Group 8: Justin Parsons and Nikul Patel
Research Project
ECON115004: Statistics Lab
Fall 2020

Correlations between Stock Exchanges and Indexes

Introduction:

The purpose of this research paper is studying the relationships between stock exchanges and indexes. Therefore, we will try to find some type of correlative or causal relationship between stock exchanges like the NASDAQ and New York Stock Exchange, and indexes such as the S&P 500 and DOW30. We will then further examine the relationships between these exchanges and indexes to see if there have been any substantial changes in their correlations over time using trading session intervals. This research is of interest because of the insight it gives about the stock market as a whole. Stock exchanges are mutually exclusive of each other because stocks are listed under one stock exchange; however, indexes are a basket of stocks that come from a variety of exchanges (e.g. the S&P 500 and Dow 30 contain stocks from both the NYSE and NASDAQ). In recent years, index funds have grown in popularity and it would be noteworthy to see if that has had an impact on the relationship between the exchanges, especially if a correlational or causal relationship establishes itself through volume of sales and daily percent change of average trading price. In summary, we learned that there is a relationship between the correlation of percent change in stock exchanges and indexes has increased over time. However because of factors such as consistently low R-squared values and outliers due to unpredictability in the market that are very prevalent, we believe this relationship is not necessarily causal but correlational.

Literature Review:

An article that inspired us to research the relationships between stock exchanges and indexes was “Correlations in Price Changes and Volatility across International Stock Markets” by Yasushi Hamao of the University of California, San Diego, Ronald W. Masulis of Vanderbilt University, and Victor Ng of the University of Michigan. This article studies price change correlations across international stock markets whereas we are studying price change correlations across various U.S. stock exchanges and indexes. They used data from daily opening and closing prices of major stock indexes across the different countries’ stock markets. To analyze, they used autoregressive conditionally heteroskedastic (ARCH family of statistical models to observe the price relationships. In their analysis, the researchers find that there were price volatility spillovers from New York to Tokyo, London to Tokyo, and New York to London but none from the other directions in the pre-October 1987 trading period.

Another article that our research took influence from was “Quantifying the Behavior of Stock Correlations Under Market Stress” headed by Tobias Preis who is affiliated with the University of Warwick, Boston University, and University College London, Dror Kenett of Boston University and Tel-Aviv University, H. Eugene Stanley of Boston University, Dirk Helbing of ETH Zurich, and Eshel Ben-Jacob of Tel-Aviv University. This article focused on how individual stocks are correlated in downturns despite many being highly diversified in what they do from each other. The publishers came to the conclusion that in downturns even the most highly diversified portfolios (in their case they used the Dow 30 as a sample) will follow a strong correlation. Our research will look at the correlations between how multiple ‘portfolios’ (indexes which are made up of stocks listed under exchanges and exchanges that contain stocks listed under them.) While they look at the correlations of stocks within a portfolio and how the portfolio moves, we will be comparing portfolios that either contain samplings from the same exchanges or are composed completely of their own exchange which are mutually exclusive to each other and seeing how their correlations change.

Data Description:

To study the relationships between stock exchanges and indexes we will be using historical price data which consists of 9016 trading day OHLCs (open, high, low, close) and trading volumes for our raw data. The main dependent variable is gonna be the correlation between the selected indexes/exchanges, which would be found by comparing the daily percent change of each pair of data sets. We will create 6 pairs of data for comparison (S&P + NASDAQ, S&P + NYSE, S&P + DOW, DOW + NASDAQ, DOW + NYSE, NASDAQ + NYSE) and see how its correlation has changed over time. We will also perform analysis of the pairs' daily volumes to see if we can find anything of interest from it (such as how markets perform and rebound in relation to each other due to black swan events/big news).

^DJI measures the number of trades in a given trading session at Dow Jones, these vary from a couple million to a couple billion as seen in Table 1. On average, there are 146069060.6 trades made of dow stocks in a day. ^GSPC measures the number of trades in a given trading session of S&P 500 stocks, these vary from tens of millions to tens of billions as seen in Table 1. On average, there are 1973191979 trades made in a day for the S&P 500. ^NYA measures the number of trades in a given trading session at the New York Stock Exchange, these vary from tens of millions to tens of billions as seen in Table 1. On average, there are 3274070814 trades made in a day at the New York Stock Exchange. ^IXIC measures the number of trades in a given trading session at the NASDAQ, these vary from tens of millions to a couple billion as seen in Table 1. On average, there are 1308463454 trades made in a day at the NASDAQ. Table 1 also displays the medians for ^DJI, ^GSPC, ^NYA, ^IXIC Volumes. For ^DJI, the median number of trades is 111895000, which is slightly less than the mean because the histogram seems to be bimodal with one of the peaks being at a lower volume of trades. ^GSPC has a median number of trades of 135185000, which is also smaller than the mean because its histogram also shows a bimodal distribution with a peak on the lower end of trade volume. ^NYA has a median number of trades of 334143000, which is larger than the mean even though the histogram shows a bimodal distribution. This is likely because the peak in the distribution occurs at a relatively large value (as compared to the other indexes and exchanges.) ^IXIC's median number of trades is 157521000 which is larger than the mean. As shown by the histogram, even though there is a large peak at a lower volume in the bimodal distribution, it is not substantial enough to affect the influence of the other peak. For our research purposes, outliers will actually be points of interest because they indicate unusual activity in the stock market. We removed volumes that were listed as 0 as they are not necessarily possible events.

