

ISTANBUL BILGI UNIVERSITY  
INSTITUTE OF SOCIAL SCIENCE  
FINANCIAL ECONOMICS MASTER'S DEGREE PROGRAM

THE RELATIONSHIP BETWEEN BITCOIN RETURNS AND GOOGLE  
TREND: COUNTRY LEVEL EVIDENCE

UĞUR ARSLAN

116620021

DOÇ. DR. ENDER DEMİR

ISTANBUL

2020

THE RELATIONSHIP BETWEEN BITCOİN RETURNS AND GOOGLE

TREND: COUNTRY LEVEL EVIDENCE

BITCOİN FİYAT VE GETİRİSİ İLE GOOGLE ARAMA SONUÇLARI İLİŞKİSİ:

ÜLKE BAZLI İNCELEME

Uğur ARSLAN

116620021

Tez Danışmanı: Doç. Dr. Ender DEMİR  
İstanbul Medeniyet Üniversitesi  
Jüri Üyesi: Doç. Dr. Serda Selin ÖZTÜRK  
İstanbul Bilgi Üniversitesi  
Jüri Üyesi: Dr. Öğr. Fatma Didin Sönmez  
İstanbul Bilgi Üniversitesi

Tezin Onaylandığı Tarih : 02.01.2020

Toplam Sayfa Sayısı: 51

Anahtar Kelimeler (Türkçe)

- 1) Bitcoin
- 2) Google Trend Arama Verileri
- 3) Korelasyon Analizi
- 4) Tahmin Gücü Ölçü
- 5) Granger Nedenselliği

Anahtar Kelimeler (İngilizce)

- 1) Bitcoin
- 2) Google Trend Search Queries
- 3) Correlation Analysis
- 4) Determining Predictive Power
- 5) Granger Causality

## **ACKNOWLEDGEMENTS**

I would like to thank some people whose precious support and love motivated me through this study.

To begin with, I would like to express my sincere thanks to my thesis advisor Prof. Ender Demir for his patient guidance, precious feedback and continuous support.

I would especially like to my college Būşra Fenerci for her valuable contributions to this thesis.

And special thanks to my mother, Gūlbahar. Her encouragement and unconditional love, all my achievements are dedicated to her.

## Table of Contents

ACKNOWLEDGEMENTS .....	iii
LIST OF ABBREVIATIONS .....	v
LIST OF FIGURES .....	vi
LIST OF TABLES .....	vii
ABSTRACT.....	viii
ÖZET .....	ix
CHAPTER 1 INTRODUCTION.....	1
CHAPTER 2 BRIEF HISTORY OF BITCOIN.....	3
CHAPTER 3 LITERATURE REVIEW.....	6
3.1. Evaluating Bitcoin as a Currency.....	6
3.2. Evaluating Bitcoin for the Efficient Market Hypothesis (EMH).....	9
3.3. Bitcoin as a Commodity and Hedging Instrument.....	12
CHAPTER 4 RESEARCH DESIGN AND METHODOLOGY .....	15
4.1. Data Collection Method.....	15
4.2. Descriptive Statistics .....	17
4.3. Correlation Analysis of Bitcoin Price and Google Trends Queries .....	20
4.4. Clustering Countries.....	24
4.5. Defining Break Point .....	26
4.6. Crating New Variables .....	27
4.7. Prediction Analysis .....	30
4.8. Stationary Tests of Variables .....	33
4.9. Granger-causality Tests.....	42
4.10. Granger-causality Tests Findings.....	43
CHAPTER 5 CONCLUSION.....	46
REFERENCES.....	47

## **LIST OF ABBREVIATIONS**

BP	Bitcoin Price
GTS	Google Trends Search
FinCEN	American Financial Crimes Enforcement Network
MSBc	Money Service Businesses
ICOs	Initial Coin Offerings
EMH	Efficient Market Hypothesis
EI	Efficiency Index
FTSE	Financial Times Stock Exchange Index
VIX	S&P 500 Volatility Index
MCC	Matthews Correlation Coefficient
ADF	Augmented Dickey-Fuller

## LIST OF FIGURES

Figure 1. Daily price evolution of Bitcoin .....	5
Figure 2. Bitcoin price and scaled Google Trends worldwide result for Bitcoin search between January 1st 2017 to September 30st 2019 .....	17
Figure 3. Box-and-whiskers Plot for Bitcoin price time between January 1st 2017 to September 30st 2019.....	18
Figure 4. Bitcoin Price Histogram .....	19
Figure 5. Selected Countries Dendrogram .....	25
Figure 6. Visualized Clusters .....	25
Figure 7. The Difference between GTS and BP .....	27
Figure 8. Stationary Test Results Taken for First Part.....	35
Figure 9. Stationary Test After First-order Integration.....	37
Figure 10. BP Stationary Test Results for Second Part .....	38
Figure 11. Stationary Test Results for GTS First Part .....	39
Figure 12. Stationary Test Results for GTS Second Part.....	41

## LIST OF TABLES

Table 1. Descriptive Statistics Results .....	17
Table 2. Moments Results for Figure 4.....	19
Table 3. Normality Test Results Sampled Countries .....	20
Table 4. GTS and Selected Countries Pearson Correlation .....	21
Table 5. GTS and BP Pearson Correlation.....	22
Table 6. GTS and Sampled Countries Spearman Correlation.....	23
Table 7. GTS and BP Spearman Correlation .....	23
Table 8. Clustered Country and Representatives. ....	26
Table 9. First Part Spearman Correlation.....	28
Table 10 Second Part Spearman Correlation .....	29
Table 11 Confusion Matrix Explanation.....	31
Table 12 Confusion Matrix and MCC Results.....	32
Table 13 Stationary Test Steps.....	34
Table 14 ADF Test Result for First Part .....	36
Table 15 ADF Test with I(I) Results for First Part .....	37
Table 16 ADF Test with I(I) Results for Second Part.....	38
Table 17 GTS, ADF Test with I(I) Results for First Part .....	39
Table 18 Stationary Summary Table for First Part .....	40
Table 19 GTS, ADF Test with First Lag Results for First Part .....	41
Table 20 GTS Stationary Summary Table for Second Part.....	42
Table 21 Price Volatility - Stationary Summary.....	42
Table 22 Granger-Causality Results for First Part.....	44
Table 23 Price Volatility - Stationary Summary.....	45

## **ABSTRACT**

### **The Relationship Between Bitcoin Returns and Google Trends: Country-Level Evidence**

This paper investigates the correlation between Google Trends search queries and Bitcoin returns at the country level over period 01 January 2017 to 30 September 2019.

We first examined the correlation between these two variables. We found that the correlation between the two variables decreased significantly after a certain period. Therefore, our data set is divided into two parts. Besides, selected twenty countries were consolidated into 5 groups based on their correlation with Google Trends search data. Moreover, the paper also runs the Granger causality test to check for the causality relationship between Google Trend search queries and Bitcoin price and volatility. Before applying the Granger-causality test, the stationarity tests are also performed. It is shown that there was a one-way causality between search data and Bitcoin price, and a two-way causality between search data and daily volatility of Bitcoin.



## ÖZET

### **Bitcoin Fiyat ve Getirisi ile Google Arama Sonuçları İlişkisi: Ülke Bazlı İnceleme**

Bu makale, Google Trends arama sorguları ve Bitcoin fiyatı arasındaki ilişkiyi seçilen 20 ülke ile 01 Ocak 2017- 30 Eylül 2019 tarihleri arasında incelemektedir. İlk olarak bu iki verinin korelasyonunu incelenmiştir. Belirlenen iki değişkenin korelasyonunun belirli bir dönem sonunda ciddi olarak azaldığını belirlenmiştir. Bu yüzden veri setimizi 2 parçaya ayrılmıştır. Ayrıca, seçilen 20 ülke, Google Trends arama verileri ile olan korelasyonlarına göre 5 grupta konsolide edilmiştir. Ayrıca, Google Trend arama sorguları ile günlük bazda Bitcoin fiyatı ve fiyat dalgalanmaları arasında nedensellik ilişkisi olup olmadığını göstermek için Granger nedensellik testi uygulanmıştır. Granger nedensellik testini uygulamadan önce her bir değişkenin durağanlık testleri yapılmıştır. Yapılan testlerinin sonucunda; arama verileri ile Bitcoin fiyatı arasında tek yönlü, gün içi fiyat dalgalanması ve aramalar arasında ise çift yönlü bir nedensellik olduğu belirlenmiştir.

## **CHAPTER 1**

### **INTRODUCTION**

Bitcoin is a digital currency created by Satoshi Nakamoto in 2009. Since its creation Bitcoin attracted much attention because of its decentralized features and the technology behind it. Bitcoin's popularity increased with an incredible jump in price and market volume. According to [coinmarketcap.com](http://coinmarketcap.com), which shows that as of 18 December 2017 the Bitcoin market capitalization exceeds 320 billion US dollars, Bitcoin is the largest market share and commonly known digital currency. Due to Bitcoin's growing popularity and excessive volatility in price, practitioners and researchers have recently begun to evaluate Bitcoin from an economic and financial perspective.

With the rapid developments in information technologies in the last decades, collecting, processing data and methods to implement these analyses into the financial markets have changed as well. Both personal and corporate investors gained profitable insights affecting their decision-making points. In addition to finance, other in industries like marketing and advertising, the information of search results from search engines like Google are also started to use frequently. According to the research done by Battelle (2005), Google search data is considered the best channel explaining users' past and present tastes and preferences. From this perspective, these data sources have been started to use in behavioral economics as well as academics, especially to predict the price estimations of finance products. The predictive power of the google search queries is examined by Choi et al. (2012). This study shows that in the short-time economic indicator are correlated with search results. Another research, Preis et al. (2010) suggest that Google search volume for a company and price of it in stock is positively correlated on both weekly and daily time scales. Google search engine data is accessible through Google Trends website and users could convert the data to use in modeling. There are many studies associated with Bitcoin price movements and google searches. The first study on Bitcoin price and google searches and twitter hashtag correlation were made by Kristoufek (2013). This study shows that Bitcoin price

and Google research have a bidirectional causal relationship. In another study, which shows Bitcoin price and Wikipedia and Google search correlation to use in algorithmic trade made by Dickerson (2018), the average correlation between Bitcoin price movements and the search results is about 95%.

The purpose of this study is to investigate the impact of Google Search result in a country based. The difference from the other research which has been done with Google search queries for worldwide, it is extended literature the see if there are any differences for the different countries. In order to show country-level differentiation, 20 different countries and worldwide search query results are chosen.

The body of this paper is organized into five major sections. This chapter outlines the Bitcoins' brief history. In the following chapter, the literature on Google search and Bitcoin relation is introduced.

Chapter four summarizes the sampling and data collection methodology as well as survey design and introduces the variables of the study as well as their operationalization. This chapter starts with descriptive findings. This section is followed by the correlation analysis, clustering data, dividing used variables in terms of correlation analysis, stationary test, and Granger-causality tests.

Finally, chapter five concludes with a discussion of the findings, an outline of the limitations of the study and suggestions for future research.

## **CHAPTER 2**

### **BRIEF HISTORY OF BITCOIN**

By the end of 2017, Bitcoin price reaches the highest of all time and it took all the attention when it broke 20,000\$. Even if it was thought to be new, Bitcoin was first created in 2009 by Satoshi Nakamoto. Nakamoto published a White paper named “Bitcoin: A Peer-to-Peer Electronic Cash System,” (Nakamoto 2008) on 31 October 2008. In this paper, Nakamoto explains the peer-to-peer (blockchain) network as a “system for electronic transactions without relying on trust,” (Nakamoto 2008). The main purposes of the creation of Bitcoin are trust, accountability and new payment instrument with a surveillance feature. It was designed to replace the current system.

By doing so, transaction agents would no longer need to have knowledge of one another's identity and transactions could be completed fully anonymously.

Satoshi Nakamoto, whose identity has remained a mystery to this date, has effectively digested the first block of the Bitcoin network, and supervising blockchain technology. Genesis Block is known as a first mined block with a reward with 50 Bitcoin. The first user except Nakamoto is Hal Finley, who downloaded the Bitcoin client and got 10 Bitcoin as a reward from Nakamoto. Nick Szabo and Wei Dai were supporters of Bitcoin after its release. Before disappearing, Nakamoto mined almost 1 million Bitcoin.

On May 2, 2010, when Laszlo Hanyecz bought two pizzas from Papa Johns for 10,000 BTC, the first recorded purchase was made with Bitcoin. This day is still referred to as Bitcoin Pizza Day. The same year, Mt. Gox, the first cryptocurrency exchange market, appeared with the name Bitcoinmarket.com.

In 2011, thanks to the open-source code of the Bitcoin, new cryptocurrencies created based on Bitcoins’ code like Litecoin. These cryptos, also referred to as altcoin, often attempt to improve the original Bitcoin design by offering higher speed, anonymity or other advantages. The same year, the first Bitcoin-related magazine called Bitcoin Magazine released.

