
Empirical data analysis (Bitcoin vs. Dow Jones 2020)

Course: Digital finance

Sebastien Trehet, Amandine Minier, Fares Khalifa, Bilel Rimani, Steve Enyegue

Efrei Paris

January 18th, 2021

Abstract

In this document, the data analysed is taken from US stock markets. We select daily data from the US stock markets, taking selected stocks as well as crypto currency, for the year 2020. Our benchmark for the US market will be the Dow Jones. We will discuss our trading strategy and then test it and analyse the results. Then applying it to several big tech companies that we carefully chose after the statistical study of our data will allow us to predict the latter. It is further assumed that the types of trends and the direction of future price movements can be represented statistically. These assumptions are tested on historical DJIA (Dow) data and confirmed. In addition, it is statistically demonstrated that a number of trends that have occurred in the near past close to the Dow can be used to predict the near future of the index. We will therefore proceed in this way on our data, after extracting them and ensuring their quality. We will study the research models, and come out with the best fitting strategy joining two world of assets.

Keywords: Data Analysis, Dow Jones, Bitcoin, Backtesting, Statistical

Contents

1	Introduction	2
1.1	Definitions	2
1.2	Correlation between Bitcoin and Dow Jones	2
1.3	Hypothesis	3
2	Development of a directional trading strategy	4
2.1	Definition : trading strategy	4
2.2	Trading objectives and ressources	5
2.3	Bollinger bands	5
2.4	Pair trading	6
2.5	Data choosen for pair trading	7
3	Statistical properties of Data	8
3.1	Analysis strategies for research	8
3.2	Forecasting of data	9
3.3	Extracting market data and validating quality	15
3.4	Time data range	17
3.5	Kind of data in statistics	18
4	Improving the trading strategy	20
4.1	Testing on a choosen asset	20
4.2	Backtesting trading strategy over a time horizon of 2020	20
5	Resume	21

1 Introduction

1.1 Definitions

Dow Jones : The Dow Jones is a stock market index that tracks the performance of the 30 largest U.S. companies. It is a value-weighted index, in that it is evaluated by the value of the stocks that make it up and not by other elements such as market capitalization. It is listed on the New York Stock Exchange (NYSE). The Dow's main objective is to provide a broad overview of the health of the U.S. stock market and even the economy as a whole. Traders use it as a benchmark against which to measure the relative performance of a stock.

For the calculation of the index, it is weighted by value, the Dow Jones price is calculated by adding the value of the shares that make it up and dividing the result by a number called the "Dow divisor":

Price weighted average = Sum of prices / Number of stocks

Originally, this divisor corresponded to the number of shares included in the index; it is now regularly adjusted to ensure that the value of the index is not negatively influenced by stock splits, changes in its composition or other modifications.

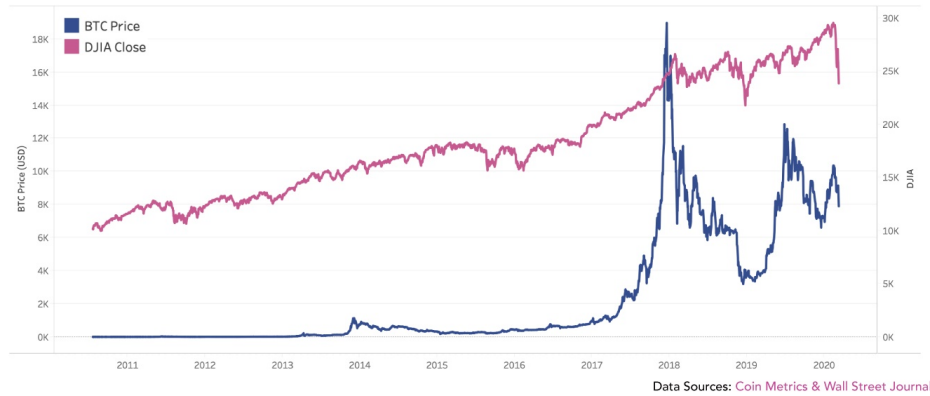
Bitcoin: Bitcoin is an immaterial currency, or crypto-money, that allows its holder to purchase goods and services on the Internet or in real life. Unlike traditional currencies, Bitcoin, like all virtual currencies, does not have a central bank or any central agency or financial institutions to regulate it. As Bitcoin is not regulated by any government or other legal entity, it is considered to be legally unclear. Instead, Bitcoin relies on a vast over-the-counter network on the Internet. Bitcoin's underlying technology is the blockchain or block chain system.

Ether: Ethereum is a blockchain protocol imagined by Vitalik Buterin, a Russian-Canadian developer. Launched in 2015, Ethereum, the second blockchain in terms of recovery after Bitcoin, will be the first to be developed.

1.2 Correlation between Bitcoin and Dow Jones

From the start of Bitcoin around 2010 until March 2020, the two prices have on the contrary a fairly strong positive correlation. That is to say, when one saw its price rise, the other followed it relatively, and similarly in decline.

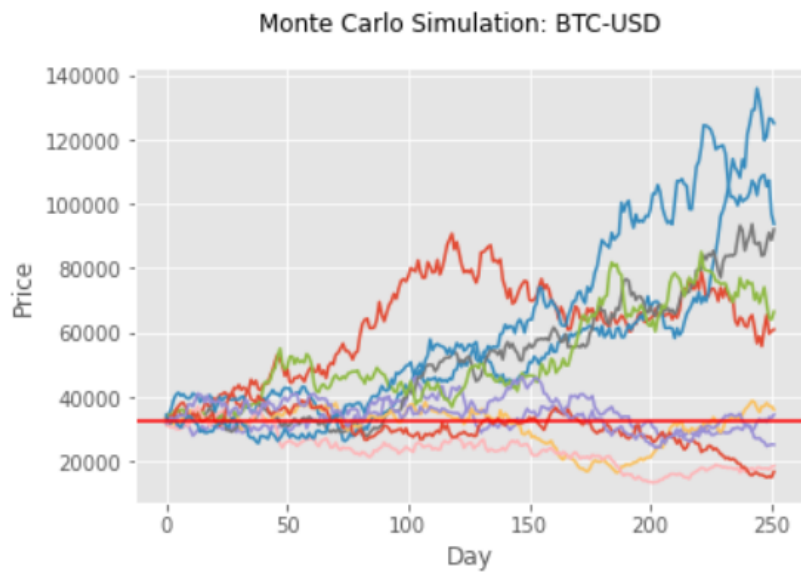
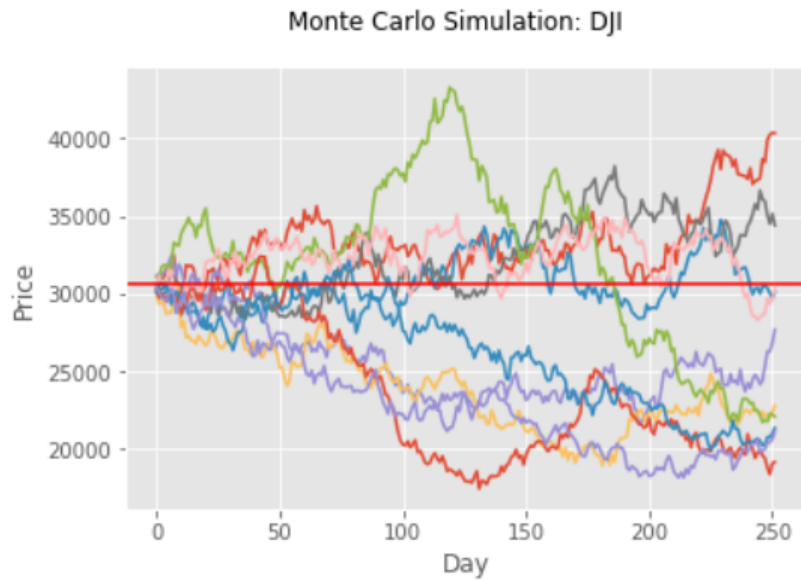
Bitcoin Price and the DJIA, 2010 – March 2020



