Global Scientific JOURNALS

GSJ: Volume 9, Issue 3, March 2021, Online: ISSN 2320-9186
www.globalscientificjournal.com

CAPSTONE FINAL REPORT

MScFE CAPSTONE PROJECT

TOPIC:

STATISTICAL CHARACTERISTICS OF MARKETS

AUTHORS:

Fatihullah Akilu (akilufatihulla@outlook.com); Theophilus Addae (adei.blesson@gmail.com); Yakubu Wurishe Issah (issahyw@yahoo.com)

Date: 26th October, 2020

GSJ: Volume 9, Issue 3, March 2021
ISSN 2320-9186

Abstract

Modern markets today are centralized places of trading, created to enable people advertise their needs and offerings on a more public domain, like the New York Stock Exchange (NYSE), London Stock Exchange (LSE), South African Stock Exchange (SASE), Ghana Stock Exchange (GSE) etc.

These markets exhibit some elements of stationarity and efficiency, while at the same time they are subject to external shocks and periodic regime of shifts. This therefore necessitates and calls for occasional reassessment of their statistical characteristics. The statistical characteristics of markets identify market trends which help measure and evaluate market potential and its success. A market success depends on identifying the target market accurately and using effective tools or methods to interpret it. These statistical methods enable collection of data, use of correct analysis and effectively producing the needed results.

Using many and varied tests of statistical and econometrics properties will draw very important insights into market behavior, and will serve as conduit for decision-making by the market participants and a reference point for researchers.

Goals sought to be achieved in this project are: To identify market datasets; Construct various tests to examine price, yield, and index properties of the markets; Implement various exploratory data analysis techniques to document the properties. The objectives sought are: Collect and collate data of prices, and yields for a representative section of instruments from stock market indexes including MTN, and JPN, spanning at least for a period not less than 10 years; Carryout extensive exploratory data analysis of these market indexes, reviewing statistical moments, correlations, and behaviors using graphical measurements; Analyze returns of these market indexes in various time periods, determine the extent to which these markets have experienced regime shifts, and document the evidence; Summarize and draw conclusions from the findings.

In our research, we have used data from two giant markets around the world like the American, Asia and African markets to ascertain how regime shifts and external shocks affect the behavior of these market trends. Furthermore, we were able to ascertain how these properties affects markets with higher daily returns like the JPN and MTN. We were also able to see how the novel COVID-19 pandemic really affected the market trends.

The use of statistical analysis has played a major role not only in financial markets setting but also in the areas of manufacturing industry, quality management process and other sectors for a very long time. To be able to analyze data reliably and get the desired results, collecting of these data was paramount. This paper primarily researched into the use of statistical tools to analyze the behavior of some specific markets, in this case, MTN, and JINFENG. The focus was on trends over a time and in this research, over 10 years, especially during the era of COVID 19 pandemic. The objective of the paper was to analyze the market behavior using data collected and identify the statistical characteristics associated with the markets for informed decisions in planning for the future.

Introduction

Since the days of barter trade, running through to today's modern, complex, regulated and technology driven financial markets, the value of goods and services have always been tied to the desire to satisfy a need or better put the demand and supply dynamics. The eventual provision of a legal tender or money as a medium of exchange, resulted in an explosion of trade in goods and services and consequently a plethora system, processes and technologies which is now known as the global financial markets.

The sophistication of the financial markets through regulation and technologies has also expanded the number of participants to include governments, central banks, investment banks, insurance and re-insurance companies, clearing houses, and dealers or intermediaries or brokers who represent other parties such as public investors. Modern day financial markets are ever-functioning, ever-expanding, and deeply interconnected, even to the extent that their interconnection is sometimes detrimental to the markets.

The advancement of the markets has been replicated in their present functions. The primary reasons participants go to the market are: to raise capital, earn profit, and manage risk. Raising capital entails competitive repayment terms i.e. lower interest rates, more favorable payment mechanisms i.e. short and long-terms, and better capital liquidity i.e. ease of access to cash. The profit incentive drives investors to put their money to best possible use. The management of risk is important because every investment carries some degree of risk, and the higher the risk, the higher the premium paid.

The current dynamics of the market has called for the development and application of mathematical and statistical tools and models for participants in the market to understand the trends and how best to stream best practices and in decision making regarding their investments.

Statistics isn't new in human development. It's been used before three hundred (300) BC as a tool for state body machinery. However, in an exceedingly restricted method. It is derived from the Latin word 'Status', in Italian 'Statista', in German 'Statistik', and in French 'Statistique', which suggests a political state. Around this point there have been some body activities geared towards the collection of information for births and deaths. Within the sixteenth century statistics was applied to the study of celestial bodies resulting in the prediction of eclipses. The emergence of important statistics was within the seventeenth century initiated by Captain John Grant from 1620-1674 in London. He introduced what was called "the systematic study of births and death statistics" (Di Sha. "Statistics: Meaning, Characteristics, and Importance). Extraordinary contributions have been made towards the great strides achieved in statistics and its application in several disciplines. They embraced Francis Galton (1822-1921) who studied multivariate analysis in biometry; Karl Pearson (1857-1936) who supported the best statistical laboratory in a European country and studied correlation analysis resulting in his picture discovery of the Chi-Square Test; and W.S. Gosset along with his t-test. For the past decades applied mathematics development has been attributed to Sir Ronald A. Fisher (1800-1962). He is honored to have applied mathematics to numerous fields as well as scientific disciplines like agriculture, biometry, education, etc. He additionally invented the estimation theory and precise little sampling distributions about analysis of variance and style of experiments.

The application of statistics and other mathematical microeconomic models are necessary to stem the shifts or movements occurring in the markets for participants to best understand when and where to put in their investments. In the financial markets there are various sub-markets in which different assets are traded. Financial markets refer to markets where entities raise funds. The capital markets are used to raise long-term finances, whereas the money markets are meant for raising short-term finances. The financial markets can also be categorized into the following types: Capital markets which embraces the stock markets and the bond markets; money markets; derivatives markets; commodities markets; futures markets; Cryptocurrency markets; interbank lending markets; and the spot market (Wikipedia: “Financial Markets”).

The capital markets consist of the stock markets which issues shares and common stocks to mobilize capital, and the bonds market which issue bonds to raise capital. The commodities markets are involved in the trading of commodities. Money markets provide short-term debt financing and investments. The derivatives markets provide instruments to manage risk, whereas the futures markets provide standardized forward contracts for trading products against the future. The cryptocurrency markets facilitate the trading of digital assets and financial technologies. The interbank lending markets is a place where banks lend money to one another for a specified period of time. The spot markets deliver the commodities traded to the clients immediately the deal is done. It differs from the futures markets which delivers its products at a future date.

Functionally, practitioners in the financial markets have broadly categorized the markets into the sell and buy sides in relation to the activities of the markets. The distinction is significant in the understanding of the behavior of the markets. The buy-side refers to the activities of the players whose primary motive is buying and holding financial investments to make profit. Investors whose primary objective is to make profit will only participate in the buy-side of the financial market. Broadly, the buy-side players consist of entities with capital to invest and are seeking to utilize it by investing in the financial markets. On the other hand, the sell-side refers to market activities that cater for the buy-side players. That is, they sell investment products and services to the buy-side, to assist in making profit.

Let’s consider the buy-side entities. The first is the Mutual Funds, which are collective investment schemes, which many investors pool their resources together to invest and share the accrued profits. The fund manager consists of a team who earn fees for their services and possibly earn a percentage of the profit accruing to the investors. Another player in the buy-side a Pension Funds and private investors. A pension fund receives pension contributions and invests the contributions to earn enough profits in order to pay contributors their pension benefits in the future. Also, insurers or insurance firms are buy-side entities.

Apparently, fund managers play an important role in the buy-side. As part of their role, they are supposed to establish and maintain reputations of good investment performance and advice. It represents a very competitive industry advisors and consultants on investment scale of performance. Sovereign wealth funds are also on the buy-side of the market. They are investment funds managed by national governments, where resources are invested so support governmental projects and programs. Hedge funds are on the buy-side, and are collective investments aimed at earning profits.

On the other hand, the sell-side entities comprise the money, equity, bond and derivative markets, where the buy-side entities invest. It is however important to mention that, institutional investors have many advantages over smaller retail investors, on the basis of economies of scale. Another reason is that, institutional investors are able to diversify to balance off some of risks and losses incurred.

Investing in shares entails a great deal of practical challenges and to overcome them, it is better to purchase the stocks indirectly, through a stock-broker. They are able to easily work through the complicated rules of the exchange. Thus, the asset manager will purchase the stocks indirectly through the stock-broker. The concept of brokerage is fundamental in the sell-side of the financial markets. Another sell-side player is the broker-dealers. They offer the sell service to potential investors and at the same are able to trade for themselves. Dealers are also players in the sell-side who buy and sell when the opportunity comes. Furthermore, Investment banks are one of the important sell-side players.

After a careful analysis of the stakeholders in the financial markets we can have a SWOT grid of the resemblance of that shown below:

SWOT GRID

STRENGTHS	WEAKNESSES
• Technology • Regulations • Professionalism • Transparency • Data Driven • Continuous Research	• Protectionism • Speculation
OPPORTUNITIES	THREATS
• Innovations • Emerging markets • Easy access to Capital • Investor confidence	• Fraud • Insider trading • Alarmists • Human errors • Technical failures

The above listed dynamics in the SWOT grid is what drives individuals, organizations and governments to making incessant attempts through technology and regulation to overcome the mistreatments of the weaknesses and threats in the markets.

Theoretical Framework

Statistics can be defined as "a mathematical science concerned with data collection, presentation, analysis, and interpretation" (Livio English Dictionary, Offline). Statistics also relates to the quantifying of data in a sample which reflects a significant characteristic of an aspect of the sample, e.g. mean or deviation.

These definitions give us a fair idea as to what our research seeks to achieve in the process. Statistical characteristics of markets is geared towards understanding the characteristics options of the monetary markets mistreatment through applied mathematics models. Several and varied applied mathematics properties or characteristics are known and systematic modeling strategies are designed to spot and use these characteristics to work out the trends and behavior of the markets. A number of these statistical characteristics are explained in the following paragraphs below.

One assumption created concerning the markets is the construct of stationarity. To create an applied mathematical inference concerning the structure of a process, method, model, or theoretical framework, on the premise of a discovered knowledge of the process, involves creating simplifying assumptions concerning the structure. The essential plan of stationarity is that the likelihood laws that govern the behavior of the method don't modify over time. That is, the method is in an exceedingly applied mathematical equilibrium. A stationary process has a joint distribution of the variables for the selected time points. A typical example of stationarity is a sequence of stand-alone and identically distributed random variables, usually termed as White noise.

However, trends at the markets as delineated by applied mathematical knowledge may be elusive. A similar statistic may be viewed otherwise by totally different analysis. This is often true of non-stationary statistic, that states that any statistic while not a continuing mean over time is non-stationary. Models of the form, wherever could be a non-constant mean operate and could be a zero mean in an exceedingly stationary series.

Another statistical characteristic of markets is the construct of improvement or optimization. Optimization is a useful antidote for decision making and for the understanding of physical systems.

According to Cont (2001), the markets are auto-correlated in absolute values. That is, absolutely the price of returns series tends to be auto-correlated. However, this may not happen with ordinarily distributed variable. He conjointly highlights the leverage result of the markets. That is, there tends to be correlational statistics between the volatility returns of an asset. Moreover, he states a correlation between volume and volatility i.e. commercial volumes tends to be correlative with any active volatility. Also, the markets exhibit uneven behavior over totally different time scales i.e. volatility measures supported low frequency knowledge with higher at predicting those of frequency knowledge than the rests. Other observable statistical characteristics of the markets is related to continuous random variables and their distributions. Most of the time financial engineers are faced with return series on assets of different sorts and these returns are appropriately modeled as continuous random variables that can take on any value within their range. One exception, however, is tick-by-tick data or an asset, because they are fundamentally discrete observations which require a completely different approach. We describe a continuous random variable X by its cumulative distribution function (CDF). The function postulates the probability that a random draw of the variable will be less than or equal to some value q, which is called a quantile. The main feature of the CDF is that, it is a non-decreasing function between 0 and 1 as expressed in the equation below:

$$\lim_{q \to \infty} F_x(q) = 0$$

$$\lim_{q \to \infty} F_x(q) = 1$$

$$F_x(q) \geq F_x(q) for all\ q_i > q_j$$

When the CDF is to be differentiated, it describes the randomness in the variable with the first derivative of the CDF, which we call the probability density function (PDF) represented mathematically as:

$$F_x(q) = \frac{\delta F_x(q)}{\delta_q}$$

The PDF measures the marginal likelihood of an observable data q.

Above all is the normal distribution which stands central in statistics. Various central limit theorems exist, showing that under some regularity conditions, estimation like sample means converge in distribution to normally distributed random variables. Another type of distribution is the t-distribution, which is essentially a distribution used for inference about some estimated mean, such as regression coefficient, when we estimate the variance. This additional uncertainty indicates that t-distribution has fatter tails i.e. it puts more probability weight on extreme events than the normal distribution. They are also called parametric distributions because they are fully described with a small number of parameters and a known functional form.

On the other hand, empirical distribution also addresses the question of how close any data series is to a given parameter. It is a statistical description of the randomness in a sample that places the least structure on the data. Also, joint, marginal and conditional distributions address the financial engineer’s interest in the co-movement of asset prices. That is, the lower the correlation between two or more assets, the greater the opportunity of reducing the risk of a portfolio invested in multiple- assets.

In addition, the standard multivariate normal distributions define the point of evaluation, the vector of means, and the covariance matrix. On the other hand, the multivariate t-distribution describes the randomness in tow correlated variables that have t-distributions as marginal distribution.

GSJ: Volume 9, Issue 3, March 2021
ISSN 2320-9186

The Support Vector Machine regression model

The SVM is a data-classification algorithm for predictive analysis, which assigns new data elements to one of the labeled groups. In most instances SVM is a binary classifier; the data in question is considered to have two potential target values. as a regression model, it uses a method known as a kernel trick to convert the data and then seeks an optimal boundary between potential outputs depending on these transformations.

