S.B.Samgat , A.S.Kenzhakhimov

Kazakh British Technological University

Almaty, 050000, Kazakhstan Republic

Predicting Bitcoin prices by calculating the Kullback-Leibler divergence between halving periods.

Abstract

Both large companies and the largest investment banks are showing interest in digital assets. Why institutional investors are turning their attention to crypto assets and how this might affect the global financial system. Therefore, we decided to make an analysis of bitcoin prices from the moment of its popularization to the present day, in order to understand the regularity of bitcoin prices, in order to predict its future prices. Our research will be based on the frequency of Bitcoin hawing. There is a common opinion among crypto traders that every 4 years the Bitcoin chart is repeated due to its halving. In this study, we will try to prove this by calculating the Kullback Leibler correlation and divergence coefficient. Dividing the Bitcoin charts into periods between halvings, we got 3 charts, 2 of which are full-fledged, the period of which has already ended. Comparing these graphs on the metrics of the correlation coefficient and divergence of Kullback Leibler, we got a stable dependence of the two graphs. It is noteworthy that the indicator for the beginning of the period for the market is the previous halving date and not the upcoming one, as no one can predict the exact date of the Bitcoin halving. Judging by our research, we came to the conclusion that the most successful and low-risk point of buying bitcoin is the spring of 2023 and the sale of these coins in the summer of 2025, since the peak on both charts is reached about 430 days from the halving.

Introduction

In the first half of 2021, events have taken place that indicate the adoption of Bitcoin as an alternative asset class. For example, Tesla conducted cryptocurrency transactions, buying $ 1.5 billion bitcoins in February and selling 10% of digital coins at the end of March. The automaker managed to record a profit of $ 101 million, and thanks to this, break the revenue record for the first quarter.

The largest US investment banks have also started using bitcoin. For example, JPMorgan recently announced plans to create the first actively managed Bitcoin fund. And Goldman Sachs promised its clients to provide an opportunity to invest in cryptocurrencies in the second quarter of this year.

Cryptocurrency, namely Bitcoin, is the same variable in the financial market, which has dependencies, so that it can be predicted. This work aims to foresee the best time to buy and sell bitcoin for ordinary people who want to hold cryptocurrency.

Divergention

Divergence (English divergence — "divergence") is the strongest signal of indicator analysis. Its essence lies in the formation of a new extreme in the price in the direction of the dominant trend when this extreme is not updated in the indicator. The price continues its movement along the trend, while the indicator says that the trend has weakened and continues to move only by inertia, and the dominant group of participants is already exhausted, which means that there is a high probability of changing the direction of movement.

Kullback-Leibler divergention

The Kullback—Leibler divergence (Kullback–Leibler divergence), RCL, information divergence, distinguishing information, information gain, relative entropy[1] is a non-negative-valued functional that is an asymmetric measure of the distance from each other of two probability distributions [2] defined on a common space of elementary events. It is often used in information theory and mathematical statistics.

Arthur Hobson proved that the Kullback-Leibler distance is the only measure of the difference between probability distributions that satisfies some desirable properties, which are canonical extensions for entropy appearing in frequently used characterizations.[5] Therefore, mutual information is the only measure of mutual dependence that is subject to certain related conditions, since it can be defined in terms of the RCL. There is also a Bayesian characterization of the Kullback - Leibler distance.

What is correlation

Correlation is the similarity or relationship between two things, people, or ideas. Means the similarity or equivalence that exists between two hypotheses, situations or things. In statistics and mathematics, correlation refers to a measure between variables (two or more) related to each other. The word correlation is a feminine noun, derived from the Latin correlatinene ("cum" (simultaneously) + "relatio" (relation)), read as "correlatione" and means "ratio" or "relationship". The word "correlation" can be replaced by synonyms, such as: connection, dependence, correlation, interconnection, interdependence and mutual correspondence.

What is correlation analyze

The purpose of the correlation coefficient is to determine the intensity of the relationship that exists between known datasets or other known information. The value of the correlation coefficient can vary from -1 to 1, and the result determines whether the correlation is negative or positive. To interpret the coefficient, it is necessary to know that 1 means that the correlation between the variables is completely positive, and -1 means that it is completely negative. If the coefficient is 0, then the variables are independent of each other.