1.3 Hypothesis

Monte Carlo Simulation: The use of the Monte Carlo approach in the preparation of financial forecasts makes it possible to considerably reduce the uncertainty of the assumptions and to measure more precisely the risk associated with some of the variables. The purpose of Monte Carlo simulations in option valuation is to:

- be able to simulate random price trajectories according to different parameters (interest rate, volatility, exercise price) and variables (strike, option maturity)
- to deduce possible spot levels at maturity
- calculate the expectation, i.e. "each time the final spot is in the currency, calculate the payoff to average it over the totality of the simulations
- Finally to update the whole thing on the departure date, since the payoff will only be available at the end of the term



2 Development of a directional trading strategy

2.1 Definition : trading strategy

A trading strategy is a plan of action for all trades made on the financial markets. Each trader defines beforehand strict investment rules in order to stick to them. Improvisation has no place in trading, it must be structured. Most of the time, a trading strategy consists of two parts: long-term trading

objectives and the means to achieve them.

2.2 Trading objectives and ressources

Our group's objective today is quite simple: Try to achieve a return of 5 to 10 percent by 2020. If this strategy is successful, we will be able to repeat it in 2021.

In terms of the means to achieve this, we will essentially rely on two methods, Bollinger's strips and pair trading strategy. After many tests, we decided to focus on pair trading, we will explain the reasons.

2.3 Bollinger bands

Bollinger Bands were invented by John Bollinger in the 1980s. In practical terms, they are a tool for measuring swings that indicate whether the market has high or low volatility and whether trends are overbought or oversold. They consist of two lines that are both above and below a central moving average, encompassing the price. The purpose of this indicator is to highlight how prices are distributed around an average value. The sidebands react to the action of market prices. They widen when volatility is high and narrow when volatility is low.

To calculate these three bands :

- Middle line: 20-day simple moving average (SMA)
- Upper band: $20\text{-day SMA} + (20\text{-day standard deviation} \times 2)$
- Lower band: $20\text{-day SMA} - (20\text{-day standard deviation} \times 2)$

Concerning the interpretation of these bands, if the moving average price is not within the corridor formed by the two bands, one can assume that the price is probably abnormal.

- If it exceeds the upper band, it means that we have an abnormally high price.
- If it falls below the lower band, it is assumed to be abnormally low.

In both cases, a return to the average is expected, allowing us to anticipate a buy or sell position.

If this analysis is very useful in the classical markets and is interesting to use in the Dow Jones, this technique is much more difficult to apprehend for bitcoin and crypto-currency in general. Indeed, Bitcoin does not have clearly defined metrics like in the classical financial market with financial reports, quarterly performance results etc. There is a lack of defined parameters to use this technique.

Instead, we turned to the pair trading technique which can be used on the Dow Jones but also on Bitcoin.

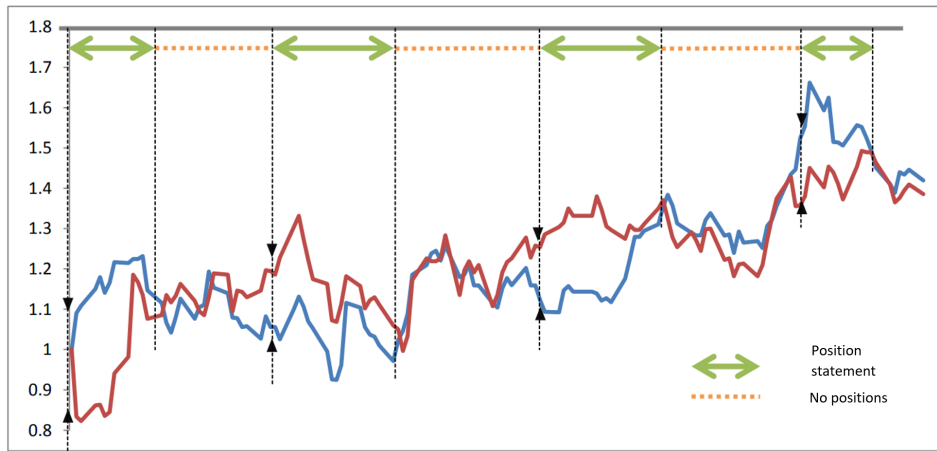
2.4 Pair trading

Pair trading is a trading method that has been used for more than 20 years on various financial markets. This quantitative method is a long/short strategy that aims to find pairs of assets that move together in a market. There are two possible analyses:

- When the gap widens between the two assets, we can hypothesize that, in the long run, the prices of these two assets will eventually meet again because this has been the case for a long period of time. From this observation, the trader can sell the assets on the upside to buy the assets on the downside. Since the two assets will converge, the investor will be able to close his positions and benefit from the realized spread.

- A second interesting point to evaluate in a pair trading strategy is the weight of an asset pair in a market. For example, it appears that when the price of Bitcoin and Ether increases significantly, it creates a pull effect on the crypto-money market. Detecting the increase signals of Bitcoin and Ether would allow to anticipate purchase orders on the other crypto-currencies.

