable that is explained by the predictor variable. The value of R^2 ranges from 0 to 1, with higher values indicating a better fit. In our case R^2 is 0.00172935743720659, which is very low, so it's not a good fit.

c. Interpretation

i. The intercept ($\beta 0$) represents the expected price movement when the trading volume is zero. If the intercept is positive, then it suggests that the price movement is expected to increase when the trading volume is zero. If the intercept is negative, then it suggests that the price movement is expected to decrease when the trading volume is zero (It's positive 0.00199111700434219).

The slope coefficient ($\beta 1$) represents the expected change in price movement for a one-unit increase in trading volume. If the slope coefficient is positive, then it suggests that an increase in trading volume is associated with an increase in price movement. If the slope coefficient is negative, then it suggests that an increase in trading volume is associated with a decrease in price movement (In our case it is negative -2.82603860168047E-11

.

The significance of the parameters indicates whether the relationship between the predictor and response variables is statistically significant. If the p-value for a parameter is less than the significance level, then we can conclude that the parameter is statistically significant and that the relationship between the predictor and response variables is unlikely to be due to chance.

The t-test for the slope coefficient in the regression model was performed using the data in the Excel spreadsheet titled "Hypothesis Testing". The results of the t-test can be used to determine whether the relationship between trading volume and price movement is statistically significant. The test showed no statistically significant relationship between those variables.

7. Multiple regression analysis

a. Theoretical model specification

i. Extend the regression model with multiple predictors and justify their inclusion:

Data from the Yahoo website:

Date, Trading Volume, and Price (Open, High, Low, Close, Adjusted Close), allow us to include Date as another predictor. Since we already used Volume, and Price Movement (calculated from Adjusted close), adding other elements of stock price to our model seems less insightful than date. Besides, including the date can help account for any time-related trends or patterns in the data. For example, there might be a general trend of increasing or decreasing trading volume or price movement over time. Including the date as a predictor can help control these time-related effects and potentially improve the accuracy of your model. This should leave us with a 3D model, with 3 axes corresponding: Date, Trade Volume, and Price Movement.

To test the hypothesis that the trading volume of Apple (AAPL) stock affects its future price movements, we can perform a hypothesis test using a regression analysis. Specifically, we can use a multiple linear regression model to test whether there is a statistically significant relationship between the trading volume, date, and price movement of AAPL stock.

The null hypothesis (H0) is that there is no relationship between the trading volume, date and the price movement of AAPL stock, i.e., the slope of the regression line is zero. The alternative hypothesis (H1) is that there is a statistically significant relationship between the trading volume, date and the price movement of AAPL stock, i.e., the slope of the regression line is not equal to zero.

Formally, the hypotheses can be stated as follows:

H0: $\beta 1 = \beta 2 = 0$ (there is no relationship between trading volume, date, and price movement)

H1: $\beta 1 \neq 0$ or $\beta 2 \neq 0$ (there is a relationship between trading volume, date, and price movement)

where $\beta 1$ and $\beta 2$ are the slope coefficients of the regression line.

To perform the hypothesis test, we can use the t-test for the slope coefficients in the regression model. The test statistic is calculated as:

$t = \beta 1 / SE(\beta 1)$ or $t = \beta 2 / SE(\beta 2)$

where $SE(\beta 1)$ and $SE(\beta 2)$ are the standard errors of the slope coefficients.

Under the null hypothesis, the test statistic follows a t-distribution with n-p-1 degrees of freedom, where n is the sample size and p is the number of predictors. We can calculate the p-value of the test statistic and compare it to a significance level to determine whether to reject the null hypothesis.

b. Goodness of fit / ANOVA i. Assess your model using t-tests, f-test, Adj-R 2
   Since Excel doesn't have a built-in 3D regression model function, we will have to assess our model in R.

