STATISTICAL INVESTIGATION OF THE POSSIBLE RELATIONS BETWEEN BITCOIN PRICES AND NASDAQ INDEX

Avtandil Gagnidze

Professor, Doctor of Sciences in Physics and Mathematics

East European University

Maksim Iavich

Professor, Doctor of Sciences in Mathematics

Caucasus University

Abstract

With the improvement of data technology advances and the sharp addition of web customers number since the 90s, numerous computerized monetary standards are presented. the most popular among them is Bitcoin. It was decided to investigate the possible relations between the most popular cryptocurrency Bitcoin price dynamics and global Nasdaq index dynamics using Mathematical and Statistical methods. The main question is: Are the Bitcoin prices somehow related with Nasdaq Composite Index? We use both, Quantitative and Qualitative data analysis methods to answer this question: Namely, the Regression model and Non-Parametric testing. According to Quantitative methods, it was found that there exists a correlation and the regression equation is not bed: it seems that it is possible to explain about 60% of changes in Bitcoin Prices by changes in the Nasdaq Index. According to Qualitative methods, it was found that these two variables are independent. In this case, the Qualitative conclusion is more likely to be right, and the correlation is most likely because of coincidence.

Keywords: *Statistical Dependence, Correlation, Cryptocurrency*

STATISTICAL INVESTIGATION OF THE POSSIBLE RELATIONS BETWEEN BITCOIN PRICES AND NASDAQ INDEX

Avtandil Gagnidze

Professor, Doctor of Sciences in Physics and Mathematics

East European University

Maksim Iavich

Professor, Doctor of Sciences in Mathematics

Caucasus University

Introduction

To begin, the history of Bitcoin is very interesting. While looking at different ideas on how to investigate if Bitcoin is really the new kind of currency, the idea is that if Bitcoin is „money", then it could be somehow related to the Real Economy and its performance. Computerized cash is a virtual cash, thusly it doesn't have a physical structure, it's unlimited, decentralized system and can displace the cash in any trade. In addition, it is interesting:

- What is the effect of the cryptographic types of cash on the economy of the country and cash subject to existing monetary techniques?
- How might we have the option to think about compensation from computerized cash in the country's finished economy?

Cryptocurrency

The first cryptocurrency was created in 2008 by an unknown person, or gang, known as Satoshi Nakamoto (Marr, 2016). However, there have been times when people have tried to make virtual money that would be cryptographically protected. For example, „Bit-Gold" and „B-Money". They were brought to an early stage, but not fully developed so that anyone could consume them. In 2009, a system was developed to make Bitcoin available to the public. „Mining" was introduced as one of the ways to generate cryptocurrencies. At this time the user will gain new bitcoins and transactions will be written and verified in blockchain. The main purpose of blockchain technology is to create a secure digital identity based on the right combination of open and closed cryptographic keys. The combination of these keys is treated as consent and represents a digital signature. In turn, this digital signature provides strong ownership control. It seems that the Bitcoin is safe, and this is the advantage.

In our previous paper (Gagnidze, 2021) we investigated the possibility of the forecasting Bitcoin price dynamics using Time-Series Analysis. We found that there is no more or less good model for forecasting. So far no one traded in real economy using Bitcoins. That is why it was impossible for anyone to price their cryptocurrency units. However, the time has come for this and Bitcoin needs to be evaluated. Initially, one customer exchanged 10,000 units of BTC in two pizzas. At today's prices, it reaches one hundred million dollars. Later, when many had already heard about bitcoin and reached a certain level of popularity, other cryptocurrencies appeared to offer improved conditions for consumers. For example, more protection, speed, etc. There are more than 1,000 cryptocurrencies in the world today and over time they are being created. Therefore it is not clear for now if Bitcoin can replace the real currency. (Chaim, 2019)

Nasdaq Index

The index was launched in 1971, with a starting value of 100. Over the years, the index has soared tremendously despite multiple periods of decline (Google, n.d.). On July 17, 1995, the index closed above the 1,000 mark for the first time. It made steady gains in the following years to reach 2,000 points by 1998, then began

to accelerate significantly. This process mushroomed in late 1999. The index closed that year at 4,069.31 points. On March 10, 2000, the index finally peaked at an intra-day high of 5,132.52, and closed at an all-time high of 5,048.62.