The example below illustrates a pair trading strategy.



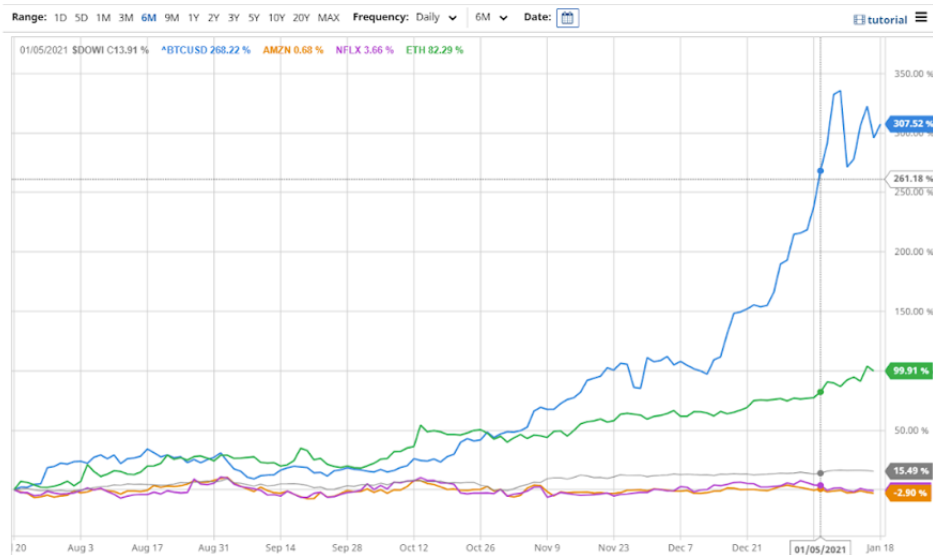
As soon as the spread between the two normalized prices exceeds a defined threshold (represented here by the spread between two black triangles), the investor takes the long/short positions described earlier. When the two prices return to an equivalent level (the curves intersect), the positions are closed and the investor waits for the next threshold to be exceeded. In this example, there were 4 positions that were closed before the end of the trading period. At the end of the period, all positions are closed, regardless of the price situation. Similarly, if the spread continues to grow and exceeds a predefined threshold, the positions are closed (stop-loss).

We will now try to find a pair on the crypto-currency market and on the Dow Jones to test this trading strategy.

2.5 Data choosen for pair trading

In order to test our trading strategy, we decided to select assets that belong to the same domain, namely Tech. Indeed, after making inquiries, we found that the shares of the US Tech had a higher correlation with the Dow Jones. Thus, we chose Netflix and Amazon as our assets were selected because they evolve together. For example, recently, because of the health crisis linked to Covid-19 in 2020, containment has driven up the share price of Netflix and Amazon. Their respective streaming platforms have experienced a strong craze during this period. It was noted that the performance for Netflix and Amazon was respectively 52% and 68% since the beginning of the year 2020. Similarly, in November 2020 when Pfizer's vaccine was announced, Netflix recorded a decline of 8.6%, and the same time for Amazon, Jeff Bezos' firm is down 5.1%.

Evolution over 6 months:



As for crypto-currencies, they are volatile and cryptomoney only obeys the law of supply and demand. The demand is constituted by the people who launch out in the crypto investment. The more reputable a crypto is, the more its demand increases. Today, because of its success, Bitcoin has seen its value soar. So much so that traditional financial institutions are starting to turn to these new assets. However, these assets are unregulated and operate on a decentralized system. Whatever the price, the cryptos market is completely free and remains open 24 hours a day, 7 days a week. Its price is therefore highly influenced by the level of speculation.

The influence of Ethereum and Bitcoin on the cryptocurrency market can be explained by the place they occupy and the role they play. A true spearhead of cryptocurrency, Bitcoin today is the most valuable virtual currency. If it is not really used as a currency of exchange, Bitcoin finds its usefulness as a growth investment or as a digital store of value. On the contrary, Ether finds a concrete use in the real world. Digital payments, intelligent contracts, applications based on the Ethereum blockchain, many applications are attributed to it. Their popularity and the enthusiasm they generate is a definite benefit to all crypto-currencies, which see some of their prices boosted during the growth period of Bitcoin and Ether.

3 Statistical properties of Data

3.1 Analysis strategies for research

We have undertaken two different strategies that allow us to accurately analyze the Dow Jones and also Bitcoin. The first one is based on a historical analysis of the stock market prices of the different American companies. In fact, we have chosen to include companies from the Tech. After making inquiries, we found that the correlation between the Dow Jones and Bitcoin was more important for this type of company. The other strategy is to predict future trends and compare them with historical data.

Indeed, the two elements do not have the same characteristics, especially in terms of date, so it is preferable to forecast these two elements. One, the Dow Jones is the American index following the prices of the 30 largest American stocks and the other the largest cryptocurrency on the market, there are divergences. The Bitcoin digital currency was launched in January 2009. However, active trading and the collection of historical price data came later. It is clear here that a way must be found to compare these two asset classes. Moreover, the two asset classes have very different risk profiles and returns. In order to make a strategy regarding Dow Jones v. Bitcoin, we will often look back at the historical of those two assets. Regarding the historical analysis on Python, we have selected two companies: Apple and Tesla in order to include them in a portfolio that also includes Bitcoin. Then, we selected the length of time they wanted. In our case we chose to take the

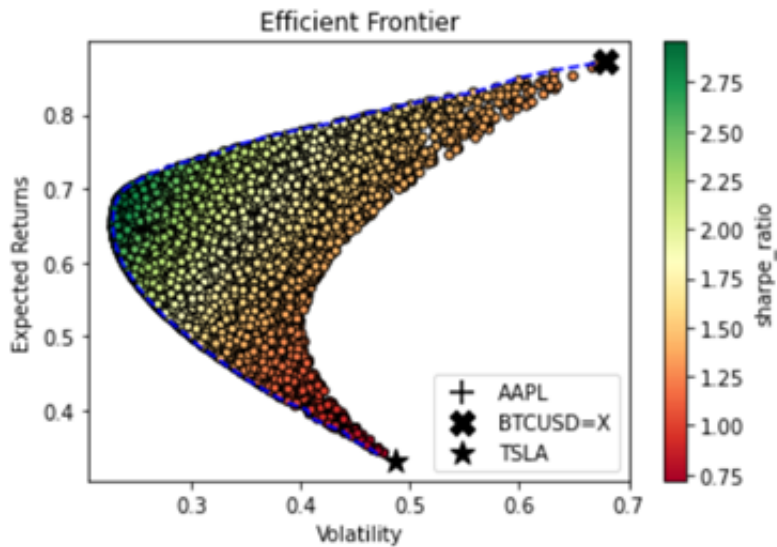
year 2019 (to have all the data available for Bitcoin).

