# Bitcoin and Retail Investor Attention

Bachelor Thesis

at the Chair of International Finance, Prof. Dr. Stefan Rünzi

University of Mannheim

Author:

Malte Brunst

Ginsterweg 2

71254 Ditzingen

Student ID: 1517488

Advisor: Fabian Brunner | Auditor: Prof. Dr. Stefan Rünzi

# Table of contents

## List of tables



## List of figures

## List of symbols

| | |
|---|---|
| $r$ | Pearson correlation coefficient |
| $x_t, y_t, z_t$ | vectors of time series data |
| $t$ | point in time: day or week |
| $n$ | sample size of a given time series |
| $\Sigma$ | sum over a set of following terms |
| $\bar{x}, \bar{y}$ | mean of time series $x, y$ |
| $s_x, s_y$ | standard deviation of time series $x, y$ |
| $\alpha$ | parameter |
| $\beta$ | regression coefficient |
| $x_t, y_t \sim \mathrm{I}(1)$ | vectors $x_t, y_t$ are integrated of order one |
| $v$ | vector of intercept terms |
| $p$ | order/lag length of a given VAR model |
| $A_t$ | coefficient matrix at time $t$ |
| $u_t$ | vector of error terms |
| $\rightarrow$ | denotes a cause and effect relationship |

## List of abbreviations

| | |
|---|---|
| ADF test | augmented Dickey-Fuller test |
| BTC | ticker symbol of the currency Bitcoin |
| e.g. | exempli gratia / for example |
| i.e. | id est / that is |
| KPSS test | Kwiatkowski-Phillips-Schmidt-Shin test |
| USD | United States dollar |
| VAR model | vector autoregressive model |
| variables of empirical work | see Table 2 |

# 1 Introduction

In 2016 I bought my first bitcoin. Since my family and friends had very little or no knowledge about this internet currency, I was soon considered an expert in the field in my direct surroundings. As a result of this imposed role as the go-to-guy for all questions concerning Bitcoin, I could experience vividly how both the price of Bitcoin and the interest in it increased significantly during 2017. Questions and discussions became a lot more frequent, friends wanted me to invest money for them or bought bitcoins themselves after asking me how and where to do so. All the while the Bitcoin price continued to surge seemingly interminably. Eventually, in the beginning of 2018, a rather harsh end to the price rally was set accompanied by what seems like a cooling off in interest of my peers.

These personal observations led to the research questions of this bachelor thesis: Are the developments in the Bitcoin market and the attention of retail investors, such as my friends and myself, interrelated?

To answer these questions, the research of this work examines the relationship between retail investor attention based on Google searches and Wikipedia pageviews and the price and trading volume of Bitcoin. Firstly, a descriptive representation of the data and the contemporaneous relationships of the investigated variables are exhibited. Furthermore, existing dynamic relationships are uncovered.

The exploration of the relationship between attention and the Bitcoin market is of particular interest since the internet currency has no intrinsic value. Buying and selling decisions can therefore not be based on fundamental indicators which determine a fair value. Hence, public attention could be a decisive factor impacting the Bitcoin market.

The application of Google searches and Wikipedia pageviews as well as the methodology used in my paper are derived from Kristoufek’s paper on the drivers and behavior of the Bitcoin price (see Kristoufek, 2013). I emulate part of his research and add to his work by incorporating the Bitcoin trading volume to the investigated variables and by applying data of a longer and more recent time period.

# 2 Data

## 2.1 Data on retail investor attention

### 2.1.1 Google search volume

The first data set related to retail investor attention consists of the weekly volume of search queries for the search term “bitcoin” on Google. The time range of the used data covers the recent five years, from 04/14/2013 to 04/14/2018.

I decided to analyze data from the search engine www.google.com for two reasons: First and foremost because it is by far the most widely used web search engine. According to statista.com Google’s global market share for web searches amounted to 87,1% in December 2017 (Statista, 2018); statcounter.com states a share of 91,25% for March 2018 (StatCounter, 2018). This dominant position indicates that Google carries out the majority of web searches worldwide and their search data should therefore very accurately reflect the attention of people. Secondly, Google’s search volume data is free of charge and easily accessible when using their service Google Trends (trends.google.com). The search term which most precisely reflects an interest in Bitcoin and is most suitable for assessing attention is determined with the help of the Google Trends tool “correlate” (google.com/trends/correlate), following the example of, among others, Dimpfl and Jank (2016).