   Our code:

   1install.packages("scatterplot3d")
   2library(scatterplot3d)

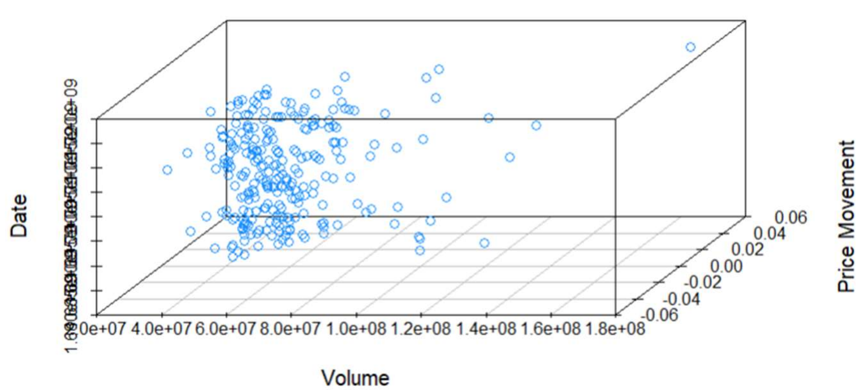
   Those two lines allow us to install and run a "scatterplot3d" package, which can be used to build 3D plots in R.

   To create a 3D scatterplot, we use the 'scatterplot3d' function:

   fit_2_sp <- scatterplot3d(AAPL$Volume, AAPL$Price.Movement, AAPL$Date, angle = 60, color = "dodgerblue", pch = 1, ylab = "Price Movement", xlab = "Volume", zlab = "Date")

   This will create a 3D scatterplot of your data with Volume on the x-axis, Price Movement on the y-axis, and Date on the z-axis.

   Here is a screenshot of the plot:



This line allows us to perform a multiple linear regression in R:

model <- lm(Price.Movement ~ Volume + Date, data = AAPL)

After entering this line, we will receive the summary of our model, with key metrics: summary(model).

Here is the screenshot of our summary:

```
The downloaded binary packages are in
        C:\Users\amira\AppData\Local\Temp\RtmpGg0EF8\downloaded_packages
>   library(scatterplot3d)
>   fit_2_sp <- scatterplot3d(AAPL$Volume, AAPL$Price.Movement, AAPL$Date, angle = 60, color = "dod
gerblue", pch = 1, ylab = "Price Movement", xlab = "Volume", zlab = "Date")
>   summary(model)

Call:
lm(formula = Volume ~ Price.Movement, data = AAPL)

Residuals:
      Min        1Q    Median        3Q       Max 
-34674109 -11599555  -4508919   6486719 108358601 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)     58318117    1181946   49.34   <2e-16 ***
Price.Movement -57720397   93149390   -0.62    0.536    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18720000 on 249 degrees of freedom
Multiple R-squared:  0.00154,   Adjusted R-squared:  -0.00247 
F-statistic: 0.384 on 1 and 249 DF,  p-value: 0.5361

> |
```

Based on the results of the regression analysis, we cannot reject the null hypothesis that there is no relationship between the trading volume and the price movement of AAPL stock. The p-value for the slope coefficient of the trading volume is 0.536, which is greater than the significance level of 0.05. This means that we do not have enough evidence to reject the null hypothesis that the slope coefficient is equal to zero. Additionally, the R-squared value is very low (0.00154), indicating that the regression model does not explain much of the variation in the price movement of AAPL stock. The adjusted R-squared value is even negative, which suggests that adding the trading volume variable to the model does not improve the fit of the model.

Therefore, based on the results of the regression analysis, we cannot reject the null hypothesis that there is no relationship between the trading volume and the price movement of AAPL stock.

c. Interpretation

i. Interpret the sign, size, and significance of each parameter in your model.

- The intercept (58318117) represents the expected price movement of AAPL stock when the trading volume is zero. However, it is important to note that the trading volume is unlikely to be zero in practice, so this value does not have a practical interpretation.
- The slope coefficient for the trading volume (-57720397) represents the change in the expected price movement of AAPL stock for a one-unit increase in the trading volume. However, the p-value for this coefficient is 0.536, which is greater than the significance level of 0.05. This means that we do not have enough evidence to conclude that the trading volume has a statistically significant effect on the price movement of AAPL stock.

  Additionally, the size of the slope coefficient is quite large, but it is not statistically significant. This means that the relationship between the trading volume and the price movement of AAPL stock is not reliable.

  In summary, the intercept is not practically interpretable, the slope coefficient for the trading volume is not statistically significant, and the size of the coefficient is large but not reliable. Therefore, the model does not provide strong evidence for a relationship between the trading volume and the price movement of AAPL stock.