3.2 Forecasting of data

[*****100%*****] 3 of 3 completed

	Adj Close			Close			High			Low			Open
	AAPL	BTCUSD=X	TSLA	AAPL	BTCUSD=X	TSLA	AAPL	BTCUSD=X	TSLA	AAPL	BTCUSD=X	TSLA	AAPL
Date													
2018-12-31	38.205803	3742.699707	66.559998	39.435001	3742.699707	66.559998	39.840000	3725.867188	67.842003	39.119999	3868.742920	65.052002	39.632500
2019-01-01	NaN	3843.519531	NaN	NaN	3843.519531	NaN	NaN	3707.231201	NaN	NaN	3850.914062	NaN	NaN
2019-01-02	38.249401	3943.408936	62.023998	39.480000	3943.408936	62.023998	39.712502	3817.410156	63.026001	38.557499	3947.981201	59.759998	38.722500
2019-01-03	34.439476	3836.741943	60.071999	35.547501	3836.741943	60.071999	36.430000	3826.222168	61.880001	35.500000	3935.684570	59.476002	35.994999
2019-01-04	35.909672	3857.718018	63.537998	37.064999	3857.718018	63.537998	37.137501	3783.854248	63.599998	35.950001	3865.934082	60.546001	36.132500

Each share has a defined weighting to compose a portfolio. Thus, we can assess the profitability and volatility of the various portfolios (constructed with different weights for each share). Then we model these results using efficient frontier to evaluate the expected return versus risk for each share (AAPL, TSLA, BTCUSD). Our time frame evaluates the simulation of 10000 portfolios from January 1st 2019 to December 31st 2019. This graph also allows us to compare its values against the Sharpe Ratio. The Sharpe Ratio shows the difference between the return of the share and its Risk-free rate on the standard deviation, i.e. its volatility.



About prediction analysis, we worked on RStudio software. We use several packages and libraries such as: prophet, rlang, Rcpp, tidyverse. Con-

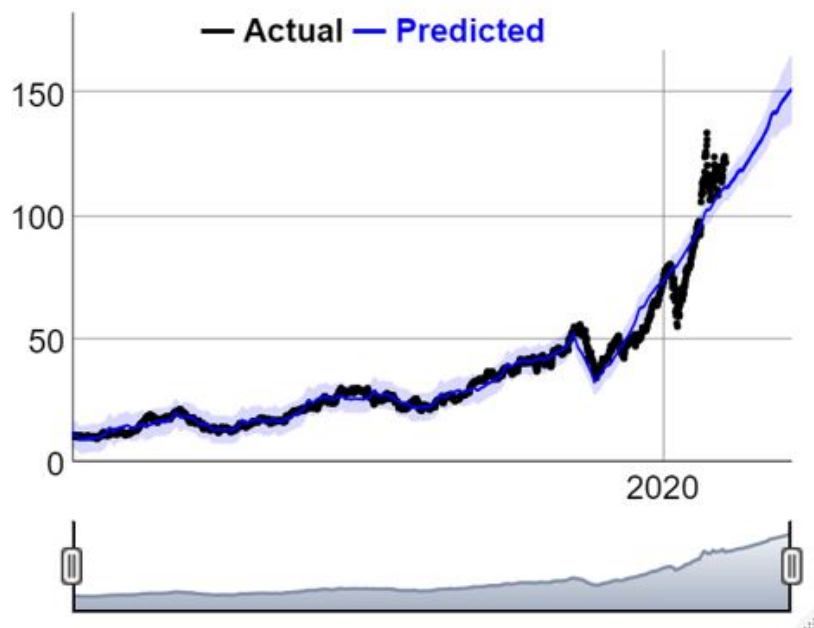
cerning the existing models, several allow us to perform our analysis. However, after our research, two tools have been distinguished: Prophet and DeepAR. The first one, the Prophet tool developed by Facebook, seemed to us the most complete and precise tool to carry out forecasting.

point of comparison	Prophet	DeepAR
Documentation	+	-
Processing multiple time series	-	+
modeling complex seasonality	+	+
distribution of exit probabilities	+	+
interpretability	+	-
training stability	+	-

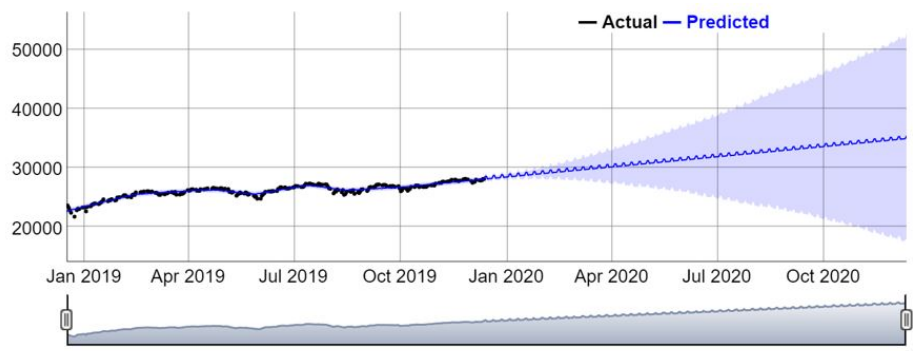
The prophet model is based on the additive model. The additive model is a statistical model. This model fuses the properties of the generalized linear model with those of the additive model. Knowing that the techniques of generalized linear models and those of regression on least squares estimate the parameters of the model to optimize the fit of the latter. The least squares technique minimizes the sum of the squared errors in order to obtain maximum likelihood estimates of the parameters. Generalized linear models are used to obtain maximum likelihood estimates of parameters using an iteratively reweighted least squares algorithm.

First, we modeled the Dow Jones over several years so that we could observe its performance in relation to the economy. On the graph are represented the current values in dotted lines and the forecasts. The latter have a "maximum" and "minimum" bound. Moreover, we may be led to wonder about the events of March 2021. Indeed, we can see that there is an estimate of a decrease: Dow Jones, perhaps a crisis in March 2021.

We said to ourselves that we should try to look at the future of a Dow Jones company and we took Apple which seems to only increase in the near future.