Pearson Correlation Coefficient

In statistics, the Pearson Correlation Coefficient (r-Pearson), also called the Pearson Product Moment Correlation Coefficient (or PPMCC, or PCC), measures the relationship between two variables on the same metric scale.

Goals

1.split datasets by halving periods

2.calculate the coefficient of bitcoin correlations

3.try to predict , prove or disprove the correlation between the 4-year intervals between halving bitcoin

4.calculate the divergention between bitcoin's halvings

5.Prove or disprove the divergention between the 4 year intervals between halving bitcoin

Research methods

Halving

Halving occurs every 210,000 blocks and decreases the reward by 50% each time exponentially. In our project, we analyze whether there is a price divergention between bitcoin halvings periods. As we all know, bitcoin halving occurs every 4 years and represents the maximum increase in the price of bitcoin.

Divergention process and it’s meaning

The Kullback-Leibler divergence (KL) is a measure that allows you to determine how much the information entropy of one distribution differs from the entropy of another distribution. To determine KL, you need to calculate the information entropy of both distributions and find their difference. However, this cannot be done simply head-on, since KL is calculated for one distribution (Q) relative to another distribution (P), therefore, in the first half of the entropy (Q) calculation formula, the probability is used from the distribution (P) relative to which the calculation is performed: . In our experiment, we use three halving ranges, they are called "2012 - 2016", "2016 - 2020" and "2020-2021"

Correlation values and it's meaning

The correlation interval is equal to the range from 0 to 1. The closer the result of correlations is to 1, the higher the dependence between halvings of bitcoin. In our experiment, we use three halving ranges, they are called "2012 - 2016", "2016 - 2020" and "2020-2021"

We calculate the correlation between "2012-2016" and "2016-2020" halvings, then save the result. After that, we calculate the correlation between the "2016-2020" and "2020-2021" halvings, then save the result again

Risks into experiment

The most common type of risks that occurs in most cases:

1)Lack of data

Since cryptoboom is a new phenomenon, we can use datasets starting from 2012. And if we compare halvings, then we can fully compare only two halvings, these are "2012-2016" and "2016-2020". And when we compare the following dates such as "2016-2020" and "2020-2021", and already at this stage there is a shortage of data, since we need a dataset until 2024 before a full-fledged experiment. Because of this, the value of correlations may have some inaccuracy.

2)Dataset inaccuracy

Also, if there are copyright errors in the datasets, it will be very difficult to find out and fix it. This phenomenon can also lead to little inaccuracy and is unpredictable

Process

Splitting datasets

We did not find on the Internet a complete chart of bitcoin prices from the moment of its appearance until today, so we used two datasets of bitcoin prices from 2014 to the present day, and from 2010 to 2018, combining these datasets, we will get an almost complete dataset of bitcoin prices.

In the history of bitcoin, halving took place only three times on 11/28/2012, 07/09/2016 and 05/11/2020

We will compare prices from 11/28/2012 to 07/09/2016 and 07/09/2016 to 05/11/2020 as well as for experiment from 07/09/2016 + and from 05/11/2020+. For this, I divided the dataset into three parts according to the periods between halvings(on picture 1.).

```python
import csv
import datetime
with open('btcusd.csv', encoding='utf-8') as r_file:
    file_reader = csv.reader(r_file, delimiter = ",")
    count = 0
    file_writer = csv.writer(open('btc1st.csv', mode="w", encoding='utf-8'), delimiter = ",", lineterminator="\r")
    file_writer.writerow(['date','open','high','low','close','volume'])
    for row in file_reader:
        if count == 0:
            print(f'{", ".join(row)}')
        else:
            date = datetime.datetime.strptime (row [0],"%Y-%m-%d")
            if date>=datetime.datetime.strptime ("2012-11-28", "%Y-%m-%d") and date<datetime.datetime.strptime ("2016-07-11", "%Y-%m-%d"):
                file_writer.writerow([row[0],row[1],row[2],row[3],row[4],row[5]])

        count += 1
    print(f'Всего в файле {count} строк.')
```

Picture 1.(Python code for splitting datasets)

The graphs between halvings

In the next step we add Bitcoin prices to the arrays and plot their charts(by code on picture 2.).