As we can see by analyzing the histogram of Percent Change (Histogram 1), the data is approximately normally distributed around the mean for each index and exchange, with more dramatic outliers on the negative side.

We decided to look into volume of sales as a possible explanatory variable for percent change, however it is clearly noticeable in Scatterplot 1 below that they show no strong correlation between adjusted volume and percent change for any of the stock exchanges or indexes. The correlation between volume of sales and percent change for the Dow Jones index was -0.05514409355582702. For the S&P 500 index, it was -0.009227276265940665. Finally, for the NASDAQ exchange and New York Stock Exchange, they were -0.012898151579222921 and -0.02468305759820111 respectively.

We then decided to see how time has an effect on the correlation between our relationships. We did this by creating a dataset of rolling correlations in 25 trading session intervals (i.e. for the 9016 trading sessions we looked at, we computed the value of each correlation using the trailing 25 trading sessions). The correlation of the relationship was determined as a function of percent change and to form a linear regression model we used time.

Empirical Strategy:

To answer our question, we want to estimate the following regressions. We believe they are important because we've identified strong correlations between all of the exchanges and indexes which have likely changed over time as the stock market has evolved. All of these relationship correlations will be used as a dependent variable with time being the independent variable at 25 trading session intervals(TS), in addition to the volume (ST) of sales for our multivariate regression.

Y = B0 + B1X + U

Regression Model #1: DOW-S&P = 0.9243 + 0.000003(TS)

Regression Model #2: DOW-NASDAQ = 0.7146 + 0.000015(TS)

Regression Model #3: DOW-NYSE = 0.8086 + 0.000015(TS)

Regression Model #4: S&P-NASDAQ = 0.7911 + 0.000018(TS)

Regression Model #5: S&P-NYSE = 0.8552 + 0.000013(TS)

Regression Model #6: NASDAQ-NYSE = 0.615 + 0.000031(TS)

Y = B0 + B1X1 + B2X2 + U

Regression Model #7: DOW-S&P = 0.9244 + 0.000002753(TS) + 0.000000000008343(ST)

Regression Model #8: DOW-NASDAQ = 0.7147 + 0.00001346(TS) + 0.0000000000364(ST)

Regression Model #9: DOW-NYSE = 0.8088 + 0.00001286(TS) + 0.00000000005021(ST)

Regression Model #10: S&P-NASDAQ = 0.8039 + 0.00000687(TS) + 0.00000000001865(ST)

Regression Model #11: S&P-NYSE = 0.8665 + 0.000002929(TS) + 0.00000000001646(ST)

Regression Model #12: NASDAQ-NYSE = 0.6162 + 0.00002767(TS) + 0.00000000001004(ST)

Our independent variables are time (trading session) and sale volume (amount of shares traded) as we are looking for how our dependent variables, which are the correlations between all of the indexes and exchanges have changed. We are testing to see the effect of growing investing practices, like algorithmic trading, have affected the overall correlations between certain indexes and exchanges. We will also be observing how correlations are affected by black swan events and market crashes. Through just observing the sheer numeric significance, we have found that volume as a variable is insignificant in our research.

We expect that over time our variables will have become more correlated due to the adoption of more passive investing strategies such as index investing by retail investors and algorithmic trading by financial institutions. We also believe that despite this, unexpected changes in the market will cause them to lose their correlations as it becomes unpredictable for investors. When conducting hypothesis tests we will assume that there is no correlation between certain indexes and exchanges, to act like a base or control group. The alternative hypothesis will be that there is a positive or negative correlation between indexes and exchanges.

Results and Analysis:

Regression Model #1:

$H0: B1 = 0$

$H1: B1 \neq 0$

Our t-statistic was (0.000003033 - 0)/(0.000000259) = 11.71 for the Dow Jones-S&P 500 regression. We will reject the null hypothesis at a 5% confidence level if the value is smaller than -1.96 or greater than 1.96. Our test statistic is not so therefore we reject the null. This means that time has an effect whether it be positive or negative on the correlation between the Dow Jones and S&P 500.

