An Empirical Risk and Return Analysis of Select Stocks in NASDAQ 100

Arindam Banerjee[1]

[1] Associate Professor and Assistant Dean of GMBA & MGB programs at S. P. Jain School of Global Management, Dubai, P.O Box 502345, United Arab Emirates

Correspondence: Dr. Arindam Banerjee, Associate Professor and Assistant Dean of GMBA & MGB programs at S. P. Jain School of Global Management, Dubai, P.O Box 502345, United Arab Emirates. E-mail: arin_006@yahoo.com; arindam.banerjee@spjain.org

Received: January 17, 2022 Accepted: March 10, 2022 Online Published: March 24, 2022

doi:10.5430/afr.v11n2p1 URL: https://doi.org/10.5430/afr.v11n2p1

Abstract

Stock market indices are considered to be a powerful economic indicator. These indices can be classified based on the methodology of weight allocation for each stock and the rules governing the entry, retention and exit criteria of various stocks in the index. This paper presents a descriptive and an exploratory analysis carried out on the daily returns data of NASDAQ 100 (^NDX) index and shortlist of 20 stocks in the index. Random sampling was conducted at the sector level strata of all stocks that make up the index. This approach was followed to avoid selection bias and that stocks from the varied sectors are represented equally for this analysis. R-squared values and correlation coefficients were used to determine the explain-ability and relationship between the stock returns and the index returns respectively. The paper applied descriptive univariate analysis on daily returns at an individual stock level and at an aggregated sector level. Inter-relationship between stocks and the index returns was carried out by computing Pearson's correlation coefficient across the different combinations of stocks and index return values. Linear regression was carried out identify the explain ability of the variance in the returns of from the index to the returns from the stocks. All analysis was carried out using the python and the stats-models library. As anticipated, the returns of randomly picked 20 stocks were able to explain ~85 % of the variance of the returns of index. One of the primary focus of the paper was to explore whether NASDAQ-100 index can explain the variability of the technology stocks relatively more than the stocks that belong to other sectors in its portfolio owing to the nature of most stocks that make up the index.

Keywords: NASDAQ, linear regression, multivariate and univariate distribution, kernel density estimate (KDE), python programming

JEL Classification: C6, C2, D53, E4, G3

1. Introduction

The National Association of Securities Dealers Automated Quotations (NASDAQ) is a stock exchange based in New York, USA. It was founded in 1971. (Nasdaq, 2021). Based on the market capitalization of shares traded, it is the world's second largest stock exchange, next to the New York Stock Exchange (NYSE). The Nasdaq 100 index (^NDX) is a basket of 100 largest most liquid stocks listed in the Nasdaq stock exchange. This index, unlike the Nasdaq Composite features companies from variety of industries except the industry that constitutes financial institutions. The industries that constitute the Nasdaq 100 are retail, biotechnology, industrial, technology, healthcare among others. The index is built using a modified capitalization weighting approach. The weight of the stocks in the index are based on their market capitalizations, ensuring checks and measure to cap the influence of its largest components. This is accomplished by quarterly reviews and weight re-adjustment activities if the distributions requirements are not met. A large portion of the index covers the technology sections accounting up to 56% of the index's weight followed by consumer services, healthcare, telecommunications and other industries. In this study, a sample of 20 stocks were selected, in the US markets, stocks are less correlated to the overall market more than they do elsewhere, an ideal number of stocks in any portfolio would be around 20 to 30. Existing research conducted in the area revealed that prior to the emergence of online investing the number hovered in the range of 20 to 30.

Zaimovic (2021). The study is of relevance considering the time frame considered and would provide avenues of further research in a post COVID scenario.

2. Literature Survey

Inspite of several research done in the area, very limited research can be found considering the time frame considered in this paper applying to the NASDAQ listed stocks. Gautami (2018) studied the fluctuations in share prices of selected Indian companies. Trading stocks provides free float of shares coupled with transparent assessment through stock market transactions. The study explored the risk and return analysis of chosen stocks in india. Risk might be described as the variance in real return and return can be defined as the addition in the worth of stock. The profit from an investment portfolio helps a financial investor to assess the monetary performance of their respective investment.

Sushma & Vikas (2019) evaluate the risk and return of the eight NSE listed companies alongside a secondary target and studied their volatility prior and after the demonetization. The techniques and methods utilized for examination were mean, standard deviation, beta, relationship, covariance and T-test. Examination was finished by utilizing the closing prices of every month for every selected companies.

According to Sonia & Ganesh (2021), the Indian Financial Services industry is diverse. This development of the financial services area drove numerous financial investors to redirect their investment towards the stock market. To build an alluring portfolio, the singular financial investor has to perform out a risk and return investigation well ahead.

Balaji (2018) conducted studies on the risk and return of selected company stocks in auto industry and analyzed the performance of five Indian auto giants: The information analyzed for the review is 5 years, i.e., January 1, 2012–March 31, 2017. Techniques utilized for the review are risk, return, positioning strategy, graphical strategy.

Modelling daily returns using the linear regression approach requires some basic assumptions on the data points to be validated to ensure authenticity and reliability of the results. A thorough literature review was carried out to understand the previous work carried out to validate the distribution assumptions of the daily stock return data. Models previously proposed by (Bachelier 1900) assume that stock prices movements are normally distributed. (Osborne 1959) shows that logarithms of changes in the stock prices are mutually independent with a common probability distribution. Further, suggestion that stock prices will need to follow a normal distribution was then put forth. (Mandelbrot 1967) proposed that stock returns follow a stable Paretian distributions due their flatter tails. (Fama 1965) further supported the claims with a demonstration of confirming the flat tails of the stock prices and put forth that they have higher peaks than normal distribution. (Praetz 1972) examined the weekly data from Sydney stock exchange for a period of 8 years and concluded that Student-t distribution can be used an alternative to explain stock price behavior. (Blattberg & Gonedes 1974) used daily and weekly returns of stock of Dow Jones industrial (DJI) and estimated that Student-t distribution performs better than the normal distribution, but normality cannot be rejected from the monthly return data. (Hagerman 1978) proposed a mixture of normal distribution and Student-t distribution can be alternative to representing characteristics of stock return data. (Borowski, 2018) research of 65 stock indices concluded that distribution of daily returns can be normal only in short time intervals.