In 2012, the Bitcoin Foundation was founded. The purpose of the Foundation is to focus on the standardization, protection and promotion of Bitcoin.

In 2013, the first Bitcoin ATM used in Canada. However, everything did not go perfect for Bitcoin. At that year, Bitcoin faced with exchange and swaps incidents. Bitcoin chain split into two sub-chain which results in some delays in transactions because of a lack of capacity which led to a sharp decrease in price.

If something took that much attention regarding money, it is not surprising to see governments attempt to control whatever it is even if they do not have enough information about it. That's why 2013 can be called a regulatory year for Bitcoin. American Financial Crimes Enforcement Network (FinCEN) established rules set for the decentralized currency. With the regulation, citizens of the U.S get Bitcoin from mining classified as Money Service Businesses (MSBs).

This year was also famous for some of the negative news in terms of legal processes. Ross William Ulbricht, the founder of Silk Road (black market), was arrested in 2013 by the FBI with 26,000 Bitcoin. The use of Bitcoin by financial institutions was banned by the People's Bank of China this year. Foreign Exchange Administration and Policy Department of Thailand announced Bitcoin does not fit any legal framework, using Bitcoin classified as an illegal activity.

In 2014 was an important year for Bitcoin to be accepted as a means of exchange. Many major companies, like Microsoft and Dell, have accepted

Bitcoin as an instrument of payment. Furthermore, Bitcoin-based derivatives mainly over-the-counter swap products started to be traded in 2014.

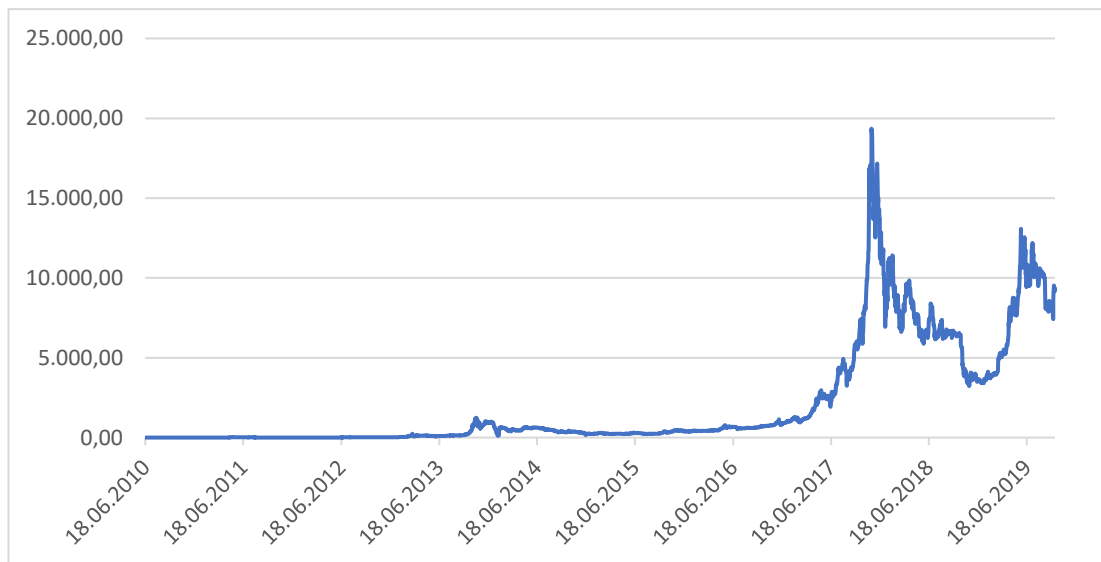
In 2014 MT. Gox, the largest Bitcoin market, declared bankruptcy by stating that 744,000 Bitcoins were stolen. Another market faced the same problem that year. Although Bitstamp, second-biggest Bitcoin market, was hacked and 19,000 Bitcoin stolen, decided to continue their operations.

Following years Bitcoin's reputation and recognition increased all around the world. For example, in March 2016, Bitcoin accepted as a medium of exchange by

Japan authorities. Bidorbuy, North Africa's largest online retail store, has accepted Bitcoin as a payment instrument for both buyers and sellers.

2016 was an important year for cryptocurrencies. We witnessed the Ethereum platform challenging Bitcoin in this year. To facilitate blockchain-based smart contracts and applications, Ether was used. With the help of Ether blockchain there were too many Initial Coin Offerings (ICOs) existed. Many of these blamed to be a scam or Ponzi scheme. As can be seen from Figure 1, next year Bitcoin reach its historical high with almost 20,000\$ and currently in 2019 price of a Bitcoin is about 7,500\$.

Figure 1. Daily price evolution of Bitcoin



Source: <https://coinmarketcap.com/>

## CHAPTER 3

### LITERATURE REVIEW

In this section of the study, Bitcoin is investigated in three different ways. First of all, we discuss whether Bitcoin can be accepted as a currency. Secondly, we collected reviews on whether Bitcoin meets the efficient market hypothesis as an asset. Lastly, we searched for literature on the use of Bitcoin as a hedging instrument against both its price volatility and other investment instruments.

#### **3.1. Evaluating Bitcoin as a Currency**

Money, referred to as currency recognized as a medium of exchange for transaction purposes. It provides a reduction in transaction costs. In order to be useful, a currency should be;

- Fungible
- Durable
- Portable
- Recognizable
- Stable

As mentioned above, money's main function is a medium of exchange. However, it has also secondary functions listed below.

- **Unit of Account** is one of the most important characteristics of the money. It means something that can be used to value goods and services, record debt and make the calculation. In another word, it is a measurement of value. The unit of account also has three main characteristics. Those are being;
  - Divisible
  - Fungible
  - Countable
- **Store of Value** is another important characteristic of the money. Simply under no inflation assumption, putting money in a bank or a strongbox, and

take it out later, does not change the value. This property is useful for future savings.

- **Standard of Deferred Payment**, basically it is anytime that people receive something but only pay for it later, and the most common form of this is debt. With the “legal tender” term banks and institutions credit people or other institutions with the determined type of money for a while. For example, for the mortgage, people borrow from banks in any kind of money with a legal obligation, that is they have to repay their debt.

After giving classical properties of the currency, in this part, it is evaluated whether Bitcoin meets those properties will be discussed.

There is an ongoing discussion about whether Bitcoin is primarily an alternative currency or just a speculative asset (European Central Bank, 2012). In order to determine the values of currency, several theoretical models have been established like Interest Rate Arbitrage, Interest Rate Parity, etc. However, Bitcoin price is determined by users who exchange it on different markets. Because of the limited supply of Bitcoin, an increase in the number of Bitcoin participants is related to the price increase.

In order to evaluate the basic properties of the money for Bitcoin, starting with the medium of exchange. Bitcoin has no intrinsic value and worth of money depending on its usefulness for the consumer. Although the increasing number of merchants who accept Bitcoin as a payment method like Overstack.com, the Bitcoin market gap for the daily exchange is not worthy. Lielacher (2019) claims that the total number of Bitcoin wallets is 32 million. Furthermore, 2.3 million people used Bitcoin to make the payment from its first introduction until August 2018. Compared with the world population, which is about 7 billion, the total number of usages for the purchasing product or service appears extraordinarily negligible. In 2014, Bitcoin’s daily usage for the purchasing good and services is like 15,000 (Yermack, 2015) Furthermore, according to Coinbase co-founder Fred Ehrsam which is the most popular and used Bitcoin market, there were 24,000 merchants



are registered Coinbase. In brief, the Bitcoin market is a shallow and merchant number who accept Bitcoin as a medium of exchange scarce.

Another drawback for Bitcoin arises from the supply of it. Unless a consumer is a miner of Bitcoin, he or she must get it from an online exchange and then need to store it safely. Except for major exchange, liquidity is insufficient, and many exchange markets do not accept credit cards or other payment methods. Besides, bid-ask spread is also high.

Another obstacle can be said that for many consumers making their transactions with a credit card which allows postponing their repayments with zero cost for one month. Bitcoin-based credit card doesn't exist yet and this option is not available.

Lastly, transferring Bitcoin from one wallet to another requires a verification process, which takes on average 10 to 20 minutes. This delay is caused by the mining process. Many merchants and customers in today's standards vulnerable to wait this time lag.

Another feature of the money is being a unit of account. To measure goods and services in terms of Bitcoin, the price level should be stable in time. However, compared to other currencies Bitcoin price change is too volatile. Retailers who accept Bitcoin as a currency need to re-calculate price levels on a daily base which is costly for them and confusing processes for the consumers.

Bitcoin divided into 100 million Satoshi, and the price level for the merchants and the consumer is not calculated in a traditional way. Prices are in the right digit and the length of the decimal is long tailored. This results in making hard to compute price level of the good for costumer and according to Thomas et al. (2009), many people even well-educated would fail to reach calculation of the prices in terms of Bitcoin

The last property of the currency is the store of value. Throughout history, people tend to store their currency to spend in the near future. In order to protect their currency against theft, people put their money into Bank accounts which guarantee and insure against the theft. For Bitcoin, there are no options like Banks or legal

protection. Bitcoin can be stored either in exchange markets' "digital wallets" or "cold storage". However, storing in the market exchange wallet might be too risky because in the short history of Bitcoin there have been several failures of the exchange markets like Mt. Gox. Cold storage is another option for storing but costs are high compared to the traditional Bank services. Even if there were an option to store Bitcoin safe and cheap, people face managing volatility risk in the long run. For example, if anyone bought Bitcoin in December 2017, for today, there is at least a 50% percent decline in the price. Because of the exchange rate of the Bitcoin is too volatile and holding Bitcoin is quite risky, this undermines the ability of Bitcoin as a unit of account.

To sum up, being disconnected from the banking and payment system, decentralized structure (not issued by any state), not protected by legal authority, limited supply and high volatility make it hard to accept as a currency in a short time. Moreover, Bitcoin is not accepted as currency in China that announced to prohibit the use of Bitcoin as a currency for financial institutions (Ruwitch and Sweeney, 2013).

### **3.2. Evaluating Bitcoin for the Efficient Market Hypothesis (EMH)**

Before starting our assessment, basic information about the efficient market hypothesis will be given. As the basis for modern portfolio theory, the efficient market theory is the basis of financial decision making. The theory of effective markets, which continued to be accepted in the field of finance, was put forward by the Nobel Prize-winning economist Fama (1960). In the efficient market hypothesis, prices include and reflect all rational expectations and available information. The underlying logic that random walk of the all subsequent prices which reflect random relation of previous prices. Since all available information is immediately reflected in asset pricing, tomorrow's prices are independent of today's prices and will only reflect news and information about tomorrow. Therefore, both novice and expert investors with diversified portfolio products will receive comparable returns regardless of their different levels of expertise.

There are three forms of the EMH summarized as;

- Weak form: Investors cannot use past information to predict future returns. However, because of the weak form, investor's returns could be higher than the market return average because the return series follows a random walk.
- Semi-strong form: Some information is not readily available to the public. Technical or fundamental analysis is not working to get more return compared with the market average. That is past information is not working for analysis. Yet, investors can pass the market return average.
- Strong form: All the information is reflected in the current price and for the future returns investors cannot go beyond the market return because all the information about the price given.'

With the given information, because Bitcoin is a crawling baby for academic research, some articles are summarized for Bitcoin and EMH.

In order to show whether Bitcoin is a speculative commodity or a currency Urquhart (2016) claims that it should provide efficient market hypothesis rules (FAMA). Even if in many articles' Bitcoin was initially discussed in terms of safety, ethical and legal aspects, recent literature has examined Bitcoin from an economic viewpoint. For the EMH, future prices should not be foreseeable and price should follow the random walk.

To test this assumption in his work, data which includes daily price of Bitcoin between 2010 and 2016 divided into two parts. Whole data (between 2010 and 2013) divided parts tested with 5 different econometric tests. These tests aimed to determine autocorrelations, unit roots, nonlinearity, and long-range dependence in Bitcoin returns. Results show that Bitcoin returns do not provide EMH rules and the market is inefficient. However, for the second part of the data R/S Hurst statistic indicate Bitcoin returns have a semi-strong form of the EMH. The result compared with the emerging market for showing the maturity effect on asset return and EMH relation. He is claiming that with the time market or asset getting older, assets return tends to behave more efficient.

- Urquhart's (2016) study has influenced many people and it brings a new perspective for Bitcoin and EMH relations. After him, many articles were

written by different academicians like Nadarajah and Chu (2016). In their article, they specify EMH as “returns of a financial asset follow a stochastic process memoryless with the given information”. In this article the same data period was chosen to test Urquhart’s results. However, in their work, instead of using Bitcoin daily returns to test EMH efficiency, they used odd integer power of Bitcoin returns. Their claim powering of the returns does not influence any information loss. They extend Urquhart's works with more tests (8 different) and results show that the Bitcoin returns with calculated odd integer power fail to reject for many tests. Their results show that Bitcoin market is not as inefficient as Urquhart claims.