The SVM has three hyper-parameters worthy of consideration.

i) The Kernel

A kernel lets one find a high-dimensional hyperplane without increasing computing costs. In general, the expense of processing increases as the data dimension increases.

ii) The Hyperplane

This is basically a line that isolate between two data classes in SVM. But this is the line that will be used to estimate the continuous performance in Support Vector Regression,

iii) The Decision boundary

A judgment limit can be regarded as a demarcation line with positive instance lying on one side and the other side lies the negative instance. On the very same line, the instances can be categorized as either positive or negative. The same concept of SVM will also be applied to support vector regression.

Support Vector Regression (SVR) operates under similar concepts to the description of the Support Vector Machine (SVM). We may assume that since the dependent variable is numerical rather than categorical, SVR is the modified variant of SVM. A big advantage of using SVR is that it is a process that is non-parametric. The performance function from the SVR does not rely on the distributions of the underlying dependent and independent variables, unlike other models. The SVR approach relies on kernel functions instead. Another advantage of SVR is that it allows a non-linear model to be built without modifying the explanatory variables, helping to further understand the resulting model. As long as the error (εi) is less than a certain value, the fundamental concept behind SVR is not to worry about the forecast. It is known as the principle of maximum margin.

This notion of maximal margin makes it possible to see SVR as a problem of convex optimization. Using a cost parameter, the regression can also be penalized, which becomes useful to prevent over-fit. In terms of the distribution of underlying variables, the relationship between independent and dependent variables and the regulation of the penalty term, the SVR is a valuable tool that provides the user with high flexibility.

The Buildup of SVR Model

1. Collect your training data $\{X, Y\}$
2. Choose a kernel and parameter and regularization if need be
3. Create the correlation matrix

$$k_{ij} = \exp\left(|\sum_k \theta_k |x_k^i - x_k^j|^2 + \epsilon\delta_{ij} \right)$$

4. Train our data by using the key part of the algorithm to get a contraction coefficient approximately or accurately.

$$\bar{K}\vec{\alpha} = \vec{y}$$

Where $\vec{y}$ the vector of the values corresponding to the training dataset

$\vec{\alpha}$ is the set of unknowns we need to solve for, thus;

$$\vec{\alpha} = \bar{K}^{-1}\vec{y}$$

5. Use the coefficient to create an estimator. However, , we need to compute the correlation vector $\vec{k}$ to estimate y^* value for a test $\vec{x}^*$ point

$$y^* = \vec{\alpha} . \vec{K} ; \quad K_i = \exp\left(\sum_k \theta_k |x_k^i - x_k^*|^2 \right)$$

In linear regression, as the objective is to minimize the error between the forecast and the results, the objective in SVR is to ensure that the errors do not surpass a given threshold.

Support Vector Regression

SVR is one of linear regression models used to minimize the total of squared errors. Using ordinary least spread as an example predictive equation is as follows:

$MIN \sum_{i-1}^{n} (Y_i - W_i X_i)^2$

SVR-FTW

SVR has room for defining the error limit acceptable in the model and will find the hyperplane in higher dimension to fit the data. The SVR is used to minimize the coefficients but not the square root error. In doing so, the error limit is constrained to an absolute error limit of less than or equal to a determined margin called the maximum error or Epsilon (Ɛ). The Epsilon can be regulated to achieve the desired accuracy in the data. The equation is given as follows:

Minimize:

$MIN ½ \left| |W| \right|^2$

Constraints:

$\left| Y_i - W_i X_i \right| \leq \varepsilon$

We can also use slack to minimize to minimize any hyper-parameter in the variable which is outside the Epsilon. That deviation is given as a slack ($\boldsymbol{\xi}$). The equation can be given as:

Minimize:

$$MIN\tfrac{1}{2}\left|\,|W|\,\right|^2 + C\sum_{i=1}^{n}|\xi_i|$$

Constraints:

$$\left|Y_i - W_i X_i\right| \leq \varepsilon + \left|\boldsymbol{\xi_i}\right|$$

Thus, SVR is a power mathematical and computational tool used to determine the tolerant error margin (Ɛ) or outside the acceptable error margin ($\boldsymbol{\xi}$).

Volatility Measurement

Volatility is the rate at which the price of a security with a given series of returns rises or decreases. Volatility is calculated by measuring the standard deviation for a fixed time span with annualized returns. This shows the degree to which the price of a defense can rise or decrease Stochastic volatility modeling is applicable when the variance of a stochastic process is randomly distributed. They are used to analyze derivative securities, such as options, in the field of mathematical finance. The name derives from the model's treatment of the volatility of the underlying security as a random mechanism, controlled, among others, by state variables such as the underlying security price level, the propensity of volatility to return to any long-run mean value, and the variance of the volatility process itself.

One approach to overcoming a flaw in the Black-Scholes model is stochastic volatility models. In fact, Black-Scholes-based models presume that the underlying variance is constant over the existence of the derivative and is unaffected by changes in the underlying security price level. These models, however, do not justify the long-observed characteristics of the implied volatility disk, such as the smile and skew of volatility, which suggest that implied volatility appears to differ in terms of strike price and expiry. It becomes possible to model derivatives more effectively by assuming that the variance of the underlying price is a stochastic phenomenon rather than a constant. Popular volatility modelling techniques include; Heston model, CEV model, SABR volatility, 3/2 model and Garch model. This paper will only focus on the Garch model.

Heteroscedasticity

As a set of observations chosen independently and from similar distributions, we may think about an idealized time series. It is as simple as estimating the parameters of the distribution to understand this sequence. The notion behind stationarity is this. We can learn the parameters of distribution and forecast once a time series is stationary, but it is rarely that simple. This kernel is about time series that show heteroscedasticity, specifically heteroscedasticity conditionality, and the models we use to predict them.

Heteroscedasticity is when the variance of the underlying distribution used to construct our time series changes as a function of time if we return to the idealized time series. A common appearance of heteroscedasticity is a time series whose variance, growing outward in a conal pattern, increases with time. A Box-Cox transformation may often mask small amounts of heteroscedasticity, which makes the distribution of the data more natural. Not only are our variance shifts too high to mask in our MTN and JIFENG forecasts, but the variance at step x_t is conditional on the variance of past time steps, x_{t-1}. This condition is termed conditional heteroscedasticity or volatility clustering.

Volatility clustering

The Garch Model

The GARCH model is an economic concept created in 1982 by Robert F. Engle, economist and recipient of the Nobel Prize of the Memorial for Economics in 2003. An approach to estimating volatility in financial markets is defined by GARCH.

By having the variation to depend on its own lags and the lags of the squared residuals, GARCH extends ARCH. GARCH will identify larger shifts, such as volatility increasing and decreasing.

$$\sigma_t{}^t = \alpha_0 + \alpha_1\mu_{t-}^2 + \alpha_2\mu_{t-2}^2 + \cdots + \alpha_q\mu_{t-q}^2 + \beta_1\sigma_{t-1}^2 + \cdots + \beta_p\sigma_{t-p}^2$$

For GARCH (q, p) we have two parameters:

q: Number of variances on lag
p: Number of residual errors with lag

The parameters that we are using here are compatible with the ARCH python library, but the switched notation is used more frequently.

When there's heteroscedasticity and volatility clustering, GARCH always suits data better than ARCH. For both GARCH and ARCH, we will do an analysis and explain the findings, although there are also differences in the GARCH model that add new criteria for unique volatility behaviors.

Model Fit

We want to check to determine how well our ARCH / GARCH model fit our data:

- Autocorrelation in the Standard residuals, we should use the Ljung-Box test again.
- We would use Engle's ARCH test on the residuals for ARCH effects (conditional heteroscedasticity) in the residuals.
- We will use the Shapiro-Wilk test, Q-Q map, and, if larger, the n-Jarque-Bera test to see if our data approaches normality in the normalized residuals.

Engle's ARCH Test

H_o: Squared residuals are a series of white noise-homoscedastic residuals.
H_a: It was not possible to match the squared residuals for a linear regression model and show heteroskedasticity.
To decide if our ARCH model has captured the conditional heteroscedasticity of our time sequence, we will use the ARCH Engle test in python, which is a hybrid of Ljung-Box and a Lagrange Multiplier test. The array of residuals and a max latency to use would be transferred to it.

Test Shapiro-Wilks

A data sample is assessed by the Shapiro-wilks test and quantifies how likely it is that the data was drawn from a Gaussian distribution.
Ho: The data is normally distributed
Ha: The data is not normally distributed

Test Jarque-Bera

A type of lag range multiplier test for normality is the Jarque-Bera test. It is generally used with large data sets and when n is large, Shapiro-Wilk is not accurate with n greater than 2,000, other normality checks are not accurate. In particular, Jarque-Bera mathematizes skewness and kurtosis to a normal distribution.

In conclusion, using statistical analysis in different fields enables informed decisions to be made. Accurate analysis and interpretation of data has been used for future predictions and forecasting. Statistical analytical techniques have been used in the area of research and solving complex problems relating to the economy and other vital areas.

Methodology

We focused our research on two market players, namely MTN and JINSENG. These two organizations are floated at various stock markets around the world and their market index were accessed through the internet for this project. The data we collected embraced historical data spanning 10 years to observe and analyze their trends using statistical tools to make insights into the statistical characteristics and behavior exhibited in the markets.

Very robust data analysis software and applications were used to analyze our data. Specifically, we used Python applications to analyze historical data from these markets on the two market players, namely MTN and JINSENG. The historical data spanned from a period of ten (10) years.

The analysis targeted the trends such as stationarity, regression modelling, volatility modelling, coupled with other observable characteristics and assumptions necessary to give us the desired results.

We basically collected the data from the internet, specifically from data sources like Yahoo finance and various indexes, using different devices like desktops, laptops, tablets, and smartphones. We also relied heavily on secured and stable internet connectivity using data bundles.

Results

The table below gives a view of the first 5 cases of the dataset downloaded from Yahoo Fin

```
#viewing mtn dataframe
mtn.head()
```

Date	Open	High	Low	Close	Adj Close	Volume
2006-06-30	10.17950	10.17950	9.81410	10.07690	8.897742	4536194
2006-07-03	9.92949	10.00000	9.67949	9.85256	8.699651	6530715
2006-07-04	9.83974	9.96154	9.61538	9.65385	8.524195	8103962
2006-07-05	9.61538	9.64744	9.01923	9.23077	8.150622	9460688
2006-07-06	9.22436	9.40385	9.11538	9.32051	8.229862	9436396

```
#viewing jifeng dataframe
Jifeng.head()
```

Date	Open	High	Low	Close	Adj Close	Volume
2006-06-30	10.17950	10.17950	9.81410	10.07690	8.897742	4536194
2006-07-03	9.92949	10.00000	9.67949	9.85256	8.699651	6530715
2006-07-04	9.83974	9.96154	9.61538	9.65385	8.524195	8103962
2006-07-05	9.61538	9.64744	9.01923	9.23077	8.150622	9460688
2006-07-06	9.22436	9.40385	9.11538	9.32051	8.229862	9436396

Figure 1. Display of MTN and JIFENG data set

Moving forward, we explore the descriptive statistics of both our data frame, this is shown in the table below

```
#mtn data frame descriptive statistics
mtn.describe()
```

	Open	High	Low	Close	Adj Close	Volume
count	3590.000000	3590.000000	3590.000000	3590.000000	3590.000000	3.590000e+03
mean	104.984415	106.249981	103.698841	104.994861	95.669072	3.363535e+05
std	76.955390	77.721595	76.164999	76.929716	75.703295	2.391296e+05
min	15.250000	16.080000	14.760000	14.800000	12.126607	3.540000e+04
25%	44.200001	44.902501	43.502500	44.279999	36.404659	1.922500e+05
50%	66.494999	67.240002	65.974998	66.445000	56.861877	2.771500e+05
75%	165.965000	169.777496	163.532501	166.862495	159.112286	4.042500e+05
max	298.950012	302.760010	296.809998	301.420013	284.857513	3.219400e+06

```
# JIfeng dataframe descriptive statistics
Jifeng.describe()
```

	Open	High	Low	Close	Adj Close	Volume
count	3473.000000	3473.000000	3473.000000	3473.000000	3473.000000	3.473000e+03
mean	8.535370	8.711044	8.372236	8.546128	7.959184	9.266162e+06
std	3.556526	3.651931	3.459879	3.556710	3.075555	8.231778e+06
min	3.130770	3.292310	3.130770	3.238460	3.219353	0.000000e+00
25%	6.400000	6.500000	6.300000	6.392310	6.079314	3.914303e+06
50%	8.044870	8.169230	7.907690	8.061540	7.572840	6.705148e+06
75%	9.846150	10.038500	9.679490	9.853850	9.140792	1.192121e+07
max	23.352600	23.352600	21.955099	22.423100	19.959593	7.675498e+07

Figure 2. Descriptive statistics of MTN and JIFENG data set

Next we plot the close price of both our stocks to view the dynamics of price action. It could be seen that the close price of MTN rises gently upwards and start experiencing a sharp fall toward the end of 2018 and rises slowly into the 2020 before experiencing a sharper fall in the early part of 2020 which could be attributed to the corona negative impact on the market. Also, the JIFENG close price plot gives an uncertainty attributes to its market. The graph depicts varying levels of sharp rises and sharp falls spanned across the timeframes. We could see the sharp rise in prices from the early part of 2017 towards the end of the year where it experienced a reversal trend with a steeper decline in price toward the beginning of 2009, from then it experienced some level of consolidated price level till the later part of 2014 when it experienced a sharp increase in stock price and then falls back in 2016. From 2016 we could see a gentle drop in prices till the year 2020, when it experienced a reversal trend.

