Project: Probability and Statistics

Log-Returns for Automotive Industry Stocks
Team Name: 44 Gatti

Sahil Mehta (sm5296), Shruti Agarwal(sa4136), Filippo Mattio(fm2790)

Data Set

Our project analyzes returns of the previous 5 years of the Automotive industry. Our aim is to perform exploratory data analysis on stocks of automobile manufacturing companies to understand the change on log return over time. The market saw a lot of variation when Covid hit and now that we are proceeding towards recession, our time period (i.e. from 1st January 2017 to 1st october 2022), focuses on the log returns from before, during and after the pandemic. The data was scraped using yahoo finance. It includes the opening and closing price, volume, and adjusted volume of the tickers. We have incorporated various Python libraries such as pandas, matplotlib, sklearn, statsmodels, and seaborn to plot, model and analyze the data.

GM - General Motors Company, F - Ford Motor Company, HMC - Honda Motor Co Ltd, TM - Toyota Motor Corp, RACE - Ferrari NV

Project Goals:

The goal of the project is to analyze 5 ticker symbols belonging to the automobile industry and provide insights regarding their performance:

1. Analysis of log returns before, during and after the pandemic.
2. How change from conventional mobility to sustainable mobility, i.e shift from gas vehicles to electric vehicles affected the stock prices?
3. How is the current recession affecting the stock prices of the industry?
4. Do the stocks of race car brands (Ferrari) affect the stock prices of general automotive selling vehicle brands?

Log Return Plots:

a)

b)
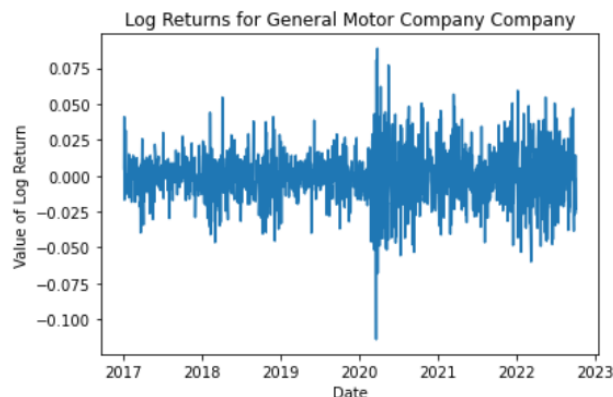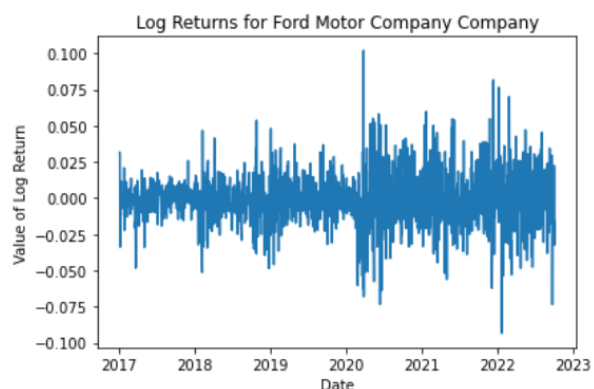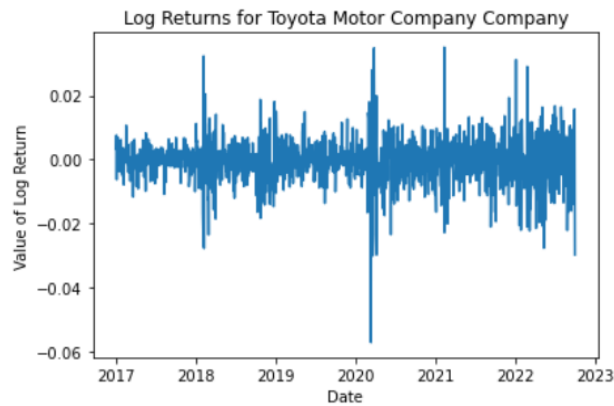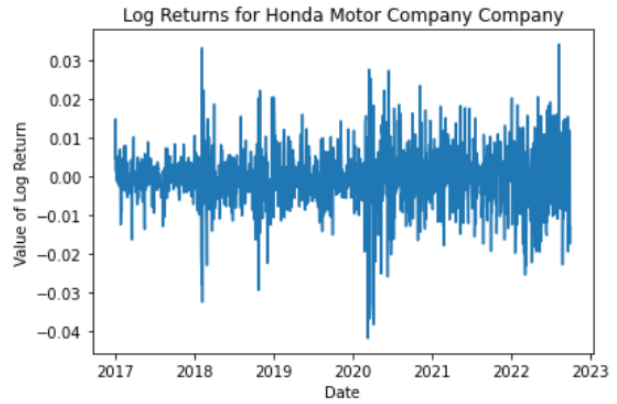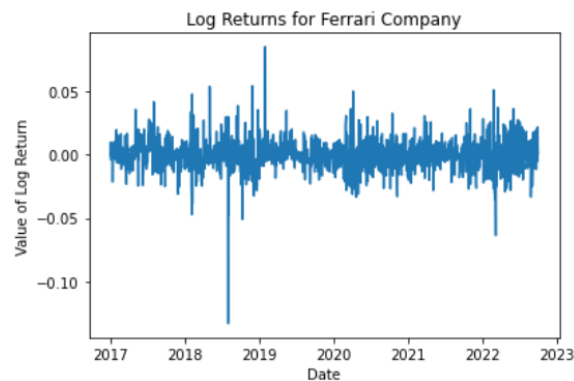
Figure 1. Log return plots for a) Gm, b) F, c) TM, d) HMC, e)RACE