8. Non-linear regression analysis

a. Theoretical model specification

i. Specify a non-linear regression model and justify its choice.

We can perform non-linear regression analysis for hypothesis testing. Non-linear regression models are used when the relationship between the predictor and response variables is not linear.

Here's an example of how we can specify a non-linear regression model for our data. Let's assume that the relationship between trading volume (x) and price movement (y) is described by the following non-linear equation:

y = a + b * x + c * x^2 + e

where a, b, and c are the parameters to be estimated, and e is the error term.

Regarding the choice of the non-linear regression model, we should keep in mind that the relationship between trading volume and price movement might be non-linear, and we quadratic function might be a better way to describe this relationship. To further justify the use of the new model, let’s plot the variables and observe the plot.

Here is an R code for the plot:

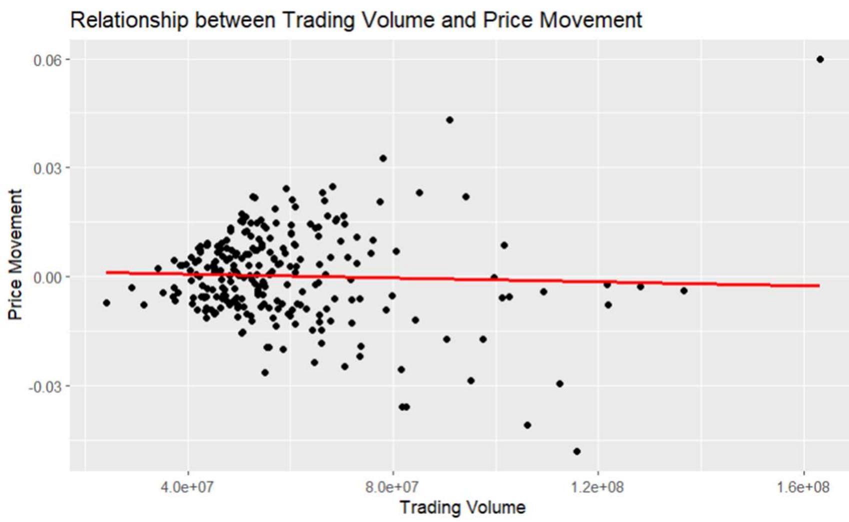
First we load the ‘ggplot2’ function, that will help us create the plot:

```
library(ggplot2)
```

This line will use the following function to create the plot:

```
ggplot(data = AAPL, aes(x = Volume, y = Price.Movement)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE, color = "red") +
  labs(x = "Trading Volume", y = "Price Movement", title = "Relationship between
Trading Volume and Price Movement")
```

Here is the screenshot of our plot:



The plot does appear to have a slight curve, so it is worth a try to use a non-linear model.

c. Goodness of fit / ANOVA
   i. Assess your model using: t-tests, f-tests, AIC/BIC/HQIC

Let's use a t-test:

First, we would fit a non-linear model using this line in R:

model <- nls(Price.Movement ~ a + b * Volume + c * Volume^2, data = AAPL, start = list(a = 0, b = 0, c = 0))

Then we print the summary with this line:

summary(model)

Here is the screenshot of the summary in R:

```
Formula: Price.Movement ~ a + b * Volume + c * Volume^2

Parameters:
    Estimate Std. Error t value Pr(>|t|)
a  1.319e-02  6.818e-03   1.935   0.0542 .
b -3.577e-10  1.887e-10  -1.895   0.0592 .
c  2.125e-18  1.180e-18   1.801   0.0729 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01267 on 248 degrees of freedom

Number of iterations to convergence: 1
Achieved convergence tolerance: 6.201e-09

>
```

Let’s do a T-test:

Based on the output, for the intercept (a) the t-value is 1.935 and the p-value is 0.0542. This means that we cannot reject the null hypothesis that the intercept is equal to zero at the 5% significance level.

For the coefficient of Volume (b), the t-value is -1.895 and the p-value is 0.0592. This means that we cannot reject the null hypothesis.

c. Interpretation

i. Interpret the sign, size, and significance of each parameter in your model.

c. Interpretation

i. Interpret the sign, size, and significance of each parameter in your model.