On March 10, 2000, the NASDAQ Composite peaked at 5,132.52, but fell to 3,227 by April 17, and in the following 30 months fell 78% from its peak. The decline from this peak signaled the beginning of the dot-com bubble burst. There were multiple things contributing to this Dot-com boom and bust. Some optimists thought the internet and the World Wide Web would be more significant to business than any Industrial Revolution in the past, possibly enabling us to achieve Technological Singularity. The index declined to half its value within a year, and finally hit the bottom of the bear market trend on October 10, 2002, with an intra-day low of 1,108.49. While the index gradually recovered since then, it did not trade for more than half of its peak value until May 2007. The 2000s (decade) brought a mix of pessimistic news stemming from the Early 2000s recession, the September 11 attacks, the impending Afghan War and the 2003 invasion of Iraq.

Research Question

It is interesting to answer the following question using statistical methods: Are the Bitcoin prices somehow related with Nasdaq Composite Index?

Quantitative Data Analysis

For the Quantitative Analysis usually using Correlation Analysis and Regression Analysis is appropriate. In statistics, **correlation** or **dependence** is any statistical relationship, whether causal or not, between two random variables or bivariate data. In the broadest sense **correlation** is any statistical association, though it commonly refers to the degree to which a pair of variables are linearly related.

A **correlation** exists between two variables when higher values of one variable consistently go with higher values of another variable or when higher values of one variable consistently go with lower values of another variable. (Bennet, 2014) Correlations are useful because they can indicate a predictive relationship that can be exploited in practice. However, in general, the presence of a correlation is not sufficient to infer the presence of a causal relationship (i.e., correlation does not imply causation) (Bennet, 2014). The most familiar measure of dependence between two quantities is the“ Pearson’s correlation coefficient“, commonly called simply „the correlation coefficient“. Mathematically, it is defined as the quality of least squares fitting to the original data. It is obtained by taking the ratio of the covariance of the two variables in question of our numerical dataset, normalized to the square root of their variances. Correlation coefficient attempts to establish a line of best fit through a dataset of two variables by essentially laying out the expected values and the resulting Pearson’s correlation coefficient indicates how far away the actual dataset is from the expected values. Depending on the sign of our Pearson’s correlation coefficient, we can end up with either a negative or positive correlation if there is any sort of relationship between the variables of our dataset.

Perhaps the most important caution about interpreting correlations is one we’ve already mentioned: ***Correlation does not necessarily imply causality.*** In general, correlations can appear for any of the following three reasons.

1. The correlation may be a *coincidence*.
2. Both correlation variables might be directly influenced by some *common underlying cause*.
3. One of the correlated variables may actually be a *cause* of the other. But note that, even in this case, it may be just one of several causes. (Bennet, 2014)

In any **regression model,** the variable to be predicted is called the **dependent variable** or **response variable**. The value of this is said to be dependent upon the value of an **independent variable**, which is sometimes called an **explanatory variable** or a **predictor variable**. To investigate the relationship between variables, it is helpful to look at a graph of the data. Such a graph is often called a **scatter diagram** or a **scatter plot**.

Simple Linear Regression Model for the population is the equation in the following form:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i,$$

where Y_i are values of a dependent variable, β_0 is the population intercept, β_1 is the population slope coefficient, and X_i are values of an independent variable, and ε_i are values of a random error term.

Because of for the finding of the regression line limited sample data is used, in reality the constructed regression equation is a little bit different and is given in the following form:

$$\hat{Y}_i = b_0 + b_1 X_i$$

where $\hat{Y}_i$ is the predicted value for observation *i*, b_0 is the regression intercept and represents the estimation of the population intercept, b_1 is the regression slope and represents the estimation of the population slope, X_i is the value of independent variable for observation *i*.

The individual random error terms $e_i = (Y_i - \hat{Y}_i) = Y_i - (b_0 + b_1 X_i)$ have mean of zero.

Slope and intercept of the above mentioned regression equation are calculated in order to minimize the Sum of Squared Errors (SSE):

$$minSSE = min\sum e_i^2 = min\sum (Yi - (b_0 + b_1 X_i))^2$$

It's evident that calculations by hands using Differential Calculus are too long and complicated. (Newbold, 2013). Therefore Excel or other Statistical software could be used.