Observations:

The main observation that can be drawn from the above plots are:

1. All stocks were affected in the first quarter of 2020 except RACE (Ferrari), when the novel coronavirus started to spread and the countries were imposing lockdowns.
2. There were a lot of variations in the log return plots after 2022 started and it is evident from the fact that the market is declining due to inflation. The frequent increase in interest rates by the FED is slowing down the economy which can end up spurring a recession.
3. There is a similar variation of log returns for the American players (GM and F) and Japanese Players (TM and HMC). The similarity of such pairs will be tested while conducting equal mean test, correlation and regression analysis for pairwise ticker symbols.

4. The log return plot of RACE is not similar with any of the other ticker symbols which can be because Ferrari doesn't share the same segment of the market as the other companies. Its target audience mostly includes the high end buyers and not the lower or middle class society. It was further observed that RACE's log returns weren't much affected by Covid either.

Analysis:

Normalized Plots:

We plot a histogram to verify the Central limit theorem which says that if a population with mean μ and standard deviation σ has a large number of random samples, then the distribution of the sample means will be approximately normally distributed. As expected, histogram plots of means of log returns for all the ticker symbols (Figure 1 a) & b)) were approximately normal. We further stack the normal distribution plot over the histogram which verifies that the log returns follows normal distribution.
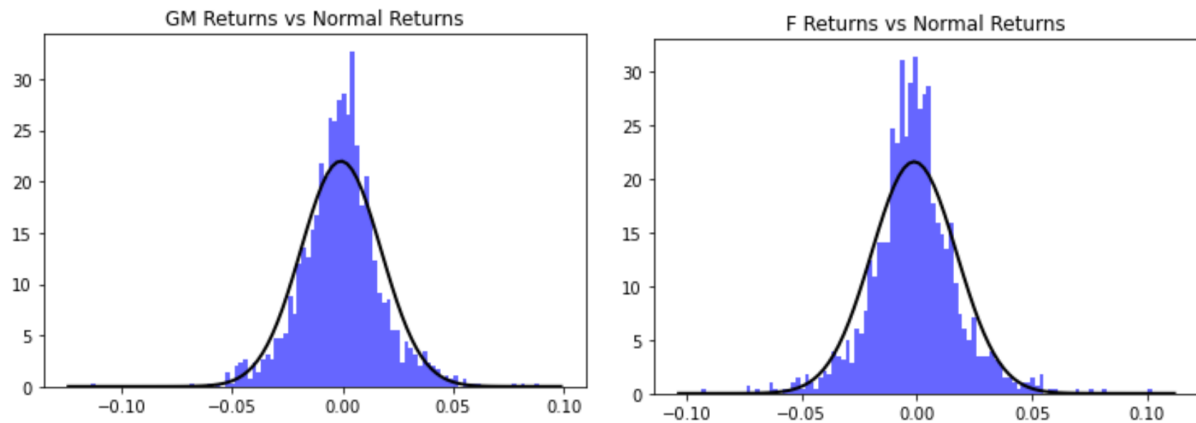