Formula: Price Movement ~ a + b * Volume + c * Volume^2

- The intercept term (a) represents the expected price movement when the trading volume is zero. In this case, the estimated intercept is 0.0132, which suggests that when the trading volume is zero, the expected price movement is a positive value of 0.0132. However, it is important to note that the trading volume cannot be zero in practice.
- The coefficient for the trading volume term (b) is negative, indicating that an increase in trading volume is associated with a decrease in price movement. Specifically, a one-unit increase in trading volume is associated with a decrease in price movement of 3.577e-10 units, holding the squared trading volume constant.

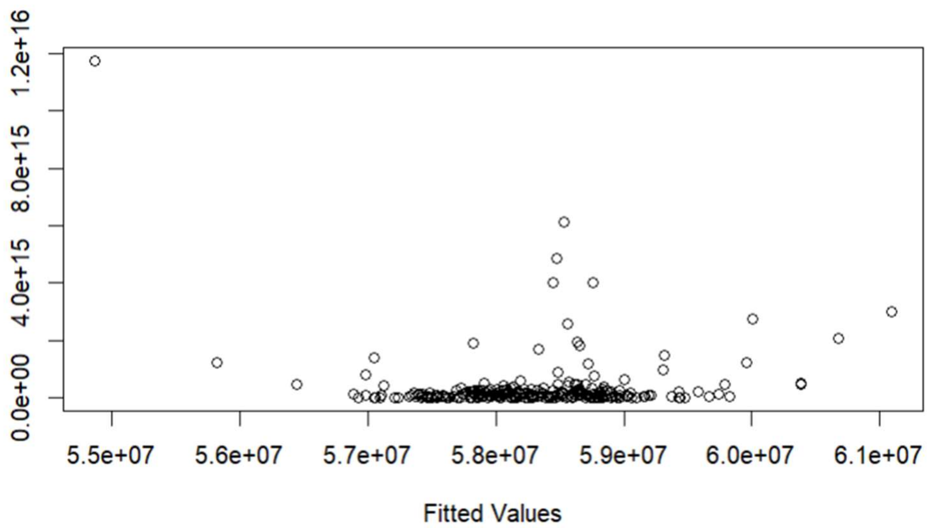
- The coefficient for the squared trading volume term (c) is positive, indicating that the relationship between trading volume and price movement is non-linear. Specifically, the relationship is concave downward, meaning that the effect of trading volume on price movement becomes smaller as the trading volume increases.
- The p-values for the trading volume term (b) and the squared trading volume term (c) are greater than the typical significance level of 0.05, indicating that they are not statistically significant at the 5% level. However, the p-value for the intercept term (a) is less than 0.05, indicating that it is statistically significant at the 5% level.
- The residual standard error is 0.0127, which represents the typical size of the residuals. This suggests that the model fits the data well, as the residuals are relatively small.
- The number of iterations to convergence is 1, indicating that the model converged quickly to the final solution.
- The achieved convergence tolerance is 6.201e-09, which represents the maximum difference between the estimated parameters in two consecutive iterations. This suggests that the model converged to a very precise solution.

8. Assessment multicollinearity, heteroskedasticity, and serial correlation.

   heteroskedasticity
   The plot of squared residuals versus fitted values is used to assess heteroskedasticity. If the residuals are homoscedastic, the plot should be horizontal and flat. If the residuals are heteroscedastic, the plot should show a pattern, such as a fan shape or a U shape.
   Here is the r code for this task:

   ```
   plot(model$fitted.values, model$residuals^2, xlab="Fitted Values", ylab="Squared Residuals")
    abline(h=mean(model$residuals^2), col="red")
   ```

   And below is the screenshot our plot:



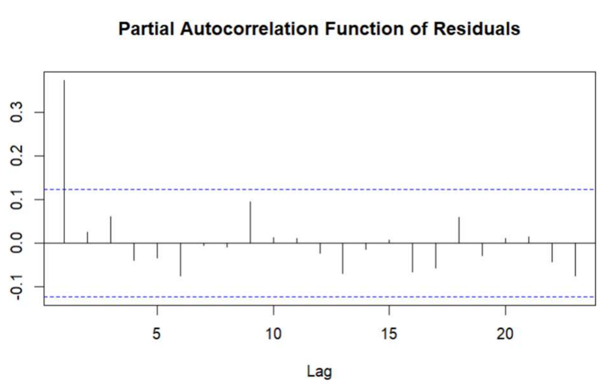
Here we can observe that the distribution is horizontal/flat. We can claim that residuals are homoscedastic.