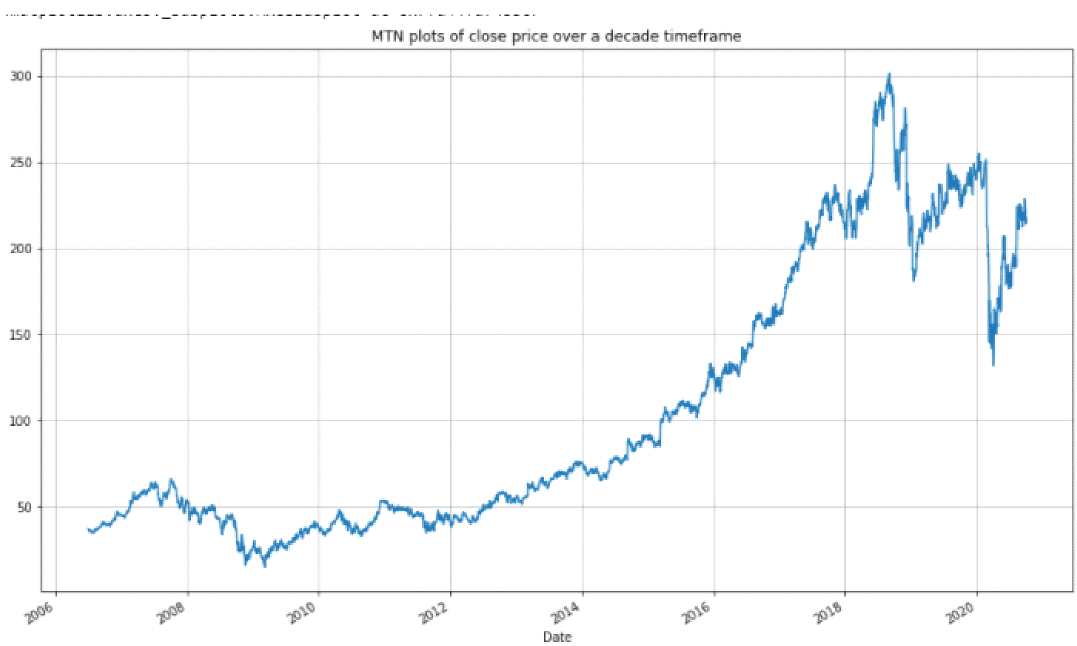


Figure 3. Plot of MTN rise over the past decade

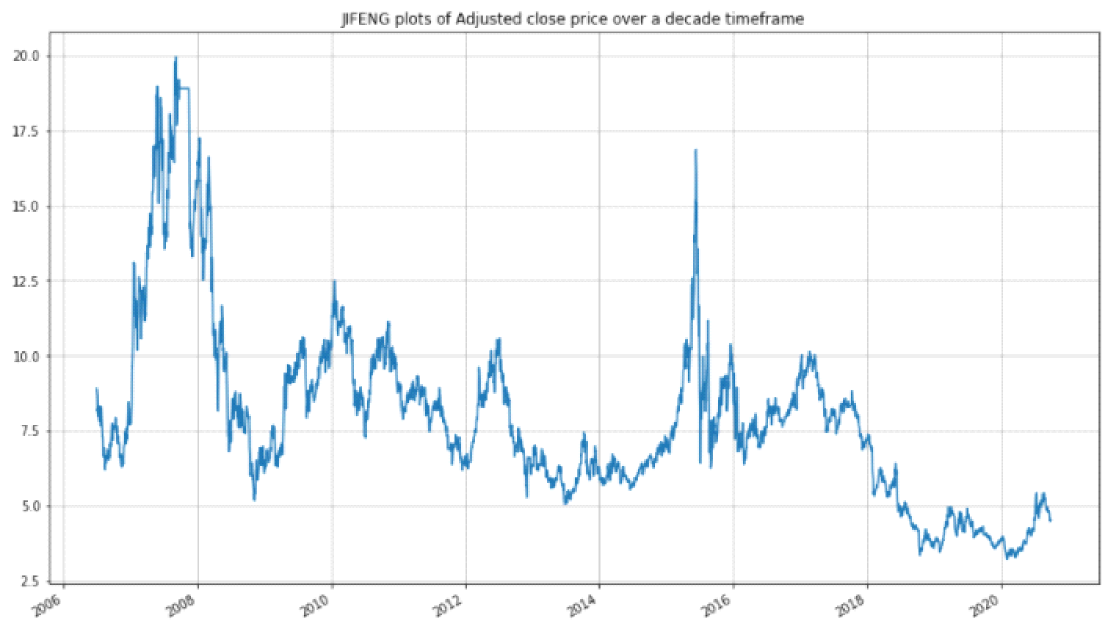


Figure 4. plot of JIFENGS Close Price over the past Decade

GSJ: Volume 9, Issue 3, March 2021
ISSN 2320-9186

Even though the plots above give a fair idea of the dynamics of price action we are limited because not all the features of our dataset is included in the plot above, we thus would want a graph that incorporate the Open, High, Low and Close features of our datasets and no plots does this better than the candlestick plot. Below is the candlestick plot of MTN and JIFENG over the past decade

Figure 5. MTN candlestick

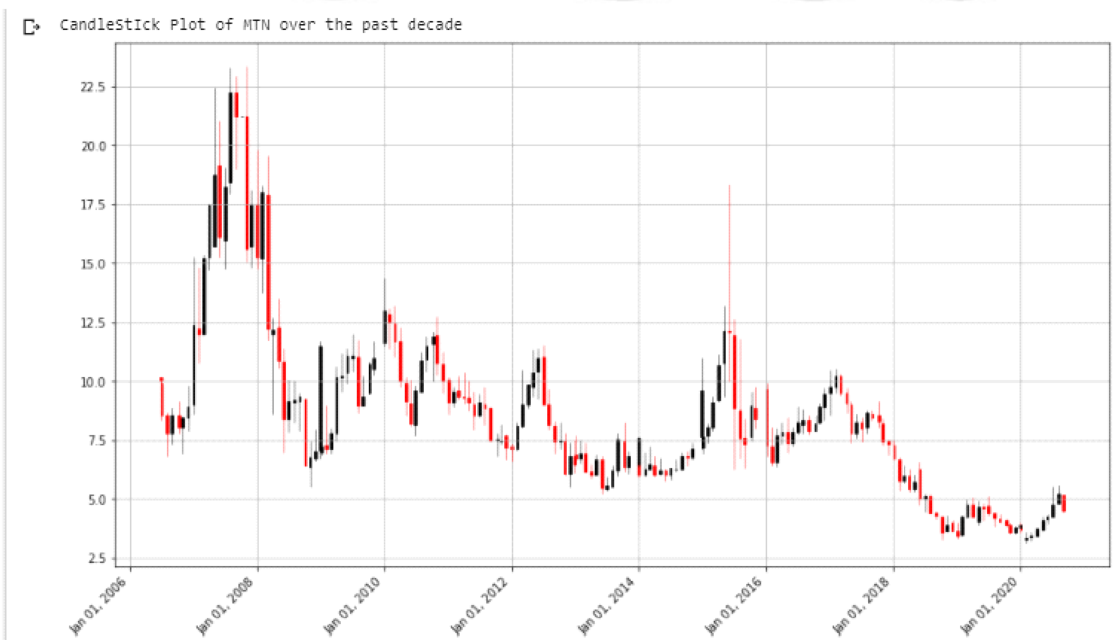
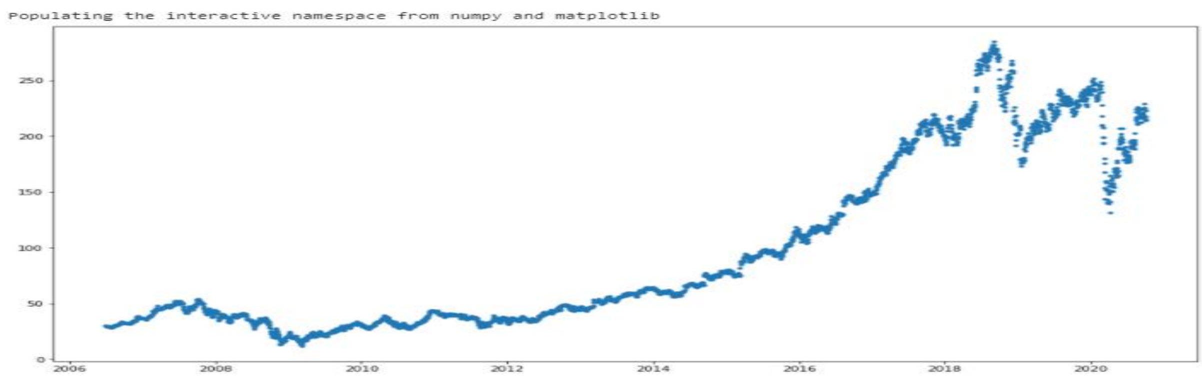


Figure 6. JIFENG candlestick plot

Problems of financial data

Putting aside all the comprehensive mathematical description, what stationarity means is that the statistics of the underlying signals (MTN and JIFENG) are stable over time. Indeed, the concept of stationarity though is, but not a major problem in the field of finance, by applying a difference we can attain stationarity. However, the biggest issues in the field of finance are rather; structural breaks, missing data and outliers. The data used for this work has been tested for missing or null data points and we fortunately don't have missing data points. Also, from the scatter plot below there is no evident of a data point completely separated from the trends hence we could conclude of the absence of extreme outliers and acknowledging the presence of mild outliers which could be considered when modelling.

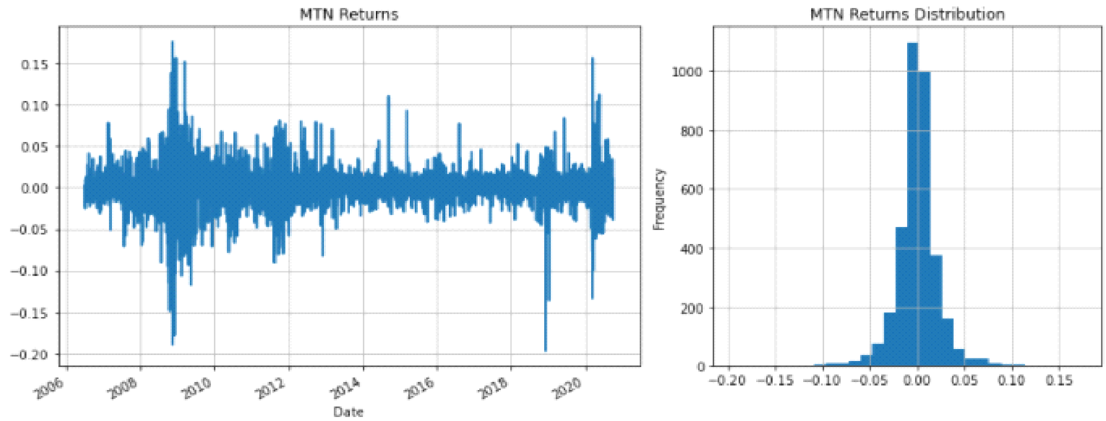


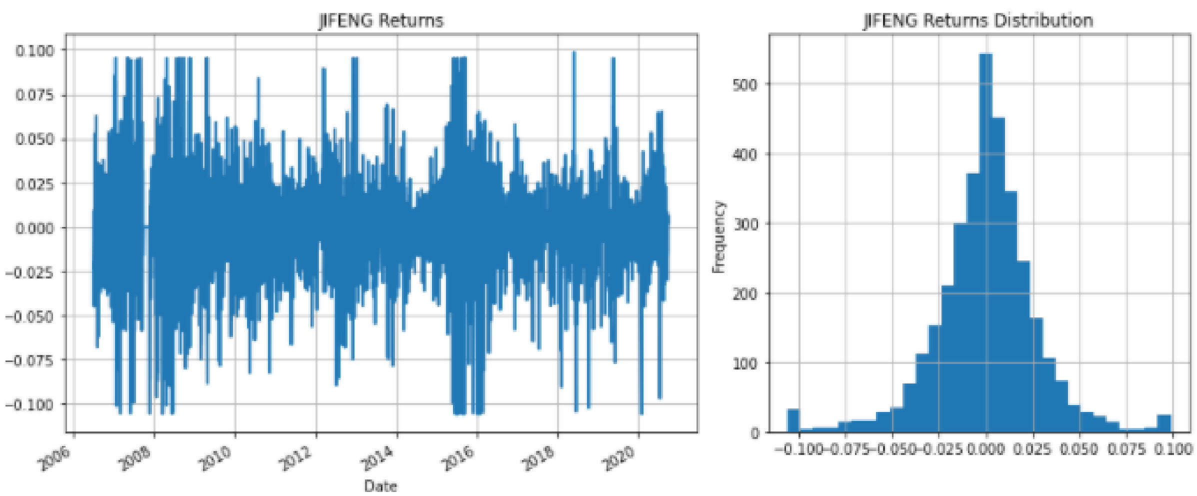
Log returns of MTN and JIFENG

Log returns are a statistical method that seeks to differentiate between log values at time $t\ -1$ and$t$, which is why it is often referred to as log-Differencing.

$$rt = \frac{\ln(p(t))}{\ln(p(t-1))} = \mathrm{LN}\big(p(t)\big)\,/(p(t-1))$$

The aim of this process is to generate stationary signals suitable for Machine learning and other classical statistics. Evidence of stationarity of the log returns is seen from the return plots OF MTN and JIFENG below, one could see that our Returns plot has a constant mean around zero and a more constant standard deviation.