- For now, all research for Bitcoin and EMH tested for Bitcoin’s dollar price. Kristoufek (2018) extend this works with both dollar and yuan markets. Also, in this study, there is a new method namely Efficiency Index (EI) which calculated by Kristoufenk and Vosvrda in order to combine fractional dimension, entropy and long-time dependency. For the two different markets (USD and Yuan) Efficiency Index calculated and results show that distance to the efficient market situation. In this article, after a sharp increase or decrease period in Bitcoin price like the second half of 2016, the Bitcoin market seems efficient according to the efficiency index. This is explained with market volume and price fluctuation. It is claimed that with the regulations and big corporate player entrance and deepening the market volume would change current inefficiency.
- Another article to testing EMH for Bitcoin returns written by Bariviera (2017). In that article, it is claimed that Bitcoin cannot be a hedging instrument because of nonlinearities in the return-volume relationship for predicting returns. Different from Urquhart’s article, in order to test EMH, long memory of the time series tested with Hurst (1951) and extended version of Mandelbrot and Wallis (1968) findings to assess efficiency throughout time. This is also helping to determine the volatility of the daily returns for long-range memory. DFA results show to the necessity to divide data into two subsegments. The results of tests show that the second part of

the data can be evaluated compatible with EMH assumptions and the trend goes toward efficiency.

- Bitcoin price efficiency reconsidered in terms of long-range dependence and time variation in Tiwari et al. (2017) study. Efficiency index reconstructed with the long-time dependency and time variation perspective. The article suggests that under the different assumption and for an exceptional time period BTC market is efficient
- In order to investigate the existence of long-term dependence in the Bitcoin market and gap Katsiampa (2017) introduced the time-varying inefficiency of Bitcoin. The study shows long-term memory analyzed with R/S and variance ratio test. With the help of a rolling window approach determining in long term memory in Bitcoin to test dynamic efficiency of it, market return series presents an inefficient. Different from Urquhart's or other results, dividing data into different samples does not help to make the market efficient.

To sum up, EMH for Bitcoin returns is a new and hot topic for many researchers. Many of them claim in current market prices Bitcoin behave weak or semi-strong efficiency. In addition, with the increasing number of participant and volume, for the recent time period Bitcoin seem more efficient. In the near future, with the help of more individual and corporate participants, accepted as a medium of exchange and widening derivatives Bitcoin returns could be more efficient.

### **3.3. Bitcoin as a Commodity and Hedging Instrument**

In order to evaluate an asset for fitting financial property to invest, it should be that we can count, liquidity, correlation with other assets and hedge capability. In recent years, Bitcoin and other cryptocurrencies have been evaluated within asset classes, such as stock and bonds. However, its popularity is mainly due to high volatility and exponentially increasing price and market volume. This dramatic increase in price attracts not only individual investors but also Wall Street banks and high-frequency traders. For example, with the peak level of Bitcoin price at the end of

2017, Chicago Board Options Exchange and Chicago Mercantile Exchange started to trading Bitcoin futures.

Bitcoin volatility is quite high, which has been addressed by many research like Bariviera (2017), Blau (2017), Katsiampa (2017), Gkillas and Katsiampa (2018). To eliminate loss caused by volatility, Bitcoin futures are investigated in Sebastiano et al. (2019) research. To show whether Bitcoin futures are a good hedge option or not GARCH model was used in their study. Results show that Bitcoin futures are an effective hedging tool at least a daily horizon.

Furthermore, there are several studies about Bitcoin and different stock markets return correlation. Dyhrberg (2016) indicates that Bitcoin is not uncorrelated with the assets in Financial Times Stock Exchange Index (FTSE) on average. Daily return of Bitcoin is not affected by changes in FTSE assets which creates an opportunity for investors to hedge some of the market risks. Yet, for Zhang et al. (2018) claims Bitcoin price changes are related to the economic fundamentals and market conditions. Vasilis et al. (2017) research supporting this by saying Bitcoin and asset prices are correlated in a shorter span.

Estrada (2017) shows the relation between Bitcoin price volatility and S&P 500. The study extended with realized volatility in Bitcoin and S&P 500 Volatility Index (VIX). Vector autoregressive models used in the study to proof writer assumptions. In order to test the relationship and correlation of the variables, Granger- causality test is used. Writer results show that volatility in Bitcoin Granger-causes S&P 500 series at 5% significance level. For the VIX, granger-causality exists in two directions. That is the reciprocal effect which is both Bitcoin price volatility and VIX changes affect another.

Bouri et al. (2017) extent assumption with testing whether Bitcoin is a hedging instrument for different assets and commodities like stocks, gold, oil and Us dollar index. The main argument in this study to see whether Bitcoin behaves as a diversifier. In this study, the bivariate DCC model is used to test authors' assumptions with daily and weekly time series. Results show that Bitcoin can be a good diversifier for Japanese and Asian Pacific stocks under adverse movement,

however, for S&P 500 and other stocks it has no diversifier feature.

To sum up, Bitcoin can be classified as an asset. As much research states that Bitcoin can be hedged with its future, it is a liquid and positive or negative correlation with other commodities or assets.

## CHAPTER 4

### RESEARCH DESIGN AND METHODOLOGY

This chapter of the study focuses on the research methodology. The sample selection and data collection processes are described; variables of the study and their operationalization are introduced, and data analysis methods are explained.

#### 4.1. Data Collection Method

For this research, Google Trends search (GTS) data are used. Data are gathered from <https://trends.google.com> and it shows the relative popularity for “Bitcoin” keyword. This website does not provide exact search volumes, but it provides scaled trends for a keyword in a specified date range for selected variables.

As Google Trends explained as in “<https://ahrefs.com/blog/how-to-use-google-trends-for-keyword-research>”, data show the popularity ratio of Bitcoin to all other possible searches and it is scaled on a range of 0 to 100 for each different query.

For this study, the daily date obtained for the period from January 2017 to September 2019. In order to overcome Google’s limitation on a query for a long time period, the last 3 years’ data gathered from 4 different windows which have the same time interval. In order to re-scale and obtain daily trend data, the whole sample gathered weekly, and weekly data readjusted and converted into daily data with 4 different time windows. In summary; the following steps are constructed to build daily scaled trends data for 3 years:

1. *Weekly* Bitcoin trends data are collected from Google Trends.
2. *Daily* Bitcoin trends data are fetched from Google as 4 parts because of limitations.
3. To use 4 different parts as 1-part, daily coefficients are used to convert weekly trends data to daily data.

Google Trends gives you the freedom to choose keyword search results by country, city, or worldwide. In our study, the data are gathered for worldwide and 20 selected countries. These countries were chosen in accordance with their rank on worldwide search. For the first 15 countries, selection criteria are determined with their rank



in worldwide results. Remaining 5 countries selected by their population size. That is to say, except for the first 15 countries, the most populous countries are selected. In order to normalize the scaled data, all values are compressed between 0 and 1 according to maximum and minimum values. By multiplying the compressed results by 100, we have achieved the final Google Trends results worldwide and by country. The whole process for normalization of the data is made in SAS Enterprise Miner and SAS Enterprise Guide. In this part, to normalize our data “PROC STDIZE” procedure is used.

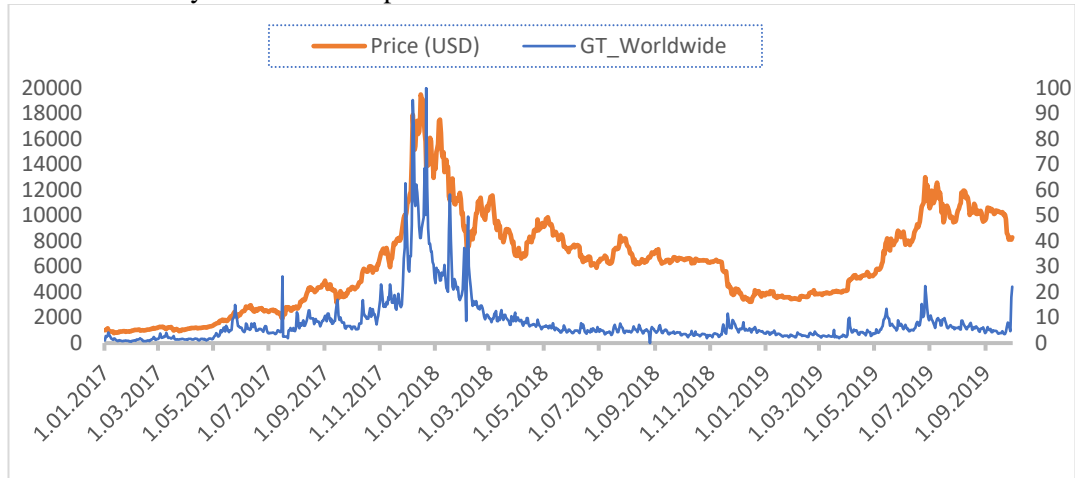
Normalization method used as a given equation;

$$Z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$

Sample of the data includes daily data of Bitcoin in terms of USD rate (XBTUSD) which provides the price levels that are robust in our research. Data sample time period from January 2017 to September 2019. The time span of the data covers market stress events that exist in Bitcoin prices as can be seen in Figure 2. We obtain the transaction data via the web-based Bitcoin data provider <https://coinmarketcap.com> which provides up-to-the-minute updates for all the market data. Our dataset comprises High, Low and Closed prices as well as exchange volumes in BTC a daily basis. In this study, the daily prices of Bitcoin were determined in accordance with Google Trends dates which started from January 2017 to September 2019.

Figure 2. Bitcoin price and scaled Google Trends worldwide result for Bitcoin search between January 2017 to September 2019.

Figure 2. Bitcoin price and scaled Google Trends worldwide result for Bitcoin search between January 1<sup>st</sup> 2017 to September 30<sup>st</sup> 2019



## 4.2. Descriptive Statistics

Data analysis of this study was made by using SAS Enterprise Miner and SAS Enterprise Guide.

Normalized GTS data and Bitcoin prices were combined to create the final data. Our final data includes 1003-day Google Trends information for worldwide and selected 20 countries as well as Bitcoin closing price, daily lowest and highest level. Firstly, descriptive statistics for Bitcoin prices are listed below:

Table 1. Descriptive Statistics Results

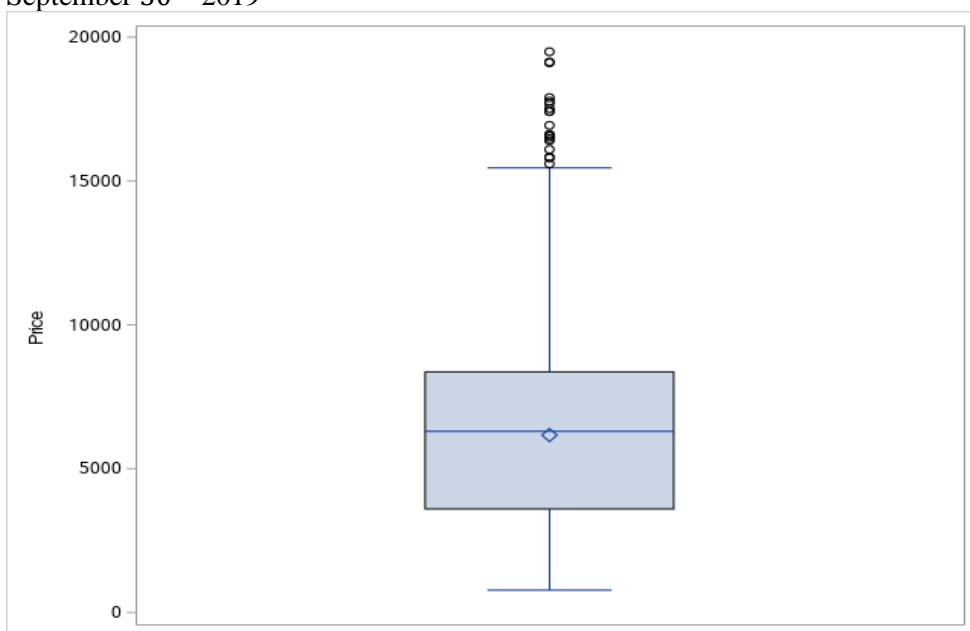
N=1003	Descriptive Statistics Results
Mean	6169.07
Std Dev	3645.62
Minimum	777.76
Maximum	19497.4
Mode	1179.97
Median	6297.57
Range*	18719.64
5th Pctl**	1042.9
95th Pctl**	11941.97
* This is the difference of highest and lowest price	
** This is the lowest 5% threshold level	
*** This is the highest 95% threshold level	

Table 1 shows that the average price of Bitcoin is around 6,150\$ and standard deviation is around 3600\$. Moreover, the table shows the minimum and maximum price of Bitcoin as well as the difference of them namely “Range” which our scaling procedure depends on.

In our data, we don’t have any missing variable problem. However, as can be seen from the table, there is an outlier effect on our data and Bitcoin price is not distributed normally. In order to observe outliers, box plot is drawn which displays the range, interquartile range and mean of a variable. [Krzywinski et al.\(2014\)](#) claim that using box plots in the analysis to observe outliers is too useful because it characterizes a sample using the 25th, 50th and 75th percentiles and the interquartile range, which covers the central 50% of the data.