There has been a large body of research that has been carried out to further understand the distribution of stock returns distributions which would further make it easy to carry out parametric modelling on the underlying data. The timeframe, nature of the economy, market scenario (bear vs bull) and the return calculation time intervals being considered to identify returns determine the distribution assumptions that can be taken forward. These distributional assumptions are crucial for deciding pricing strategies, to be able to understand the behavior of the of stock at various levels and its indices to aid decision making during stock selections for entry and exist. A critical understanding of the distribution of stock returns and the correlations with the market indices can be used collectively to create a reasonable framework driven by statistical analysis that can drive scientific investment.

Patel and Surti (2020) Srivastava (2020) took the 10 years of data on monthly basis to correlate the directional movement of FMCG and the Nifty 50 index using correlation, regression, and ANOVA. They observed a strong positive correlation between the Nifty 50 and Nifty FMCG sector which means any change in the Nifty 50 index would result in a similar proportionate change for Nifty FMCG and related FMCG companies. Selected FMCG companies were Procter & Gamble (P&G), ITC, Hindustan Unilever Limited (HUL), and Godrej Consumer Products. The coefficient of correlation between Nifty 50 and Nifty FMCG was found 0.94 confirming the statement of strong correlation. Further regression analysis for Nifty 50 as the dependent variable and Nifty FMCG as the independent variable was performed to cross-verify the statistical significance. The R square value came out 0.89 (rounded to 2

digits) which means 89% of the Nifty FMCG index can be explained by the Nifty 50 index. As a next step to evaluate the relation between both indexes and each FMCG stock, regression analysis was performed by taking these companies as dependent variables. R-square value for Nifty 50 and P&G was 0.94 which shows strong relation and de-scribes that 94% of the P&G stock closing price can be explained by the Nifty 50 index. R-square values for the rest of FMCG stocks were above 0.8 hence it was concluded that FMCG stocks are strongly correlated with both indexes as well as they are statistically significant.

The study conducted by Rane and Gupta (2021) presented the outcome of regression analysis to predict the relationships between stock prices and various financial ratios for the Nifty Bank index which is a sub-index of Nifty 50. The duration for data extraction for analysis was taken from 2010-2019. Financial ratios related to the banking sector such as Net NPA ratio, capital adequacy ratio (CAR), net interest margin (NIM), return on equity (ROE), earning per share (EPS), dividend payout ratio, and net profit margin (NPM), was chosen in such a way that they can predict the company's performance. The aim of the study was to develop a model using a panel data regression model instead of a simple linear regression model due to the multi-dimensional nature of the data. With a 95% confidence level, four ratios (NPA, NPM, EPS, and ROE) out of the selected seven ratios are significant. With an adjusted R square value of 74.3%, which tells that four significant ratios out of seven can predict stock prices with 74.3% accuracy. The regression equation showed a negative coefficient for NPA that means if the NPA ratio (higher the ratio, lower the bank credibility) increases stock price will go down.

Srivastava, H. (2020) studied theoretical and empirical relationships between different sectoral indices listed at National Stock Exchange (NSE). Selected indexes were NIFTY IT, NIFTY Bank, NIFTY Media, NIFTY FMCG, NIFTY Auto, NIFTY Realty, NIFTY Metal, NIFTY Financial Seer-vices, and NIFTY Pharma. Data collection for correlation and regression analysis was from January 2012 to April 2018 on weekly basis. They categorized into three parameters: weakly correlated, moderately correlated, and strongly correlated to understand the correlation strength among the different indices based on R-square value. It was found that various indices are moderately or strongly correlated with each other except a few such as NIFTY Metal with NIFTY IT and NIFTY Media, NIFTY Bank with NIFTY Media, and NIFTY Pharma.

Biswas (2018) has performed multiple regression analysis, dropping variable analysis, Volatility, and Granger causality test using the ARCH (autoregressive conditionally heteroscedastic) mod-el. The period for the data analysis was from 2008 to 2018. The objective of the study was to find out how IT (Information Technology) stock prices affect the Nifty 50 index. From multiple regression analysis, they showed that except Wipro, all other IT stocks are statistically significant. Even after dropping the Wipro from regression, still, all IT stocks were highly significant. Granger causality test which is used to check the relation between two-time series data, predicted that IT stocks like Wipro, TCS, and Infosys independently do not affect Nifty 50 rather there is a joint significance of these stocks on the Nifty 50. They also found that Nifty 50 Granger cause Tech Mahindra and Wipro stock price which means if Nifty 50 gets affected, both stocks get affected. The third model tested, the ARCH model predicted that Infosys serves as one of the external causes for the volatility of Nifty 50 other than internal causes.

3. Methodology - Sampling

10-year, daily, adjusted closing price data between 2010-2020 for selected tickers were queried using the Yahoo finance API via python. A shortlist of 20 tickers was carried out independently using a stratified random sampling approach. The sampling activity was stratified at the sector level to ensure equal representation of across the industries represented in the index. The sample were picked from a list of companies listed on the NASDAQ stock exchange and a part of the NASDAQ 100 (^NDX) index. Daily returns were calculated and further analyzed.

Below is a table capturing the stocks that were selected based using the technique of stratified random sampling. The stratification was carried out at a sector level. A cut-off to the IPO year was also applied to be on or before 2010 to ensure sufficient 10-year stock return data can be gathered for the selected stocks. The choice of 20 stocks provide a healthy sample size for a robust study.