Regression Model #2:

$H0: B1 = 0$

$H1: B1 \neq 0$

Our t-statistic was (0.00001468 - 0)/(0.000000623) = 23.56 for the Dow Jones-NASDAQ regression. We will reject the null hypothesis at a 5% confidence level if the value is smaller than -1.96 or greater than 1.96. Our test statistic is not so therefore we will reject the null. This means that time has an effect whether it be positive or negative on the correlation between the Dow Jones and NASDAQ.

Regression Model #3:

$H0: B1 = 0$

$H1: B1 \neq 0$

Our t-statistic was (0.00001455 - 0)/(0.000000319) = 45.61 for the Dow Jones-NYSE regression. We will reject the null hypothesis at a 5% confidence level if the value is smaller than -1.96 or greater than 1.96. Our test statistic is so therefore we will reject the null. This means that time has an effect whether it be positive or negative on the correlation between the Dow Jones and NYSE.

<u>Regression Model #4:</u>

H0: B1 = 0

H1: B1 ≠ 0

Our t-statistic was (0.00001792 - 0)/(0.000000331) = 54.14 for the S&P 500-NASDAQ regression. We will reject the null hypothesis at a 5% confidence level if the value is smaller than -1.96 or greater than 1.96. Our test statistic is so therefore we reject the null. This means that time has an effect whether it be positive or negative on the correlation between the S&P 500 and NASDAQ.

<u>Regression Model #5:</u>

H0: B1 = 0

H1: B1 ≠ 0

Our t-statistic was (0.00001268 - 0)/(0.000000218) = 58.17 for the S&P 500 - NYSE regression. We will reject the null hypothesis at a 5% confidence level if the value is smaller than -1.96 or greater than 1.96. Our test statistic is so therefore we reject the null. This means that time has an effect whether it be positive or negative on the correlation between the S&P 500 and NYSE.

<u>Regression Model #6:</u>

H0: B1 = 0

H1: B1 ≠ 0

Our t-statistic was (0.00003081 - 0)/(0.000000557) = 55.31 for the NYSE-NASDAQ regression. We will reject the null hypothesis at a 5% confidence level if the value is smaller than -1.96 or greater than 1.96. Our test statistic is so therefore we reject the null. This means that time has an effect whether it be positive or negative on the correlation between the NYSE and NASDAQ.

<u>Regression Models #7-12:</u>

As noted before, we found after doing the regression in our code, that sale volume as a variable was insignificant. We felt that Regression Models #1-6 showed strongly that the correlation between indexes and exchanges does indeed exist.

<u>Confidence Intervals:</u>

We have found the 95% confidence interval, meaning 95% of interval observations are accounted for, which is for the DOW and S&P to be from 2.53e-06 to 3.54e-06 correlation coefficients/trading sessions, for the DOW and NYSE from 1.39e-05 to 1.52e-05 correlation coefficients/trading sessions, for the DOW and NASDAQ from 1.35e-05 to 1.59e-05 correlation coefficients/trading sessions. For the S&P and NYSE it is from 1.23e-05 to 1.31e-05 correlation coefficients/trading sessions, for the S&P and NASDAQ from 1.73e-05 to 1.86e-05 correlation coefficients/trading sessions. For the NYSE and NASDAQ it is from 2.97e-05 to 3.19e-05 . These intervals are very tight and close to 0 which was expected given the large sample size and the fact that many of these exchanges and indexes require a substantial movement in capital to have a noticeable change in value, that if were to occur would not be exclusive.

<u>Regression Analysis:</u>

For regression model #1, the y-intercept or B0 was 0.9243. This means with no additional trading session interval the correlation between the DOW and S&P would be 0.9243 (usually on trading day one.) The slope of the regression or B1 was 0.000003, this means with an additional trading session added, the correlation increased by 0.000003 correlation coefficients/trading sessions. For regression model #2, the y-intercept or B0 was 0.7146. This means with no additional trading session interval the correlation between the DOW and NASDAQ would be 0.7146. The slope of the regression or B1 was 0.000015, this means with an additional trading session added, the correlation increased by 0.000015 correlation coefficients/trading sessions. For regression model #3, the y-intercept or B0 was 0.8086. This means with no additional trading session interval the correlation between the DOW and NYSE would be 0.8086. The slope of the regression or B1 was 0.000015, this means with an additional trading session added, the correlation increased by 0.000015 correlation coefficients/trading sessions. For regression model #4, the y-intercept or B0 was 0.7911. This means with no additional trading session interval the correlation between the S&P and NASDAQ would be 0.7911. The slope of the regression or B1 was 0.000018, this means with an additional trading session added, the correlation increased by 0.000018 correlation coefficients/trading sessions. For regression model #5, the y-intercept or B0 was 0.8552. This means with no additional trading session interval the correlation between the S&P and NYSE would be 0.8522. The slope of the regression or B1 was 0.000013, this means with an additional trading session added, the correlation increased by 0.000013 correlation coefficients/trading sessions. For regression model #6, the y-intercept or B0 was 0.615. This means with no additional trading session interval the correlation between the NASDAQ and NYSE would be 0.615. The slope of the regression or B1 was 0.000031, this means with an additional trading session added, the correlation increased by 0.000031 correlation coefficients/trading sessions.