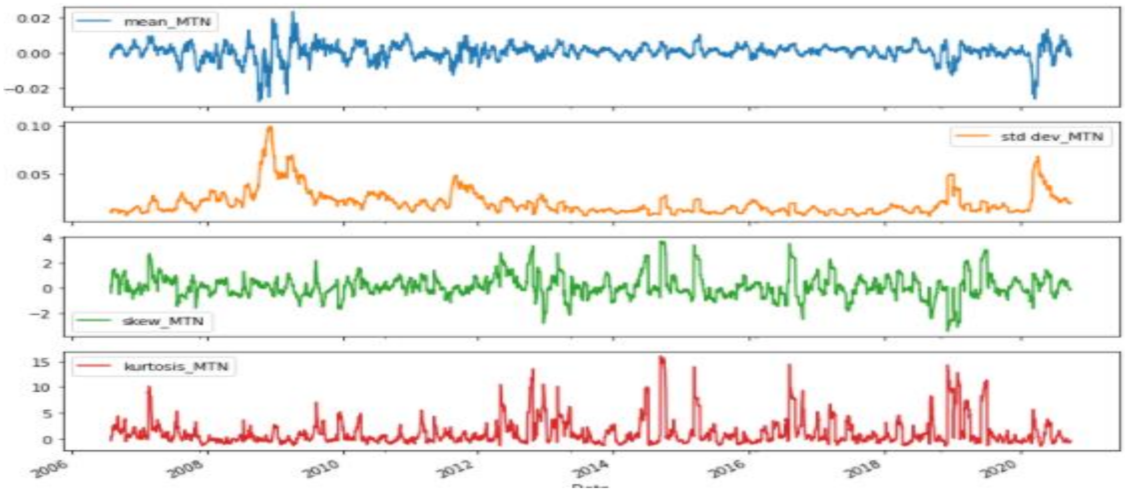




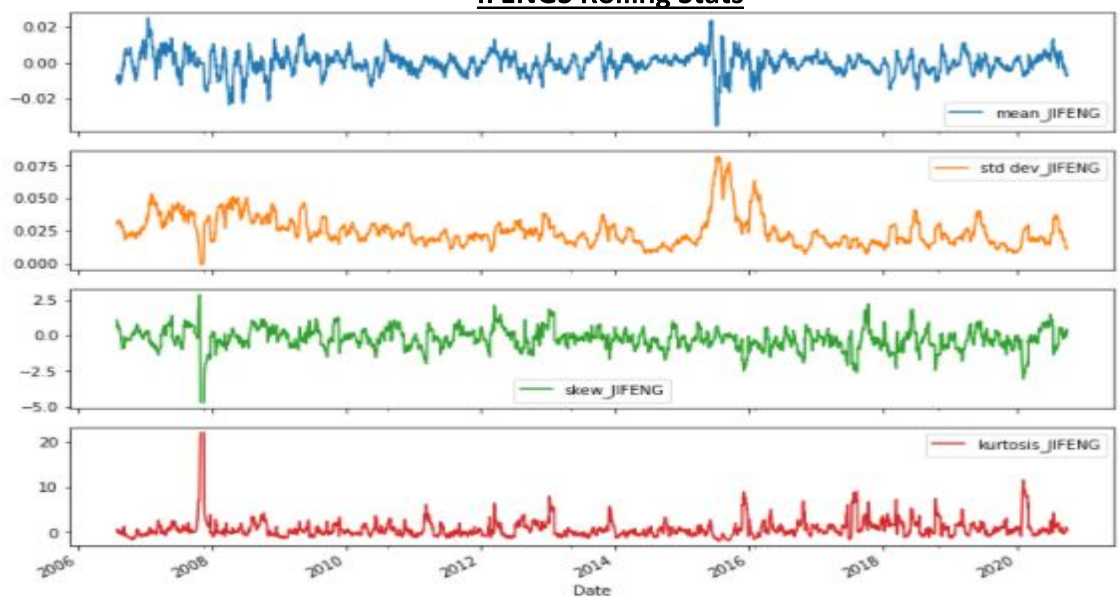
Rolling Statistics of our returns

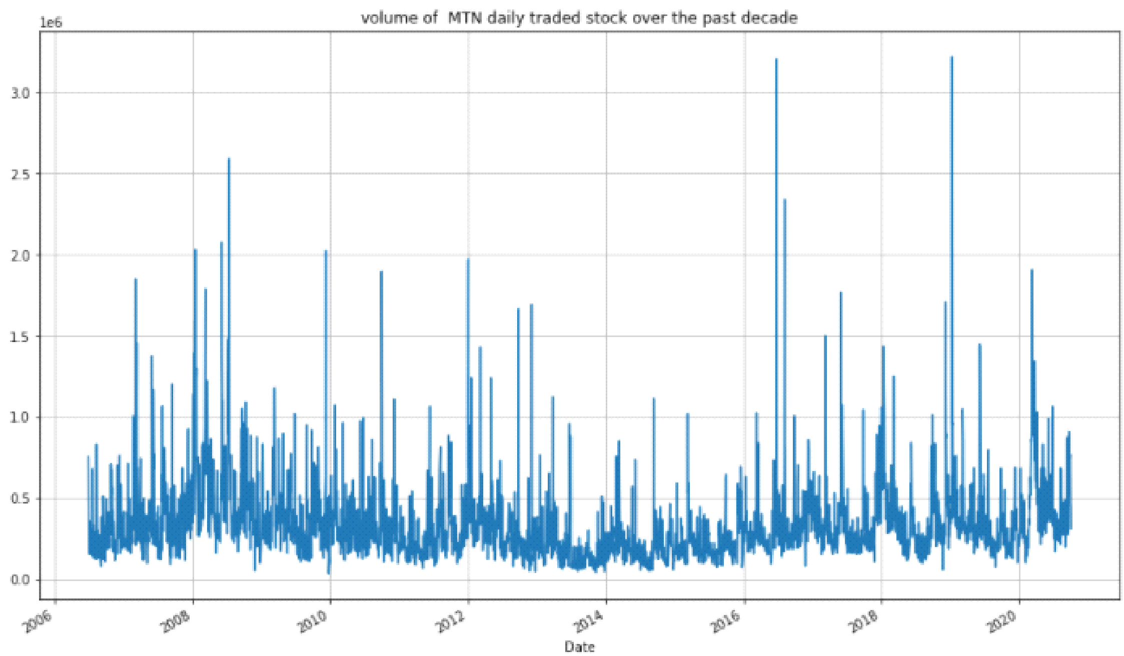
We have shown that log returns, with some indications of stationarity, obey classic normal distribution. By looking at the rolling statistics of our returns over a short term window, we wish to have a deeper dive into the market insights. From both plots below, we can clearly see the stationarity displayed by the moving average, volatility, skew and kurtosis. This confirms our confidence in the stationarity of both our returns

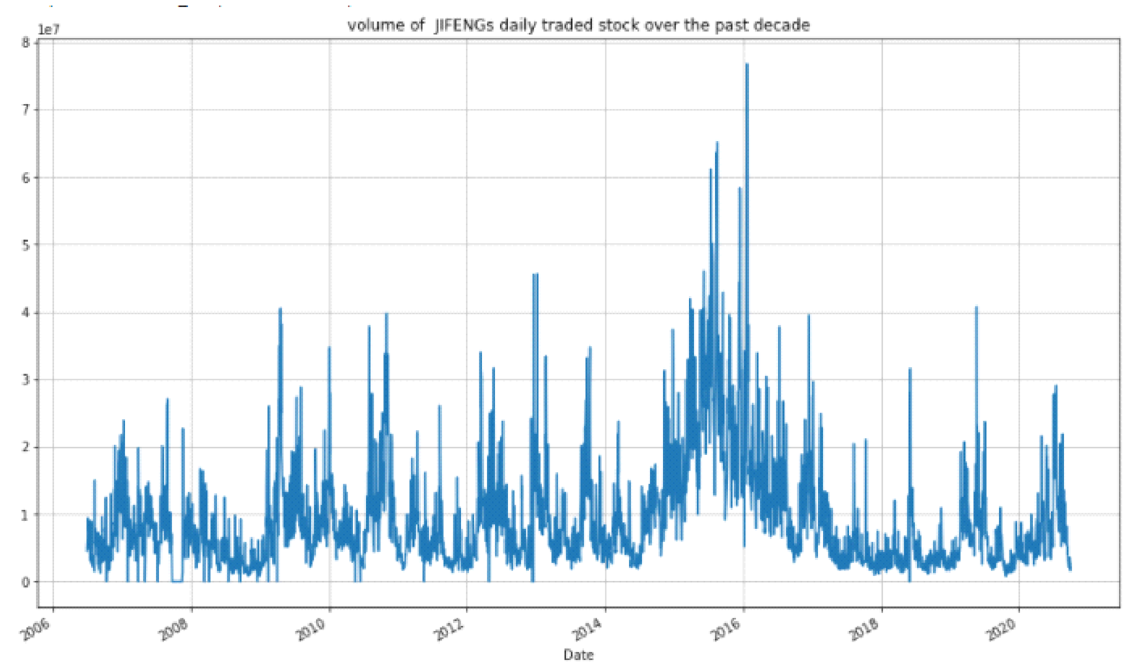
MTN Rolling Stats



IFENGS Rolling Stats



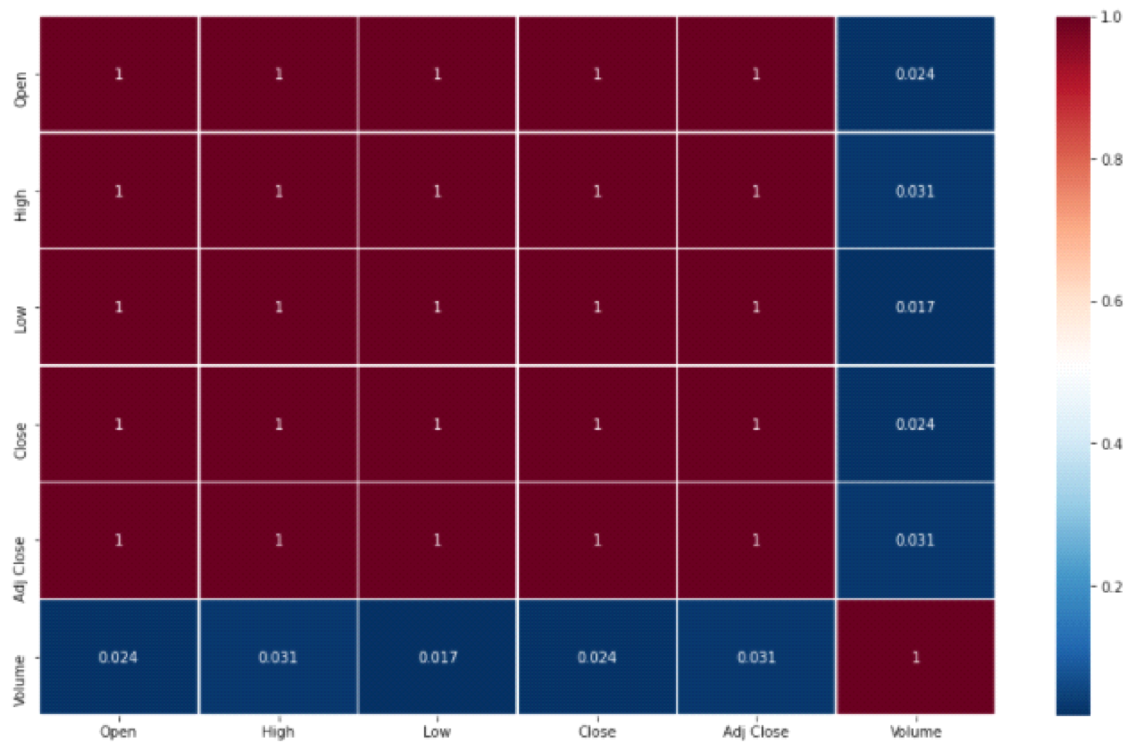

volume of MTN daily traded stock over the past decade
1e6
3.0
2.5
2.0
1.5
1.0
0.5
0.0
2006
2008
2010
2012
2014
2016
2018
2020
Date


volume of JIFENGs daily traded stock over the past decade
1e7
8
7
6
5
4
3
2
1
0
2006
2008
2010
2012
2014
2016
2018
2020
Date

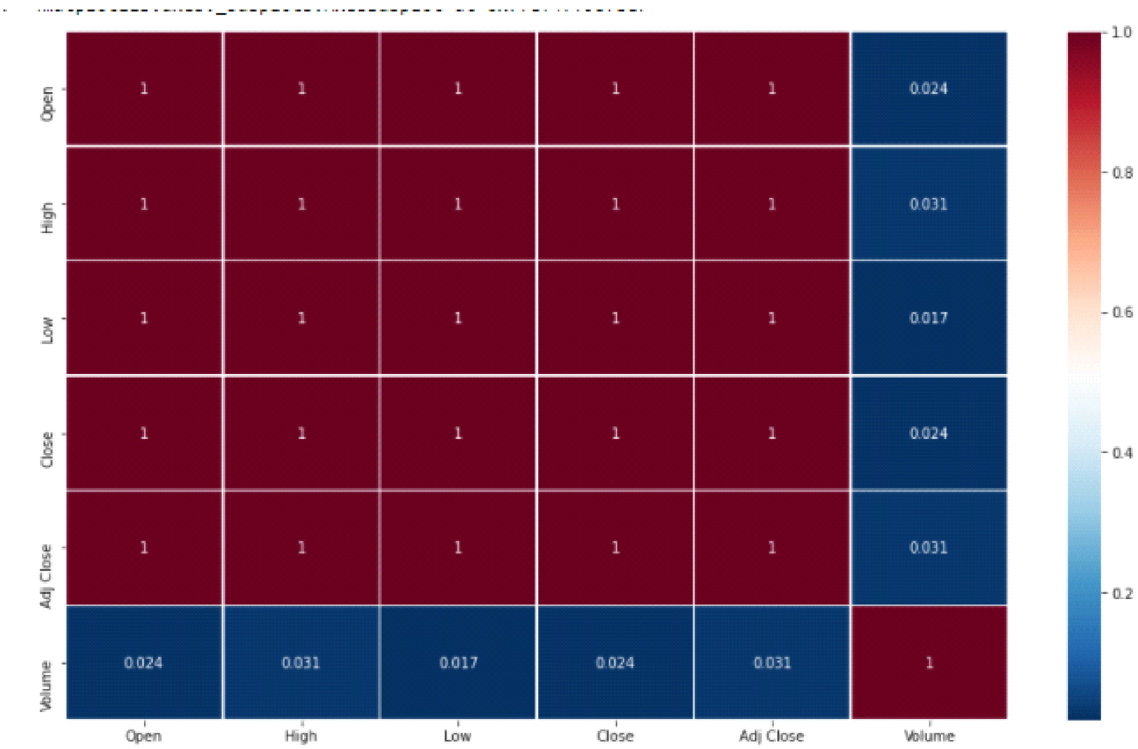
GSJ: Volume 9, Issue 3, March 2021
ISSN 2320-9186