Table 1. List of 20 shortlisted stocks for the analysis

Company	Ticker	IPO Year	GICS Sector	GICS Sub-Industry
Baidu	BIDU	2005	Communication Services	Interactive Media & Services
Activision Blizzard	ATVI	1993	Communication Services	Interactive Home Entertainment
T-Mobile US	TMUS	2007	Communication Services	Wireless Telecommunication Services
Ross Stores	ROST	1985	Consumer Discretionary	Apparel Retail
eBay	EBAY	1998	Consumer Discretionary	Internet & Direct Marketing Retail
Starbucks	SBUX	1992	Consumer Discretionary	Restaurants
Mondelēz International	MDLZ	2001	Consumer Staples	Packaged Foods & Meats
PepsiCo	PEP	1984	Consumer Staples	Soft Drinks
Costco	COST	1986	Consumer Staples	Hypermarkets & Super Centers
Vertex Pharmaceuticals	VRTX	1991	Health Care	Biotechnology
Illumina	ILMN	2000	Health Care	Life Sciences Tools & Services
Cerner	CERN	1986	Health Care	Health Care Technology
Paccar	PCAR	1984	Industrials	Construction Machinery & Heavy Trucks
Copart	CPRT	1994	Industrials	Diversified Support Services
Cintas	CTAS	1983	Industrials	Diversified Support Services
Applied Materials	AMAT	1972	Information Technology	Semiconductor Equipment
Cadence Design Systems	CDNS	1987	Information Technology	Application Software
Paychex	PAYX	1983	Information Technology	Data Processing & Outsourced Services
Xcel Energy	XEL	1984	Utilities	Multi-Utilities
American Electric Power	AEP	1985	Utilities	Electric Utilities

3.1 Research Design - Analysis

Descriptive univariate analysis was carried out on daily returns at an individual stock level and at an aggregated sector level where aggregation was carried out by considering the average return for the stock belonging to the industry being considered. Inter-relationship between stocks and the index returns was carried out by computing Pearson's correlation coefficient across the different combinations of stocks and index return values. Linear regression was carried out identify the explain ability of the variance in the returns of from the index to the returns from the stocks. All analysis was carried out using the python and the stats-models library (Seabold, Skipper & Perktold, 2010)

3.2 Descriptive Analysis

T-Mobile US, Vertex Pharmaceuticals, and Illumina seem to have the highest daily movements over the period of 2010 to 2020. The seems to be a signal of stronger positive movement in returns of stocks as compared to the negative movement.

Table 2. Descriptive statistics of ten years daily returns of the 20 shortlisted stocks

Statistics	Count	Mean	Standard Deviation	Min	25%	50%	75%	Max	Kurtosis	Skewness	Variance
BIDU	2518	0.02%	2.41%	-19.79%	-1.24%	-0.02%	1.30%	12.47%	5.40	-0.27	0.06%
ATVI	2518	0.05%	1.88%	-14.14%	-0.91%	0.10%	1.04%	15.88%	7.97	-0.01	0.04%
TMUS	2518	0.04%	2.51%	-57.70%	-0.95%	0.11%	1.11%	26.89%	119.78	-4.64	0.06%
ROST	2518	0.09%	1.47%	-10.50%	-0.66%	0.08%	0.87%	9.64%	6.28	-0.14	0.02%
EBAY	2518	0.04%	1.78%	-14.22%	-0.82%	0.05%	0.96%	12.22%	8.22	-0.17	0.03%
SBUX	2518	0.08%	1.49%	-11.37%	-0.65%	0.09%	0.85%	9.03%	7.25	-0.31	0.02%
MDLZ	2518	0.05%	1.18%	-6.93%	-0.56%	0.06%	0.65%	7.56%	4.13	0.02	0.01%
PEP	2518	0.04%	0.89%	-4.81%	-0.45%	0.04%	0.55%	4.54%	2.53	-0.21	0.01%
COST	2518	0.07%	1.13%	-9.39%	-0.51%	0.08%	0.67%	5.92%	7.00	-0.62	0.01%
VRTX	2518	0.03%	2.67%	-19.40%	-1.29%	0.00%	1.31%	38.24%	37.23	2.53	0.07%
ILMN	2518	0.06%	2.65%	-46.91%	-1.01%	0.09%	1.19%	31.66%	67.28	-3.14	0.07%
CERN	2518	0.04%	1.60%	-12.19%	-0.79%	0.08%	0.86%	11.72%	6.61	0.01	0.03%
PCAR	2518	0.03%	1.64%	-11.95%	-0.81%	0.05%	0.88%	7.50%	3.84	-0.40	0.03%
CPRT	2518	0.08%	1.41%	-15.45%	-0.59%	0.09%	0.83%	10.51%	13.05	-0.74	0.02%
CTAS	2518	0.09%	1.30%	-8.55%	-0.53%	0.13%	0.73%	8.94%	6.92	0.24	0.02%
AMAT	2518	0.05%	1.91%	-9.16%	-0.94%	0.08%	1.14%	12.14%	2.81	-0.14	0.04%
CDNS	2518	0.08%	1.67%	-11.53%	-0.72%	0.11%	0.96%	13.32%	6.00	0.03	0.03%
PAYX	2518	0.05%	1.11%	-5.29%	-0.51%	0.09%	0.69%	5.65%	2.31	-0.38	0.01%
XEL	2518	0.05%	0.98%	-5.83%	-0.49%	0.07%	0.67%	4.51%	2.06	-0.51	0.01%
AEP	2518	0.05%	1.01%	-5.51%	-0.50%	0.10%	0.66%	4.61%	1.99	-0.38	0.01%
^NDX	2518	0.06%	1.09%	-6.51%	-0.41%	0.11%	0.64%	5.80%	3.16	-0.47	0.01%

Commercial Services and Healthcare sectors seems to have the highest variability in stock returns indicating possible volatility.

Table 3. Descriptive statistics of the ten year daily returns of the 20 shortlisted stocks averaged at the sector level

Statistics	Count	Mean	Standard Form	Min	25%	50%	75%	Max	Kurtosis	Skewness	Variance
Commercial Services	2518	0.04%	1.62%	-21.00%	-0.78%	0.10%	0.94%	10.37%	13.17	-0.98	0.03%
Consumer Discretionary	2518	0.07%	1.19%	-7.20%	-0.54%	0.09%	0.74%	5.98%	2.73	-0.3	0.01%
Consumer Staples	2518	0.05%	0.85%	-5.00%	-0.39%	0.07%	0.53%	3.64%	2.81	-0.44	0.01%
Health Care	2518	0.04%	1.70%	-16.30%	-0.78%	0.06%	0.97%	13.00%	9.86	-0.37	0.03%
Industrial	2518	0.07%	1.18%	-8.60%	-0.51%	0.11%	0.70%	5.73%	3.43	-0.47	0.01%
Information Technologies	2518	0.06%	1.28%	-6.30%	-0.61%	0.11%	0.80%	5.53%	1.72	-0.32	0.02%
Utilities	2518	0.05%	0.96%	-5.70%	-0.48%	0.09%	0.64%	4.35%	2.05	-0.48	0.01%
^NDX	2518	0.06%	1.09%	-6.50%	-0.41%	0.11%	0.64%	5.80%	3.16	-0.47	0.01%

3.3 Correlation between Stock Returns and the Index

The correlation matrix below shows the Pearson's correlation coefficient for the different combination of the ten-year stock and index return values. From the table, it is evident that there is a positive correlation between the return of the NASDAQ – 100 index (^NDX) and the returns from the 20 stocks. Stocks belonging to the information technology and industrial sector has the strongest positive correlation with returns of the index relative to the other industries represented in the sample. There are signs of high correlation in returns among stocks that are in the same sector and industry as well.