Figure 2. Mean of log returns of a) General Motor (GM) b) Ford Motor Company(F)

Confidence Intervals:

After confirming that the distribution is approximately normal, we carried out statistical testing on the means and sample variance of the samples of log return values after setting different levels of significance (Table 1). The confidence intervals for means are calculated using the normal distribution (due to the fact that our sample size is very large) and the variances are calculated using the chi square distribution.

	95						98					
	Mean			Variance			Mean			Variance		
	lower	Upper	Range	Lower	Upper	Range	Lower	Upper	Range	Lower	Upper	Range
F	-0.0027	-0.00017	0.00253	0.00032	0.00037	0.00005	-0.00225	7.57E-06	2.26E-03	0.00031	0.00037	0.00006
GM	-0.00179	7.44E-05	0.001864394	3.10E-04	3.50E-04	0.00004	-1.97E-03	2.50E-04	2.22E-03	3.00E-04	3.60E-04	0.00006
HMC	-0.00083	-3.27E-05	0.000797346	5.69E-05	6.59E-05	8.9367E-06	-0.00091	4.27E-05	9.53E-04	5.62E-05	6.68E-05	1.06155E-05
RACE	-0.00064	0.00068	0.00132	0.00015	0.00017	0.00002	-0.00076	0.0008	1.56E-03	0.00015	0.00018	0.00003
TM	-0.0008	-7.93E-05	0.000720704	4.58E-05	5.30E-05	7.1948E-06	-8.70E-04	-1.17E-05	8.58E-04	4.52E-05	5.38E-05	8.5464E-06

Table 1 Upper and lower values of the means of log returns for 95% and 98% confidence

The data in the above table proves that the range of the confidence interval increases as we increase the significance of confidence from 95% to 98%. Also, it can be seen that the mean of log returns are maximum for Ford and General motors which can be deduced from the fact that America is the primary market for F and GM and since we are analyzing the stocks based on NYSE (American Stock Exchange) the expected value for their stock returns can be high.

Regression Analysis:

Regression analysis was performed on all the ticker symbols with Time as the independent feature and Log returns as the dependent variable that we are trying to predict. The results ended up giving very small values of R-square, as shown in the table below, which suggests that the log returns of the stocks were not affected with change in time. To confirm this, a scatter plot of log returns vs time was obtained that gave a flat line, implying that log returns were not dependent on dates.

Ticker Symbol	R Square Values
F	4.68E-05
GM	4.72E-06
HMC	2.83E-06
RACE	6.99E-04
TM	1.34E-03

Table 2 of r square values of log returns vs time plot of stocks log returns vs time

Two Stock Analysis:

In this section we will check how the change in returns of one ticker affects the change in returns of another. We start by comparing the means of log returns of pairwise ticker symbols by hypothesis testing and then we do correlation and regression analysis in order to get more insights on the effects of one ticker on the other.