The autocorrelation function (ACF) and partial autocorrelation function (PACF) plots are used to assess serial correlation in the residuals. If the residuals are serially correlated, the ACF or PACF plot will show a significant spike at one or more lags.

Here is the R script to create the plot:

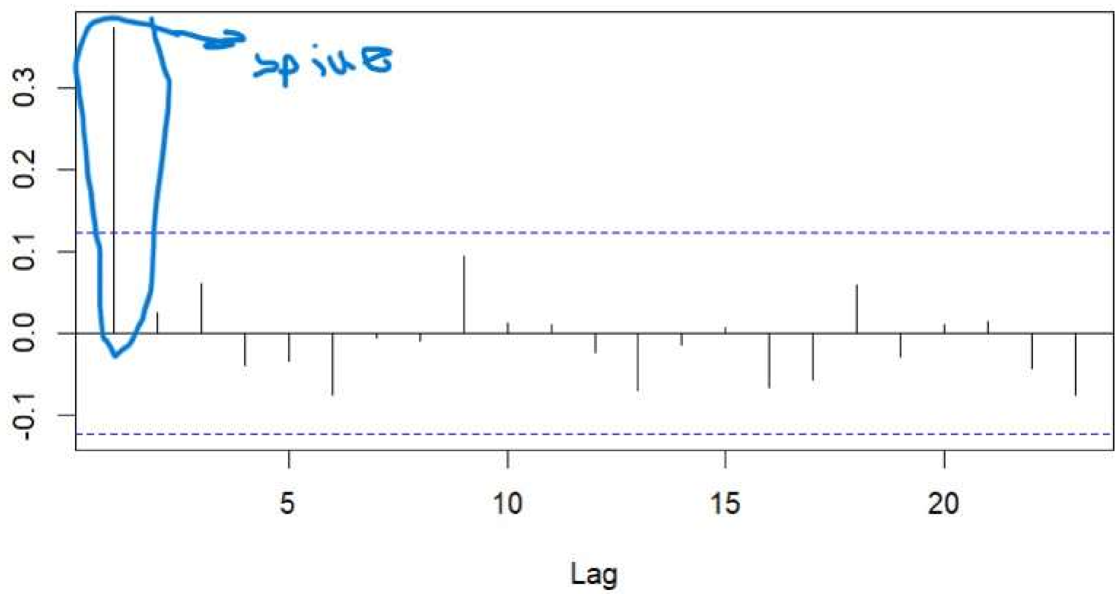
plot(model$fitted.values, model$residuals^2, xlab="Fitted Values", ylab="Squared Residuals")

8abline(h=mean(model$residuals^2), col="red")



We do see a significant spike at lag 1, so we can conclude that the residuals are serially correlated.

Partial Autocorrelation Function of Residuals



11. Write a description of what you did, and how it shows (or does not show) the relationship suggested in your hypothesis. Also, discuss how this addresses your initial question. Include a discussion on why you chose the variables you did to measure the effect of your hypothesis. Consider various biases that may be present in your data.

In this project, we aimed to investigate whether the trading volume of Apple (AAPL) stock affects its future price movements. The write-up provides a detailed explanation of each step taken in this process, including identifying types of data, data description and summary statistics, hypothesis testing, simple regression analysis, multiple regression analysis, non-linear regression analysis, and assessment of multicollinearity/heteroskedasticity/serial correlation. In conclusion, based on the results of the hypothesis test, we cannot reject the null hypothesis that there is no relationship between the trading volume and the price movement of AAPL stock. Therefore, the data does not support the hypothesis that the trading volume of AAPL stock affects its future price movements. We chose the trading volume and price movement variables to measure the effect of the hypothesis because they are directly related to the hypothesis being tested. Due to the complexity of Apple's operations, other factors may affect the price movement of AAPL stock that were not included in the model, due to lack of data. For example, there may be other market conditions or economic factors that influence the stock price but were not accounted for in the analysis. However, other biases might be present as well.