	BIDU	ATVI	TMUS	ROST	EBAY	SBUX	MDLZ	PEP	COST	VRTX	ILMN	CERN	PCAR	CPRT	CTAS	AMAT	CDNS	PAYX	XEL	AEP	^NDX
BIDU	1.00																				
ATVI	0.32	1.00																			
TMUS	0.25	0.21	1.00																		
ROST	0.25	0.24	0.20	1.00																	
EBAY	0.37	0.30	0.26	0.30	1.00																
SBUX	0.34	0.31	0.28	0.40	0.36	1.00															
MDLZ	0.22	0.30	0.22	0.32	0.29	0.36	1.00														
PEP	0.19	0.24	0.22	0.29	0.24	0.34	0.54	1.00													
COST	0.24	0.26	0.21	0.44	0.31	0.39	0.41	0.42	1.00												
VRTX	0.23	0.23	0.20	0.19	0.26	0.24	0.24	0.18	0.19	1.00											
ILMN	0.27	0.25	0.22	0.19	0.28	0.25	0.21	0.17	0.22	0.31	1.00										
CERN	0.31	0.27	0.25	0.31	0.36	0.40	0.34	0.31	0.35	0.30	0.29	1.00									
PCAR	0.38	0.32	0.33	0.38	0.43	0.42	0.38	0.35	0.38	0.29	0.32	0.42	1.00								
CPRT	0.30	0.30	0.26	0.31	0.34	0.35	0.30	0.28	0.34	0.26	0.27	0.37	0.45	1.00							
CTAS	0.33	0.36	0.35	0.37	0.39	0.44	0.39	0.38	0.37	0.29	0.31	0.43	0.54	0.46	1.00						
AMAT	0.39	0.39	0.28	0.31	0.42	0.37	0.33	0.28	0.30	0.30	0.32	0.37	0.50	0.38	0.46	1.00					
CDNS	0.37	0.39	0.31	0.31	0.40	0.41	0.31	0.30	0.36	0.27	0.30	0.42	0.46	0.43	0.49	0.51	1.00				
PAYX	0.35	0.36	0.33	0.43	0.43	0.51	0.45	0.48	0.46	0.30	0.29	0.43	0.56	0.46	0.56	0.48	0.50	1.00			
XEL	0.12	0.18	0.21	0.23	0.17	0.28	0.43	0.49	0.32	0.15	0.14	0.27	0.25	0.24	0.35	0.19	0.26	0.39	1.00		
AEP	0.12	0.18	0.20	0.21	0.17	0.27	0.42	0.48	0.31	0.14	0.14	0.26	0.26	0.23	0.34	0.21	0.25	0.38	0.85	1.00	
^NDX	0.57	0.54	0.41	0.49	0.59	0.59	0.49	0.46	0.52	0.45	0.46	0.57	0.66	0.55	0.65	0.68	0.65	0.68	0.34	0.33	1.00

Chart 1. Correlation matrix of 10 year returns (2010-2020) of select stocks and the NASDAQ 100 index

Note. Column/Row1: Stock ISIN Code. Columns/Rows 2-21 Correlation Coefficients

Table 4. Correlation matrix of ten years (2010-2020) daily returns of stocks from the commercial service sector

	BIDU	**ATVI**	**TMUS**	**^NDX**
BIDU	1.00	0.32	0.25	0.57
ATVI	0.32	1.00	0.21	0.54
TMUS	0.25	0.21	1.00	0.41
^NDX	0.57	0.54	0.41	1.00

Table 5. Correlation matrix of ten years (2010-2020) daily returns of stocks from the consumer discretionary service sector

	ROST	**EBAY**	**SBUX**	**^NDX**
ROST	1.00	0.30	0.40	0.49
EBAY	0.30	1.00	0.36	0.59
SBUX	0.40	0.36	1.00	0.59
^NDX	0.49	0.59	0.59	1.00

Table 6. Correlation matrix of ten years (2010-2020) daily returns of stocks from the consumer staples sector

	MDLZ	PEP	COST	^NDX
MDLZ	1.00	0.54	0.41	0.49
PEP	0.54	1.00	0.42	0.46
COST	0.41	0.42	1.00	0.52
^NDX	0.49	0.46	0.52	1.00

Table 7. Correlation matrix of ten years (2010-2020) daily returns of stocks from the healthcare sector

	VRTX	ILMN	CERN	^NDX
VRTX	1.00	0.31	0.30	0.45
ILMN	0.31	1.00	0.29	0.46
CERN	0.30	0.29	1.00	0.57
^NDX	0.45	0.46	0.57	1.00

Table 8. Correlation matrix of ten years (2010-2020) daily returns of stocks from the industrial sector

	PCAR	CPRT	CTAS	^NDX
PCAR	1.00	0.45	0.54	0.66
CPRT	0.45	1.00	0.46	0.55
CTAS	0.54	0.46	1.00	0.65
^NDX	0.66	0.55	0.65	1.00

Table 9. Correlation matrix of ten years (2010-2020) daily returns of stocks from the information technology sector