| Search term | Correlation | Search volume |
|---|---|---|
| **bitcoin** | **1** | **100%** |
| bitcoin price | 0,9573 | 27% |
| btc | 0,9437 | 16% |
| buy bitcoin | 0,9454 | 4% |
| bitcoin rate | 0,9784 | 1% |
| how bitcoin works | 0,9746 | <1% |

Table 1 – Comparison of highly correlated search terms with Google correlate.
(Correlations as of 04/20/2018; Search volume refers to the chosen five year time period)

The search terms shown in Table 1 are very highly correlated which indicates that all of them contain the same information. As shown in the table the correlated search term with the second highest search volume accomplishes only 27% of the volume of the search term “bitcoin” for the investigated five-year time period. Furthermore, Google Trends reveals that all search terms except for the term “bitcoin” exhibit at least one week with a search volume of less than 1%. Hence, to achieve the best possible results from the data analysis, the search term with the highest search volume was selected: “bitcoin”. It should be noted that a Google search is not case sensitive meaning that searches for variations of the word “bitcoin”, such as “Bitcoin”, are included in the data as well. It is reasonable to assume that retail investors predominantly prefer to use a domestic exchange with their domestic currency for trading. Since I use Bitcoin data from an US-American exchange (see: *2.1.2 Data on Bitcoin*), I limit the retrieved data to Google searches which were run in the USA. The underlying assumption is that investors who are trading at the chosen exchange are predominantly US-Americans and for valid results the measured attention should be that of Americans as well. Yet, since I cannot base this assumption on empirical findings, I apply the used methodology to the worldwide Google search volume as well. The US search volume serves as the main variable while the worldwide volume acts as a control variable. Results of the control variable are only addressed if they differ from the main variable in a statistically significant way.

Google Trends provides the search volume for a given search term on a relative scale. The scale is normalized by the date type (day, week, month, depending on the chosen time range) with the maximum search volume in the consulted time range which receives a search volume of 100%. The chosen time horizon of five years (04/14/2013 – 04/14/2018) is the maximum time range for which weekly data is provided. For longer time periods only monthly data is retrievable. Since monthly data implies a less precise analysis, I use the maximum possible time range which yields weekly data. The final dataset consists of 261 data points corresponding to 261 weeks.

### 2.1.2 Wikipedia pageviews

The second data set representing retail investor attention consists of the daily pageviews of the anglophone Wikipedia article “Bitcoin” (en.wikipedia.org/wiki/Bitcoin) for the time range from 07/01/2015 to

04/14/2018. The data is obtained from the Wikipedia/Wikimedia pageview tool tools.wmflabs.org/pageviews. The investigated time range is selected based on the fact that the earliest possible data point provided refers to the 07/01/2015. Bitcoin exchanges are operating on weekends and holidays as well which means Bitcoin data on a seven-day-per-week basis is used in my research. Thus, I also apply pageview data of all seven days of the week. The final data set consists of 1019 data points corresponding to 1019 days.

### 2.2 Data on Bitcoin

Following Kristoufek (2013), the trading data on Bitcoin is obtained from the website bitcoincharts.com for a time horizon matching the time range of the Google search volume data and Wikipedia pageviews data respectively. The data relates to the US cryptocurrency-exchange bitstamp.net. This exchange is selected because of its status and importance: Bitstamp is currently one of the most liquid exchanges for Bitcoin in the world. On April 15, 2018 it was the exchange with the largest 24-hours Bitcoin trading volume in US-Dollar (USD) according to bitcoincharts.com’s market ranking. I use the daily closing prices in USD and the daily trading volume in number of bitcoins traded and in USD from bitstamp.net for my empirical research. The analyses which are conducted together with the weekly Google search volume data also require weekly data on Bitcoin. Therefore, I calculate the weekly average closing price and weekly average trading volume using the downloaded daily data.