Like descriptive statistics, Figure 3 represents a Box-and-whiskers plot which shows the interquartile range which is between 3,000\$ and 8,000\$ and the mean is around 6,000\$. However, there are some extreme values for the closing price which is higher than 15,000\$. For some studies, extreme variables can be evaluated as outliers and they might be removed from data, yet, these variables were not excluded in our study because they are meant for the study.

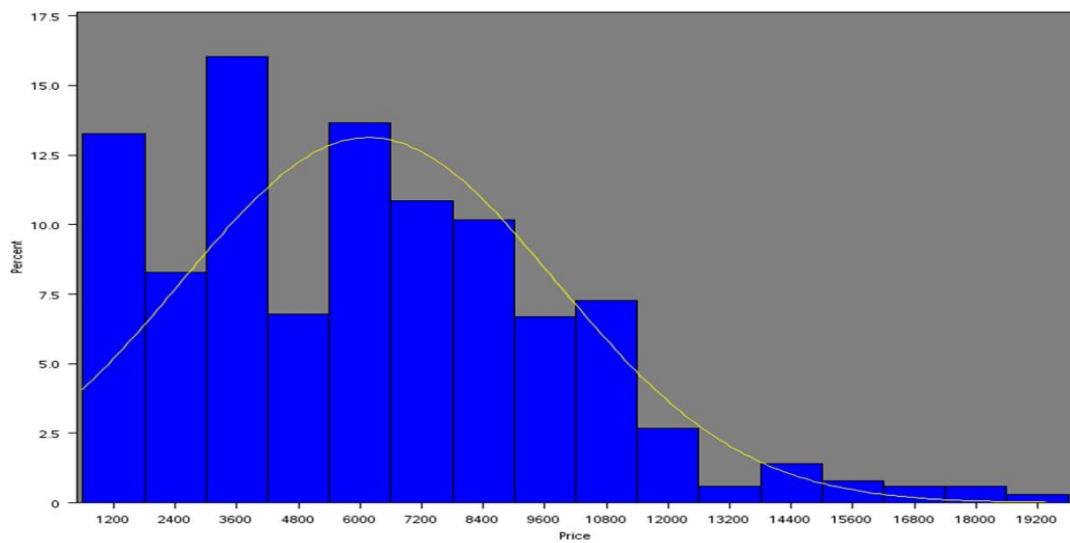
Figure 3. Box-and-whiskers Plot for Bitcoin price time between January 1<sup>st</sup> 2017 to September 30<sup>st</sup> 2019



In the next step, the histogram is drawn for Bitcoins' closing price. Additionally, some statistical results are calculated such as the mean, median, skewness and kurtosis to show that our data are not normally distributed.

As can be seen from Figure 4, the histogram of Bitcoin price is right-skewed which is also called positive-skew distribution. That's because there is a long tail in the positive direction on the horizontal axis.

Figure 4. Bitcoin Price Histogram



Furthermore, moments results reported in table 2, shows the skewness ratio is between 0.5 and 1 which means that our distribution is moderately skewed. The mean of the Bitcoin price is also at peak at the right side of the graph.

Table 2. Moments Results for Figure 4

Moments			
N	1003	Sum Weights	1003
Mean	616.906.777	Sum Observations	6187574.97
Std Deviation	36.456.178	Variance	13290529.2
Skewness	0.64621303	Kurtosis	0.31960195

In order to standardize the distribution of Bitcoin price, we tried many transformation methods discussed in Hosseinifard et al. (2009), namely lognormal, power, and square root which is often used in both applied and theoretical statistics.

Yet, for all different standardization methods, our distribution is not standardized. So, we decided to continue with non-normal data in our analysis.

Furthermore, for the selected 20 different countries and worldwide Google Trends search query results tested. As the Bitcoin price, scaled GTS queries for all countries are right-skewed and non-standardized distribution they have. Table 3 shows some selected countries to show the unstandardized distribution of GTS queries.

Table 3. Normality Test Results Sampled Countries

Country Name N=1003	Kolmogorov-Smirnov		Cramer-von Mises	
	Statistic ( <i>D</i> )*	Significance level **	Statistic ( <i>W</i> )*	Significance level **
USA	0.27	<0.01	28,3	<0.005
Germany	0.22	<0.01	17,2	<0.005
AUSTRALIA	0.25	<0.01	25,1	<0.005
AUSTRIA	0.21	<0.01	16,99	<0.005
*. This is a lower bound of the true significance.				
**. <i>D</i> and <i>W</i> are test statistics for Kolmogorov-Smirnow and Cramer-von Mises				

As it is mentioned before, the normalization method is applied for not only Google Trends query results but also Bitcoin price data. Figure 2 shows that the outliers for both Google Trends searches and Bitcoin price points overlap at a similar time period. Since extreme values are observed in the same periods and this match is significant in correlation analysis, we continue with non-standardized data.

### 4.3. Correlation Analysis of Bitcoin Price and Google Trends Queries

In this part, GTS data and Bitcoin price correlation will be examined. In general, the outlier is expected to disrupt the correlation analysis but, in our data, as can be seen from Figure 2, outliers of both series exist in a similar time period. Furthermore, selected countries' trend searches are also evaluated. According to the results, as we mentioned before, all countries and worldwide GTS have a right-skewed distribution. Moreover, for all normalized and scaled searches which are between 0 and 100, 99% of data have a lower than "60 scaled value" trend for all countries. This result shows that the outlier effect is similar to the country level.

In this step, correlation analyses are conducted for each country search and the worldwide search as well as Bitcoin price. Initially, to check the correlation between price and GTS we used Pearson correlation.

The Pearson product-moment correlation measures both the strength and the direction of a linear relationship. If the first variable is an exact linear function of the second variable, it shows that a relationship exists between these variables. According to the correlation coefficient sign which gives the direction of the correlation, it is decided whether there is a negative or positive correlation between these variables.

The formula for the population Pearson product-moment correlation, denoted  $\rho_{xy}$  is;

$$\rho_{xy} = \frac{COV(x,y)}{\sqrt{V(x)V(y)}} = \frac{E((x-E(x))(y-E(y)))}{\sqrt{E(x-E(x))^2 E(y-E(y))^2}} \quad (2)$$

Table 4 given below, shows the correlation between the worldwide and country-level search examples.

Table 4. GTS and Selected Countries Pearson Correlation

First Variable	Second Variable	Correlation
USA	Worldwide	99.4%
United Kingdom	Worldwide	99.2%
Nigeria	Worldwide	51.8%

It shows that some countries like the USA or the United Kingdom are highly correlated with worldwide, whereas some countries like Nigeria have low correlation. This means that the popularity of Bitcoin for the USA and the United Kingdom appears similar whereas Nigeria GTS behaves differently within the given time period. In our study, we assume that down to %70 correlation, can be enough to accept it as a high correlation. Schober et al. (2018) claim that more than %70 correlation can be classified as a "strong correlation" which supports our threshold level in our study.

Furthermore, we checked the correlation between GTS and Bitcoin price. Table 5 shows the correlation between the Bitcoin closing price and Google Trends searches

of countries on a daily base. The most correlated country in our data is Turkey and the least correlated one is Russia. Therefore, correlation with closing price significantly changes for each country level.

Table 5. GTS and BP Pearson Correlation

First Variable	Second Variable	Correlation
Worldwide	Closing Price	62.1%
Turkey	Closing Price	71.1%
Brazil	Closing Price	67.1%
Russia	Closing Price	46.3%

According to some studies like Myers et al (2006) and Chok (2010) Pearson's correlation works in data that have a normal distribution. Myers et al. (2006) study it is claimed that for Pearson correlation, data should have the standard normal distribution. If data are not normally distributed, then Spearman's rank-correlation coefficient should be used in the analysis. As we mentioned before, our Bitcoin price data is not normally distributed. Therefore, we calculate the correlation with Spearman's rank-correlation in order to calculate the correlation of variables more accurately. Spearman's rank-correlation is a non - parametric measurement of association that is based on the rank of the data values. The correlation range in Spearman is the same as with the Pearson which is from -1 to 1. Unlike Pearson's correlation, Spearman correlation is robust to outliers and it does not expect that the data are normal. It applies to the rank and provides a measure of the monotonic relationship between given variables. Spearman's correlation formula is given below.

$$\theta = \frac{\sum_i ((R_i - \bar{R})(S_i - \bar{S}))}{\sqrt{\sum_i (R_i - \bar{R})^2 \sum_i (S_i - \bar{S})^2}} \quad (3)$$

In this formula  $R_i$  is the rank of one variable,  $S_i$  is the rank of another variable,  $\bar{R}$  is the mean of the  $R_i$  values, and  $\bar{S}$  is the mean of the  $S_i$  values.

Table 6 represents Spearman's correlation of sampled country's searches between worldwide searches.

Table 6. GTS and Sampled Countries Spearman Correlation

<b>First Variable</b>	<b>Second Variable</b>	<b>Correlation</b>
USA	Worldwide	97.5%
United Kingdom	Worldwide	95.3%
Nigeria	Worldwide	37.4%

Compared with Pearson’s results, the USA and the United Kingdom are still the highest correlated countries with worldwide, and Nigeria is the lowest one. However, for both correlated and uncorrelated countries, their correlation level is decreased. Nigeria is a good example to show Pearson’s and Spearman’s correlation difference which drops from 52 to 37.

Table 7 represents the Spearman’s correlation of closing price and sampled countries. In this table, different than the countries and worldwide correlation, Spearman’s correlation shows bettering relationship for worldwide and price, whereas Brazil and Russia’s correlation between closing price worsening.

Table 7. GTS and BP Spearman Correlation

<b>First Variable</b>	<b>Second Variable</b>	<b>Correlation</b>
Worldwide	Closing Price	64.7%
Turkey	Closing Price	79.7%
Brazil	Closing Price	64.3%
Russia	Closing Price	34.0%
Japan	Closing Price	27.0%

After all, we observe some countries GTS is highly correlated with worldwide GTS some are not. Furthermore, Bitcoin’s daily closing price level correlated with worldwide GTS (around 65%) and some selected countries’ GTS. However, correlation does not imply causality. Since, in some cases, an underlying causal relationship might not exist According to Bianchi (1995), correlation itself can provide misleading inference about the underlying relationship of causality. In order to test the high correlation and causality, the Granger Causality test will be performed in the following parts.



#### 4.4. Clustering Countries

In the previous chapter, we explain that most of the countries are highly correlated with worldwide and each other. Therefore, in this part variable clustering method is used to cluster these similar countries.

The variable clustering feature is used for data reduction and it selects the best variables which represent all other variables in the same cluster. In order to select different behaviors for selected countries in our data, we used this methodology and we grouped 20 countries in 5 different clusters. SAS Enterprise Miner “Variable Cluster Node” is used for grouping the countries in terms of their scaled Google Search time series. This process is summarized below.

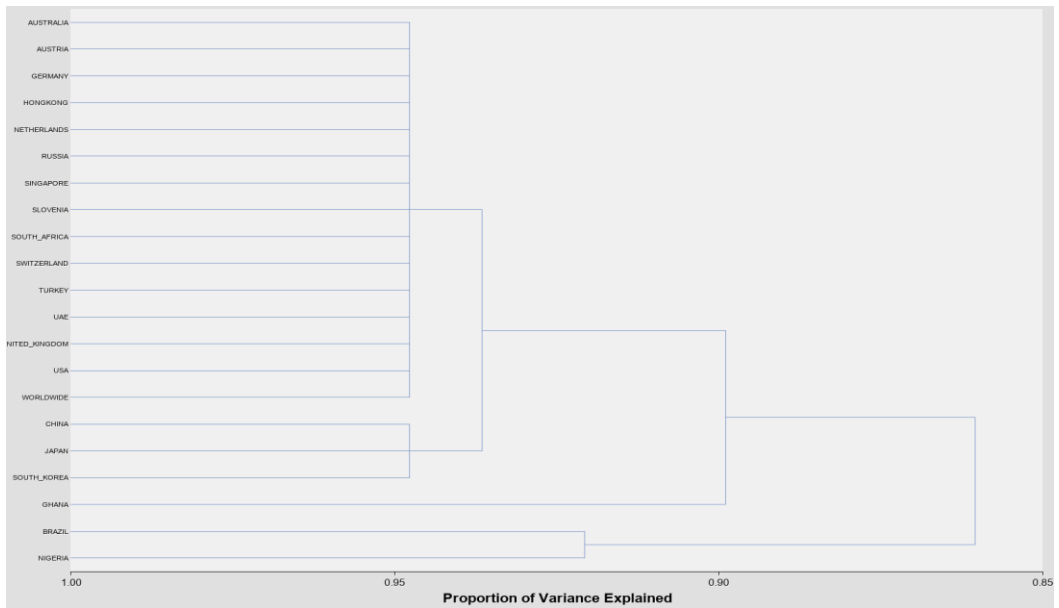
Variable Clustering node takes all data as one single cluster. Then it follows the below steps:

- 1) Clustering Variable Node chooses a cluster for splitting. This cluster might have the smallest variation or have the highest eigen value.
- 2) Chosen cluster is split into two clusters and 2 principal components are selected as the best variables.
- 3) It assigns all the other variables into these two clusters according to their higher correlation.
- 4) Each variable re-assigned to a cluster again to maximize the variables.
- 5) It stops when each cluster has one eigenvalue which is greater than 1.