	AMAT	CDNS	PAYX	^NDX
AMAT	1.00	0.51	0.48	0.68
CDNS	0.51	1.00	0.50	0.65
PAYX	0.48	0.50	1.00	0.68
^NDX	0.68	0.65	0.68	1.00

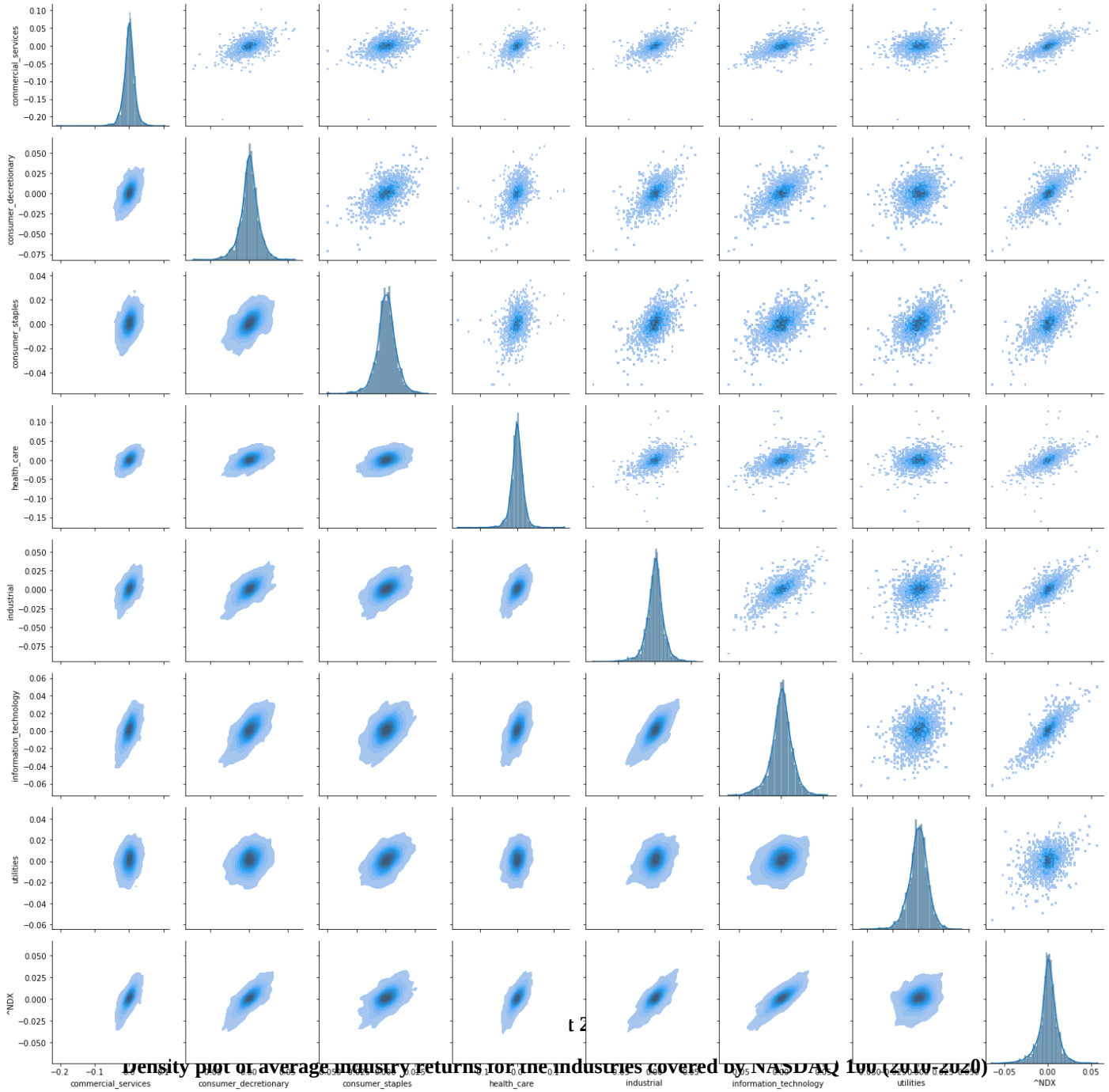
Table 10. Correlation matrix of ten years (2010-2020) daily returns of stocks from the utilities sector

	XEL	AEP	^NDX
XEL	1.00	0.85	0.34
AEP	0.85	1.00	0.33
^NDX	0.34	0.33	1.00

From the above sector level correlation matrices, it is evident that all the stock that have been identified from the pool of eligible stocks have a positive correlation with the NASDAQ-100 index. Daily returns of the stocks belonging to the information technology sector have the higher correlation with the daily returns of index in discussion. This agrees in with the fact that, like the NASDAQ Composite index, Nasdaq-100 (^NDX) Index is heavily weighted towards technology companies. (Tretina, 2021)