The R^2 value measures how close the data is to the regression line of the scatter plot. One closer to 0 indicates a weak relationship and one closer to 1 indicates a strong relationship. For regression model #1, the R^2 value is 0.015. For regression model #2, the R^2 value is 0.015. For regression model #3, the R^2 value is 0.188. For regression model #4, the R^2 value is 0.246. For regression model #5, the R^2 value is 0.273. For regression model #6, the R^2 value is 0.254.

Conclusion:

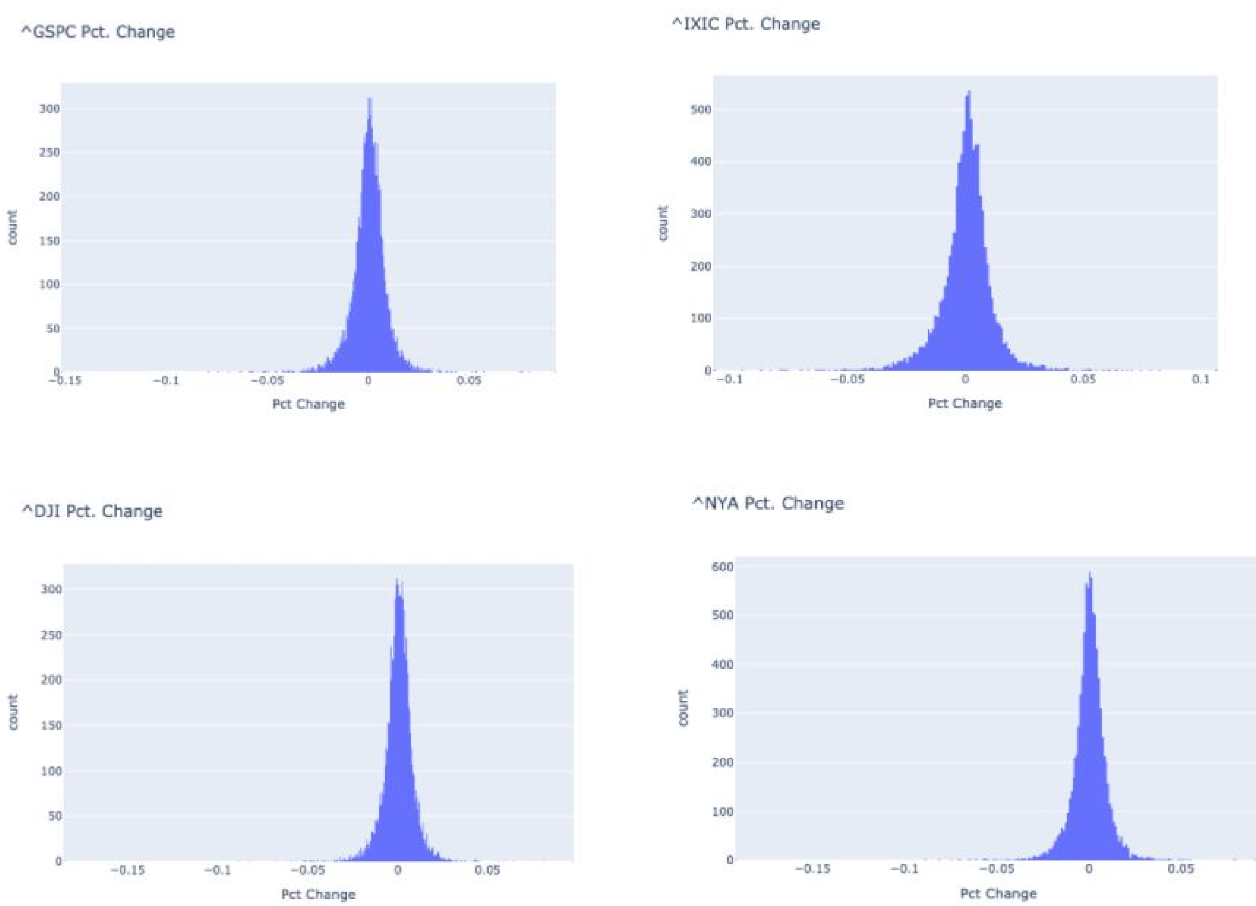
We were able to discover multiple findings through our research. The first of which is that there is no relationship between volume and the correlation between any stock exchanges or indexes, which is clearly seen in the abnormal shape of the scatterplots. We then decided to look at the correlation between stock indexes and exchanges to see how they've changed over time. Based on our regression analysis of the dependent variable (correlation between exchanges/indexes in intervals of 25 trading sessions) and the independent variables (time and volume of sales), every combination of pairs showed a positive correlation with time, however, the change is very small. Furthermore, due to the low R^2 value, we don't believe we can say this is a causal relationship that can be turned into a general statement, but a relationship that is present in our sample data. We can conclude that over time the correlation between the combinations of indexes and exchanges has increased in our dataset, but we can't assume this will continue to happen because of the unpredictability of the stock market. To try and find explanations for the increase in correlations we looked at how the market evolved over time. The only pair that was considered not to have a strong correlation at B0 and showed the greatest increase was between the S&P 500 and NASDAQ which increased by .279 over the 9016 trading sessions observed. Overall the correlation between them grew from 0.615 in 1985 to .894 in 2020. This makes sense because the NASDAQ is known as a tech heavy exchange, and since the start of our observation period, tech, which is mostly listed on the NASDAQ, has grown to become a significant portion of the S&P 500. Compared to both pieces of literature we discussed earlier, our research came to the same conclusion: over time, correlations between the index-exchange combinations had an effect on each other.The government could be a potential contributor to the correlational relationship between indexes and exchanges. In the COVID era, the federal reserve has shifted toward a policy of printing more money which has driven interest rates down. This is critical because as interest rates decrease, so do the yield of bonds, which is thought to lead investors into buying riskier assets such as stocks. When this occurs the price all securities will move in the same direction (up) and therefore be more correlated. Government policies can also result in inflation which would have the same effect on stock prices. The increase in correlation can also be explained by fewer companies taking up a bigger share of the market. With added knowledge of the stock market in the future, our research could be fine tuned. The data is also one that changes over time so it would be interesting to look back at the current COVID-19-effected stock market and compare it to this research in the future.