The data of the Dow Jones from 2018 to 2019 to be able to predict 2020 we notice that what we have with the markets have nothing to show at all. So you have to pay close attention to the data taken

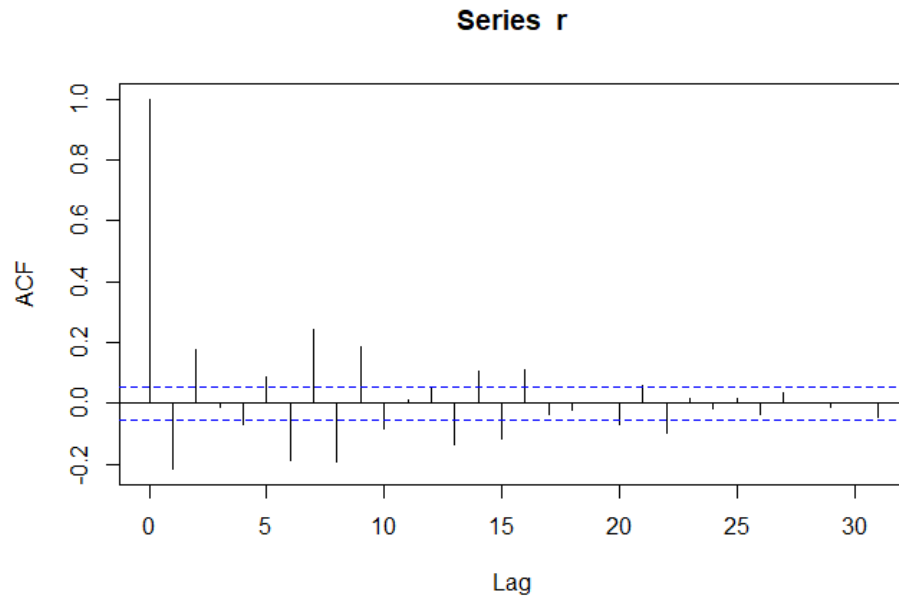


Then we proceeded to different calculations in order to analyze the data. The modeling of the financial series presents some constraints. One must take into account :

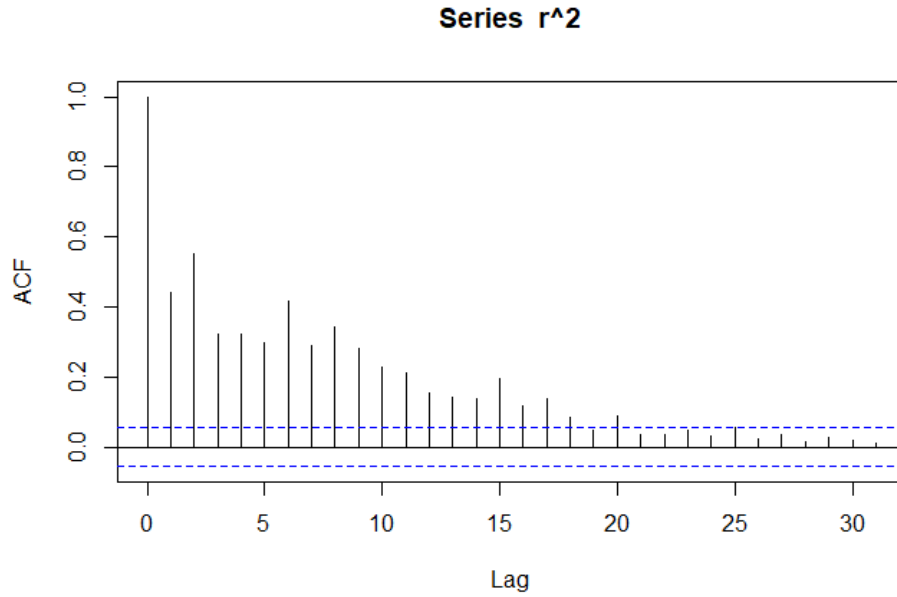
- The non-correlation of returns but autocorrelation of squares
- The non-stationarity of prices

First of all, we decided to calculate the autocorrelation of our series. We compare the series to itself according to different "lag" shifts. The dotted

horizontal line from the "ACF" function indicates the critical threshold beyond which the autocorrelation is considered significant. It can be observed that with a "lag = 0", the function is 1. We can therefore see that the autocorrelations are well within the 95 range, i.e. around zero. The formula applied is as follows:



Date of the Bitcoin from 20/11/2015 to 24/11/2020

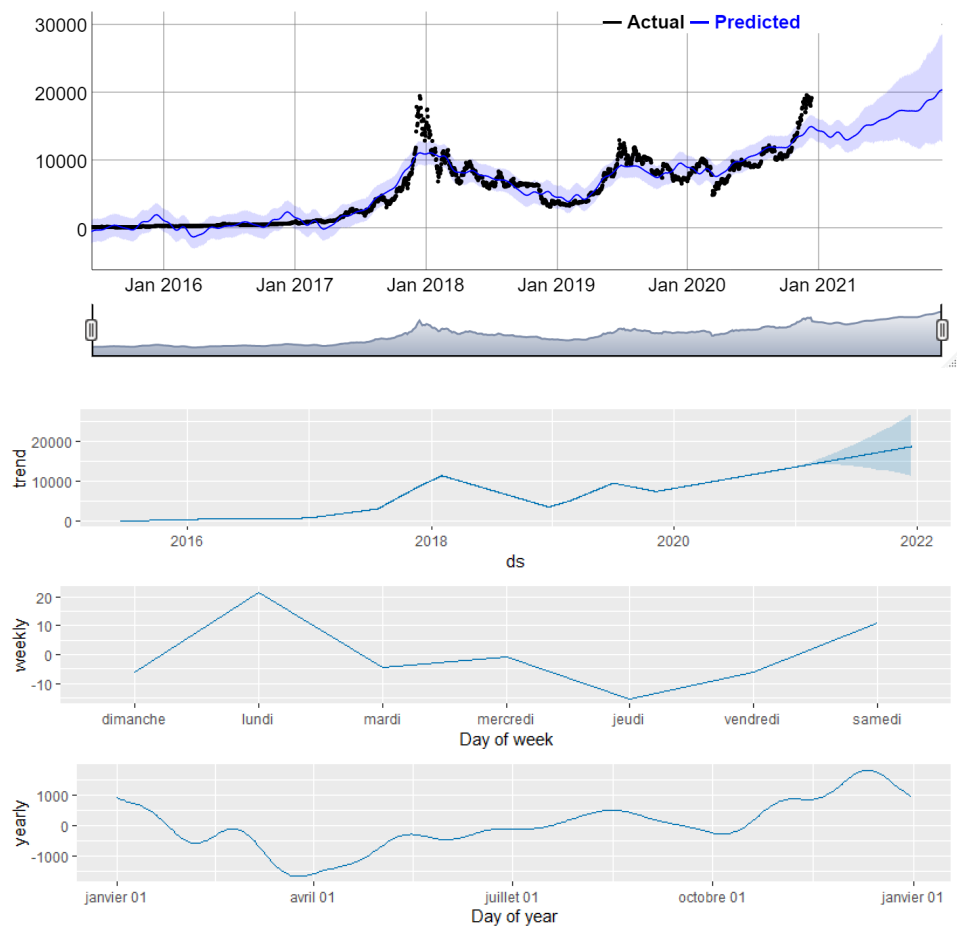


Date of the Bitcoin from 20/11/2015 to 24/11/2020 R^2

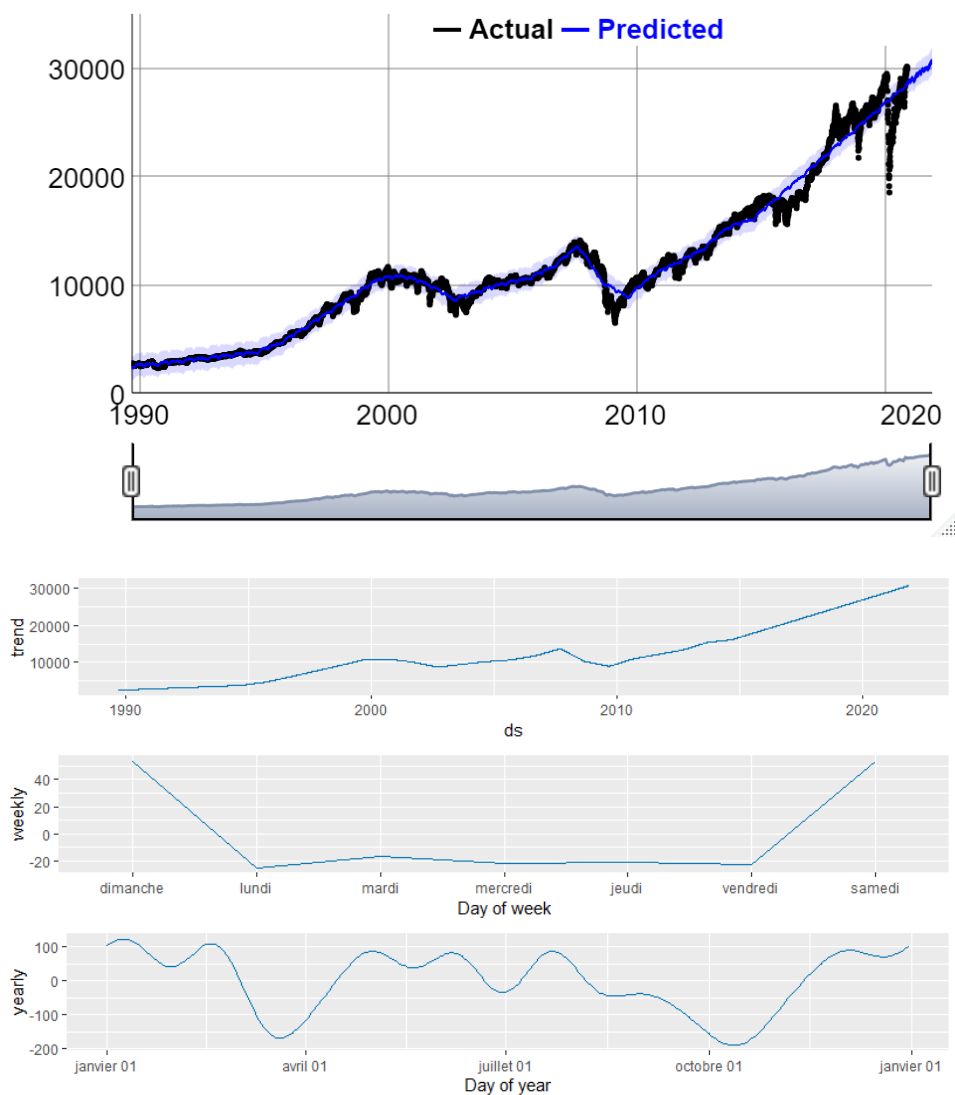
Then, we did the same process again but with the yield squared. We note that the results are slightly different in some places. Nevertheless, we can say that the autocorrelations are still well within the 95 percent range, despite the higher ACF for offsets up to 7.

In addition, we have used the Garch model. To be checked: positivity of the coefficients and stationarity condition, considering the error of estimate. Advantages of the GARCH model:

- More realistic than BS
- Non-constant conditional volatility
- Weak but not strong white noise
- Leptokurticity of the yield distribution
- Much more reactive volatility predictions



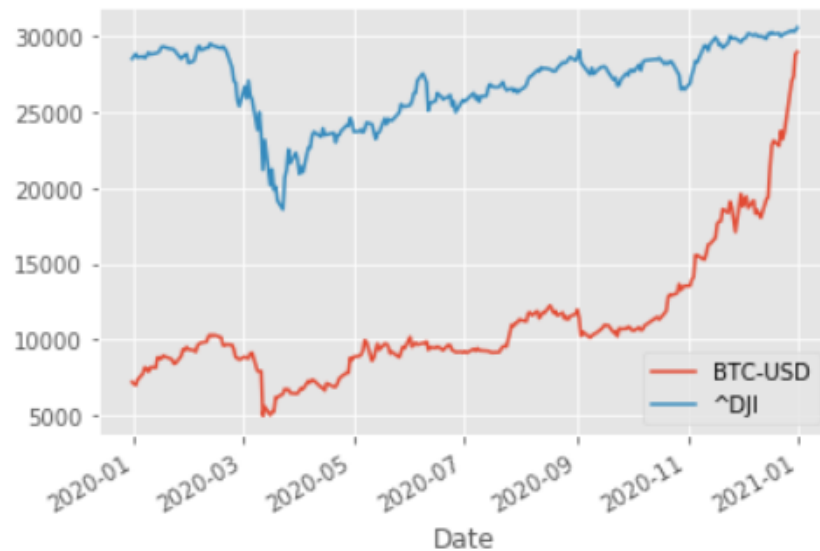
Prediction based on trend seasonality of Bitcoin over 1 year data date from 20/11/2015 to 15/12/2020 prophet plot components(Model 1, Forecast 1)



Prediction based on trend seasonality of Dow Jones over 1 year data date from 20/11/2015 to 15/12/2020 prophet plot components (Model 1, Forecast 1)

3.3 Extracting market data and validating quality

We load data for Bitcoin and Dow Jones. We now decided to compare Bitcoin and Dow Jones daily returns.