In our study, correlation is used to cluster 20 countries and worldwide Google Search series together. Additionally, we detected principal components which are the best representation variables for our clusters. The best variable in each cluster is the variable that has the highest  $R^2$  Ratio.

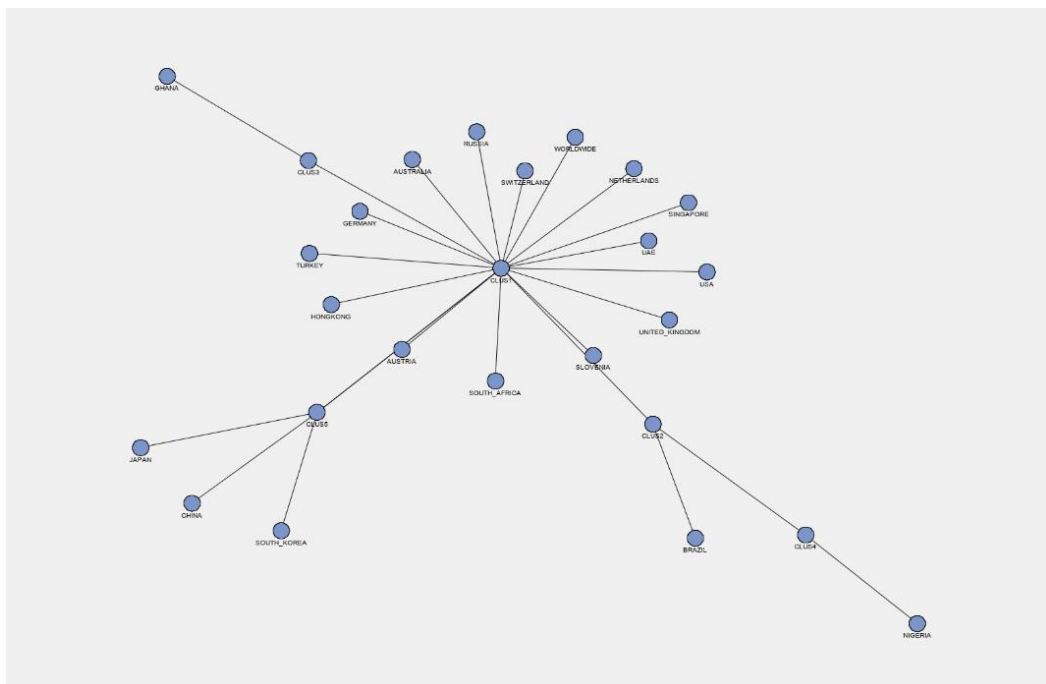
The following dendrogram plot shows cluster organization and clusters which are determined with their variance. Figure 5 shows the dendrogram plot, 5 cluster is enough to keep distances sufficiently.

Figure 5. Selected Countries Dendrogram



After 5 clusters are chosen, countries Google Trends data is clustered according to their correlation and Figure 6 shows the center and nearness of clusters for each country as below.

Figure 6. Visualized Clusters



The first cluster contains 15 countries and worldwide, which represents these countries with 0.88 R<sup>2</sup>, which is in the center of the first cluster. Brazil, Ghana, and Nigeria are represented in 3 different clusters. This is because the Google Trends searches queries of these countries do not any relation with the data from other countries.

For the last cluster, it contains 3 countries which are Japan, China and South Korea. Japan is the representative country for this cluster.

The table shows the representative countries for each cluster and their representative power in those clusters.

Table 8. Clustered Country and Representatives.

Number of Clusters	Representative Country	R <sup>2</sup>	Correlation with Cluster	Correlation Coefficient with Other Clusters
Cluster 1	Worldwide	0.992	Cluster 5	0.887
Cluster 2	Brazil	1	Cluster 1	0.584
Cluster 3	Ghana	1	Cluster 1	0.519
Cluster 4	Nigeria	1	Cluster 2	0.447
Cluster 5	Japan	0.932	Cluster 1	0.868

#### 4.5. Defining Break Point

As can be seen from Figure 2, the correlation between GTS queries for worldwide and Bitcoin price corrupt somewhere. In order to determine the breaking point, the following analysis is made.

In this analysis scaled closed price and scaled worldwide trend data is used. The goal is to find the farthest point between closing prices and the Google search data. For each day in our data, the difference between these 2 variables was calculated as in the following formula.

$$Diff = \text{abs}(\text{Closing\_Price} - \text{Worldwide}) \quad (4)$$

Then the maximum point was selected as the breakpoint. Figure 7 shows the differences between the closing price and worldwide searches.

Figure 7. The Difference between GTS and BP



Since, the maximum difference appears on the 6<sup>th</sup> of January 2018, this point is selected as the breaking point. After deciding on the break point, the whole working data is divided into two parts. The data before 6 January 2018 are determined as 1st part and the data after this date are determined as 2nd part. Subsequent analysis was performed with these parts instead of all data.

#### 4.6. Crating New Variables

In this section, in order to test Granger-causality in the following section, we are creating new variables with gathered data from <https://coinmarketcap.com>. Data includes “daily low” and “daily high” variables, which are the lowest and highest price levels in a day. With the help of these variables, two new variables were calculated.

These are:

- **Daily Volatility:** It shows the percentage of the difference between daily high and low price.

$$\text{Volatility} = \frac{\text{Daily\_High} - \text{Daily\_Low}}{\text{Daily\_Low}} * 100 \quad (5)$$

- **Daily Change in Price:** It shows the daily change in the closing price in terms of the previous day.

$$\Delta \text{ Price} = \frac{\text{Price}_t - \text{Price}_{t-1}}{\text{Price}_{t-1}} * 100 \quad (6)$$

Firstly, worldwide GTS queries and these two new created variables' correlation have been calculated for the whole time period. Spearman correlation method is used for these analyses and the correlation coefficient for volatility and change in price are 0.64 and 0.006 respectively. These correlation ratios show that worldwide searches and daily volatility has a statistically significant relationship because their correlation is around %64. However, for daily change in price ( $\Delta$  Price) correlation coefficient is close to "0" which means there is no significant relationship between worldwide searches and  $\Delta$  Price. For the following test, the daily change in price variable is not used anymore. However, 64% correlation is meaningful to test our hypothesis for the following sections.

Daily volatility, for each part of our study data which divided into 2 parts at the previous section, is subjected to correlation analysis. In addition, correlation coefficients were calculated for 4 representative countries and worldwide because in the previous section, 20 countries clustered together in 5 different groups with the clustering method.

For the first part which is between 01 January 2017 and 06 January 2018 results listed below.

Table 9. First Part Spearman Correlation

<b>Spearman Correlation Coefficients, N = 371</b>					
<b>Variables Name</b>	<b>Worldwide</b>	<b>Brazil</b>	<b>Ghana</b>	<b>Nigeria</b>	<b>Japan</b>
<b>Closing Price</b>	0.93185	0.86243	0.69514	0.2376	0.8253
<b>Daily Volatility</b>	0.58245	0.5361	0.38933	0.27522	0.57184

When the results compared with the whole data correlation coefficient, the daily volatility coefficient has decreased from 64% to 58% for the first part of the data.

Therefore, it is expected that this variable has a stronger relationship in the second part.

Table 9 shows 5 different clusters of Google Trends series and their correlation between daily volatility and closing price. The highest correlated cluster is worldwide and the lowest one is Nigeria and their correlation coefficients are 58% and 27% respectively.

Table 10 shows the correlation results for the second part of the data which includes all the variables between the January 6, 2018, and September 30, 2019 time interval.

Table 10 Second Part Spearman Correlation

<b>Spearman Correlation Coefficients, N = 371</b>					
<b>Variables Name</b>	<b>Worldwide</b>	<b>Brazil</b>	<b>Ghana</b>	<b>Nigeria</b>	<b>Japan</b>
<b>Closing Price</b>	0.77807	0.52432	0.27593	0.33846	0.56464
<b>Daily Volatility</b>	0.67049	0.64942	0.36818	0.53136	0.39846

It is clear that for the second part of the data, daily volatility and worldwide search correlation coefficient jump to 78%. On the other hand, closing price and worldwide search correlation have decreased from 93% to 67% which means a strong correlation between the closing price and worldwide search is rejected in this part.

Moreover, when the correlation coefficients of 5 different clusters are examined for each part, the following results are obtained.

- We have stated that the clusters created by Brazil, Ghana, and Nigeria contain only these countries' data. When these 3 different clusters are compared with the first cluster which contains 15 different countries and represented by the worldwide, very different behaviors are observed in terms of correlation analysis. For example, in all analyzes, it was found that the correlation coefficients of these clusters for all variables were much lower than the first clusters, and sometimes they had a meaningless relationship for the variables.

- As we said before, for the second part the data Google Trend search scaled variables are more correlated with daily volatility, instead of the closing price. However, this can be said only for the first cluster represented by worldwide. For the rest of the clusters which consist of only 5 countries, this situation is working in the opposite direction. That is, for the remaining four clusters, Google Trend search variable is somehow uncorrelated with the closing price and daily volatility or correlated just for the closing price. These differences can be observed not only for the whole data but also in 2 different parts where the data is divided.

#### **4.7. Prediction Analysis**

In this part, we wanted to test whether daily fluctuations in Bitcoin price and Google Trend searches are in the same direction. In order to test this assumption, we created two new binary variables. These are:

Change in Price (CP): Difference between the current price and the previous day price. If the result is negative then the flag is 0, and if vice versa then the flag is 1.

$$CP = \text{Price}_t - \text{Price}_{t-1}$$

Code used in SAS Enterprise Miner: if CP > 0 then 1 else 0

Change in Search (CS): Difference between Google Trend search and previous day Google Trend search. If the result is negative then the flag is 0, and if vice versa then the flag is 1.

$$CS = \text{Search}_t - \text{Search}_{t-1}$$

Code used in SAS Enterprise Miner: if CS > 0 then 1 else 0

These flags assigned to both Bitcoin price and each countries' Google Trends search queries variable for all dates. After that, these flags have been summed up for each country and confusion matrix is created.

Confusion matrix which is also known as error matrix shows precision, accuracy and misclassification rates. This is supervised learning and it is used to measure the performance.

In order to understand the calculation method easily, the sample table for flags and accuracy and misclassification calculation methods are summarized below.

Table 11 Confusion Matrix Explanation

$\Delta$ CP Flag	$\Delta$ Change in Search flag	Explanation
0	0	False Positives (FP)
0	1	False Negatives (FN)
1	0	True Negatives (TN)
1	1	True Positives (TP)

Accuracy:  $(TP + TN) / (TP+TN+FP+FN)$

Misclassification:  $(FP + FN) / (TP+TN+FP+FN)$

Furthermore, in order to interpret the confusion matrix, Matthews correlation coefficient (MCC) is used. MCC is simply a correlation coefficient between observed and predicted binary variable and it returns between -1 and +1. 1 means perfect prediction and -1 means total disagreement. MCC formula is given below.

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (7)$$

Confusion matrix and MCC are calculated for 5 different clusters for whole data.

Table 12 indicates the accuracy and the MCC ratios.

The table shows that the accuracy rates and misclassification rates are similar for 5 cluster countries. It means that these dummy variables which can be evaluated as flags did not provide differentiation between clusters. Furthermore, MCC values for all clusters are too close to 0 which means Bitcoin price change cannot be predicted by daily change in Google Trend search variable. This study shows that the prediction of the Bitcoin price with the help of GTS without any transformation or lagged variable will not give meaningful results.



Table 12 Confusion Matrix and MCC Results

Clusters	$\Delta$ CP flag	$\Delta$ CS flag	Count	1002	Accuracy	MCC	Misclassification Rate
Brazil	0	0	221	22%	48%	0.113	52%
	0	1	222	22%			
	1	0	303	30%			
	1	1	256	26%			
Ghana	0	0	232	23%	51%	0.116	49%
	0	1	211	21%			
	1	0	282	28%			
	1	1	277	28%			
Japan	0	0	213	21%	48%	0.115	52%
	0	1	230	23%			
	1	0	294	29%			
	1	1	265	26%			
Nigeria	0	0	217	22%	47%	0.112	53%
	0	1	226	23%			
	1	0	310	31%			
	1	1	249	25%			
Worldwide	0	0	271	27%	47%	0.106	53%
	0	1	172	17%			
	1	0	355	35%			
	1	1	204	20%			

The table shows that the accuracy rates and misclassification rates are similar for 5 cluster countries. It means that these dummy variables which can be evaluated as flags did not provide differentiation between clusters. Furthermore, MCC values for all clusters are too close to 0 which means Bitcoin price change cannot be predicted by daily change in Google Trend search variable. This study shows that the prediction of the Bitcoin price with the help of GTS without any transformation or lagged variable will not give meaningful results.