Graphs, Tables, Charts, Diagram:

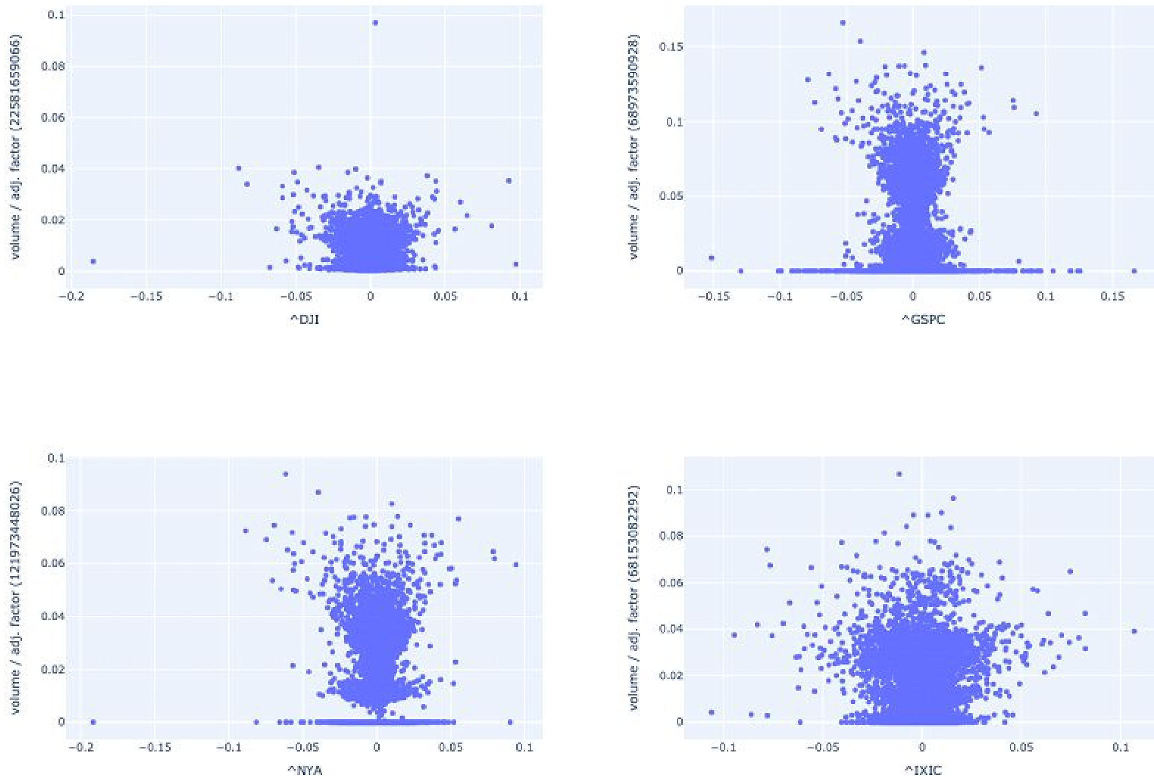
Table 1: Summary Statistics for ^DJI, ^GSPC, ^NYA, ^IXIC Volumes