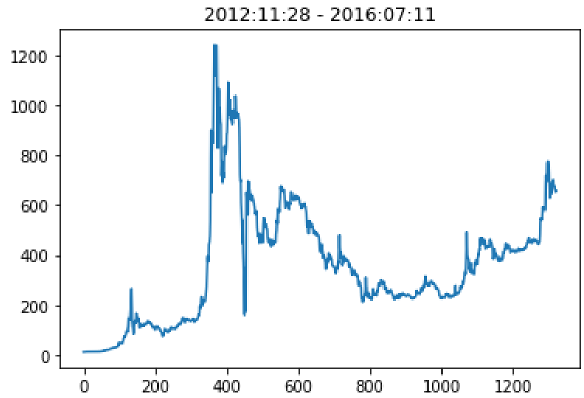
```python
import csv
import datetime
import numpy as np
import math
import matplotlib.pyplot as plt
a=list()
b=list()
bb=list()
c=list()
with open('btc1st.csv', encoding='utf-8') as r_file:
    file_reader = csv.reader(r_file, delimiter = ",")
    count = 0
    for row in file_reader:
        if count != 0:
            x=float(row[2])
            a.append(x)
        count += 1
    print(len(a))
```

```
0.5s
2
```

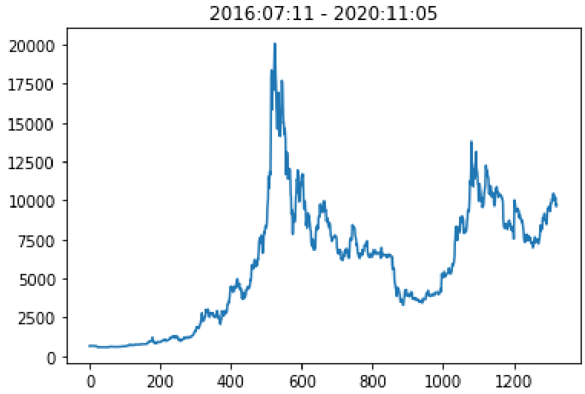
```python
plt.plot(a)
plt.title('2012:11:28 - 2016:07:11')
plt.show()
```

Picture 2.(Python code for reading datasets)

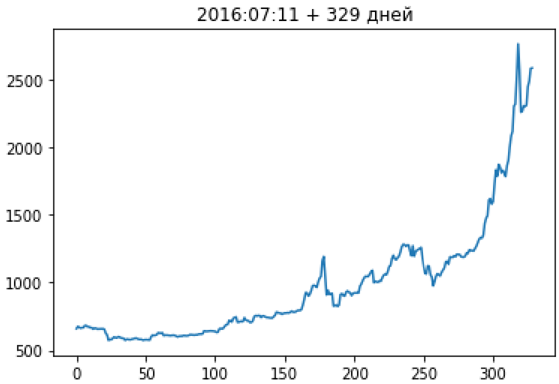
Then we build graphs for each array



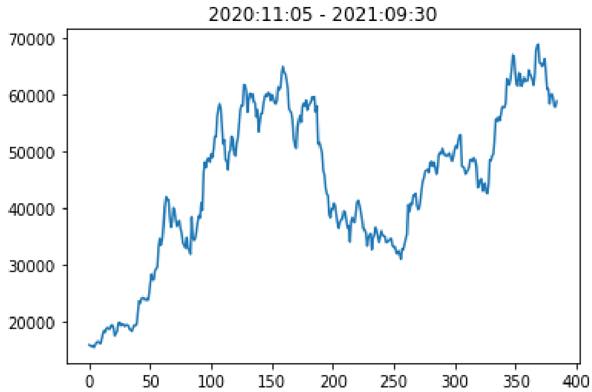
Picture 3. (Periodic Bitcoin graph1)



Picture 4. (Periodic Bitcoin graph2)



Picture 5. (Periodic Bitcoin graph3)



Picture 6. (Periodic Bitcoin graph4)

Looking at the graphs in pictures 3 and 4, you can determine the relationship, but to make sure of this, it is necessary to make calculations.