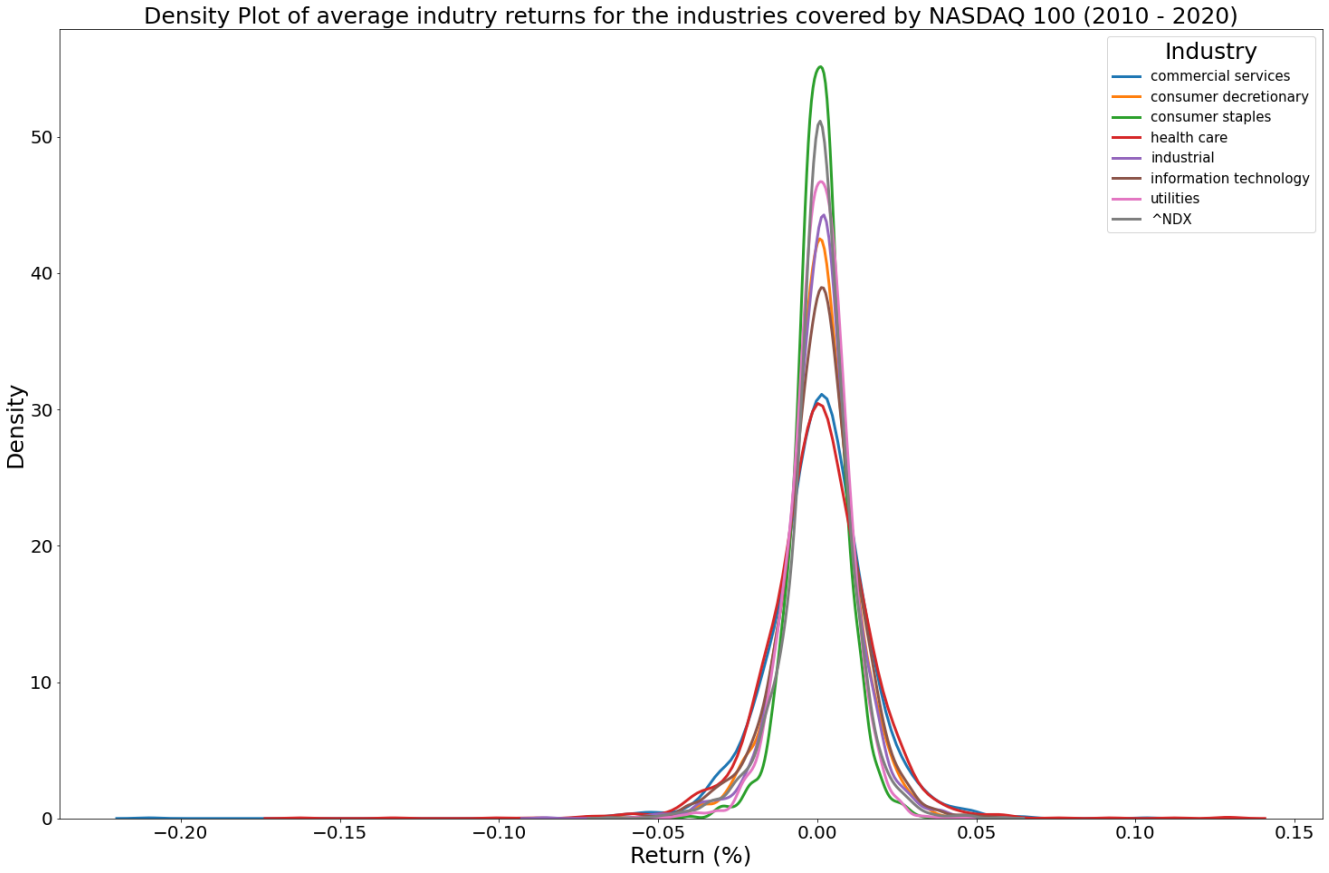
3.4 Multivariate and Univariate Density and Kernel Distribution Plots of Daily Returns of the Select 20 Stocks Averaged at A Sector Level

A kernel density estimate (KDE) plot is a method for visualizing the distribution of observations in a dataset, analogous to a histogram. KDE represents the data using a continuous probability density curve in one or more dimensions. Relative to a histogram, KDE can produce a plot that is less cluttered and more interpretable, especially when drawing multiple distributions. The below plot paints a picture of the univariate and multivariate distributions of the daily returns data of the select stocks aggregated at sector level.



Density plot of average industry returns for the industries covered by NASDAQ 100 (2010-2020)

Chart 2. Univariate and multivariate distributions of the daily returns data of the select stocks aggregated at sector level



3.5 Regression with Returns of NASDAQ-100 Index (^NDX) as the Dependent Variable

For the regression model, daily returns from the select stocks were considered as independent variable to understand the extent of variance in the return of the index that can answered by the return of the randomly samples stocks. This help is providing an understanding of the performance of index and its dependence of its returns on the individual returns of the stocks that it represents.

The regression model below, with an r-square value of 0.85 implies that the independent variable together can explain the ~ 85% of the variation of the returns in of the index in which they are represented. Owing to the nature of sampling, it is safe to say that the index is representing the returns of all the stocks across all the sectors that it covers. It is important to note that both the stocks representing the utilities sector have insignificant p-values/confidence intervals which is probably due to the under representation of stocks belonging to the utility sector in the index.

3.5.1 OLS Regression Results

Dep. Variable:	^NDX	R-squared (uncentered):	0.852
Model:	OLS	Adj. R-squared (uncentered):	0.850
Method:	Least Squares	F-statistic:	717.2
Date:	Fri, 10 Dec 2021	Prob (F-statistic):	0.00
Time:	15:10:07	Log-Likelihood:	10196.
No. Observations:	2518	AIC:	-2.035e+04
Df Residuals:	2498	BIC:	-2.024e+04
Df Model:	20		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
BIDU	0.0652	0.004	15.841	0.000	0.057	0.073
ATVI	0.0649	0.005	12.479	0.000	0.055	0.075
TMUS	0.0145	0.004	3.880	0.000	0.007	0.022
ROST	0.0374	0.007	5.443	0.000	0.024	0.051
EBAY	0.0708	0.006	12.280	0.000	0.060	0.082
SBUX	0.0681	0.007	9.498	0.000	0.054	0.082
MDLZ	0.0268	0.009	2.895	0.004	0.009	0.045
PEP	0.0592	0.013	4.710	0.000	0.035	0.084
COST	0.0677	0.009	7.228	0.000	0.049	0.086
VRTX	0.0354	0.004	10.070	0.000	0.029	0.042
ILMN	0.0286	0.004	7.965	0.000	0.022	0.036
CERN	0.0532	0.007	8.186	0.000	0.040	0.066
PCAR	0.0492	0.007	6.918	0.000	0.035	0.063
CPRT	0.0346	0.007	4.713	0.000	0.020	0.049
CTAS	0.0652	0.009	7.253	0.000	0.048	0.083
AMAT	0.1022	0.006	17.489	0.000	0.091	0.114
CDNS	0.0657	0.007	9.769	0.000	0.053	0.079
PAYX	0.0822	0.011	7.217	0.000	0.060	0.105
XEL	-0.0208	0.017	-1.237	0.216	-0.054	0.012
AEP	0.0011	0.016	0.070	0.944	-0.030	0.032

Omnibus:	188.001	Durbin-Watson:	2.052
Prob(Omnibus):	0.000	Jarque-Bera (JB):	816.445
Skew:	0.221	Prob(JB):	5.14e-178
Kurtosis:	5.754	Cond. No.	11.9

Notes:

[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.

[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

3.5.2 Regression with the Returns of the Stocks Averaged at the Sector Level as the Dependent Variable

For the regression model, returns from the index (NDX^) were considered as independent variable to understand how much of the variance in the daily returns of the stocks averaged at a sector can answered by the daily returns of the index. This is to understand the extent to which the daily returns of the index represent the daily returns of the sectors (for the stock captured in the index). Below are the results of the stepwise linear regression carried out with the ten years daily returns of the NASDAQ-100 index as an independent variable and ten years daily returns of the shortlisted stocks averaged at a sector level as dependent variable. The return of the index can explain the average returns of the stocks belonging to the industrial and information technology sectors more than those belonging to other sectors.