#### **4.8. Stationary Tests of Variables**

In the previous sections, it was detected that closing price and daily volatility is correlated with Google trend searches at first & second parts. However, this result does not guarantee that this variable can help forecasting Bitcoin price change by using Google searches. In this section, we tried to explain the causality of the results.

In order to determine the causality relationship and the direction of it, the Granger causality test can be used. As a prerequisite, working data must be stationary to apply the Granger-causality test because it is very sensitive to the choice of lag length and to the methods employed in dealing with any non-stationarity of the time series. The stationarity test is used to determine whether differencing is necessary.

Most of the time series are non-stationary because of seasonality effects. Before applying Autoregressive Moving Average (ARMA) models to time series data must be transformed into stationary. Therefore, firstly stationary test was applied to identify whether data has seasonality or not. If the data is nonstationary, it requires differencing.

The Augmented Dickey-Fuller (ADF) test can be used to test three types of data generation models for stationary. These are;

- Zero-mean stationary (zero mean),
- Nonzero-mean stationary (single mean),
- Linear time trend stationary (trend).

When there is a unit root (that is, the series is nonstationary), the sum of the autoregressive parameters is 1 and hence  $\beta=0$ . The ADF tests build on  $\beta$ . There are three kinds of tests under the ADF tests: rho ( $\rho$ ) test, tau ( $\tau$ ) test, and F test. The rho test is the regression coefficient test, which is also called the normalized bias test. The tau test is the studentized test. The F test is a joint test for unit root. Table 13 presents the null hypotheses and decision rules for the three test statistics under the three different model types.

Table 13 Stationary Test Steps

Types	Rho Test		Tau Test		F- TEST	
	$H_0$	$p$ -value	$H_0$	$p$ -value	$H_0$	$p$ -value
Zero mean	$\delta = 0$	Stationary if low	$\delta = 0$	Stationary if low	N/A	N/A
Single mean	$\delta = 0$	Stationary if low	$\delta = 0$	Stationary if low	$\delta = \alpha_0 = 0$	Stationary if low
Trend	$\delta = 0$	Stationary if low	$\delta = 0$	Stationary if low	$\delta = \Upsilon = 0$	Stationary if low

In this table  $\alpha_0$  is the vector intercept in the test regression for the single-mean model,

$$\nabla y_t = \alpha_0 + \delta y_{t-1} + \theta_1 \nabla y_{t-1} + \dots + \theta_p \nabla y_{t-p} + e_t \quad (8)$$

And  $\Upsilon$  is the parameter of  $t$  in the test regression for the trend model,

$$\nabla y_t = \alpha_0 + \Upsilon t + \delta y_{t-1} + \theta_1 \nabla y_{t-1} + \dots + \theta_p \nabla y_{t-p} + e_t \quad (9)$$

As shown in the regression for the single-mean model, when  $\delta = 0$  (that is there is a unit root), the intercept  $\alpha_0 = \delta \mu = 0$ . The F test with a joint hypothesis of zero intercepts and zero slopes can be used as a unit root test for the single-mean model. The F test for the trend model follows a similar logic. When  $\delta = 0$ ,  $\Upsilon = -\delta \beta = 0$  the F statistic for the trend model, therefore, tests the joint null hypothesis:  $\delta = \Upsilon = 0$ . When testing for unit root, the tau test is more powerful than F-test.

In order to do the stationary tests of the variable SAS Enterprise Miner ARIMA method is used.

Firstly, observation and correlation graphs were created by SAS Enterprise Miner before applying the augmented Dickey-Fuller (ADF) test. This is because we tried to understand whether our sample is stationary or not. While doing ADF tests that is necessary if the time series is non-stationary, it is essential to detect how many differences will be used and which model will be used. (zero mean, single mean, or trend).

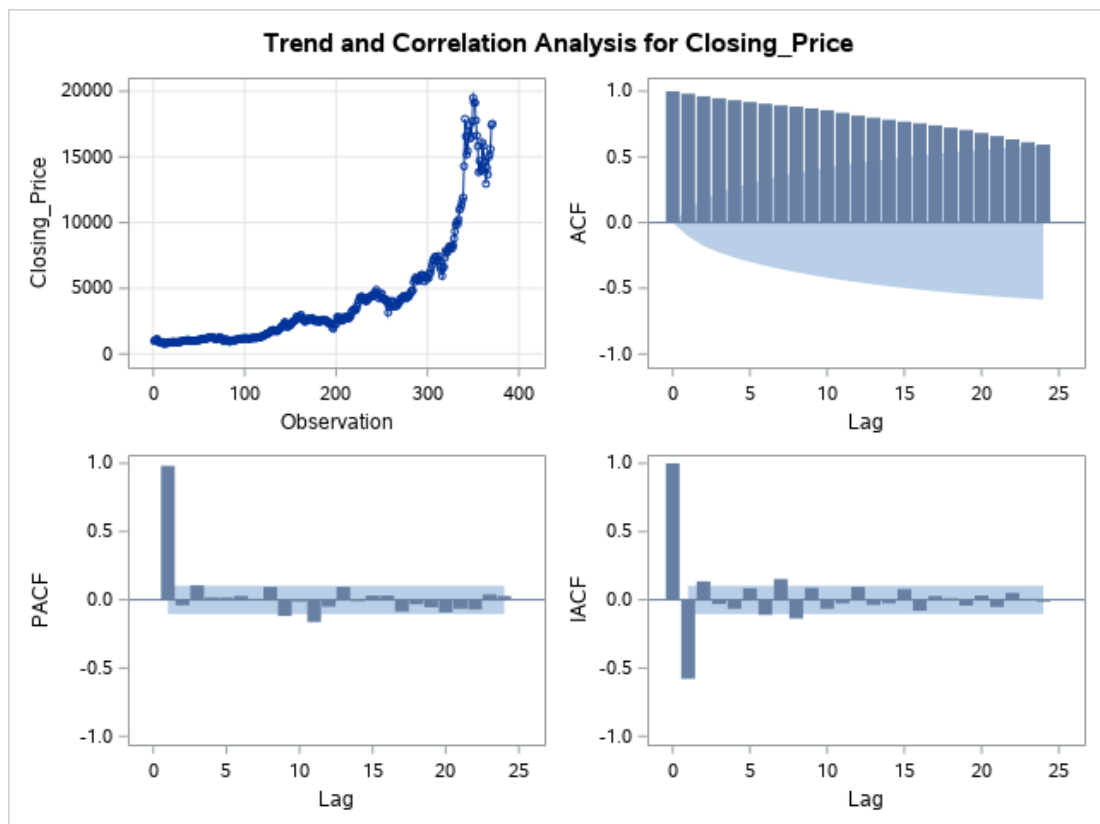
The results of the graphical series correlation analysis taken from SAS Enterprise Miner are shown in Figure 8. It shows trend and correlation results for the first part which includes data for the time between 01 January 2017 and 06 January 2018. Before we started to, some explanations are listed below to give a brief description of the figure divided into 4 windows.

There are four graphs of the closing price. These graphs are:

- The time series plot of the series
- The sample autocorrelation function plot (ACF)
- The sample inverse autocorrelation function plot (IACF)
- The sample partial autocorrelation function plot (PACF)

These autocorrelation graph (ACF) shows the degree of correlation with historic variables that is called lag variables which is 24 by default in SAS Enterprise Guide.

Figure 8. Stationary Test Results Taken for First Part



It is obvious that the closing price of Bitcoin has time - trend mean is not closed to zero. As mentioned before it is a trend data, “Trend” type will be the most important

test for this part. In table 14, the row for “Lags” columns is zero means that AR(1) model was used for this test. Tau test statistics are mostly used to decide whether to reject the null hypothesis. P-values for Tau statistics are greater than 0.01, and the null hypothesis (the series is nonstationary; that is, there is a unit root) cannot be rejected at the 1% level.

For the first part of our data, Dickey-Fuller Unit Root (ADF) is applied with one lag results is shown in Table 14.

Table 14 ADF Test Result for First Part

<b>Dickey-Fuller Unit Root Tests</b>							
<b>Type</b>	<b>Lags</b>	<b>Rho</b>	<b>Pr &lt; Rho</b>	<b>Tau</b>	<b>Pr &lt; Tau</b>	<b>F</b>	<b>Pr &gt; F</b>
Zero Mean	0	3.4263	0.9995	2.54	0.9975		
Single Mean	0	2.8975	0.9996	1.53	0.9994	3.3	0.2252
Trend	0	-1.4503	0.9816	-0.45	0.9855	2.53	0.6712

ADF Test Hypothesis

$H_0$ : Stationary

$H_a$ : Non-Stationary

Based on the p-values of all tests, the null hypothesis is rejected at the 1% level.

It is necessary to convert this data from non-stationary to stationary level. Therefore, difference (first- order integration) was calculated as Price(t) – Price(t-1) for first part of data. All process was done again with new transposed data.

As can be seen in Figure 9 partial autocorrelation functions (PACFs) decreases dramatically after first- order integration and PACF plot and ACF do not increase as much as the lag increases in ACF plot. As a result, for the first part of our data, to make our series stationary, we checked Augmented Dickey-Fuller Unit Root (ADF) again and the result is shown in Table 15. This is because even if the first plot seems that there is no trend and the mean is around zero, so at first, for the Dickey-Fuller test, we need to check the zero mean. Furthermore, the graph does not give us information about single mean. To check stationary of Bitcoin price at first order integration as we said we do the second test.

Figure 9. Stationary Test After First-order Integration

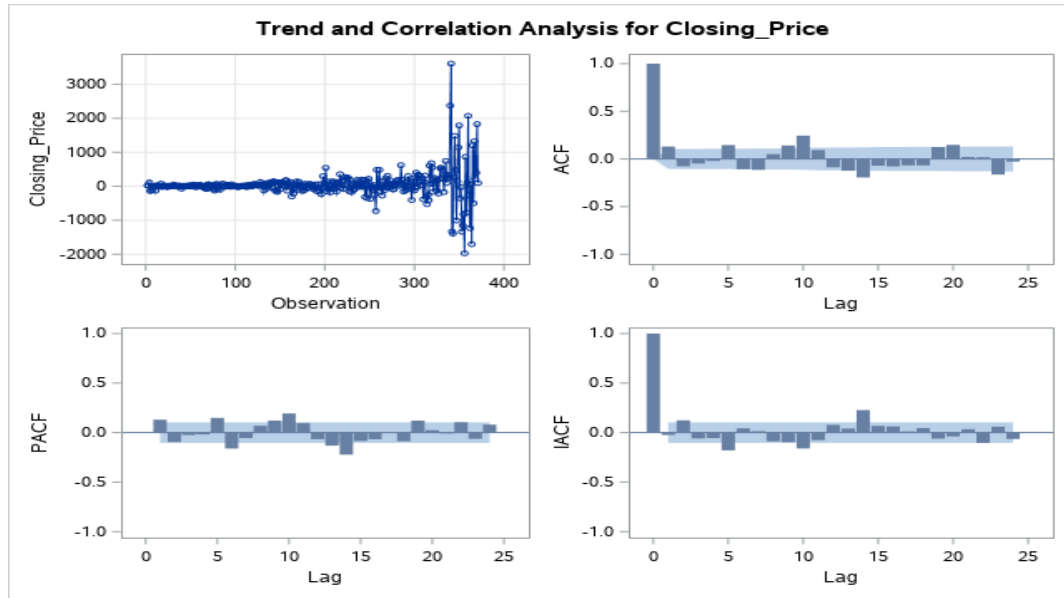


Table 15 ADF Test with I(I) Results for First Part

Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-317.27	0.0001	-16.66	<.0001		
Single Mean	0	-320.905	0.0001	-16.8	<.0001	141.18	0.001

Based on the p-values of both tests, the null hypothesis is rejected at the 1% level, indicating that the differenced time series for the first part of the data is being stationary.

All calculations are repeated for the second part of data which is between 06 January 2018 and 30 September 2019 in Figure 10.

It seems that the second part of the data has no time trend, but the mean of the Bitcoin price is not close to zero. Using first-order (I(1)) integration will be enough for the stationary because partial autocorrelation functions (PACFs) decreases dramatically after the first lag. Moreover, as can be seen in Figure 10, ACF does not increase when the lag increases in ACF plot. Therefore, first order integration is selected for the single mean stationary test. First order one lagged stationary test result listed below.