Name	Median	Mean	Std. Deviation	Min	Max	25th percentile	50th percentile	75th percentile
^DJI	111895000	146069060.6	129941322.3	2530000	2190810000	27227500	111895000	233805000
^GSPC	1351850000	1973191979	1828903144	14990000	11456230000	276905000	1351850000	3464835000
^NYA	3341430000	3274070814	1478265836	61800000	11456230000	2079340000	3341430000	4044640000
^IXIC	1575210000	1308463454	949266651.5	43670000	7279230000	277642500	1575210000	1968187500

Histogram 1: Percent Change Among Indexes and Exchanges



Scatterplot 1: Percent Change vs. Adjusted Volume



Note: the high number of 0 volume points is due to the lack of them being record pre-1985, this was adjusted for in all other calculations

Table 2: Standard Error for Regression/Confidence Intervals

	DJI	GSPC	NYA	IXIC
DJI	0	0.0006752432163	0.0009185914676	0.001663651663
GSPC	0.0006752432163	0	0.0006629320419	0.0009872065699
NYA	0.0009185914676	0.0006629401328	0	0.001670619513
IXIC	0.001663651663	0.0009871086188	0.001670619513	0

Regression Model 1: DOW-S&P 500

Dep. Variable:	Pct Change_^GSPC	R-squared:	0.015
Model:	OLS	Adj. R-squared:	0.015
Method:	Least Squares	F-statistic:	137.6
Date:	Wed, 09 Dec 2020	Prob (F-statistic):	1.49e-31
Time:	23:35:52	Log-Likelihood:	12010.
No. Observations:	8991	AIC:	-2.402e+04
Df Residuals:	8989	BIC:	-2.400e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.9243	0.001	688.716	0.000	0.922	0.927
Index	3.033e-06	2.59e-07	11.731	0.000	2.53e-06	3.54e-06

Omnibus:	6845.876	Durbin-Watson:	0.041
Prob(Omnibus):	0.000	Jarque-Bera (JB):	154520.738
Skew:	-3.506	Prob(JB):	0.00
Kurtosis:	22.061	Cond. No.	1.04e+04

Regression Model 2: DOW-NASDAQ

Dep. Variable:	Pct Change_^IXIC	R-squared:	0.058
Model:	OLS	Adj. R-squared:	0.058
Method:	Least Squares	F-statistic:	555.2
Date:	Wed, 09 Dec 2020	Prob (F-statistic):	3.58e-119
Time:	23:35:52	Log-Likelihood:	4103.9
No. Observations:	8991	AIC:	-8204.
Df Residuals:	8989	BIC:	-8190.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.7146	0.003	220.990	0.000	0.708	0.721
Index	1.468e-05	6.23e-07	23.563	0.000	1.35e-05	1.59e-05

Omnibus:	2598.557	Durbin-Watson:	0.046
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6590.487
Skew:	-1.587	Prob(JB):	0.00
Kurtosis:	5.742	Cond. No.	1.04e+04

Regression Model 3: DOW-NYSE

Dep. Variable:	Pct Change_^NYA	R-squared:	0.188
Model:	OLS	Adj. R-squared:	0.187
Method:	Least Squares	F-statistic:	2075.
Date:	Wed, 09 Dec 2020	Prob (F-statistic):	0.00
Time:	23:35:52	Log-Likelihood:	10108.
No. Observations:	8991	AIC:	-2.021e+04
Df Residuals:	8989	BIC:	-2.020e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.8086	0.002	487.623	0.000	0.805	0.812
Index	1.455e-05	3.19e-07	45.550	0.000	1.39e-05	1.52e-05

Omnibus:	4068.620	Durbin-Watson:	0.071
Prob(Omnibus):	0.000	Jarque-Bera (JB):	29839.400
Skew:	-2.028	Prob(JB):	0.00
Kurtosis:	10.950	Cond. No.	1.04e+04

Regression Model 4: S&P 500-NASDAQ

Dep. Variable:	Pct Change_^IXIC	R-squared:	0.246
Model:	OLS	Adj. R-squared:	0.246
Method:	Least Squares	F-statistic:	2938.
Date:	Wed, 09 Dec 2020	Prob (F-statistic):	0.00
Time:	23:35:52	Log-Likelihood:	9799.0
No. Observations:	8991	AIC:	-1.959e+04
Df Residuals:	8989	BIC:	-1.958e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.7911	0.002	460.944	0.000	0.788	0.794
Index	1.792e-05	3.31e-07	54.203	0.000	1.73e-05	1.86e-05

Omnibus:	3466.119	Durbin-Watson:	0.069
Prob(Omnibus):	0.000	Jarque-Bera (JB):	17650.349
Skew:	-1.802	Prob(JB):	0.00
Kurtosis:	8.842	Cond. No.	1.04e+04