We explore the correlation between the features within our dataset

	Open	High	Low	Close	Adj Close	Volume
Open	1.000000	0.999788	0.999756	0.999593	0.998499	0.024013
High	0.999788	1.000000	0.999628	0.999771	0.998916	0.030613
Low	0.999756	0.999628	1.000000	0.999809	0.998442	0.017486
Close	0.999593	0.999771	0.999809	1.000000	0.998891	0.023650
Adj Close	0.998499	0.998916	0.998442	0.998891	1.000000	0.031036
Volume	0.024013	0.030613	0.017486	0.023650	0.031036	1.000000

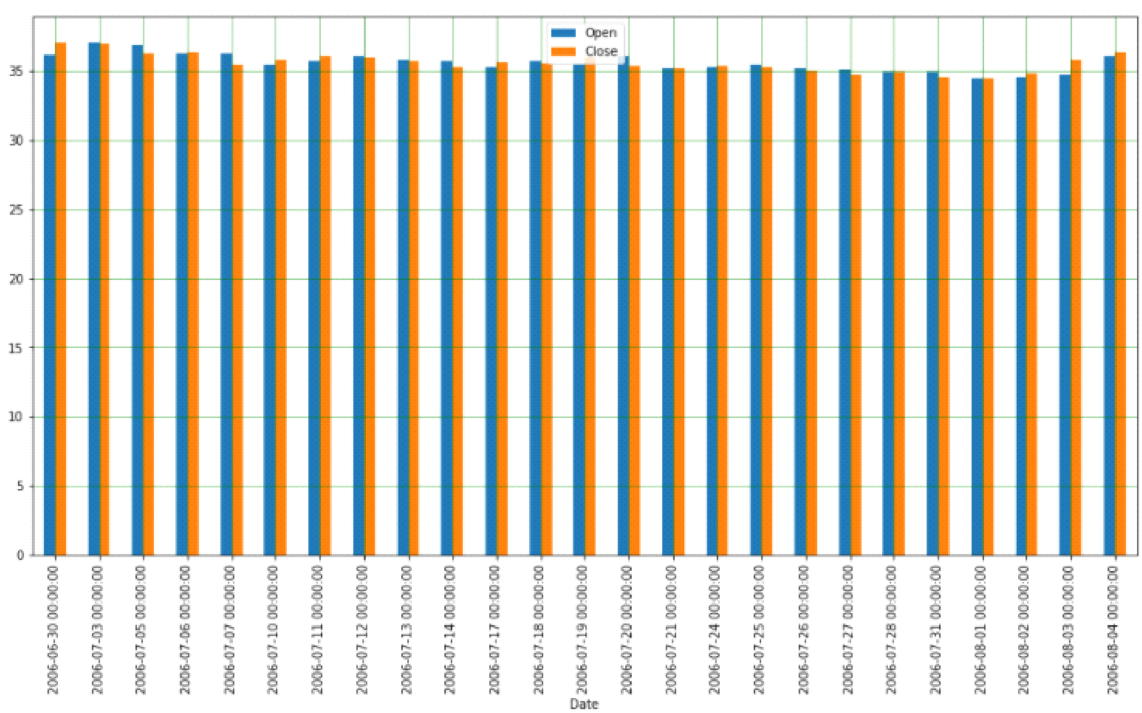


	Open	High	Low	Close	Adj Close	Volume
Open	1.000000	0.999788	0.999756	0.999593	0.998499	0.024013
High	0.999788	1.000000	0.999628	0.999771	0.998916	0.030613
Low	0.999756	0.999628	1.000000	0.999809	0.998442	0.017486
Close	0.999593	0.999771	0.999809	1.000000	0.998891	0.023650
Adj Close	0.998499	0.998916	0.998442	0.998891	1.000000	0.031036
Volume	0.024013	0.030613	0.017486	0.023650	0.031036	1.000000

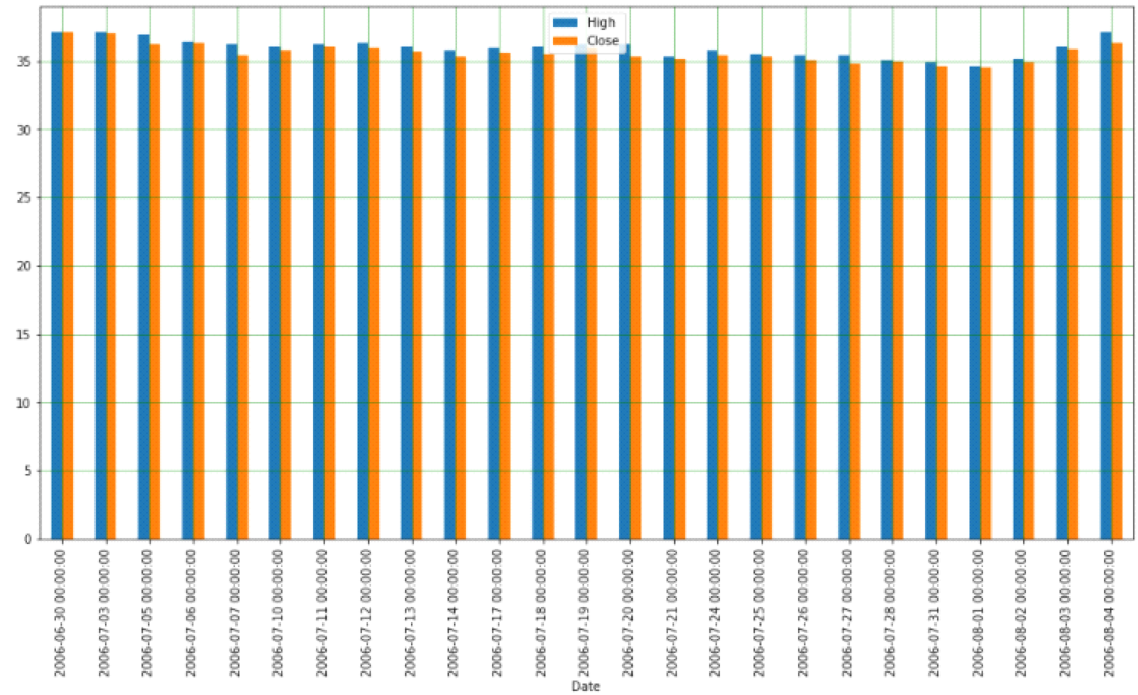


Assessing the visual comparison of dependent variable (Close Price) against the independent variable.

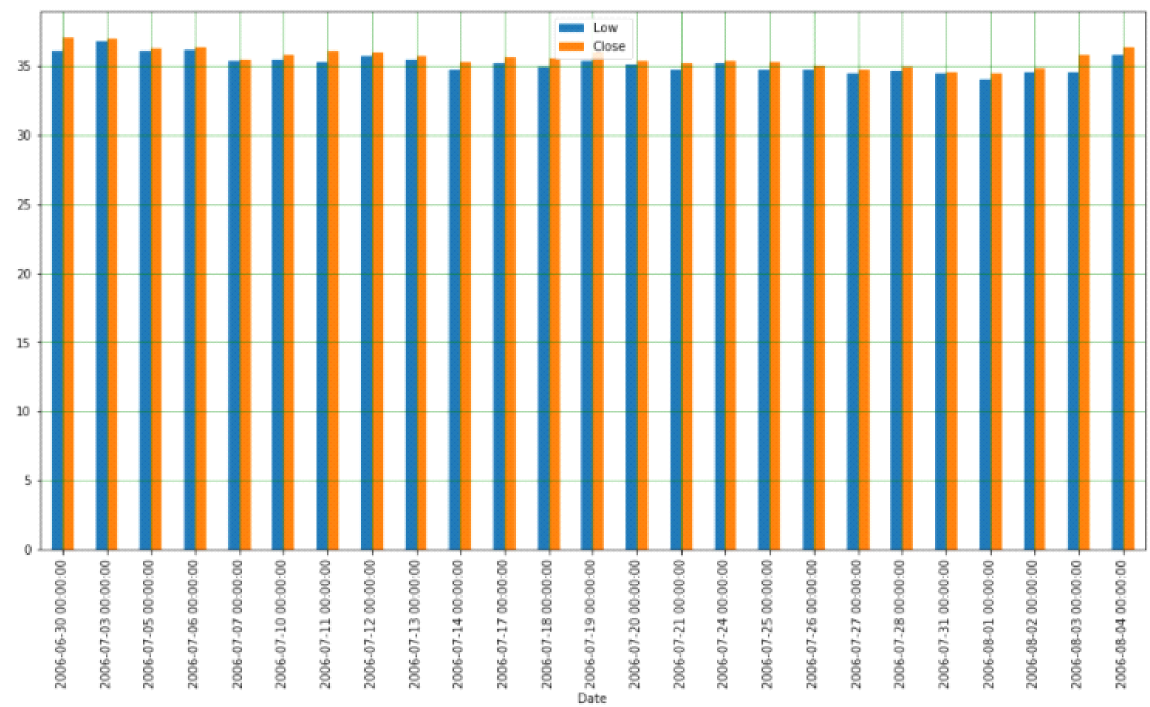
<u>Open vs. Close (MTN)</u>



High vs. Close (MTN)

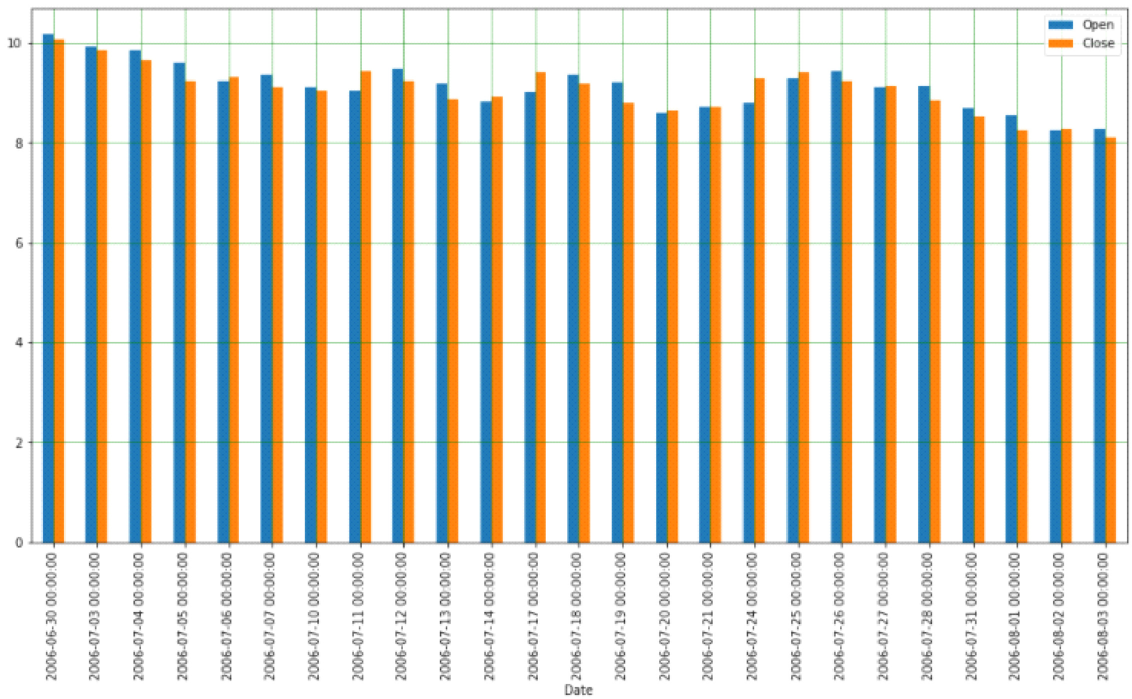


Low vs. Close (MTN)

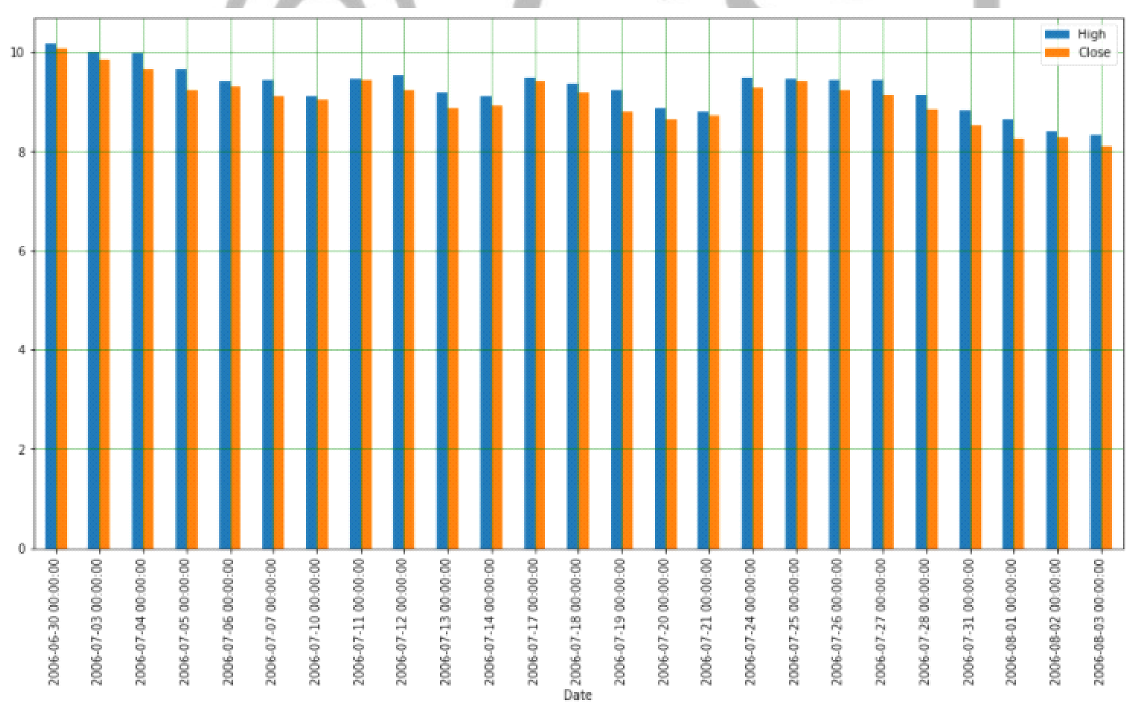


GSJ: Volume 9, Issue 3, March 2021
ISSN 2320-9186

Open vs. Close (JIFENG)