3.5.3 OLS Regression Results

Dependent variable: averaged returns of the shortlisted stocks belonging to commercial service sector

Dep. Variable:	commercial services	R-squared (uncentered):	0.495
Model:	OLS	Adj. R-squared (uncentered):	0.494
Method:	Least Squares	F-statistic:	2463.
Date:	Tue, 14 Dec 2021	Prob (F-statistic):	0.00
Time:	17:08:52	Log-Likelihood:	7671.7
No. Observations:	2518	AIC:	-1.534e+04
Df Residuals:	2517	BIC:	-1.534e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
^NDX	1.0381	0.021	49.627	0.000	0.997	1.079

Omnibus:	1274.512	Durbin-Watson:	2.052
Prob(Omnibus):	0.000	Jarque-Bera (JB):	85768.469
Skew:	-1.559	Prob(JB):	0.00
Kurtosis:	31.421	Cond. No.	1.00

Notes:

[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.

[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

3.5.4 OLS Regression Results

Dependent variable: averaged returns of the shortlisted stocks belonging to consumer discretionary

Dep. Variable:	consumer-discretionary	R-squared (uncentered):	0.554
Model:	OLS	Adj. R-squared (uncentered):	0.554
Method:	Least Squares	F-statistic:	3131.
Date:	Tue, 14 Dec 2021	Prob (F-statistic):	0.00
Time:	17:08:52	Log-Likelihood:	8597.3
No. Observations:	2518	AIC:	-1.719e+04
Df Residuals:	2517	BIC:	-1.719e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
^NDX	0.8104	0.014	55.951	0.000	0.782	0.839

Omnibus:	248.595	Durbin-Watson:	2.081
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1689.837
Skew:	0.170	Prob(JB):	0.00
Kurtosis:	6.999	Cond. No.	1.00

Notes:

[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.

[2] Standard Errors assume that the covariance matrix of the errors is correctly specified

OLS Regression Results

Dependent variable: averaged returns of the shortlisted stocks belonging to consumer staples

Dep. Variable:	consuner-staples	R-squared (uncentered):	0.382
Model:	OLS	Adj. R-squared (uncentered):	0.382
Method:	Least Squares	F-statistic:	1557.
Date:	Tue, 14 Dec 2021	Prob (F-statistic):	1.47e-265
Time:	17:08:52	Log-Likelihood:	9020.1
No. Observations:	2518	AIC:	-1.804e+04
Df Residuals:	2517	BIC:	-1.803e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
^NDX	0.4832	0.012	39.463	0.00	0.459	0.507

Omnibus:	102.547	Durbin-Watson:	2.059
Prob(Omnibus):	0.000	Jarque-Bera (JB):	307.906
Skew:	-0.081	Prob(JB):	1.38e-67
Kurtosis:	4.706	Cond. No.	1.00

Notes:

[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.

[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

3.5.5 OLS Regression Results

Dependent variable: averaged returns of the shortlisted stocks belonging to health care

Dep. Variable:	health-care	R-squared (uncentered):	0.420
Model:	OLS	Adj. R-squared (uncentered):	0.419
Method:	Least Squares	F-statistic:	1820.
Date:	Tue, 14 Dec 2021	Prob (F-statistic):	9.34e-300
Time:	17:08:52	Log-Likelihood:	7364.9
No. Observations:	2518	AIC:	-1.473e+04
Df Residuals:	2517	BIC:	-1.472e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
^NDX	1.0081	0.024	42.664	0.000	0.962	1.054

Omnibus:	679.895	Durbin-Watson:	1.990
Prob(Omnibus):	0.000	Jarque-Bera (JB):	50094.591
Skew:	-0.262	Prob(JB):	0.00
Kurtosis:	24.845	Cond. No.	1.00

Notes:

[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.

[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

3.5.6 OLS Regression Results

Dependent variable: averaged returns of the shortlisted stocks belonging to industrial sector

Dep. Variable:	industrial	R-squared (uncentered):	0.528
Model:	OLS	Adj. R-squared (uncentered):	0.582
Method:	Least Squares	F-statistic:	3506.
Date:	Tue, 14 Dec 2021	Prob (F-statistic):	0.00
Time:	17:08:52	Log-Likelihood:	8710.0
No. Observations:	2518	AIC:	-1.742e+04
Df Residuals:	2517	BIC:	-1.741e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
^NDX	0.8201	0.014	59.211	0.000	0.793	0.847

Omnibus:	255.509	Durbin-Watson:	1.975
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1866.226
Skew:	-0.143	Prob(JB):	0.00
Kurtosis:	7.208	Cond. No.	1.00

Notes:

[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.

[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

3.5.7 OLS Regression Results

Dependent variable: averaged returns of the shortlisted stocks belonging to information technology sector

Dep. Variable:	information technology	R-squared (uncentered):	0.671
Model:	OLS	Adj. R-squared (uncentered):	0.670
Method:	Least Squares	F-statistic:	5122.
Date:	Tue, 14 Dec 2021	Prob (F-statistic):	0.00
Time:	17:08:52	Log-Likelihood:	8796.5
No. Observations:	2518	AIC:	-1.759e+04
Df Residuals:	2517	BIC:	-1.759e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
^NDX	0.9578	0.013	71.569	0.000	0.932	0.984

Omnibus:	172.906	Durbin-Watson:	1.985
Prob(Omnibus):	0.000	Jarque-Bera (JB):	733.355
Skew:	0.179	Prob(JB):	5.67e-160
Kurtosis:	5.620	Cond. No.	1.00

Notes:

[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.