### 2.3 Variable overview

All variables used in the following empirical work are summarized in Table 2 below.
For the bivariate time series analyses of this research, the variables form the following logical variable pairs: Google SV/Bitcoin WAP, Google SV/Bitcoin WAV, Wikipedia PV/Bitcoin DP, Wikipedia PV/Bitcoin DV.

| Variable Name | Explanation |
|---|---|
| **Google SV** | Google Search Volume: Relative Google search volume for "bitcoin" on a weekly basis. The main variable refers to the search volume in the US. Worldwide search volume is used as a control variable. |
| **Wikipedia PV** | Wikipedia Pageviews: Absolut pageviews of the anglophone Wikipedia article "Bitcoin" on a daily basis. |
| **Bitcoin WAP** | Bitcoin Weekly Average Price: Average Bitcoin closing price in USD on a weekly basis. |
| **Bitcoin DP** | Bitcoin Daily Price: Closing price of Bitcoin in USD on a daily basis. |
| **Bitcoin WAV** | Bitcoin Weekly Average Volume: Average Bitcoin trading volume in USD or as the number of bitcoins traded (BTC) on a weekly basis. |
| **Bitcoin DV** | Bitcoin Daily Volume: Trading volume for Bitcoin in USD or as the number of bitcoins traded (BTC) on a daily basis. |

Table 2 – Overview of investigated variables

# 3 Methodology

## 3.1 Summary

In order to achieve a general overview of the historical development and behavior of the examined variables, I plot the data on retail investor attention (Google SV and Wikipedia PV) jointly with the corresponding Bitcoin price data (Bitcoin WAP and Bitcoin DP) and Bitcoin trading volume data (Bitcoin WAV and Bitcoin DV). Furthermore, the summary statistics of all variables are considered. After that, the interdependence between the different variables is investigated. I do that by calculating the Pearson correlation coefficients and by inspecting for a cointegration between the time series of all associated variable pairs. Finally, I estimate vector autoregressive (VAR) models of the variable pairs and derive the impulse response functions from them. The VAR models allow to determine (Granger-) causality between retail investor attention and the Bitcoin price or trading volume, while the impulse response functions permit to observe the magnitude and manifestation of these causal connections.

## 3.2 Detailed realization

I use Lütkepohl's work "New Introduction to Multiple Time Series Analysis" (see Lütkepohl, 2005) throughout my empirical work for guidance and to ensure methodical correctness. Additional sources are used as well, wherever indicated in the text.

Since I am working with times series data, I investigate the stationarity of the data for all variables before any statistical methods are applied. Removing potential trends and/or seasonality from my data ensures constant statistical properties and guarantees that the effect of spurious correlation and, in the consecutive VAR models, the effect of spurious regression can be avoided (see, for example, Yule, 1926 and Phillips, 1986). I test the stationarity of the original data sets, as well as that of the log-transformation, the first differences and the first logarithmic differences. Following the example of Kristoufek (2013), I apply two unit root tests to all transformations to examine stationarity: An augmented Dickey-Fuller test (ADF test) and a Kwiatkowski-Phillips-Schmidt-Shin test (KPSS test). The two tests have opposing null hypotheses: The ADF test verifies a unit root in the null hypothesis, the KPSS test conversely checks for stationarity in the null hypothesis. This renders them a particularly suitable pair to test stationarity. The results of the tests are summarized in Appendix A. For all six variables the first differences and first logarithmic differences are stationary at a 5% significance level according to both tests. Next to the stationarity, I investigate the distribution of the data to help me decide which transformation to use in the following statistical methods. You can find the mean, kurtosis and skewness of the original time series and all transformations in Appendix B. Ultimately, I select the first logarithmic differences of all variables for application in the correlation analyses and the VAR models. Besides stationary, the logarithmic differences are also substantially closer to a normal distribution when compared to the first differences. On top of that, employing first logarithmic differences allows for an easier interpretation of the estimated VAR models since the regression coefficients $\beta$ are elasticities: A one percent increase in the independent variable $x$ leads to a (all other variables held constant) $\beta\%$ change in the dependent variable $y$. For the impulse response functions this implies that the shocks and responses of the variables can be interpreted as relative changes as well.

In a first step, I use the Pearson correlation coefficient to determine the correlation between two time series, $x$ and $y$:

$$(1) \qquad r_{xy} = \frac{\Sigma x_t y_t - n\bar{x}\bar{y}}{(n-1)\, s_x s_y}$$

The coefficient $r$ is calculated for the original time series of the variable pairs which are non-stationary (Wikipedia PV is the only exception to that, see Appendix A) and for the logarithmic first differences which are stationary time series. This way the effect of trend or seasonality on the relationship between the variables can be eliminated and I can determine if and to what degree a contemporaneous correlation is left after adjusting for it.