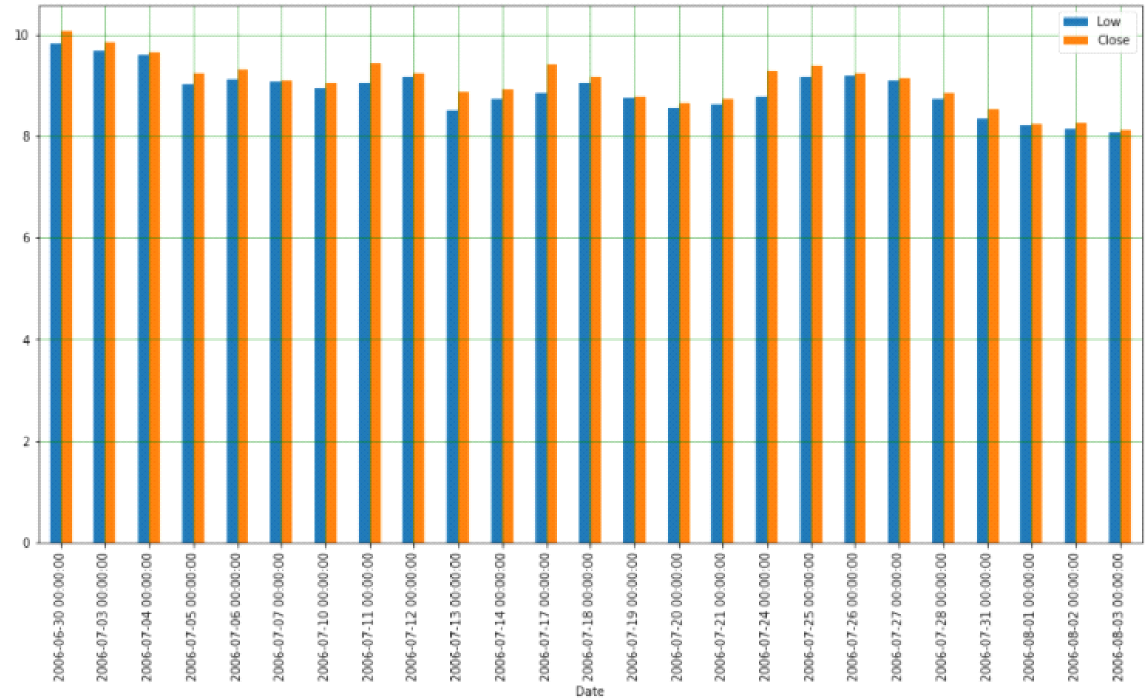


High vs. Close (JIFENG)



GSJ: Volume 9, Issue 3, March 2021
ISSN 2320-9186

Low vs. Close (JIFENG)



The Modeling Phase
The Support Vector Machine regression model

Preparing our final datasets for Regression modeling

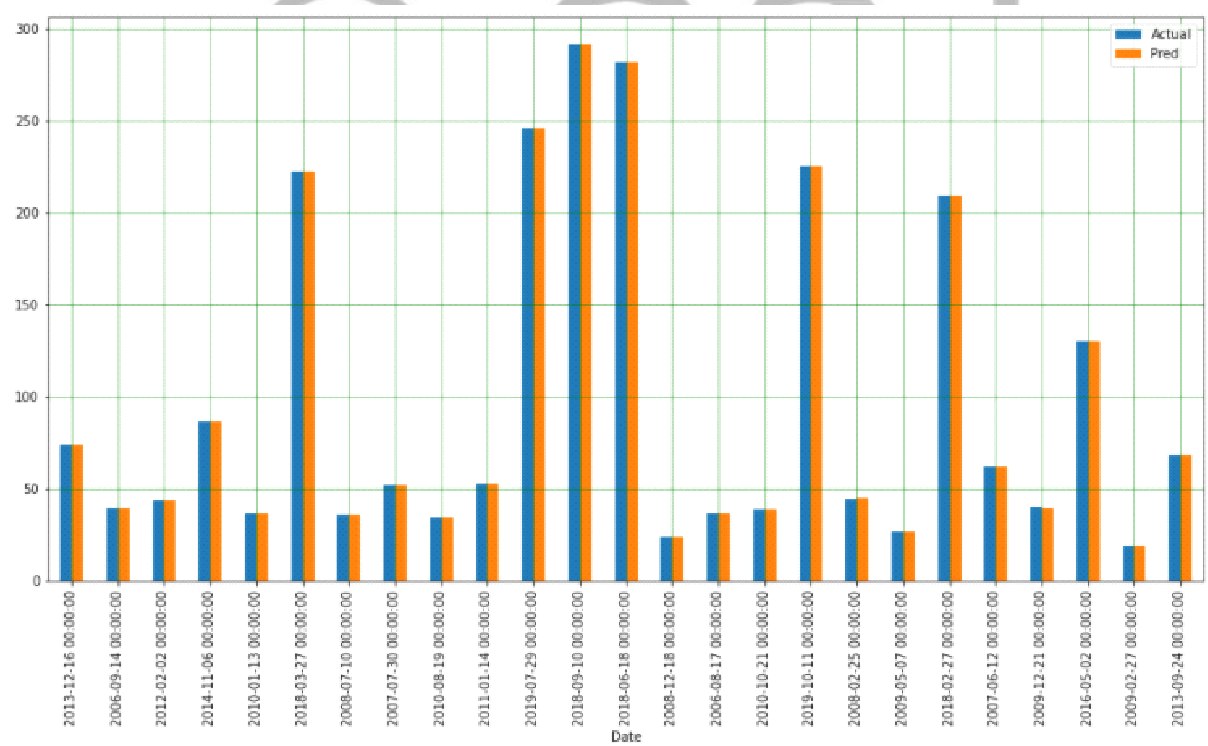
Final Dataset for regression(MTN)

Date	Day	Month	Year	High	Open	Low	Close
2006-06-30	30	6	2006	37.099998	36.200001	36.049999	37.099998
2006-07-03	3	7	2006	37.150002	37.049999	36.770000	37.000000
2006-07-05	5	7	2006	36.970001	36.930000	36.119999	36.259998
2006-07-06	6	7	2006	36.410000	36.290001	36.200001	36.349998
2006-07-07	7	7	2006	36.220001	36.220001	35.369999	35.450001
2006-07-10	10	7	2006	36.000000	35.450001	35.430000	35.799999
2006-07-11	11	7	2006	36.200001	35.700001	35.310001	36.060001
2006-07-12	12	7	2006	36.270000	36.099998	35.730000	35.980000
2006-07-13	13	7	2006	36.060001	35.840000	35.419998	35.680000
2006-07-14	14	7	2006	35.799999	35.750000	34.750000	35.310001

Final Dataset for regression Model(JIFENG)

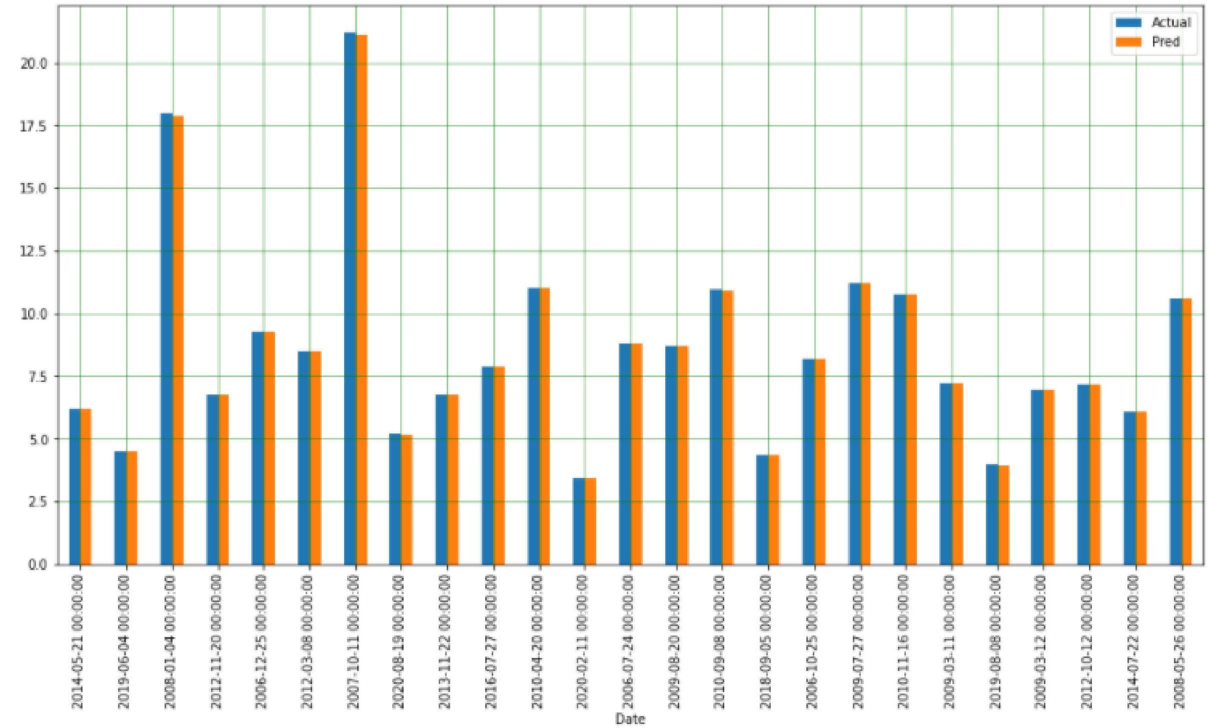
Date	Day	Month	Year	High	Open	Low	Close
2006-06-30	30	6	2006	10.17950	10.17950	9.81410	10.07690
2006-07-03	3	7	2006	10.00000	9.92949	9.67949	9.85256
2006-07-04	4	7	2006	9.96154	9.83974	9.61538	9.65385
2006-07-05	5	7	2006	9.64744	9.61538	9.01923	9.23077
2006-07-06	6	7	2006	9.40385	9.22436	9.11538	9.32051
2006-07-07	7	7	2006	9.44231	9.35897	9.07051	9.10256
2006-07-10	10	7	2006	9.10256	9.10256	8.94231	9.04487
2006-07-11	11	7	2006	9.44872	9.05128	9.05128	9.43590
2006-07-12	12	7	2006	9.51923	9.49359	9.17949	9.23077
2006-07-13	13	7	2006	9.19231	9.19231	8.51282	8.86538

SVM Predictions Vs. Actual Dataset (MTN)



GSJ: Volume 9, Issue 3, March 2021
ISSN 2320-9186

SVM Predictions Vs. Actual Dataset (JIFENG)



Also, the SVM Regression Test Statistics of Prediction is shown in appendix 1.

Time series modeling

For our GARCH Model we laid off all the variables in our dataset with the exception of the Close Price and we went ahead to use the close price to create the percentage change in close price which is the variable needed for our modeling. Below is a view of the new dataset needed for this phase of our analysis for both market.

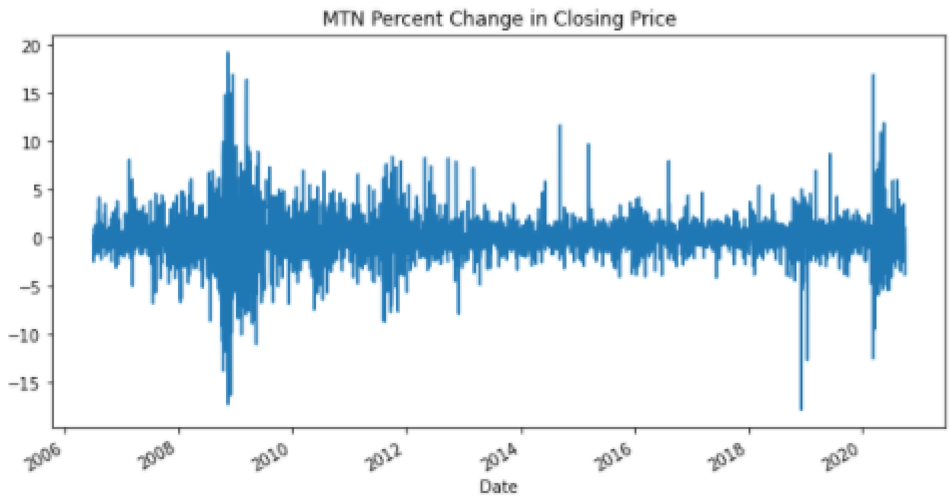
Date	Close	pct_change
2006-07-03	37.000000	-0.269538
2006-07-05	36.259998	-2.000005
2006-07-06	36.349998	0.248208
2006-07-07	35.450001	-2.475922
2006-07-10	35.799999	0.987302

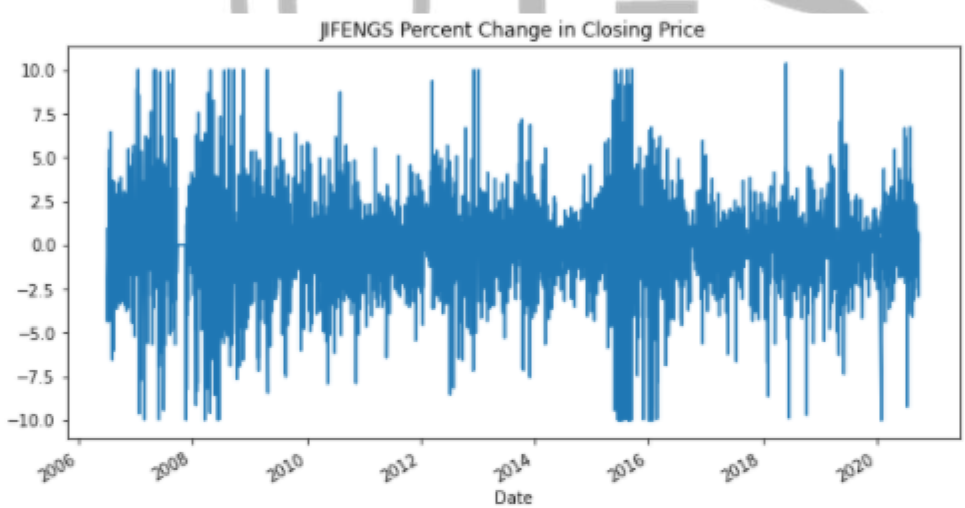
MTN Dataset 1

Date	Close	pct_change
2006-07-03	9.85256	-2.226275
2006-07-04	9.65385	-2.016841
2006-07-05	9.23077	-4.382495
2006-07-06	9.32051	0.972181
2006-07-07	9.10256	-2.338390

jifeng Dataset

We plot the percentage change in close price for both our stocks, it could be seen that each plot exhibits the property volatility clustering: large changes in prices tend to cluster and creating a long span of price changes. However, JIFENG exhibits a higher level of volatility clustering in comparison to MTN