Calculating the Kullback-Leibler divergence

To calculate this divergence, we used this formula:

$$KL(P\|Q) = -\sum P(x)\log Q(x) + \sum P(x)\log P(x)$$

Python realization of Kullback-Leibler divergence:

```python
def kl(a,b):
    pa =list()
    for i in range(0,len(a)):
        q=a[i]/sum(a)
        pa.append(q)
    Shannon1 = -np.sum(pa*np.log2(pa))
    print(Shannon1)
    pb =list()
    for i in range(0,len(b)):
        q=b[i]/sum(b)
        pb.append(q)
    Shannon2 = -np.sum(pb*np.log2(pb))
    print(Shannon2)
    return(Shannon1-Shannon2)
kl(b,a)
```

Picture 7.(Kullback-Leibler divergence calculation code part)

We calculate the entropy of bitcoin prices from 2012-2016 (9.850013386880555) and 2016-2020 (9.951596262837509)

Subtract from each other, and we get 0.1015828759569537, which indicates a small discrepancy. We also calculated the divergence 2016 + 329 days and 2020-present.

Entropy 2016 (8.43301469804441), 2020 (8.570900984770606), the divergence turned out to be 0.13788628672619652, which also indicates a small discrepancy. Which proves our assumption about the dependence of the two arrays.

Calculating the correlation coefficient

We wrote the code for calculating the Pearson correlation coefficient using the formula that was indicated above. By translating our bitcoin prices into an array, we calculated their correlations.

Corr. coef between 2012-2016 and 2016-2020 Bitcoin prices is: 0.49213954439533764

Corr. coef between 2016-.. and 2020.. Bitcoin prices is: 0.2261482923780638

In the first case, we have a small correlation, but we cannot say that it is stable, in the second case, we have a very weak correlation, most likely this is caused by a lack of information, since we do not have data later than today.

```python
def average(x):
    assert len(x) > 0
    return float(sum(x)) / len(x)

def pearson_def(x, y):
    assert len(x) == len(y)
    n = len(x)
    assert n > 0
    avg_x = average(x)
    avg_y = average(y)
    diffprod = 0
    xdiff2 = 0
    ydiff2 = 0
    for idx in range(n):
        xdiff = x[idx] - avg_x
        ydiff = y[idx] - avg_y
        diffprod += xdiff * ydiff
        xdiff2 += xdiff * xdiff
        ydiff2 += ydiff * ydiff

    return diffprod / math.sqrt(xdiff2 * ydiff2)
```

Формулы коэффициентов корреляции

$$r = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]\left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}}$$

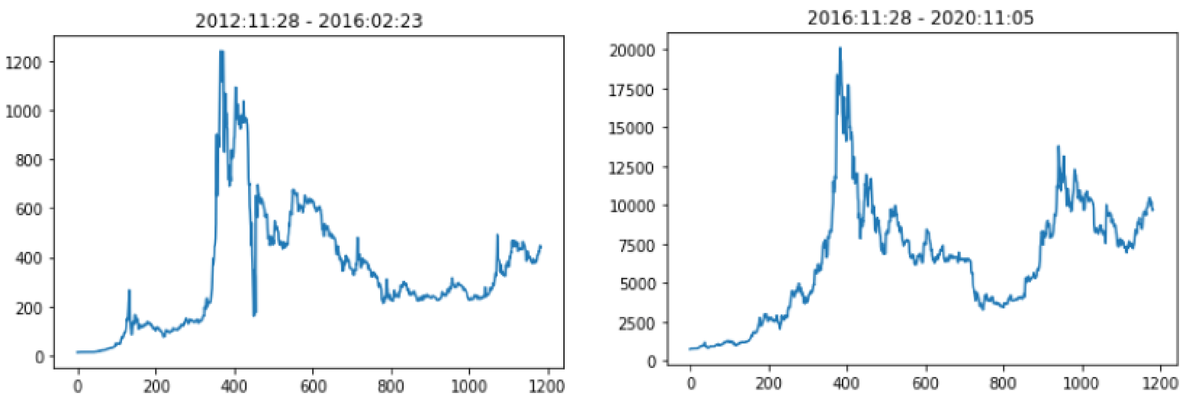
$$r = \frac{\overline{xy} - \overline{x} * \overline{y}}{\sigma_x \sigma_y}$$

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}}$$

Picture 8.(Correlation calculate code) Picture 9.(Correlation coefficient) formula)

Since our correlation was not confirmed, we decided to slightly move the dates of the beginning of the charts, to the day and month of the previous halving, i.e. if we calculated the start dates of the charts on 11/28/2012 and 07/09/2016, then now we will consider the start dates of the charts on 11/28/2012 and 11/28/2016, since the halving date cannot always be predicted accurately and the crypto community relies on the previous halving date and so that the size of the arrays is the same we fitted them to each other. We will not count 2020 - 2021 due to lack of data.