Additionally, I test for a connection between the corresponding variables with the augmented Engle-Granger two-step cointegration test. A cointegration test allows to investigate if a long run link between non-stationary variables exists (see, for example, Dickey et al. 1994). This means that time series which contain a unit root can be employed here. I therefore use the initial non-transformed and non-stationary data of all variables for these calculations. The approach of cointegration circumvents spurious regression by looking at the linear combination $z_t$ of the vector $y_t$ and $\alpha x_t$, ($y_t$ and $x_t$ are the time series of given variables, see list of symbols) whereby both $y_t$ and $x_t$ are integrated of order one, $\mathrm{I}(1)$ (see Lütkepohl, 2005, p.245):

$$(2) \qquad z_t = y_t - \alpha x_t \qquad\qquad x_t, y_t \sim \mathrm{I}(1)$$

Integrated of order one means that the vector is stationary after differencing one time. Cointegration is also possible with lower orders of integration. In the empirical study of this work however, all variables are first-order integrated. $y_t$ and $x_t$ are said to be cointegrated if a $z_t$ exists which is stationary, meaning of order $\mathrm{I}(0)$. To describe it differently: While $y_t$ and $x_t$ are non-stationary processes with changing statistical properties, a linear combination of the two vectors is stationary and stable over time. We can describe cointegration therefore as a "*long-run equilibrium relation*" (Lütkepohl, 2005, p.244) of the variables. Technically, the cointegration test estimates a parameter $\alpha$ for which it holds true that vector $z_t$ is $\mathrm{I}(0)$, by using a linear regression model (e.g. ordinary least squares method) and a unit root test (e.g. the ADF test) (see, for example, Lütkepohl, 2005 and Engle and Granger, 1987). In the case of the applied Engle-Granger cointegration test the null hypothesis declares that the time series of a inspected variable pair are not cointegrated. I reject the null hypothesis at the 5% significance level.

Eventually, to analyze the dynamic structure of my system of variables, I apply vector autoregressive (VAR) models. Vector autoregressive models became popular after noble laureate Christopher A. Sims' paper "Macroeconomics and Reality" (see Sims, 1980). In this paper, the author successfully suggested VAR models as the new standard macroeconomic framework for estimating relationships between variables. Nowadays, these models are commonly used for the analysis of multivariate time series in an economic or financial context. A VAR model is a regression model which is based on the assumption that every investigated variable $y_t$ is dependent on its own lagged values of $p$ –prior periods, as well as on the lagged values of all other exogenous variables. In this sense, all of the variables are exogenous and endogenous at the same time.

Mathematically a VAR model of order $p$ can be written as (notation based on Lütkepohl, 2005):

$$y_t = v + A_1 y_{t-1} + \ldots + A_p y_{t-p} + u_t \tag{3}$$

Where, in the case of this research, $y_t$ is a $(2 \times 1)$ - vector, consisting of the two simultaneously investigated variable pairs at time $t$. $A$ is the corresponding $(2 \times 2)$ - coefficient matrix; $v$ is the $(2 \times 1)$ - vector of intercept terms and $u_t$ is a $(2 \times 1)$ - vector of zero-mean error terms. Starting at $t = 0$, this leads to the following bivariate-model of order $p$:

$$y_1 = v_1 + a_{11} y_{1,-1} + a_{12} y_{2,-1} + \cdots + A_p y_{-p} + u_1 \tag{3.1a}$$

$$y_2 = v_2 + a_{21} y_{2,-1} + a_{22} y_{2,-1} + \cdots + A_p y_{-p} + u_2 \tag{3.1b}$$