Regression Model 5: S&P 500-NYSE

Dep. Variable:	Pct Change_^NYA	R-squared:	0.273
Model:	OLS	Adj. R-squared:	0.273
Method:	Least Squares	F-statistic:	3382.
Date:	Wed, 09 Dec 2020	Prob (F-statistic):	0.00
Time:	23:35:52	Log-Likelihood:	13543.
No. Observations:	8991	AIC:	-2.708e+04
Df Residuals:	8989	BIC:	-2.707e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.8552	0.001	755.683	0.000	0.853	0.857
Index	1.268e-05	2.18e-07	58.155	0.000	1.23e-05	1.31e-05

Omnibus:	4860.262	Durbin-Watson:	0.091
Prob(Omnibus):	0.000	Jarque-Bera (JB):	56832.896
Skew:	-2.351	Prob(JB):	0.00
Kurtosis:	14.384	Cond. No.	1.04e+04

Regression Model 6: NASDAQ-NYSE

Dep. Variable:	Pct Change_^NYA	R-squared:	0.254
Model:	OLS	Adj. R-squared:	0.254
Method:	Least Squares	F-statistic:	3064.
Date:	Wed, 09 Dec 2020	Prob (F-statistic):	0.00
Time:	23:35:52	Log-Likelihood:	5115.9
No. Observations:	8992	AIC:	-1.023e+04
Df Residuals:	8990	BIC:	-1.021e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.6152	0.003	212.933	0.000	0.610	0.621
Index	3.081e-05	5.57e-07	55.354	0.000	2.97e-05	3.19e-05

Omnibus:	2580.214	Durbin-Watson:	0.072
Prob(Omnibus):	0.000	Jarque-Bera (JB):	7479.765
Skew:	-1.504	Prob(JB):	0.00
Kurtosis:	6.304	Cond. No.	1.04e+04

Regression Model 7: DOW-S&P 500 (including Sale Volume)

Dep. Variable:	Pct Change_^GSPC	R-squared:	0.015
Model:	OLS	Adj. R-squared:	0.015
Method:	Least Squares	F-statistic:	69.53
Date:	Wed, 09 Dec 2020	Prob (F-statistic):	1.08e-30
Time:	22:21:04	Log-Likelihood:	12011.
No. Observations:	8991	AIC:	-2.402e+04
Df Residuals:	8988	BIC:	-2.399e+04
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.9244	0.001	688.559	0.000	0.922	0.927
Index	2.753e-06	3.49e-07	7.888	0.000	2.07e-06	3.44e-06
Volume	8.343e-12	6.97e-12	1.197	0.231	-5.32e-12	2.2e-11

Omnibus:	6849.981	Durbin-Watson:	0.041
Prob(Omnibus):	0.000	Jarque-Bera (JB):	154568.070
Skew:	-3.509	Prob(JB):	0.00
Kurtosis:	22.061	Cond. No.	3.92e+08

Regression Model 8: DOW-NASDAQ (including Sale Volume)

Dep. Variable:	Pct Change_^IXIC	R-squared:	0.059
Model:	OLS	Adj. R-squared:	0.058
Method:	Least Squares	F-statistic:	280.1
Date:	Wed, 09 Dec 2020	Prob (F-statistic):	1.01e-118
Time:	22:21:04	Log-Likelihood:	4106.3
No. Observations:	8991	AIC:	-8207.
Df Residuals:	8988	BIC:	-8185.
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.7147	0.003	221.023	0.000	0.708	0.721
Index	1.346e-05	8.41e-07	16.006	0.000	1.18e-05	1.51e-05
Volume	3.64e-11	1.68e-11	2.167	0.030	3.48e-12	6.93e-11

Omnibus:	2624.978	Durbin-Watson:	0.046
Prob(Omnibus):	0.000	Jarque-Bera (JB):	6736.442
Skew:	-1.597	Prob(JB):	0.00
Kurtosis:	5.789	Cond. No.	3.92e+08

Regression Model 9: DOW-NYSE (including Sale Volume)

Dep. Variable:	Pct Change_^NYA	R-squared:	0.191
Model:	OLS	Adj. R-squared:	0.190
Method:	Least Squares	F-statistic:	1058.
Date:	Wed, 09 Dec 2020	Prob (F-statistic):	0.00
Time:	22:21:04	Log-Likelihood:	10125.
No. Observations:	8991	AIC:	-2.024e+04
Df Residuals:	8988	BIC:	-2.022e+04
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.8088	0.002	488.518	0.000	0.806	0.812
Index	1.286e-05	4.3e-07	29.888	0.000	1.2e-05	1.37e-05
Volume	5.021e-11	8.6e-12	5.840	0.000	3.34e-11	6.71e-11

Omnibus:	4062.355	Durbin-Watson:	0.072
Prob(Omnibus):	0.000	Jarque-Bera (JB):	29427.963
Skew:	-2.028	Prob(JB):	0.00
Kurtosis:	10.880	Cond. No.	3.92e+08