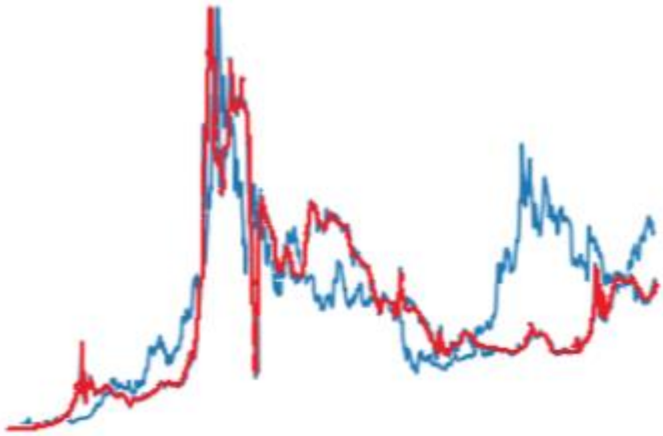
Now our graphs look like this:



Picture 10.(graph1 and graph2)

Calculating their correlation coefficient, we got this: 0.7192444980334539. Which is an indicator of good correlation. Which proves our assumption about the dependence of the two arrays.

Overlay of two graphs on top of each other.



Picture 11.(Compared graphs)

Conclusion

The theory that there is a pattern between the periods between halving has been confirmed, as a result of which some conclusions and predictions can be made. In our opinion, there is no reason for a deviation from the pattern in the next 10 years, since the cryptocurrency bubble, although it exists, is not inflated enough to burst, and this pattern will continue and the growth will be higher with each halving period in the next 10 years.

Our prediction to next halving:

To buy: the spring of 2024

To sell: the summer of 2025

Since the minimum value in the charts was the start point, and the maximum was about 400 days after the point with the day and month of the previous halving, but with the year of the new halving.

The correlation between bitcoin halvings has been confirmed and proved by us using the datasets of the range from 2012 to 2021 inclusive.

And thanks to the divergention, we have made a forecast with confirmation of halving by the calculation of Kullback-Leibler divergention that has a little amount of value means that divergention is small .

The chances of missing bitcoin halvings should be reduced with the help of this article and of course this is an achievement in favor of the development of trading

References:

1. TRADING BOT (Chattan Singh MAY 2019)
2. Algorithmic Trading Bot (Medha Mathur, Satyam Mhadalekar- 2021)
3. Information Entropy and Measures of Market Risk (Department of Statistics and Econometrics March 2017)
4. Bitcoin halvings (Rect. Capital Sep. 2019)
5. Секреты дивергенций (Кучера С окт. 2014)
6. Тонкости дивергенций (Вадим трейд июнь 2019)
7. Эпоха криптовалют (Майкл Кейси)
8. Маркет фокусы криптовалют (Руслан Акст,2015)
9. Биткоин talk (Crypto forum)
10. Coindar (Crypto forum)
11. Coin market cap (Crypto forum)
12. Clockchain (Gonkin Micheev)
13. Altcoinder (Crypto forum)
14. Official bitcoin web-site(https://bitcoin.org)
15. Bitcoin halving(https://www.investopedia.com/bitcoin-halving-4843769)
16. Mastering Bitcoin: Unlocking Digital Cryptocurrencies(Andreas M. Antonopoulos,2014)
17. Blockchain Revolution: How the Technology Behind Bitcoin is Changing Money, Business, and the World(Don Tapscott, Alex Tapscott,2016)
18. The Age of Cryptocurrency: How Bitcoin and the Blockchain Are Challenging the Global Economic Order(Paul Vigna, Michael Casey, 2015)
19. Blockchain: Blueprint for a New Economy(Melanie Swan,2015)
20. Value Web(Chris Skinner,2016)

Dataset of project – Google Диск-
https://drive.google.com/drive/folders/14Y71Pe3DYuxLVsj272-1804p14QdsPIC?usp=sharing