Jifeng Garch(1,1) 1

Constant Mean - GARCH Model Results

Dep. Variable:	pct_change	**R-squared:**	-0.000
Mean Model:	Constant Mean	**Adj. R-squared:**	-0.000
Vol Model:	GARCH	**Log-Likelihood:**	-8080.76
Distribution:	Normal	**AIC:**	16169.5
Method:	Maximum Likelihood	**BIC:**	16194.1
		No. Observations:	3472
Date:	Sat, Oct 24 2020	**Df Residuals:**	3468
Time:	05:00:21	**Df Model:**	4

Mean Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
mu	-1.8171e-03	3.804e-02	-4.777e-02	0.962	[-7.636e-02,7.273e-02]

Volatility Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
omega	0.1107	7.607e-02	1.455	0.146	[-3.843e-02, 0.260]
alpha[1]	0.0610	2.458e-02	2.480	1.313e-02	[1.279e-02, 0.109]
beta[1]	0.9234	3.390e-02	27.238	2.292e-163	[0.857, 0.990]

Covariance estimator: robust

MTN Garch(1,1)

Constant Mean - GARCH Model Results

Dep. Variable:	pct_change	**R-squared:**	-0.000
Mean Model:	Constant Mean	**Adj. R-squared:**	-0.000
Vol Model:	GARCH	**Log-Likelihood:**	-7468.37
Distribution:	Normal	**AIC:**	14944.7
Method:	Maximum Likelihood	**BIC:**	14969.5
		No. Observations:	3589
Date:	Sat, Oct 24 2020	**Df Residuals:**	3585
Time:	05:41:17	**Df Model:**	4

Mean Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
mu	0.1297	3.026e-02	4.285	1.829e-05	[7.035e-02, 0.189]

Volatility Model

	coef	std err	t	P>|t|	95.0% Conf. Int.
omega	0.0787	3.243e-02	2.428	1.519e-02	[1.517e-02, 0.142]
alpha[1]	0.0668	1.622e-02	4.120	3.781e-05	[3.504e-02,9.863e-02]
beta[1]	0.9164	2.060e-02	44.491	0.000	[0.876, 0.957]

Covariance estimator: robust

Here are the results of our MTN and JIFENG GARCH (1,1) model:
Omega is the baseline variance for the formula, so the square root of omega is the standard dev iation of returns, 3 percent. MTN's return of 0.1 has a standard deviation of 3%, whereas, Jifeng's return as 0 has a standard deviation of 7%. These figures are very low to be considered unpredictable. Also, refer to appendix 6.

Next, our coefficients alpha and beta. Alpha tests how much a shock of volatility feeds into volatility in the coming period. In our model, volatility will be moved to the next day for 9% of previous periods. Beta is our persistence parameter, if beta reaches 1, it contributes to a positive feedback loop that can deliver flat variance in small shocks. The sum of alpha and beta is the rate at which our variance decreases and if alpha plus beta is equal to 1, then our model is constantly volatile and we will want to analyze other models, such as IGARCH (Integrated Garch).
The text notes that the normal range of alpha in a stable market is $0.05 < \alpha<0.1$ and beta $0.85 < \beta < 0.98$. The market risk estimation is typically a stable market. These are not entirely applicable to a single stock estimate, however, it does give us some insight into what kind of values we might see. Next, we have our t-statistics and p-values. T is our estimate divided by the standard error and is used to calculate our p-value. The normal null hypothesis is that we have no effect, but will refuse zero if our p-value is less than a n-alpha (0.0.5). It is helpful to pay attention to the p-value as we start to get models with more parameters.

Over there we carried out some exercise test, which indicated that we should deny the null so that we begin to look for better model parameters as all the tests suggest. We can see that at least 25^{th} lag is obviously self-correlated and that uniform remnants are not like white sound. In order to find a model ideally suited for variance, we can a grid search for wide range of p and q.

Discussion

From Yahoo Finance we downloaded data on MTN and JinFeng date starting 2006-06-30 and ending 2020-10-30, viewing their data frame on the following variables: open, high, low, close,

adj. close, and volumes. We did a descriptive-statistics of the two companies including the following:

i) Plotted their adjusted close price over a decade time frame.
ii) Candle stick plotted their stocks over the past decade.
iii) Stationarity and Normality in financial data by plotting the price and price distributions and returns and returns distributions.
iv) Plot of volumes of daily traded stocks over a decade.
v) Correlation of our data features
vi) Assessing the visual comparison of dependent variables against the independent variables.
vii) Regression models using Machine Learning (ML) approach.
viii) We explored three different techniques of regression models, namely: Linear Regression Model; K-nearest Neighbor Regression Model; and Support Vector Machine-regression Model.
ix) Kernel – this is used to find a high dimensional hyperplane without increasing cost of computing.
x) Hyperplane – this is a line that isolates two data classes in SVM
xi) Decision-Boundary
xii) Modeling for the two markets
xiii) Test statistics for modeling.
xiv) Root Mean Square Error (RMSE).
xv) Application of time series models, including time series analysis plots, autocorrelation, partial autocorrelation, QQ plot, and probability plot.
xvi) Auto-regression conditionally heteroskedastic modeling – ARCH (p).
xvii) Generalized auto-regression conditionally heteroskedastic models – GARCH (p<q)
xviii) Alpha and Beta values are parameters or weights used in the GARCH calculations to determine the presence and magnitude of volatility. Like all population parameters, they are theoretical–we don't know their true values unless the experiment is administered.
xix) Alpha measures the probability of Type I error in any hypothesis test–incorrectly claiming statistical significance while Beta is the probability of Type II error in any hypothesis test–incorrectly concluding no statistical significance. (1 – Beta is power).

xx) If a manager has had a high alpha, but also a high beta, investors might not find that acceptable, because of the chance they might have to withdraw their money when the investment is doing poorly.

xxi) The value of an alpha is interpreted to be a measure of past change in volatility denoted as the standard deviation, (small alpha=small impact of innovation), while beta measures the impact of past value of volatility on today's volatility.

xxii) Volatility is the tendency to change abruptly to the worse, in finance, volatility, degree of variation of a stock price time series, which is the standard deviation of log returns.

xxiii) Historical volatility measures the time series of past market. Implied volatility measures or looks forward in time, being derived from the market price of a market traded derivative (in particular an option).

xxiv) Looking at the results, we had alpha to be as high as 0.9048 and 1.0 for both MTN and Jinfeng respectively, which shows there is a very strong correlation among the data and therefore high volatility of the market prices. Hence the impact of implied volatility will be high. Therefore, there is a volatility clustering among them.

xxv) Rolling GARCH model is a model that can be used to determine volatility and variance over a given period of time

xxvi) Support Vector Regression

xxvii) SVR is one of linear regression models used to minimize the total of squared errors. Using ordinary least spread as an example predictive equation is as follows:

$$MIN \sum_{i-1}^{n} (Y_i - W_i X_i)^2$$

xxviii) SVR-FTW

xxix) SVR has room for defining the error limit acceptable in the model and will find the hyperplane in higher dimension to fit the data. The SVR is used to minimize the coefficients but not the square root error. In doing so, the error limit is constrained to an absolute error limit of less than or equal to a determined margin called the maximum error or Epsilon (Ɛ). The Epsilon can be regulated to achieve the desired accuracy in the data. The equation is given as follows:

Minimize:

$$MIN½ \left| \, |W| \, \right|^2$$

Constraints:

$$\left| Y_i - W_i X_i \right| \leq \varepsilon$$

xxx) We can also use slack to minimize to minimize any hyper-parameter in the variable which is outside the Epsilon. That deviation is given as a slack (**ξ**). The equation can be given as:

Minimize:

$$MIN½\left|\,|W\,|\right|^2 + C\sum_{i=1}^{n} |\xi_i|$$

Constraints:

$$\left|Y_i - W_iX_i\right| \leq \mathcal{E} + \left|\boldsymbol{\xi_i}\right|$$

xxxi) SVR is a power mathematical and computational tool used to determine the tolerant error margin (Ɛ) or outside the acceptable error margin ($\boldsymbol{\xi}$).

MTN

From the first graph, it can be seen that MTN prices went up exponentially from 2006 until it peaked in later part of 2018, where it dropped a little in 2019 and then went up again until early part of 2020. The price is seen dropping from early days of 2020 which indicates the period of COVID-19, and rise as COVID cases subside. Again looking at the price distribution, it is asymmetrical, which indicates the mean is not constant and standard deviation more than one. Looking at the time series plot of returns, returns went as far in 2009 which became normal until 2019 and 2020. Returns distribution too is normal with a mean around zero and standard deviation approximately equal to 1.

A lag k **autocorrelation** is the correlation between values that are k time periods apart. The **ACF** is a means to measure the linear relationship between an observation at time t and the observations at previous time. The ACF correlogram shows that the value of the function of the lag is constant and the PACF correlogram had two spikes corresponding to two lags.

The **Q-Q plot** of MTN stock is not normally distributed whose data is sparingly distributed away from the mean.

As explained before, that the lag value is not full as in other variations and correlations. The PACF correlogram had two spikes corresponding to first two lags. This shows there is a significant correlation at lag 1 and 2 followed by correlations that are not significant. This pattern indicates an autoregressive term of order 2.

Jinfeng

For Jinfeng, the price increased from 2006 and peaked at latter part of 2007 and started declining to a lower level in the later part of 2008. It went up again in 2010 and stabilized more or less till 2014 which again increased in 2016 and later declined exponentially to an all-time low in 2020 which can be attributed to the COVID-19 pandemic. The graph is seen moving up during the latter part of 2020 which indicates COVID-19 pandemic cases going down. The price distribution too is not normal but asymmetric. Time series plots of the return indicates variance throughout the 10-years period, but distribution of the returns is normal which indicates a mean of zero and standard deviation of 1. There is a stationarity in the variables with respect to the mean, variance and covariance.

The ACF correlogram shows that the value of the function of the lag is declining and the PACF correlogram also had two spikes corresponding to two lags. The ACF correlogram shows that the value of the function of the lag decays geometrically which indicates auto correlation decays over time and the PACF correlogram also had two spikes corresponding to two lags which also shows correlation is strong at these lags but insignificant afterwards.

On a whole, Outliers were determined using a scatter plot. This was due to some data differing significantly from the mean or the median. Their presence affects the effectiveness of data for analysis, making future predictions very difficult. This result in heavy tails and skewness of data display and this is caused by error during data collection. During working with high data of samples like the data we worked with, outliers cannot be totally eliminated.

Fortunately, there is no missing data or unresponsiveness in the stocks and therefore our data can be used to draw a reliable conclusion that, after the COVID, and this regime shift, prices will rise to the normal.

Again, there is no Structural Breaks or stability which is the regression model factor that determines the vulnerability of the model and hence resulting in forecast error.

The **Q-Q plot** of Jinfeng stock is normally distributed and the data is not distributed away from the mean as in the MTN data.

The concept of stationarity though is, but not a major problem in the field of finance, by applying a difference we can attain stationarity. However, the biggest issues in the field of finance are rather; structural breaks, missing data and outliers. The data used for this work has been tested for missing or null data points and we fortunately don't have missing data points. Also, from the scatter plot below there is no evident of a data point completely separated from the trends hence we could conclude of the absence of extreme outliers and acknowledging the presence of mild outliers which could be considered when modelling.

Conclusion

The Stationarity of financial data indicates the returns plot has a constant mean around zero and a more constant standard deviation. Putting aside all the comprehensive mathematical description, what stationarity means is that the statistics of the underlying signals (MTN and JIFENG) are stable over time. Indeed, the concept of stationarity is a major problem in the field of finance. For instance, considering the price signal in the figure below for MTN and JIFENG, a robust statistical test is not required to ensure the average price does not stay constant over time. Thus, one could see that our Returns plot has a constant mean around zero and a more constant standard deviation.

Putting aside all the comprehensive mathematical description, what stationarity means is that the statistics of the underlying signals (MTN and JIFENG) are stable over time. Indeed, the

Secondly, the Rolling Statistics of our returns have shown that the log returns show some indications of stationarity and obey classic normal distribution. By looking at the rolling statistics of our returns over a short term window, we wish to have a deeper dive into the market insights. From both plots, we can clearly see the stationarity displayed by the moving average, volatility, skew and kurtosis. This confirms our confidence in the stationarity of both our returns.

Our test statistics of prediction of the linear regression model, namely, KNN Regression model and SVM Regression model are shown in appendix 1.

It is evident that our QQ plot is right skewed, however a majority of the data point fall close to the normality line and thus conclude that it supports normality and so is our probability plot which also tells us that the specified theoretical distribution is not a good fit due to the right skewedness nature of the plots, however a larger portion of the data points falling close to the straight lead us to believe otherwise.

The Jarque Bera test statistic for MTN and JIFENG Close Price reads 508.77416464325773 and 2 141.5634330749203 with both p value less than 0.05(i.e. 0). Refer to appendix 3. We thus reject the null hypothesis that both our dataset is not normally distributed. In order words, we conclude that the departure from normality as measures by the test statistics is statistically significant.