[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

3.5.8 OLS Regression Results

Dependent variable: averaged returns of the shortlisted stocks belonging to utilities sector

Dep. Variable:	utilities	R-squared (uncentered):	0.122
Model:	OLS	Adj. R-squared (uncentered):	0.121
Method:	Least Squares	F-statistic:	348.8
Date:	Tue, 14 Dec 2021	Prob (F-statistic):	5.36e-73
Time:	17:08:52	Log-Likelihood:	8294.4
No. Observations:	2518	AIC:	-1.659e+04
Df Residuals:	2517	BIC:	-1.658e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
^NDX	0.3051	0.016	18.676	0.000	0.273	0.337

Omnibus:	159.752	Durbin-Watson:	2.008
Prob(Omnibus):	0.000	Jarque-Bera (JB):	391.208
Skew:	-0.371	Prob(JB):	1.12e-85
Kurtosis:	4.782	Cond. No.	1.00

Notes:

[1] R^2 is computed without centering (uncentered) since the model does not contain a constant.

[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.

4. Conclusion

The objective of the report was to understand the relationship between the daily returns of the index and the daily returns of the stocks that it represents. It is evident that due to nature in which the index is constructed, it fulfills the role of encapsulating the performance of the cohort of stocks it represents.

In case of an index like NASDAQ-100 which is meant to represent a stock of companies belonging to a broader sector, it is evident from the evidence above that the returns of the index tend towards the performance of the stocks belonging the sector that is strongly represented. A large portion of the index covers the technology sector accounting up to 56% of the index's weight, which is evident in relatively higher R-squared values between the returns of the index (^NDX) aggregated returns information technology sector. The theme of strong relationship of the index returns with the averaged sector returns of industrial and information technology sectors is also back by relatively strong correlation between the daily returns of index and daily returns of the shortlisted stocks. When the daily returns of the stocks represented in the index(^NDX) is modelled using the daily returns of the index, returns of the index explain the variability of the averaged returns of the stocks from the companies belonging to the industrial and the information technology sectors relatively higher than those stocks of companies belong to a different sector. The above analysis confirms the strong relation between the performance of the index and the performance of the individual stocks it represents. This characteristic of index is what allows investors and economists to rely on it profoundly for decision making.

Acknowledgements

The author would like to acknowledge the contribution of Mr. Prajwal Gowda and Mr. Ajay Singh Parihar, MBA students at SP Jain School of Global Management, Dubai., for their research inputs.

References

Bachelier, L. (1990). Théorie de la spéculation [Speculation theory]. *Annales scientifiques de l'École Normale Supérieure, 3*(17), 21-86. https://doi.org/10.24033/asens.476

Balaji CH, Kusuma GDV, Kumar BR. (2018). Risk and return analysis of equity shares of selected companies in automobile industry. *Asian J Mult-Disciplinary Res, 4*(2), 37-42. https://doi.org/10.20468/ajmdr.2018.02.09

Biswas, A. (2018). Effect of IT stock prices on Nifty 50 - an empirical analysis on Indian stock. *International Journal of Science and Research (IJSR)*, 1394-1401.

Blattberg, R. C., & Gonedes, N. J. (1974). A comparison of the stable and student distributions as statistical models for stock prices. *The Journal of Business, 47*(2), 244-280. https://doi.org/10.1086/295634

Borowski, K. (2018). Testing 65 equity indexes for normal distribution of returns. *Journal of Economics and Management, 34*(4). https://doi.org/10.22367/jem.2018.34.01

Fama, E. F. (1965). The Behavior of Stock-Market Prices. *The Journal of Business, 38*(1), 34–105. https://doi.org/10.1086/294743

Gautami, D. S., & Kalyan, D. N. B. (2018). A Comparative Study on Risk & Return Analysis of Selected Stocks in India. *International Journal of Management and Economics Invention, 4*(05), 1730-1736. https://doi.org/10.31142/ijmei/v4i5.03

Hagerman, R. L. (1978). More evidence on the distribution of security returns. *The Journal of Finance, 33*(4), 1213-1221. https://doi.org/10.1111/j.1540-6261.1978.tb02058.x

Mandelbrot, B. (1967). The Variation of Some Other Speculative Prices. *The Journal of Business, 40*(4), 393–413. https://doi.org/10.1086/295006

Nasdaq. (n.d.). *Nasdaq internet index methodology* [Fact sheet]. Nasdaq. Retrieved December 15, 2021, from https://indexes.nasdaq.com/docs/methodology_QNET.pdf

Osborne, M. F. (1959). Brownian motion in the stock market. *Operations Research, 7*(2), 145-173. https://doi.org/10.1287/opre.7.2.145

Patel, D., & Surti, D. M. (2020). A study on relationship between FMCG sector and Nifty fifty. *International Journal of Creative Research Thoughts (IJCRT)*, 1042-1050.

Praetz, P. D. (1972). The distribution of share price changes. *The Journal of Business, 45*(1), 49-55. https://doi.org/10.1086/295425

Rane, N., & Gupta, P. (2021). Impact of financial ratios on stock price: Evidence from Indian listed. Revista Geintec-Gestao Inovacao E Tecnologias, 5132-5144.

Seabold, Skipper, & Josef Perktold. (2010). statsmodels: Econometric and statistical modeling with python. *Proceedings of the 9th Python in Science Conference*. https://doi.org/10.47059/revistageintec.v11i4.2553

Sonia Lobo, & Ganesh Bhat, S. (2021). Risk Return Analysis of Selected Stocks of Indian Financial Sector. *International Journal of Case Studies in Business, IT, and Education (IJCSBE), 5*(2), 111-124. https://doi.org/10.47992/IJCSBE.2581.6942.0124

Srivastava, H. (2020). A study of the relationship between various NIFTY sectoral indices. *International Journal of Advanced Science and Technology*, 2885-2898.

Tretina, K. (2021, July 22). *What are the Nasdaq composite and Nasdaq 100 indexes?* (B. Curry, Ed.). https://10.35940/ijrte.C3850.098319

Sushma K. S., Charithra C. M., & Bhavya Vikas. (2019). A Study On Risk and Return Analysis of Selected Financial Services Companies Listed on NSE. *International Education and Research Journal (IERJ), 5*(7). Retrieved from http://ierj.in/journal/index.php/ierj/article/view/1836

Zaimovic, Azra, Adna Omanovic, & Almira Arnaut-Berilo. (2021). How Many Stocks Are Sufficient for Equity Portfolio Diversification? A Review of the Literature. *Journal of Risk and Financial Management, 14*, 551. https://doi.org/10.3390/jrfm14110551

Copyrights