Equal Mean Test:

To test whether the means of log returns are equal or not, we consider a null hypothesis that the two ticker symbols have the same expected log returns mean and an alternative hypothesis to be unequal means of the two ticker symbols.

	HMC	TM	GM	F	RACE
HMC	1	0.985	0.414	0.193	0.245
TM	0.985	1	0.412	0.19	0.227
GM	0.414	0.412	1	0.702	0.13
F	0.193	0.19	0.702	1	0.053
RACE	0.245	0.227	0.13	0.053	1

Table 3. P values calculated for the equal mean tests for different samples

By considering a 95% level of confidence, we observed that we were unable to reject the null hypothesis that the log returns of all the pairs is the same, as shown in table 3. However we can see that the p values for RACE with all the other ticker symbols (HMC, TM , GM, F) is low (especially with GM and F, in which they are on the boundary of critical regions). This is intuitive from the fact that RACE doesn't belong to the same group of automobile companies, being it focused on luxury cars. This provides us with an analysis that the means log returns can be considered the same with 95% confidence but we reject some of the null hypothesis in the case when we decrease our confidence interval to 80% or 90%.

Regression analysis for pairwise ticker symbols:

A pairwise regression analysis was conducted by taking the log returns of one ticker symbol as independent variable and the other ticker symbol as dependent variable and the R square values are provided in table 4.

	HMC	TM	GM	F	RACE
HMC	1	**0.595**	0.295	0.301	0.243
TM	0.595	1	0.273	0.291	0.271
GM	0.295	0.273	1	**0.542**	0.178
F	0.301	0.291	0.542	1	0.13
RACE	0.243	0.271	0.178	0.13	1

Table 4. R-square values for pairwise regression analysis

Correlation coefficients for pairwise ticker symbols:

A pairwise correlation coefficient was calculated and the values of correlation coefficients are provided in table 5.

	HMC	TM	GM	F	RACE
HMC	1	**0.771253**	0.543204	0.548325	0.49325
TM	0.771253	1	0.522268	0.539103	0.520726
GM	0.543204	0.522268	1	**0.736292**	0.422303
F	0.548325	0.539103	0.736292	1	0.360131
RACE	0.49325	0.520726	0.422303	0.360131	1

Table 5) Correlation coefficient of pairwise correlation coefficient

The main observations from the regression analysis and correlation analysis were:

- The regression analysis on the pair of American auto manufacturers like General Motor and Ford Motor company provided a large value of R square i.e, 0.542 and similarly the pair of Japanese brands, i.e Toyota Motors and Honda Moto gave a large value of R square of 0.595., indicating that the log returns of these pairs were highly correlated with each other. To confirm this, correlation coefficient between the two pairs was also calculated which supported the claim that American stocks and Japanese stocks are correlated amongst each other with a correlation coefficient of 0.74 and 0.77 respectively.
- No company has a good correlation with 'RACE' since 'RACE', i.e Ferrari serves a very different market and is considered a luxury brand. So there is very little correlation between the cars manufactured by GM, F, TMC, TMC and those manufactured by RACE.

Conclusions:

The automotive sector has been dealing with certain challenges with the pandemic and recession in place, which has affected both the production and sales. Further, the boom of e-mobility and the tax benefits on electric cars, provided by the government to shift users to e-mobility, has been positively affecting the sales of the conventional fuel car industry. Moreover, with recession around the corner, the sales will decline significantly as many buyers will back out of the market. However, this recession won't have the same outcome as the Great Recession of 2008. Now, there is an all time low in the inventory and if the recession weakens the demand for cars, the stock prices might go down but not as much as they did in 2008. In the 2008 recession, the supply of vehicles far exceeded the demand for new and used cars. Looking from an Investor's perspective, we believe that buying stocks at this moment might be risky for the short term, however, buying more stocks from companies that have invested huge money in innovations on e-mobility might be good as the prices are low right now. We are expecting the stock prices to shoot in the coming 5 years.