I select the lag order $p$ of the models by consulting the Akaike information criterion and the Bayesian information criterion. The minimum lag proposed by the minimum value of the two information criteria is selected. The VAR model is estimated using the ordinary least squares method. The resulting manifestations of the estimated parameters are not of direct interest to me since I am not using the model to forecast but rather for a structural analysis. More specifically, the dynamic interrelations of the variables of the VAR model are investigated using an impulse response analysis. This analysis allows to determine the response of one variable of the VAR model to an impulse or shock from another variable. I use orthogonalized impulse response functions for the analysis since a contemporaneous correlation of the residuals in every one of the estimated VAR models is present. This implies that an impulse in variable $y_1$ has a contemporaneous effect on variable $y_2$ as well as the other way around. In order to obtain the isolated, independent responses of one variable following a shock in the other variable the residuals are orthogonalized, such that the error-term covariance matrix is decomposed into a lower triangular matrix, using Cholesky decomposition. A consequence of the lower triangularity of the matrix is that the order of the variables in the VAR model has an impact on the responses. I follow the suggestion by Lütkepohl (2005) to sort the variables in the VAR model in such a way *“[…] that the first variable is the only one with a potential immediate impact on all other variables”* (Lütkepohl, 2005, p.61). This is only logical since the second variable cannot have an immediate impact in the first equation of the VAR model due to the lower triangularity of the covariance matrix. For the main results in chapter four, I estimate every VAR model so that the variable relating to Bitcoin, the Bitcoin price or Bitcoin trading volume, is the first variable. This decision has the following reason: A simple web search on Google or pageview on Wikipedia does not require a lot of time and effort and can be done instantaneously. Therefore, it is likely that a shock in the Bitcoin price or trading volume has (if any) an immediate impact on the measures of investor attention. Buying and selling of bitcoins, which translates into impact on the Bitcoin price and trading volume, is more time consuming, especially when it is done for the first time, and often postponed. I therefore assume that an impulse in investor attention does not as likely translate into an immediate impact on these variables. Nonetheless, I conduct the impulse response analysis for each variable pair twice, so that each variable is $y_1$ and $y_2$ at one point. This allows the reader to form his own opinion based on different assumptions. The impulse response functions with the reverse order of the variables, in which an immediate impact of retail investor attention on the Bitcoin price/trading volume is possible, can be found in Appendix F.

The impulse response analysis is also a test for causality because, *"[…] if there is a reaction of one variable to an impulse in another variable we may call the latter causal for the former"* (Lütkepohl, 2005, p.51). This reasoning is derived from the concept of Granger causality, which is based on the straightforward principle that *"[…] a cause cannot come after the effect"* (Lütkepohl, 2005, p.41). I verify the statistical significance of the causality between given variables with a Granger causality test based on a f-distribution of the test statistics which is applied to the estimated VAR models. The null hypothesis of the test states that there is no Granger causality between the two variables of a given VAR model.