We can see that last year both assets were kinda correlated, and in January 2021 both prices started to meet up. However, let's verify that numerically.

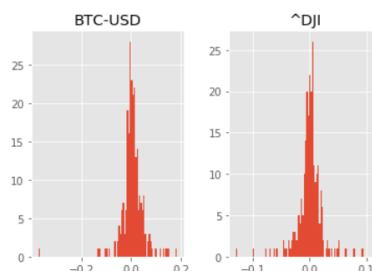
```
corr_matrix_returns = df.pct_change().corr()  
corr_matrix_returns["BTC-USD"]["^DJI"]
```

0.4101251681654658

Let's compare the main attributes of both elements and plot the distribution, the daily returns for the pair BTC/DJ and do a clustering.

	BTC-USD	^DJI
count	249.000000	249.000000
mean	0.006747	0.000552
std	0.045931	0.023200
min	-0.371695	-0.129265
25%	-0.010815	-0.006151
50%	0.004630	0.001570
75%	0.023563	0.009212
max	0.181878	0.113650

(a) Statistical elements



(b) Distribution Plot



We can see the Bitcoin fat tail is higher, and as expected Bitcoin is more volatile than Dow Jones Market. After these analyzes, we can see it doesn't really make any sense to apply pair trading on Bitcoin and Dow Jones Market directly.

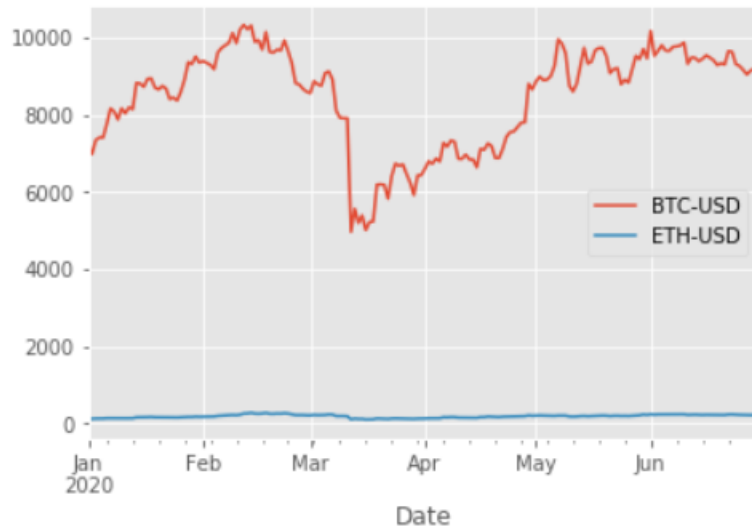
3.4 Time data range

We decided to do analyzes on 6 months periods to have harmonious results because COVID-19 impacted the market a lot so analyzing the whole year wouldn't be as relevant.

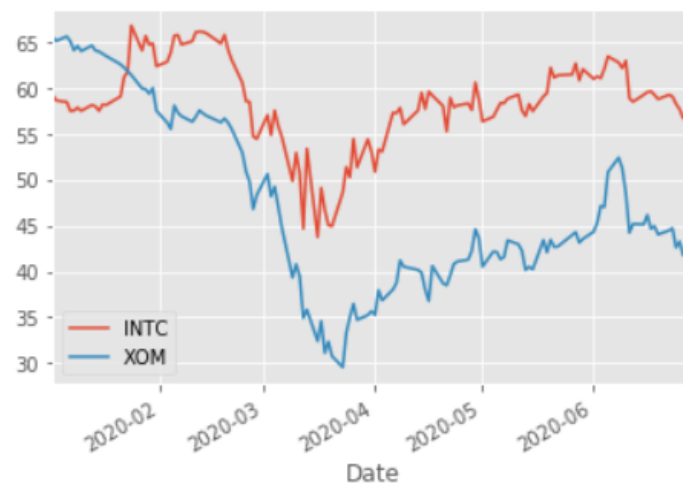
3.5 Kind of data in statistics

We finally decided to do pair trading on two assets from the same field because it made more sense.

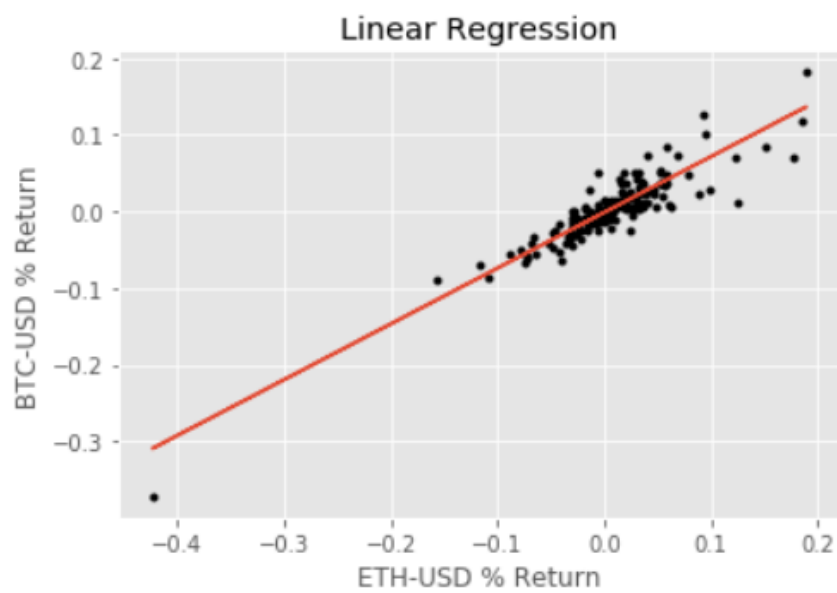
We load data for Bitcoin and Ethereum because they're the most reliable cryptocurrencies.