Figure 10. BP Stationary Test Results for Second Part

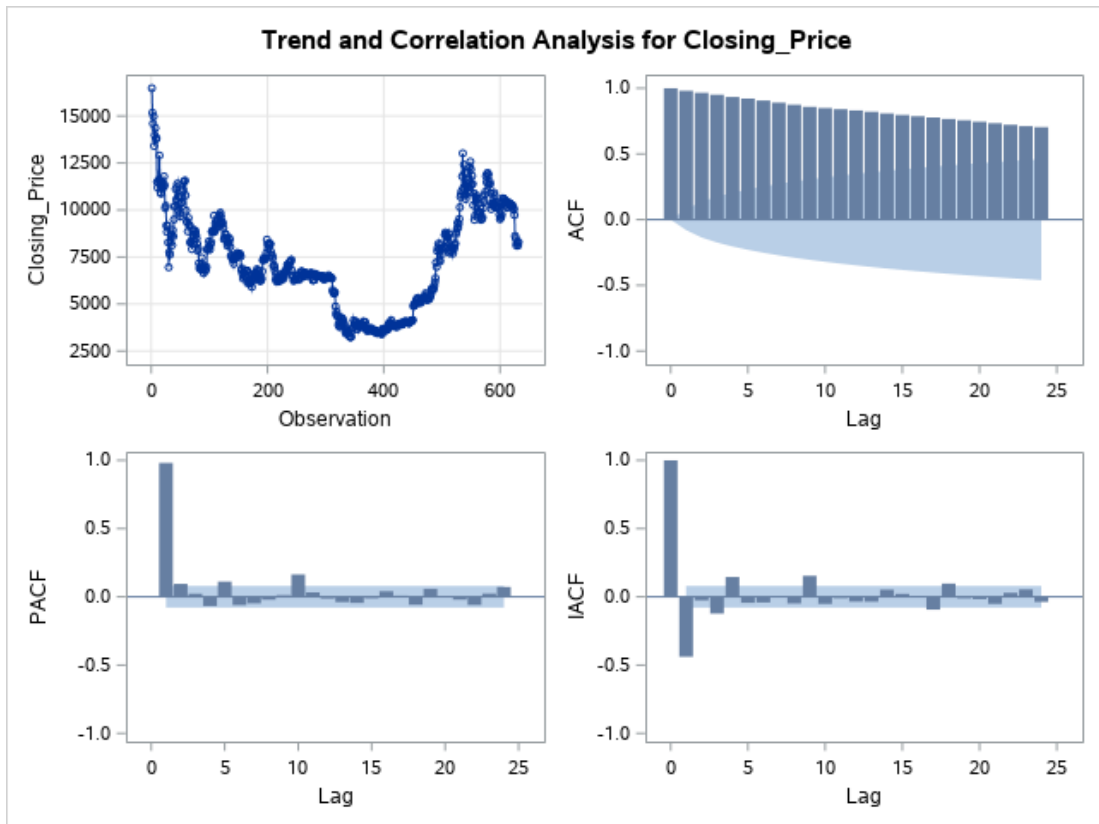


Table 16 ADF Test with I(I) Results for Second Part

Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Single Mean	0	-12.0993	0.0080	-3.55	0.0073	6.74	0.001

In table 16, p-value of the F test statistic is 0.001. This means that the joint test of time trend and nonstationary is rejected at the 1% level.

To sum up, for both parts, first-order integration is used to make time series of the first part in our data being stationary.

To check stationary of Google Trend research time series, we apply the same test with the price. For the first part of the worldwide Google Trend (cluster1) search query result tested for stationary.

Figure 11. Stationary Test Results for GTS First Part

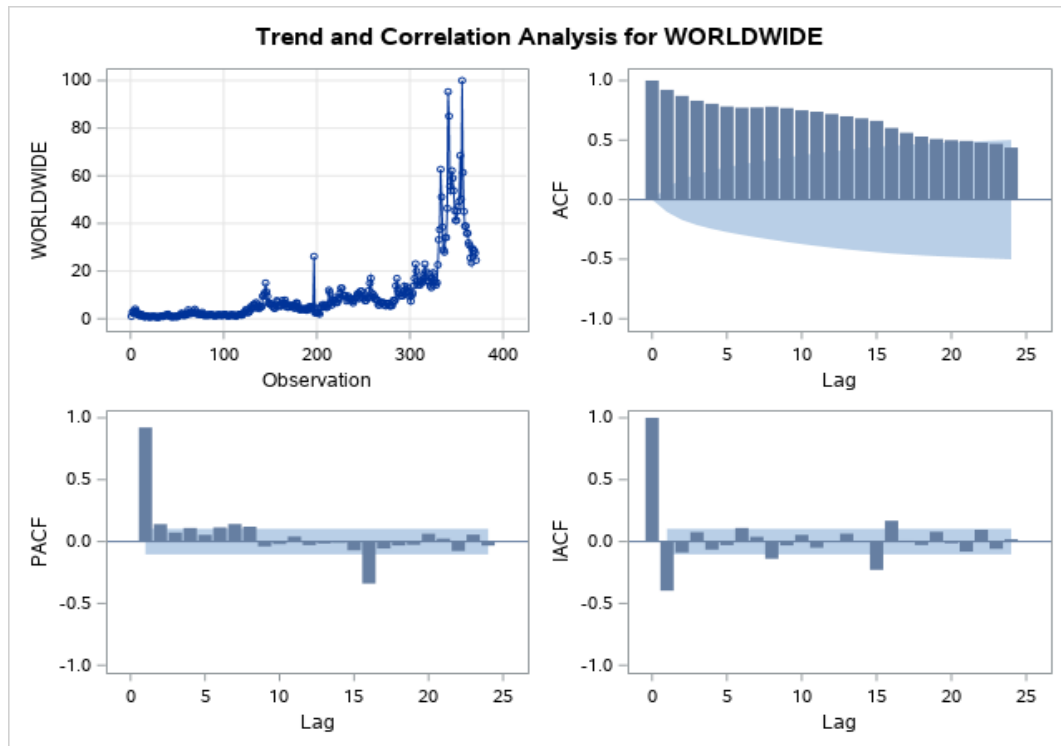


Figure 11 indicates that Google Trends searches have time - trend for the series and the mean is not closed to zero. First-order integration (difference the time series with itself using one lag) will be enough for augmented Dickey-Fuller Unit Root test, because the partial autocorrelation functions (PACFs) decreases dramatically after the first lag, as shown in the PACF plot and ACF does not increase as much as the lag increases in ACF plot. After getting first-order integration of the worldwide series results to show our time series become stationary.

Table 17 GTS, ADF Test with I(I) Results for First Part

Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	435.845	0.0001	-23.03	<.0001		
Single Mean	0	435.904	0.0001	-23.01	<.0001	264.63	0.001

According to table 16, p-value of the F test statistic is 0.001. This means that the joint test of time trend and nonstationary is rejected at the 1% level.



Stationary test studies for the remaining four clusters were repeated and results are summarized in Table 17.

Table 18 Stationary Summary Table for First Part

<b>Cluster</b>	<b>Country_Name</b>	<b>Raw_Data_Stationary</b>	<b>Differential_Equation</b>
Cluster 1	Brazil	No	First order integration I(1)
Cluster 3	Ghana	No	First order integration I(1)
Cluster 4	Nigeria	No	First order integration I(1)
Cluster 5	Japan	No	First order integration I(1)

As can be seen from table 18, row data are non-stationary for all clusters. Google Trend search values have a clear pattern in their ACF plots which makes the time series non-stationary. All clusters first-order integration is used to make our series stationary and prevent spurious relationships in the data. Additionally, after differencing for first-order integration, we again check the Augmented Dickey Fuller test with one lag to observe our data is stationary at first-order integration.

For the second part of the data, we do the same analysis for the Google Trends search queries for all clusters. Worldwide (first cluster) results are listed as an example. Figure 12 shows the result of worldwide Google Trend search queries stationary. It can be observed from figure 12 the mean has not trend pattern. Although ACF plots of the search queries decrease slowly with the lags, the GTS variables result listed below.

Table 19 shows the worldwide GTS time series is stationary with the first lag. Different from the first part, we do not use first-order integration to make our time series stationary. In this part, for the GTS variable, we cannot reject the zero mean and the single mean hypothesis for 1% significance level. As mentioned before, the mean does not have a trend pattern.

Figure 12. Stationary Test Results for GTS Second Part

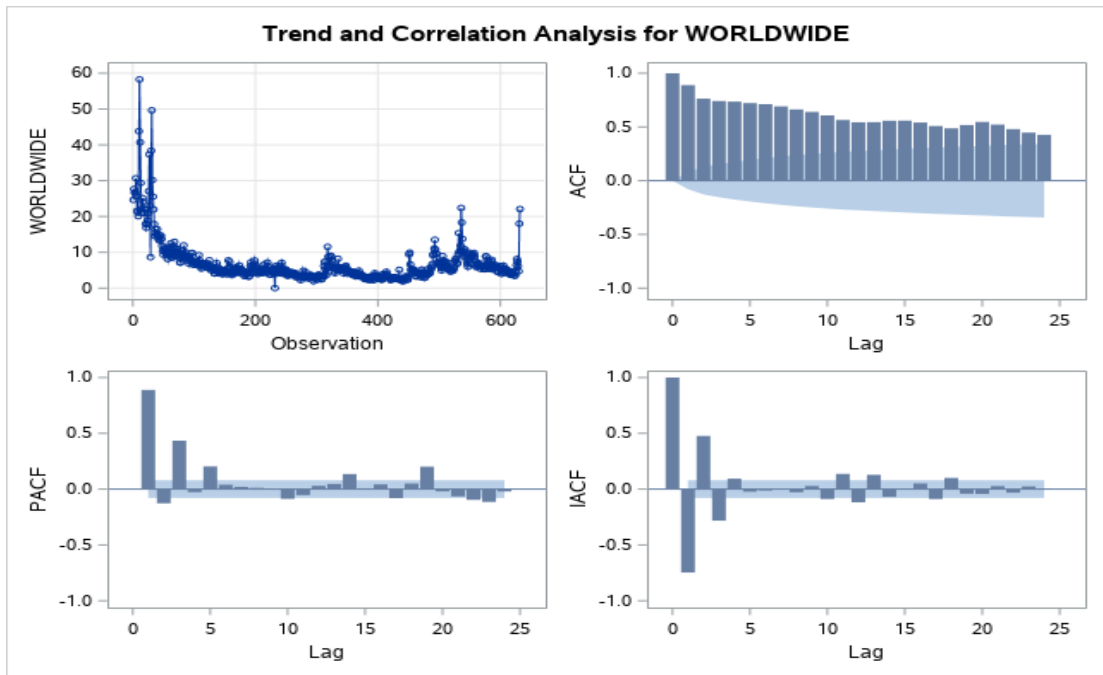


Table 19 GTS, ADF Test with First Lag Results for First Part

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-27.6816	0.0001	-3.81	0.0002		
Single Mean	0	-64.2997	0.0018	-5.87	<.0001	17.22	0.001

Stationary test studies for the remaining four clusters were repeated and results are summarized below.

Different from the first part of the data, in the second part of data except Bitcoin price all variables are stationary with first lag. For all clusters, the mean of the GTS does not have a trend and at 1% significance level all data provides the zero mean and the single mean prerequisites.

Stationary test studies for the remaining four clusters were repeated and results are summarized in Table 20.

Table 20 GTS Stationary Summary Table for Second Part

Cluster	Country_Name	Raw_Data_Stationary	Stationary with lag
Cluster 1	Brazil	Yes	First lag
Cluster 3	Ghana	Yes	First lag
Cluster 4	Nigeria	Yes	First lag
Cluster 5	Japan	Yes	First lag

As we mention before, the first and the second part of the data has a strong correlation with the closing price and daily volatility respectively. Nevertheless, the existing correlation does not imply causation.

At last, we checked the calculated price volatility time-series stationarity. For the dirs. And the second part of the data, series are stationary at first lag.

Table 21 Price Volatility - Stationary Summery

Bitcoin Price Volatility Time Series	Dickey-Fuller Unit Root Tests							
	Zero Mean		Single Mean			Trend		
	Rho	Tau	Rho	Tau	F - Test	Rho	Tau	F Test
First Part	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	N/A
Second Part	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001	N/A

#### 4.9. Granger-causality Tests

After checking the stationary status of the variables for each part and making the necessary adjustments, the Granger-Causality test is performed to verify whether Google Trends search is a good predictor of Bitcoin price. Before discussing Granger-Causality test results, some information about the test is summarized.

Granger (1969) proposed a time series approach to identify causality. In the Granger-causality x is useful for predicting y if it is a cause of y. The term “useful” means if x increases the likelihood of predicting future data of y by looking only at historical data of it.

Two different types of situations where the Granger- causality test can be applied according to Pearl (2009). These are;

- 1) Simple Granger - Causality Test: Causality tested for just two variables with their lags.

- 2) Multivariate Granger - Causality Test: Causality tested for more than two variables because it is assumed that more than one variable has a correlation with others.

Empirical results for Granger-causality test are calculated with Simple Granger - Causality Test because our purpose in this study is to show Google Trends search (GTS) queries “Granger-cause” Bitcoin price (BP) change and vice versa.