Furthermore, the test statistics of prediction applied the linear regression model, KNN regression model, and the SVM regression model. Refer to appendix 1. In addition, both the ARCH Model and GARCH Model for Jinfeng showed a constant mean and the covariance estimator was robust. Refer to appendix 2 and 5. Similarly, the GARCH Model for MTN showed a constant mean. Refer to appendix 4. Also, refer to appendix 6.

It is further clear that, SVR is a power mathematical and computational tool used to determine the tolerant error margin (Ɛ) or outside the acceptable error margin ($\boldsymbol{\xi}$).

We plotted the percentage change in close price for both our stocks, and it could be seen that each plot exhibits the property volatility clustering: large changes in prices tend to cluster creating a long span of price changes. However, JIFENG exhibits a higher level of volatility clustering in comparison to MTN.

It can also be concluded that, there is volatility clustering in the data and this can be seen from the results of the GARCH and ARMA models. Statistically, the analysis indicates a period of regime shifts every decade, and this can be seen from the candle stick plot and the Time series plot, which could be caused by either a pandemic or a financial crisis. COVID-19 indeed has indeed affected market prices, returns and the statistical characteristics of the market data from latter part of 2019 through to first two quarters of 2020 which indicate that stock market is actually affected by the global pandemic which will subsequently lead to economic contraction of companies, businesses and countries at large, hence recession of many economies around the world. There has also been a regime shift which can be observed within the past 10-year period. We can confidently deduce that after the pandemic is subsided, data price will go up again.

Next we plotted the close price of both our stocks to view the dynamics of price action. It could be seen that the close price of MTN rises gently upwards and start experiencing a sharp fall toward the end of 2018 and rises slowly into the 2020 before experiencing a sharper fall in the early part of 2020 which could be attributed to the corona negative impact on the market. It could therefore be concluded that, COVID-19 indeed has affected market prices, returns and the statistical characteristics of the market data from latter part of 2019 through to first two quarters of 2020 which indicate that stock market is actually affected by the global pandemic which will subsequently lead to economic contraction of companies, businesses and countries at large, hence recession of many economies around the world. There has also been a regime shift which can be observed within the past 10-year period.

Last but not least, is the fact that, our GARCH Model is not statistically significant because the modeling of a single stock is somehow difficult, as such we recommend for a better GARCH (p, q) Model in a further research.

You can access our source code repository at the Github link below:

https://github.com/Akilu-phat/SCM_8.git

GSJ: Volume 9, Issue 3, March 2021
ISSN 2320-9186

Appendices

Appendix 1

SVM REGRESSION MODEL TEST STATISTICS OF PREDICTION

	RMSE	R^2
MTN	0.03679421800249008	0.9999997612594144
JIFENG	0.02689172496468494	0.9999420392398427

Appendix 2

ARCH MODEL(JIFENG)

```
Optimization terminated successfully.    (Exit mode 0)
            Current function value: 6548.510640065466
            Iterations: 19
            Function evaluations: 112
            Gradient evaluations: 19
<bound method ARCHModelResult.summary of                       Constant Mean - ARCH Model Results
==============================================================================
Dep. Variable:                  Close   R-squared:                      -0.074
Mean Model:             Constant Mean   Adj. R-squared:                 -0.074
Vol Model:                       ARCH   Log-Likelihood:               -6548.51
Distribution:                  Normal   AIC:                           13103.0
Method:            Maximum Likelihood   BIC:                           13121.5
                                        No. Observations:                 3473
Date:                Sat, Oct 03 2020   Df Residuals:                     3470
Time:                        16:29:02   Df Model:                            3
                               Mean Model
========================================================================
                 coef    std err          t      P>|t|   95.0% Conf. Int.
------------------------------------------------------------------------
mu             7.5776  5.686e-02    133.277      0.000 [  7.466,  7.689]
                              Volatility Model
============================================================================
                 coef    std err          t      P>|t|      95.0% Conf. Int.
----------------------------------------------------------------------------
omega          0.0326  4.598e-03      7.090  1.345e-12 [2.359e-02,4.161e-02]
alpha[1]       1.0000  8.590e-03    116.410      0.000     [  0.983,  1.017]
============================================================================

Covariance estimator: robust
ARCHModelResult, id: 0x7fef36f638d0>
```

Appendix 3

STOCK	JARQUE BERA TEST	P-VALUE
MTN	508.77416464325773	0.0
JIFENG	2141.5634330749203	0.0

Appendix 4

Garch Model(MTN)

```
Optimization terminated successfully.    (Exit mode 0)
            Current function value: 16609.788319273925
            Iterations: 40
            Function evaluations: 259
            Gradient evaluations: 38
<bound method ARCHModelResult.summary of                      Constant Mean - GARCH Model Results
==============================================================================
Dep. Variable:                  Close   R-squared:                      -0.612
Mean Model:             Constant Mean   Adj. R-squared:                 -0.612
Vol Model:                      GARCH   Log-Likelihood:               -16609.8
Distribution:                  Normal   AIC:                           33227.6
Method:            Maximum Likelihood   BIC:                           33252.3
                                        No. Observations:                 3590
Date:                Sat, Oct 03 2020   Df Residuals:                     3586
Time:                        16:40:21   Df Model:                            4
                               Mean Model
========================================================================
                 coef    std err          t      P>|t|  95.0% Conf. Int.
------------------------------------------------------------------------
mu            44.8087      0.387    115.911      0.000 [ 44.051, 45.566]
                              Volatility Model
===========================================================================
                 coef    std err          t      P>|t|     95.0% Conf. Int.
---------------------------------------------------------------------------
omega          0.7429      0.240      3.098  1.948e-03    [  0.273,  1.213]
alpha[1]       0.9048  8.711e-02     10.386  2.857e-25    [  0.734,  1.075]
beta[1]        0.0952  9.017e-02      1.056      0.291 [-8.149e-02,  0.272]
===========================================================================

Covariance estimator: robust
ARCHModelResult, id: 0x7fef36fd2cf8>
```

GSJ: Volume 9, Issue 3, March 2021
ISSN 2320-9186

Appendix 5

Garch Model (JIFENG)

```
Optimization terminated successfully.    (Exit mode 0)
            Current function value: 6544.051212484994
            Iterations: 24
            Function evaluations: 164
            Gradient evaluations: 24
<bound method ARCHModelResult.summary of                      Constant Mean - GARCH Model Results
==============================================================================
Dep. Variable:                  Close   R-squared:                      -0.077
Mean Model:             Constant Mean   Adj. R-squared:                 -0.077
Vol Model:                      GARCH   Log-Likelihood:               -6544.05
Distribution:                  Normal   AIC:                           13096.1
Method:            Maximum Likelihood   BIC:                           13120.7
                                        No. Observations:                 3473
Date:                Sat, Oct 03 2020   Df Residuals:                     3469
Time:                        16:40:49   Df Model:                            4
                               Mean Model
========================================================================
                 coef    std err          t      P>|t|  95.0% Conf. Int.
------------------------------------------------------------------------
mu             7.5622  6.100e-02    123.977      0.000 [  7.443,  7.682]
                              Volatility Model
============================================================================
                 coef    std err          t      P>|t|      95.0% Conf. Int.
----------------------------------------------------------------------------
omega          0.0243  3.909e-03      6.225  4.807e-10 [1.667e-02,3.199e-02]
alpha[1]       0.9026  4.428e-02     20.386  2.245e-92     [  0.816,  0.989]
beta[1]        0.0974  4.471e-02      2.178  2.941e-02   [9.744e-03,  0.185]
============================================================================

Covariance estimator: robust
ARCHModelResult, id: 0x7fef36f5aeb8>
```

Appendix 6

	Largrange Multiplier p-value	F-Test p-value	Shapiro-Wilks p-value
	$7.251452093353409e^{171}$	$5.227704787659386e^{203}$	$4.548744367819772e^{-33}$
	$7.251452093353409e^{203}$	$5,22770487659386e^{203}$	$4.54874436781772e^{33}$

References

- Jonathan, D.C., Kung-Sik, C. Time Series Analysis with Applications in R. 2nd Edition. 2008. Springer Texts in Statistics.
- Jorge, N., Stephen, J.W. Numerical Optimization. 2nd Edition. 2006. Springer Series in Operations Research and Financial Engineering.
- Cont, R. 2001. Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. Quantitative Finance.
- Ruppert, D. and Matteson, D.S. 2015. Statistics and Data Analysis for Financial Engineering with R Examples. Second Ithaca, NY, USA. Springer Texts in Statistics.
- Tsay, R.S. 2010. Analysis of Financial Time Series. Wiley.
- English Dictionary. Offline. Livio. Android-dictionary-livio@googlegroups.com.
- Buy-Side Vs Sell-Side Analyst: What is the Difference? https://www.investopedia.com/articles/financialcareers/11/sell-side-buy-side
- Buy-Side Vs Sell-Side – Corporate Finance Institute. https://corporatefinancialinstitute.com/resources/careers/jobs/buy-side-vs-sell-side
- Financial Market – Wikipedia. https://en.wikipedia.org/wiki/financial-market
- Financial Markets – Overview, Types, and Timelines. https://corporatefinanceinstitute.com/resources/knowledge/trading-invest
- Di Sha, M. Statistics: Meaning, Characteristics, and Importance. https:/yourarticlelibrary.com/education/statistics/
- Spot Market -Wikipedia – https://en.wikipedia.org/wiki/spot_market
- Akeyede, I., Oyeyemi, G.M. 2016. On Performance of Some Methods of Detecting Non-Linearity, in Stationary and Non-Stationary Time Series Data. https://pdfs.semanticscholar.org/
- Artificial Intelligence Algorithms for Beginners – Regression. June 27, 2020. https://www.machineexp.com/post/regression_techniques_in_machine_learning_explanation.
- What is Decision Science? Centre for Health Decision Science > Approaches. Harvard T.H. CHAN School of Public Health.
- K Nearest Neighbors – Classification – Data Mining. https://www.saedsayad.com/k_nearest_neighbors.htm
- Time Series Analysis for Financial Data VI-GARCH Model. https://medium.com/auquan/time-series-analysis-for-finance-arch-garch-models-822f87fid755

- Garch Model Formula – Pozelar. https://pozelar.com.br/vjcftw/garch-model-formula.html
- Alswarya, A. Jan 14, 2020. Modelling Volatility for Stock Markets. https://medium.com/analytics-vidhya/modelling-volatility-for-stock-markets-32d23c502c9f.
- Python Time Series Model – Bing. 23-08-2020. https://www.windowssearch-exp.com/search?q=python+time+series+model&first=11&FORM=PERE
- Conroy, R. and Byrne, A. (2015). Portfolio Management: An Overview. Corporate Finance and Portfolio Management; CFA Program Curriculum Level I, 4, pp. 149-171.
- Langer, E. (1983). The Psychology of Control, Berverly: Sage Publications, 1983
- Markowitz, H. (1952). Portfolio Selection. Journal of Finance, 7(1), pp. 77-91.
- Hui-Shan, L; Fan-Fah, C; and Shyne-Chuan, C. 2016. International Journal of Economics and Financial Issues. http:www.econjournal.com.
- Acharya, Viral, Christian Brownlees, Robert Engle, Farhang Farazmand and Mathew Richardson, 2010, ―Measuring Systemic Risk in Acharya, Viral, Thomas Cooley, and Mathew Richardson (Eds.), *Regulating Wall Street: The Dodd-Frank Act and the New Architecture of Global Finance,* John Wiley and Sons.
- Adrian, Tobias, and Markus K. Brunnermeier, 2011, ―CoVaR, NBER Working Paper 17454.
- Adrian, Tobias and Hyun S. Shin, 2010, ―Liquidity and Leverage, *Journal of Financial Intermediation* 19(3), pp. 418-437.
- Agur, Itai and Sunil Sharma, 2013, ―Rules, Discretion, and Macroprudential Policy, IMF Working Paper, 13/65 (Washington: International Monetary Fund).
- Blancher, Nicolas, Srobona Mitra, Hanan Morsy, Akira Otani, Tiago Severo, and Laura Valderrama, 2013, ―SysMo - A Practical Approach to Systemic Risk Monitoring, IMF Working Paper 13/168 (Washington: International Monetary Fund).
- Brunnermeier, Marcus, Andrew Crockett, Charles Goodhart, Avinash D. Persaud, and Hyun Shin, 2009, ―The Fundamental Principles of Financial Regulation, The International Center for Money and Banking Studies, Geneva, Switzerland.
- Brunnermeier, Marcus, Arvind Krishnamurthy, and Gary Gorton, 2013, ―Liquidity Mismatch Measurement, in *Risk Topography: Systemic Risk and Macro Modeling*, ed. by M.K. Brunnermeier and A. Krishnamurthy (Chicago, Illinois: NBER/University of Chicago Press).
- Ntweb.deltastate.edu. (2019). *Correlation Analysis*. [online] Available at: http://ntweb.deltastate.edu/vp_academic/bmoore/correlation_analysis.htm

- Tastytrade.com - A Real Financial Network. (2019). *Correlation, Beta and R-Squared*. [online] Available at: https://www.tastytrade.com/tt/shows/the-skinny-on-options-math/episodes/correlation-beta-and-r-squared-10-09-2014

- The Economic Times. (2019). *What is Standard Deviation? Definition of Standard Deviation, Standard Deviation Meaning - The Economic Times*. [online] Available at: https://economictimes.indiatimes.com/definition/Standard-Deviation

- Famous Mathematicians and Statisticians – Statistics How To.. https://www.statisticshowto.com/probability-and-statistics/famous-mathematicians-and-statisticians-

- Volatility Clustering and GARCH/kaggle. https://www.kaggle.com/nholloway/volatility-clustering-and-garch.

- Stochastic Volatility – Wikipedia. https://en.wikipedia.org/wiki/stochastic_volatility

- Finance Stochastic Volatility – Handwiki. https://handwiki.org/financial-stochastic-volatility

- Stochastic Volatility – Wikipedia. https://en.wikipedia.org/wiki/stochastic-volatility

- (PDF) Volatility Clustering in Financial Markets: Empirical Facts and Agent-Based Models. https://www.researchgate.net/publication/228310932-Volatility-Clustering-in -Financial-Markets-Empirical-Facts-and-Agent-Based-Models.

.