Data Collection

Data was collected from official sites (See (Google, n.d.), (Coindesk, n.d.)) from December 1, 2019 to February 6, 2020. Because of Bitcoin prices in this data are given daily (every day), but Nasdaq Index data is given for working days, for the Weekend and Holiday days Nasdaq Index is assumed to equal to the most close preceding working day. The data is presented in the table and chart below:

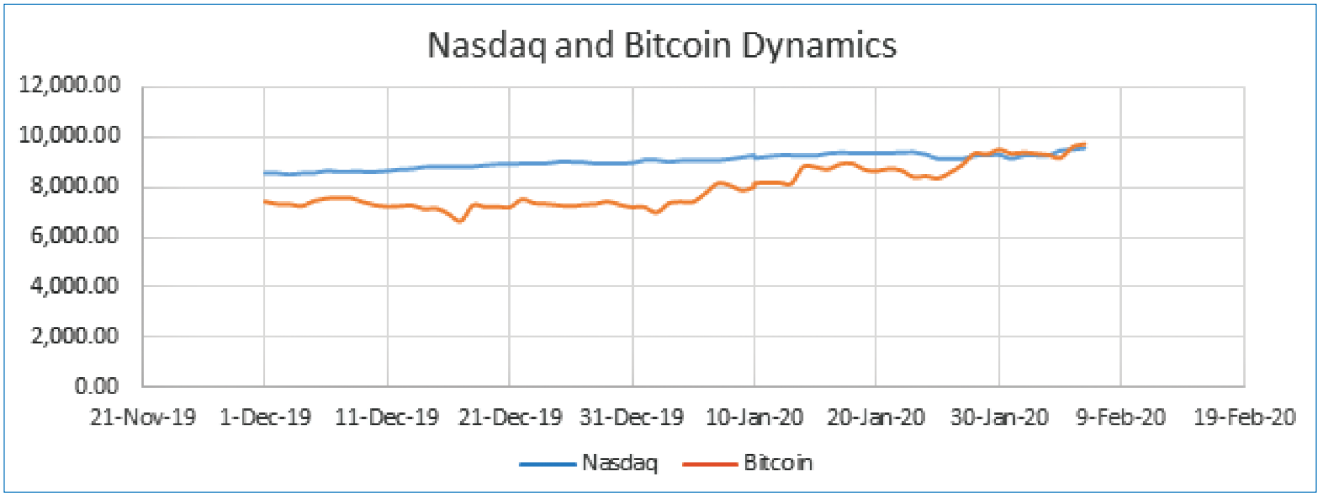
Table 1:

Bitcoin Price and Nasdaq Index Dynamics

Date	Nasdaq Close	Bitcoin Close	Date	Nasdaq Close	Bitcoin Close	Date	Nasdaq Close	Bitcoin Close
1-Dec-19	8,567.99	7424.292	**24-Dec-19**	8,952.88	7322.532	**16-Jan-20**	9,357.13	8723.786
2-Dec-19	8,567.99	7321.988	**25-Dec-19**	9,022.39	7275.156	**17-Jan-20**	9,388.94	8929.038
3-Dec-19	8,520.64	7320.146	**26-Dec-19**	9,022.39	7238.967	**18-Jan-20**	9,370.81	8942.809
4-Dec-19	8,566.67	7252.035	**27-Dec-19**	9,006.62	7290.088	**19-Jan-20**	9,370.81	8706.245
5-Dec-19	8,570.70	7448.308	**28-Dec-19**	8,945.99	7317.99	**20-Jan-20**	9,370.81	8657.643
6-Dec-19	8,656.53	7546.997	**29-Dec-19**	8,945.99	7422.653	**21-Jan-20**	9,370.81	8745.895
7-Dec-19	8,621.83	7556.238	**30-Dec-19**	8,945.99	7292.995	**22-Jan-20**	9,383.77	8680.876
8-Dec-19	8,621.83	7564.345	**31-Dec-19**	8,972.60	7193.599	**23-Jan-20**	9,402.48	8406.516
9-Dec-19	8,621.83	7400.899	**1-Jan-20**	9,092.19	7200.174	**24-Jan-20**	9,314.91	8445.435
10-Dec-19	8,616.18	7278.12	**2-Jan-20**	9,092.19	6985.47	**25-Jan-20**	9,139.31	8367.848
11-Dec-19	8,654.05	7217.427	**3-Jan-20**	9,020.77	7344.884	**26-Jan-20**	9,139.31	8596.83
12-Dec-19	8,717.32	7243.134	**4-Jan-20**	9,071.47	7410.657	**27-Jan-20**	9,139.31	8909.819
13-Dec-19	8,734.88	7269.685	**5-Jan-20**	9,071.47	7411.317	**28-Jan-20**	9,269.68	9358.59
14-Dec-19	8,814.23	7124.674	**6-Jan-20**	9,071.47	7769.219	**29-Jan-20**	9,275.16	9316.63
15-Dec-19	8,814.23	7152.302	**7-Jan-20**	9,068.58	8163.692	**30-Jan-20**	9,298.93	9508.993
16-Dec-19	8,814.23	6932.48	**8-Jan-20**	9,129.24	8079.863	**31-Jan-20**	9,150.94	9350.529
17-Dec-19	8,823.36	6640.515	**9-Jan-20**	9,203.43	7879.071	**1-Feb-20**	9,273.40	9392.875
18-Dec-19	8,827.74	7276.803	**10-Jan-20**	9,273.93	8037.538	**2-Feb-20**	9,273.40	9344.365
19-Dec-19	8,887.22	7202.844	**10-Jan-20**	9,178.86	8166.554	**3-Feb-20**	9,273.40	9293.521
20-Dec-19	8,924.96	7218.816	**12-Jan-20**	9,273.93	8192.494	**4-Feb-20**	9,467.97	9180.963
21-Dec-19	8,924.96	7191.159	**13-Jan-20**	9,273.93	8144.194	**5-Feb-20**	9,508.68	9613.424
22-Dec-19	8,945.65	7511.589	**14-Jan-20**	9,251.33	8827.765	**6-Feb-20**	9,572.15	9729.802
23-Dec-19	8,945.65	7355.628	**15-Jan-20**	9,258.70	8807.011			