The values of Google Trend searches (GTS) are used as an indicator of Bitcoin price (BP). Thus, according to Mahdavi and Sohrabian (1993), we can write the equation for Granger- causality test.

$$(GTS)_t = \alpha + \sum_{i=1}^m \beta_i (GTS)_{t-i} + \sum_{j=1}^n \tau_j (BP)_{t-j} + \mu_t \quad (10)$$

$$(BP)_t = \theta + \sum_{i=1}^p \phi_i (BP)_{t-i} + \sum_{j=1}^n \psi_j (GTS)_{t-j} + \eta_t \quad (11)$$

Depending on the coefficient of equations, four different types of correlation between GTS and BP can be said.

- 1) Unidirectional Granger-causality from GTS to BP. This case explains that Google Trends search predicts Bitcoin price changes.  $GTS \rightarrow BP$
- 2) Unidirectional Granger-causality from BP to GTS. This case is the vice versa of the first case.  $BP \rightarrow GTS$
- 3) Bidirectional (or feedback) causality. Changes in any variable BP or GTS can predict another variable. The relationship is valid for both directions.  $GTS \leftrightarrow BP$
- 4) Independence between BP and GTS. In this case, there is no Granger-causality.

#### **4.10. Granger-causality Tests Findings**

In the previous section, we granted that our time series is stationary. In this part, we perform Granger-causality tests on SAS Enterprise Miner. Results of the Granger-causality test are listed for two different parts. The first part of the data results listed in table 22.

Table 22 Granger-Causality Results for First Part

<b>First Part Granger Causality →(causes) Results</b>							
	Worldwide	Brazil	Ghana	Nigeria	Japan	Bitcoin Price	BP Volatility
Worldwide →	-	N/A	N/A	N/A	N/A	YES***	YES***
Brazil →	N/A	-	N/A	N/A	N/A	YES**	NO
Ghana →	N/A	N/A	-	N/A	N/A	YES***	NO
Nigeria →	N/A	N/A	N/A	-	N/A	YES***	NO
Japan →	N/A	N/A	N/A	N/A	-	YES***	YES*
Bitcoin Price →	NO	NO	NO	NO	NO	-	N/A
BP Volatility →	NO	NO	NO	YES***	NO	N/A	-

\*\*\* %1 significance level, \*\* 5% significance level, \*10% significance level

Granger-causality between Bitcoin price and GTS queries of clustered countries finding listed as;

- Worldwide, Ghana, Nigeria, and Japan GTS queries have unidirectional Granger-causality with Bitcoin price.
- Brazil has also unidirectional Granger-causality with Bitcoin price at 5% significance level.
- Bitcoin price volatility has unidirectional Granger-causality with worldwide GTS queries.
- Bitcoin price volatility has unidirectional Granger-causality with worldwide GTS queries at 10% significance level.
- Bitcoin price volatility is independent of Brazil, Ghana and Nigeria.

The first part results show that worldwide GTS can be used for prediction on the future price of Bitcoin.

For the second part of the data, Granger-causality test results summarized in table 23.

Table 23 Price Volatility - Stationary Summary

Second Part Granger Causality → (causes) Results							
	Worldwide	Brazil	Ghana	Nigeria	Japan	BP_Price	BP Volatility
Worldwide →	-	N/A	N/A	N/A	N/A	YES***	YES***
Brazil →	N/A	-	N/A	N/A	N/A	NO	YES***
Ghana →	N/A	N/A	-	N/A	N/A	YES***	YES***
Nigeria →	N/A	N/A	N/A	-	N/A	NO	YES**
Japan →	N/A	N/A	N/A	N/A	-	YES***	YES***
BP Price →	NO	NO	YES*	NO	YES**	-	
BP Volatility →	YES***	YES***	YES**	YES***	YES***	N/A	-

\*\*\* %1 significance level, \*\* 5% significance level, \*10% significance level

The second part of the data differs considerably from the first part. Results of the clusters, which represent chosen 20 countries, summarized as;

- Worldwide GTS queries have unidirectional Granger-causality with Bitcoin price.
- Ghana GTS series has bidirectional causality with BP which means a change in price or the GTS can predict each other. This can be said at 10% significance level. (p-value = 0.051)
- Japan is also a bidirectional causality relationship with BP at 5% significance level. (p-value=0.027)
- Nigeria GTS query and Bitcoin price is independent of each other.
- Bitcoin price volatility has bidirectional Granger-causality with all clusters. (Note that Nigeria → Bitcoin price volatility at 5% significance level)

The first part of the data, which is closer for the time window, results show instead of using Bitcoin closing price, the daily volatility of the Bitcoin price is more accurate for prediction.

## CHAPTER 5

### CONCLUSION

This paper investigates the correlation between Google Trends search queries and Bitcoin returns with a country-level over period 01 January 2017 to 30 September 2019.

We first examined the correlation between these two data. As a result of our observations, we observed that the correlation between the two variables decreased significantly after a certain period. Therefore, data is divided into 2 parts according to the breaking period. Ongoing studies have been applied separately for each part. Different from other researches 20 country GTS queries gathered to test if there are any behavioral differences between countries. Then, selected twenty countries were consolidated into 5 groups with a variable clustering method depending on their correlation with Google Trends search data.

Before starting the Granger-causality test, both BP and GTS queries checked for stationary for each part. For the first part, in general, first-order integration is used to make studied data being stationary. To be sure, all test repeated for each GTS and BP variables.

In order to examining the relationship between GTS and Bitcoin closing price The Granger-causality tests are performed. Furthermore, study extended Bitcoin volatility which is created in our study. Therefore, the one contribution of this paper is that the causality tests runs in not only for daily closing price but also daily volatility level with both worldwide GTS queries and clustered countries GTS queries.

The Granger-causality rest result show that for the first part of the data all countries have a unidirectional causality with Bitcoin closing price at 5% significance level. Furthermore, for the second part of data, unidirectional causality for Brazil and Japan is no longer acceptable.

In the study conducted with Bitcoin price volatility, for the first part, some countries are independent, and some has unidirectional Granger-causality. However, for the second part, all countries GTS and Bitcoin price volatility has bidirectional Granger

causality. Our results do not exactly match with Dastgir et al. (2018) findings. In that study, it is found Bitcoin price and GTS had bidirectional causality. This is caused by different sample time period used in their research.

To sum up, Google Trends search queries can be used to predict Bitcoin price. Main observation in this study is that, even if some countries like Ghana and Nigeria seems less correlated with rest of the world, their result also meaningful to prediction. Moreover, worldwide GTS queries has a representative power for the rest of the world.



## REFERENCES

- Bariviera, A. F. (2017). The inefficiency of Bitcoin revisited: A dynamic approach. *Economics Letters*, 161, 1-4.
- Bianchi, M. (1995). Granger causality in the presence of structural changes (No. 1995018). Université catholique de Louvain, Institut de Recherches Economiques et Sociales (IRES).
- Blau, B. M. (2017). Price dynamics and speculative trading in Bitcoin. *Research in International Business and Finance*, 41, 493-499.
- Bouri, E., Gupta, R., Tiwari, A. K., & Roubaud, D. (2017). Does Bitcoin hedge global uncertainty? Evidence from wavelet-based quantile-in-quantile regressions. *Finance Research Letters*, 23, 87-95.
- Bouri, E., Jalkh, N., Molnár, P., & Roubaud, D. (2017). Bitcoin for energy commodities before and after the December 2013 crash: diversifier, hedge or safe haven?. *Applied Economics*, 49(50), 5063-5073.
- Bouri, E., Molnár, P., Azzi, G., Roubaud, D., & Hagfors, L. I. (2017). On the hedge and safe haven properties of Bitcoin: Is it really more than a diversifier?. *Finance Research Letters*, 20, 192-198.
- Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record*, 88, 2-9.
- Chok, N. S. (2010). Pearson's versus Spearman's and Kendall's correlation coefficients for continuous data (Doctoral dissertation, University of Pittsburgh).
- Dastgir, S., Demir, E., Downing, G., Gozgor, G., & Lau, C. K. M. (2019). The causal relationship between Bitcoin attention and Bitcoin returns: Evidence from the Copula-based Granger causality test. *Finance Research Letters*, 28, 160-164.
- Dickerson, A. (2018). Algorithmic Trading of Bitcoin Using Wikipedia and Google Search Volume. Available at SSRN 3177738.

- Dyhrberg, A. H. (2016). Bitcoin, gold and the dollar—A GARCH volatility analysis. *Finance Research Letters*, 16, 85-92.
- Dyhrberg, A. H. (2016). Hedging capabilities of Bitcoin. Is it the virtual gold?. *Finance Research Letters*, 16, 139-144.
- Estrada, J. C. S. (2017). Analyzing Bitcoin price volatility. University of California, Berkeley.
- Fama, E. F. (1960). Efficient market hypothesis (Doctoral dissertation, PhD Thesis, Ph. D. dissertation, University of Chicago Graduate School of Business).
- Gkillas, K., & Katsiampa, P. (2018). An application of extreme value theory to cryptocurrencies. *Economics Letters*, 164, 109-111.
- Granger, C. W. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: Journal of the Econometric Society*, 424-438.
- Hosseini-fard, S. Z., Abbasi, B., Ahmad, S., & Abdollahian, M. (2009). A transformation technique to estimate the process capability index for non-normal processes. *The International Journal of Advanced Manufacturing Technology*, 40(5-6), 512.
- Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Trans. Amer. Soc. Civil Eng.*, 116, 770-799.
- J. Battelle. *The search: how Google and its rivals rewrote the rules of business and transformed our culture*. Nicholas Brearley Publishing, 2005
- Katsiampa, P. (2017). Volatility estimation for Bitcoin: A comparison of GARCH models. *Economics Letters*, 158, 3–6. doi: 10.1016/j.econlet.2017.06.023
- Kristoufek, L. (2013). Bitcoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific reports*, 3, 3415.

- Kristoufek, L. (2013). Can Google Trends search queries contribute to risk diversification? *Scientific reports*, 3, 2713.
- Kristoufek, L. (2018). On Bitcoin markets (in) efficiency and its evolution. *Physica A: Statistical Mechanics and Its Applications*, 503, 257–262. doi: 10.1016/j.physa.2018.02.161
- Kristoufek, L. (2018). On Bitcoin markets (in) efficiency and its evolution. *Physica A: Statistical Mechanics and its Applications*, 503, 257-262.
- Krzywinski, M., & Altman, N. (2014). Points of significance: visualizing samples with box plots.
- Lielacher, A., Lielacher, A. L. A., & Cass Business School. (2019, April 19). How Many People Use Bitcoin in 2019? Retrieved November 10, 2019, from <https://www.bitcoinmarketjournal.com/how-many-people-use-bitcoin/>.
- Mahdavi, S., & Sohrabian, A. (1993). The exchange value of the dollar and the US trade balance: An empirical investigation based on cointegration and Granger causality tests. *The Quarterly Review of Economics and Finance*, 33(4), 343-358.
- Mandelbrot, B., Wallis, J.R., 1969. Robustness of the rescaled range R/S in the measurement of noncyclic long-run statistical dependence. *Water Resources Research* 5, 967–988.
- Myers, L., & Sirois, M. J. (2004). Spearman correlation coefficients, differences between. *Encyclopedia of statistical sciences*, 12.
- Myers, L., & Sirois, M. J. (2004). Spearman correlation coefficients, differences between. *Encyclopedia of statistical sciences*, 12.
- Nadarajah, S., & Chu, J. (2017). On the inefficiency of Bitcoin. *Economics Letters*, 150, 6-9.
- Nakamoto, S. (2019). Bitcoin: A peer-to-peer electronic cash system. Manubot.

- Pearl, J. (2009). *Causality*. Cambridge university press.
- Preis, T., Reith, D., & Stanley, H. E. (2010). Complex dynamics of our economic life on different scales: insights from search engine query data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1933), 5707-5719.
- Ruwitch, J. and Sweeney, P. (2013). China bars banks from Bitcoin transactions. Reuters, <http://www.reuters.com/article/2013/12/05/us-china-Bitcoin-idUSBRE9B407L20131205> (accessed: 01-10-2019).
- Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, 126(5), 1763-1768.
- Thomas, M., & Morwitz, V. (2009). Heuristics in numerical cognition: Implications for pricing. *Handbook of pricing research in marketing*, 132-149.
- Tiwari, A. K., Jana, R., Das, D., & Roubaud, D. (2018). Informational efficiency of Bitcoin—An extension. *Economics Letters*, 163, 106–109. doi: 10.1016/j.econlet.2017.12.006
- Urquhart, A. (2016). The inefficiency of Bitcoin. *Economics Letters*, 148, 80-82.
- Yermack, D. (2015). Is Bitcoin a Real Currency? An Economic Appraisal. *Handbook of Digital Currency*, 31–43. doi: 10.1016/b978-0-12-802117-0.00002-3
- Zhang, W., Wang, P., Li, X., & Shen, D. (2018). Quantifying the cross-correlations between online searches and Bitcoin market. *Physica A: Statistical Mechanics and its Applications*, 509, 657-672.