Regression Model 10: S&P 500-NASDAQ (including Sale Volume)

Dep. Variable:	Pct Change_^IXIC	R-squared:	0.285
Model:	OLS	Adj. R-squared:	0.285
Method:	Least Squares	F-statistic:	1792.
Date:	Wed, 09 Dec 2020	Prob (F-statistic):	0.00
Time:	22:21:04	Log-Likelihood:	10036.
No. Observations:	8991	AIC:	-2.007e+04
Df Residuals:	8988	BIC:	-2.004e+04
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.8039	0.002	454.293	0.000	0.800	0.807
Index	6.87e-06	5.96e-07	11.537	0.000	5.7e-06	8.04e-06
Volume	1.865e-11	8.45e-13	22.064	0.000	1.7e-11	2.03e-11

Omnibus:	3421.314	Durbin-Watson:	0.085
Prob(Omnibus):	0.000	Jarque-Bera (JB):	18170.894
Skew:	-1.758	Prob(JB):	0.00
Kurtosis:	9.012	Cond. No.	5.70e+09

Regression Model 11: S&P 500-NYSE (including Sale Volume)

Dep. Variable:	Pct Change_^NYA	R-squared:	0.340
Model:	OLS	Adj. R-squared:	0.340
Method:	Least Squares	F-statistic:	2317.
Date:	Wed, 09 Dec 2020	Prob (F-statistic):	0.00
Time:	22:21:04	Log-Likelihood:	13976.
No. Observations:	8991	AIC:	-2.795e+04
Df Residuals:	8988	BIC:	-2.793e+04
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.8665	0.001	759.019	0.000	0.864	0.869
Index	2.929e-06	3.84e-07	7.623	0.000	2.18e-06	3.68e-06
Volume	1.646e-11	5.45e-13	30.174	0.000	1.54e-11	1.75e-11

Omnibus:	5220.755	Durbin-Watson:	0.122
Prob(Omnibus):	0.000	Jarque-Bera (JB):	78063.294
Skew:	-2.494	Prob(JB):	0.00
Kurtosis:	16.546	Cond. No.	5.70e+09

Regression Model 12: NASDAQ-NYSE (including Sale Volume)

Dep. Variable:	Pct Change_^NYA	R-squared:	0.255
Model:	OLS	Adj. R-squared:	0.255
Method:	Least Squares	F-statistic:	1540.
Date:	Wed, 09 Dec 2020	Prob (F-statistic):	0.00
Time:	22:21:04	Log-Likelihood:	5121.7
No. Observations:	8992	AIC:	-1.024e+04
Df Residuals:	8989	BIC:	-1.022e+04
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.6162	0.003	212.382	0.000	0.610	0.622
Index	2.767e-05	1.08e-06	25.656	0.000	2.56e-05	2.98e-05
Volume	1.004e-11	2.95e-12	3.400	0.001	4.25e-12	1.58e-11

Omnibus:	2645.447	Durbin-Watson:	0.072
Prob(Omnibus):	0.000	Jarque-Bera (JB):	7928.999
Skew:	-1.528	Prob(JB):	0.00
Kurtosis:	6.438	Cond. No.	3.25e+09

Note: the dependent variables for all scatter plots look at correlation of pct change over time, not correlation of percent change in general

References:

Finance , Yahoo. “Dow Jones Industrial Average (^DJI) Historical Data.” Yahoo! Finance, Yahoo!, 8 Dec. 2020, finance.yahoo.com/quote/^DJI/history?p=%5EDJI.

Finance, Yahoo. “NASDAQ Composite (^IXIC) Historical Data.” Yahoo! Finance, Yahoo!, 10 Dec. 2020, finance.yahoo.com/quote/^IXIC/history?p=%5EIXIC.

Finance, Yahoo. “S&P 500 (^GSPC) Historical Data.” Yahoo! Finance, Yahoo!, 9 Dec. 2020, finance.yahoo.com/quote/^GSPC/history?p=%5EGSPC.

Finance, Yahoo. “NYSE COMPOSITE (DJ) (^NYA) Historical Data.” Yahoo! Finance, Yahoo!, 10 Dec. 2020, finance.yahoo.com/quote/^NYA/history?p=%5ENYA.

Tobias Preis, Dror Y. Kenett, H. Eugene Stanley, Dirk Helbing and Eshel Ben-Jacob, “Quantifying the Behavior of Stock Correlations Under Market Stress,” Scientific Reports, Vol. 2 (2012): 1-5

Yasushi Hamao, Ronald W. Masulis and Victor Ng, “Correlations in Price Changes and Volatility across International Stock Markets,” The Review of Financial Studies, Vol. 3, no. 2 (1990): 281-307

Note: The code used for our research is attached separately

Acknowledgments:

We'd like to thank Gonzalo Antonio Salazar Campos for all the work he has done this semester teaching us and guiding our research. This would not have been possible without him.