Chart 1:

Bitcoin Price and Nasdaq Index Dynamics



The correlation coefficient calculated using Excel equals 0.784073, that indicates the moderate strong correlation. Using Excel Data Analysis Regression tool it is possible to derive the Linear regression model:

Table 2:

Regression summary output

Regression Statistics	
Multiple R	0.784073
R Square	0.614771
Adjusted R Square	0.608934
Standard Error	524.624
Observations	68

ANOVA

	df	SS	MS	F	Significance. F
Regression	1	28989062.57	28989063	105.3265	2.636E-15
Residual	66	18165203.26	275230.4		
Total	67	47154265.83			

	Coefficients	Stand. Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-13627.6	2106.071112	-6.47063	1.4E-08	-17832.52	-9422.7
Nasdaq	2.388713	0.232752892	10.26287	2.64E-15	1.924007	2.853419

The regression equation according to this model is:

Bitcoin Price = -13627.6 + 2.388713*Nasdaq Index

The coefficient of determination is 0.614771. It means that about 61% of changes in Bitcoin Prices could be explained by changes in Nasdaq Index. That is not too bad. At the same time, the Standard Error of the estimate is 524.624. It means that for an estimated value of the particular Bitcoin Price, say 8000, the real price is expected to be roughly between 7480 and 8520. The interval is not small.

Qualitative Data Analysis

When thinking about dependence/independence of two qualitative variables it is usual to consider two Hypothesis:

- The **null hypothesis**, H_0, states that the variables are independent (there is *no relationship* between them).
- The **alternative hypothesis**, H_a, states that there *is* a relationship between the two variables.

With the hypotheses identified, the next step in the hypothesis test is to examine the data set to see if it supports rejecting or not rejecting the null hypothesis.

Data could be displayed efficiently with a **two-way table** (also called a **contingency table**), so named because it displays two variables. (Bennet, 2014)

A **two-way table** shows the relationship between two variables by listing one variable in the rows and the other variable in the columns. The entries in the table's cells are called *frequencies*.

The basic idea of the hypothesis test is to decide whether the data provide enough evidence to reject the null hypothesis.

For the case of a test with a two-way table, the specific steps are as follows:

- Starting by assuming that the null hypothesis is true, meaning there is no relationship between the two variables. In that case, it is expected the frequencies (the numbers in the individual cells) in the two-way table to be those that would occur by pure chance. So the first step, then, is to find a way to calculate the frequencies we would expect by chance. For each cell in the two-way table, identify *O* as the observed frequency and E as the expected frequency if the null hypothesis is true (no relationship between the variables).
- The next step is comparison the frequencies expected by chance to the observed frequencies from the sample, which are the frequencies displayed in the table. To do this chi-square statistic was calculated for the sample data. In general, the larger this value is, the greater the average difference between the observed and expected frequencies in the cells.

Decision rule:

- If the calculated value of $\chi 2$ is less than the critical value, the differences between the observed and expected values are small and there is not enough evidence to reject the null hypothesis.
- If the calculated value of $\chi 2$ is greater than or equal to the critical value, then there is enough evidence in the sample to reject the null hypothesis.