We load data for Intel and ExxonMobil (stocks from the Dow Jones presenting the best returns for 2020)



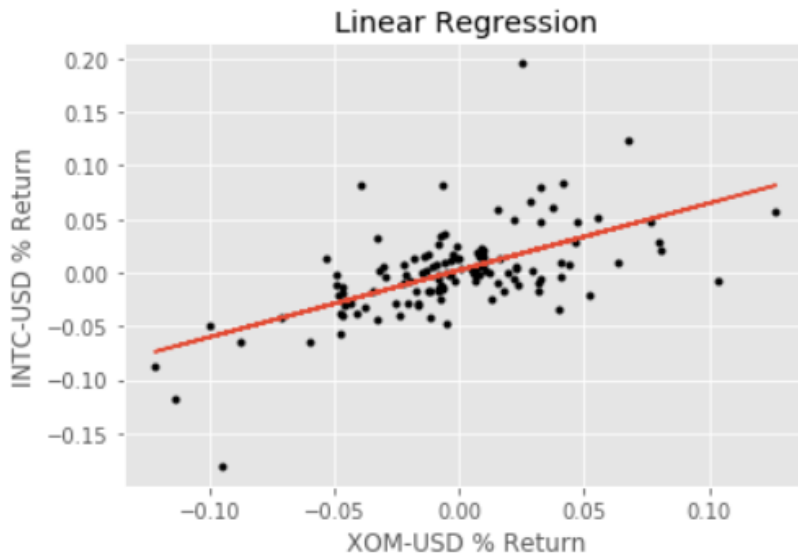
We can now analyze the daily returns of each pair and see the correlation of those with a linear regression.



```
corr_matrix_crypto = daily_crypto_1st_half.corr()  
corr_matrix_crypto["BTC-USD"]["ETH-USD"]
```

```
0.9120591260515014
```

Linear regression and coefficient correlation between Bitcoin and Ethereum



```
corr_matrix_dowjones = daily_dowjones_1st_half.corr()
corr_matrix_dowjones["INTC"]["XOM"]
```

0.5925377326175447

Linear regression and coefficient correlation between Intel and ExxonMobil

Assets' daily returns of both pairs are highly correlated. A pair trading is legitimate and interesting to apply.

4 Improving the trading strategy

4.1 Testing on a choosen asset

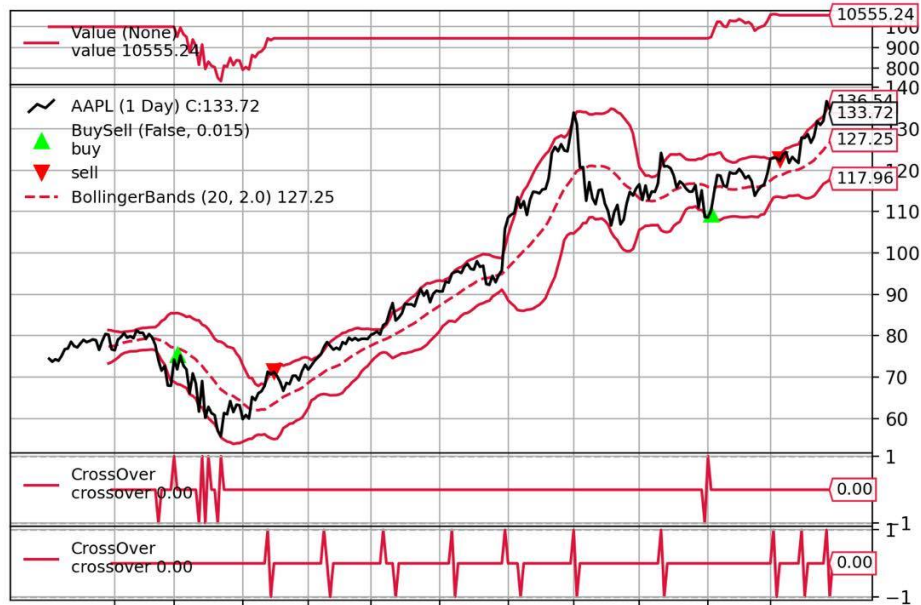
Let us try the Bollinger model on an asset from the Dow Jones. So we will consider Apple (AAPL), since from our statistical analysis it is one of the best company tech in 2020 alongside Intel, and that we analysed earlier.

4.2 Backtesting trading strategy over a time horizon of 2020

The general idea behind backtesting is to evaluate the performance of a trading strategy— built using a practical method or using technical indicators—by applying it to historical data.

By using the backtester library, we consider a basic strategy based on the SMA. As explained, at the start of the paper with Bollinger's Bands, the key points of the strategy are as follows: -When the close price becomes higher

than the 20-day SMA, buy one share. -When the close price becomes lower than the 20-day SMA and we have a share, sell it. -We can only have a maximum of one share at any given time. No short selling is allowed.



Starting Portfolio Value: 10000.00
 2020-03-03, BUY CREATED --- Size: 132, Cash: 10000.00, Open: 75.45, Close: 71.88
 2020-03-03, BUY EXECUTED --- Price: 75.45, Cost: 9959.40, Commission: 9.96
 2020-04-16, SELL CREATED --- Size: 132
 2020-04-16, SELL EXECUTED --- Price: 71.40, Cost: 9959.40, Commission: 9.42
 2020-04-16, OPERATION RESULT --- Gross: -534.60, Net: -553.98
 2020-11-03, BUY CREATED --- Size: 86, Cash: 9446.02, Open: 109.47, Close: 110.25
 2020-11-03, BUY EXECUTED --- Price: 109.47, Cost: 9414.42, Commission: 9.41
 2020-12-04, SELL CREATED --- Size: 86
 2020-12-04, SELL EXECUTED --- Price: 122.60, Cost: 9414.42, Commission: 10.54
 2020-12-04, OPERATION RESULT --- Gross: 1129.18, Net: 1109.22
 Final Portfolio Value: 10555.24

Backtesting the Bollinger model on the year 2020 for APPL: Returns are higher than we the pair trading one.

5 Resume

The idea behind this paper was to explain and use the best fitting strategy to the world of cryptocurrency and classic financial market. As we tried different techniques to maximize our profit, our research work showed that it is unwise to make a strategy based on metrics that are not the same between the two worlds. We concluded that whereas pair trading between assets of both worlds was a good way of trading, the Bollinger Band strategy was the best one concerning the profits made out of our portfolio. A conclusion

made by analysing the statistical parameters of the Bitcoin and the Dow Jones, and by extending our work to a prediction of the stocks movements, see what strategy would be the better one for the future.

References

- [1] Benjamin Graham, "The Intelligent Investor" (original:1949, revised edition: 2006)
- [2] Jack Schwager (1989), " Market Wizards"
- [3] Vernimmen (2011),"Le pair trading et la crise.", Xiayong Qin, Carole Mery
- [4] Rao Dabbeeru, Rao Babu "Investment Patterns and its Strategic Implications for Fund Managers: An Empirical Study of Indian Mutual Funds Industry" (2010)