The commonly used critical values for the usual significance level of 0.05 for 3 by 3 tables is 9.488. As for regression, calculations are long and complicated. Therefore again Statistical software is used. (Bennet, 2014)

Qualitative Data preparation

To determine qualitative links, the Nasdaq Index and Bitcoin price percentage changes were calculated:

Table 3:

Changes in Bitcoin Price and Nasdaq Index – Quantitative

Nasdaq % Change	Bitcoin % Change	Nasdaq % Change	Bitcoin % Change	Nasdaq % Change	Bitcoin % Change
0.000	-1.378	0.081	-0.450	0.080	-0.235
-0.553	-0.025	0.776	-0.647	1.063	-0.945
0.540	-0.930	0.000	-0.497	0.340	2.353
0.047	2.706	-0.175	0.706	-0.193	0.154
1.001	1.325	-0.673	0.383	0.000	-2.645
-0.401	0.122	0.000	1.430	0.000	-0.558
0.000	0.107	0.000	-1.747	0.000	1.019
0.000	-2.161	0.297	-1.363	0.138	-0.743
-0.066	-1.659	1.333	0.091	0.199	-3.161
0.440	-0.834	0.000	-2.982	-0.931	0.463
0.731	0.356	-0.786	5.145	-1.885	-0.919
0.201	0.367	0.562	0.895	0.000	2.736
0.908	-1.995	0.000	0.009	0.000	3.641
0.000	0.388	0.000	4.829	1.426	5.037
0.000	-3.073	-0.032	5.077	0.059	-0.448
0.104	-4.212	0.669	-1.027	0.256	2.065
0.050	9.582	0.813	-2.485	-1.591	-1.666
0.674	-1.016	0.766	2.011	1.338	0.453
0.425	0.222	-1.025	1.605	0.000	-0.516
0.000	-0.383	1.036	0.318	0.000	-0.544
0.232	4.456	0.000	-0.590	2.098	-1.211
0.000	-2.076	-0.244	8.393	0.430	4.710

According to the table, the maximal percentage increase for Nasdaq Index is 2.098 and the maximal decrease is 1.885. The corresponding values for Bitcoin Prices are 9.582 and 4.212.

The following criteria for the Nasdaq Index were introduced:

- If the percentage change is less than -0.1%, then Index went down;
- If the percentage change is more than 0.1%, then Index went up;
- Otherwise Index unchanged.

The following criteria for the Bitcoin Prices were introduced:

- If the percentage change is less than -1%, then Price went down;
- If the percentage change is more than 1%, then Price went up;
- Otherwise that Price is unchanged.

According to these criteria, the following Two-Way table is created:

Table 3:

Changes in Bitcoin Price and Nasdaq Index – Qualitative

NASDAQ → / BITCOIN ↓	Down	Stay	Up	Total
Down	1	7	3	11
Stay	8	11	8	27
Up	8	13	7	28
Total	17	31	18	66

After this compilation, we used Excel Add-in for conducting the χ2 test for independence with a 3 by 3 table. The Null Hypothesis is that Nasdaq Index changes and Bitcoin Price changes are independent. Testing results are given in a Table below.

Table 4:

Nasdaq vs Bitcoin Independence Testing

Observed Frequencies			
	Down	Stay	Up
Down	1	7	3
Stay	8	11	8
Up	8	13	7
Expected Frequencies			
	Down	Stay	Up
Down	2.833	5.167	3.000
Stay	6.955	12.682	7.364
Up	7.212	13.152	7.636

After relevant calculations we find that: the critical value of χ2 is 9.488, the test statics is 2.413, and the P-value of the test is 0.660. Because 2.413 < 9.488 and the P-value is too large, we conclude that the Nasdaq Index changes and Bitcoin Price changes are not related.

Conclusions

According to Quantitative methods, it was found that there exists correlation and the regression equation is not bed: it is possible to explain about 61% of changes in Bitcoin Prices by changes in Nasdaq Index. According to Qualitative methods, it was found that these two variables are independent. In this case, the Qualitative conclusion is more likely to be right, and the correlation is most likely because of coincidence.

REFERENCES

Gagnidze A, Iavich M, Time-series Analysis of Bitcoin prices, Collection of Scientific papers, East European University, 2021

Bennet B, Brigs G, Triola M, Statistical Reasoning for Everyday Life, 4th ed. Pearson education 2014

Marr B, A short history of Bitcoin and Cryptocurrency, 2016 https://bernardmarr.com/a- short-history-of-bitcoin-and-crypto-currency-everyone-should-read/

Chaim P, & Laurini, M. „Is Bitcoin a bubble?, 2019. „Statistical Mechanics and its Applications, Elsevier, vol. 517(C), pages 222-232.

CoinDesk https://www.coindesk.com/price/bitcoin/

NASDAQ composite charts, data, news https://finance.yahoo.com/quote/%5EIXIC

Newbold P, Carlson B, Thorne W. Statistics for Business and Economics, Pearson education, 2013.