# 4 Results

## 4.1 The development of Bitcoin and retail investor attention

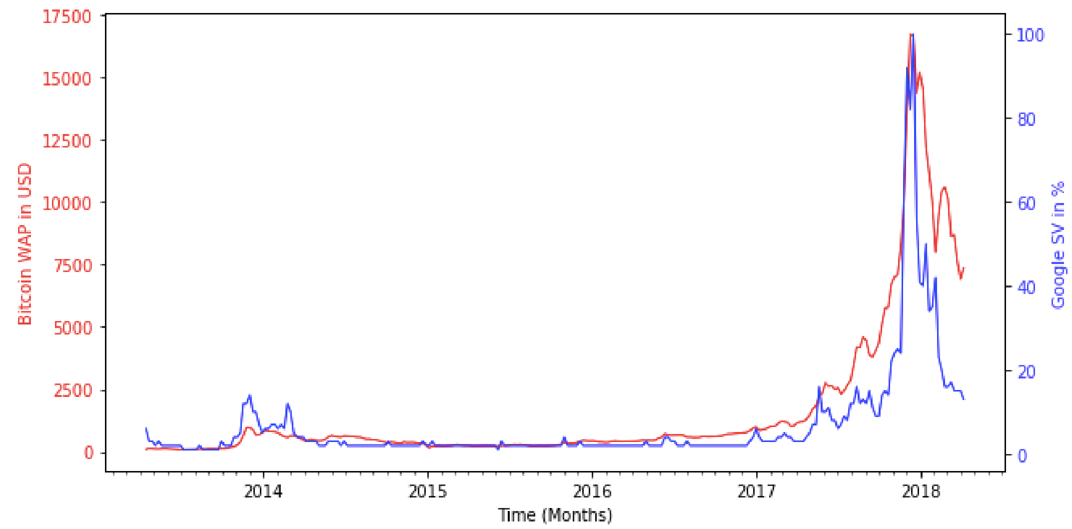



Figure 1 – Historical development of Google SV and Bitcoin WAP

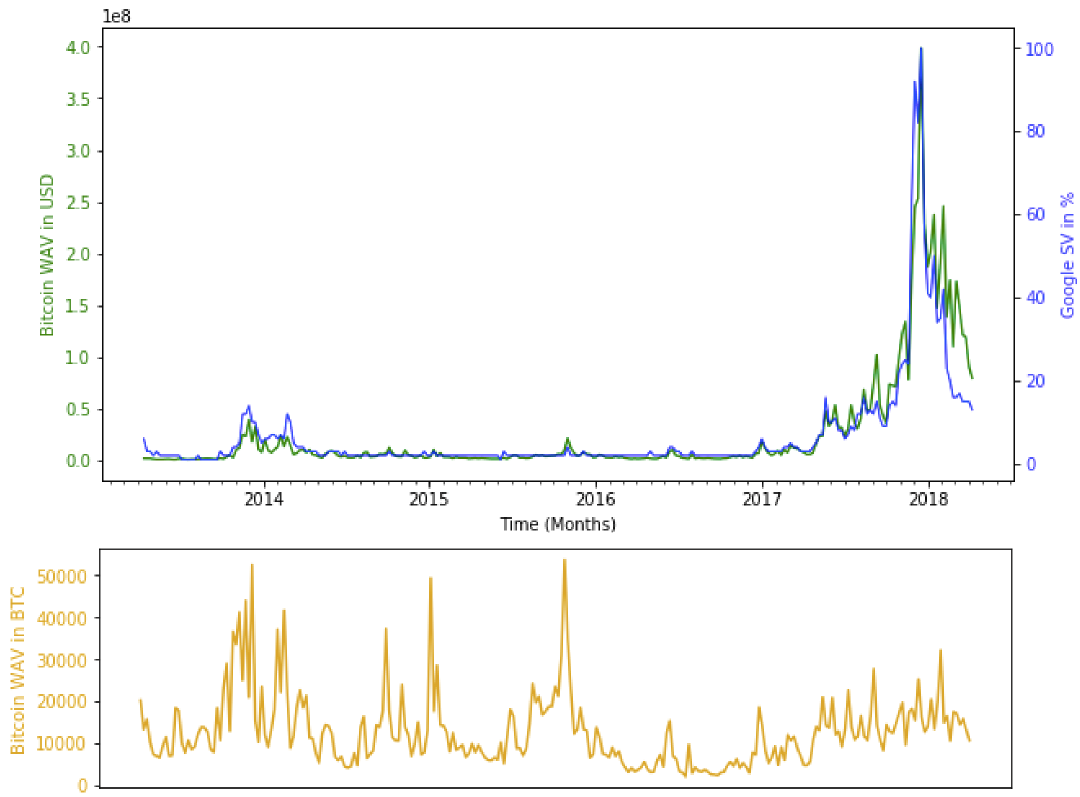



Figure 2 – Historical development of Google SV and Bitcoin WAV

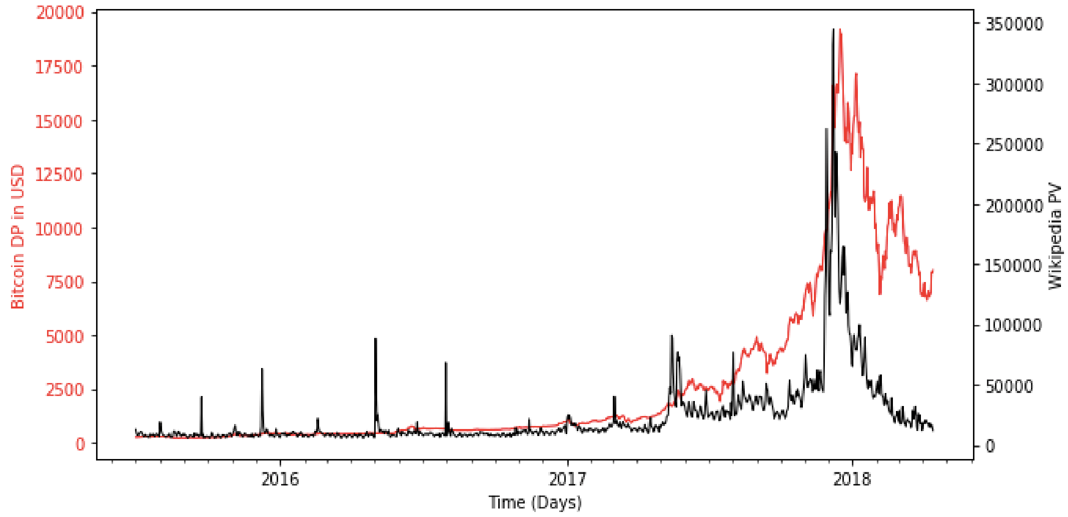



Figure 3 – Historical development of Wikipedia PV and Bitcoin DP

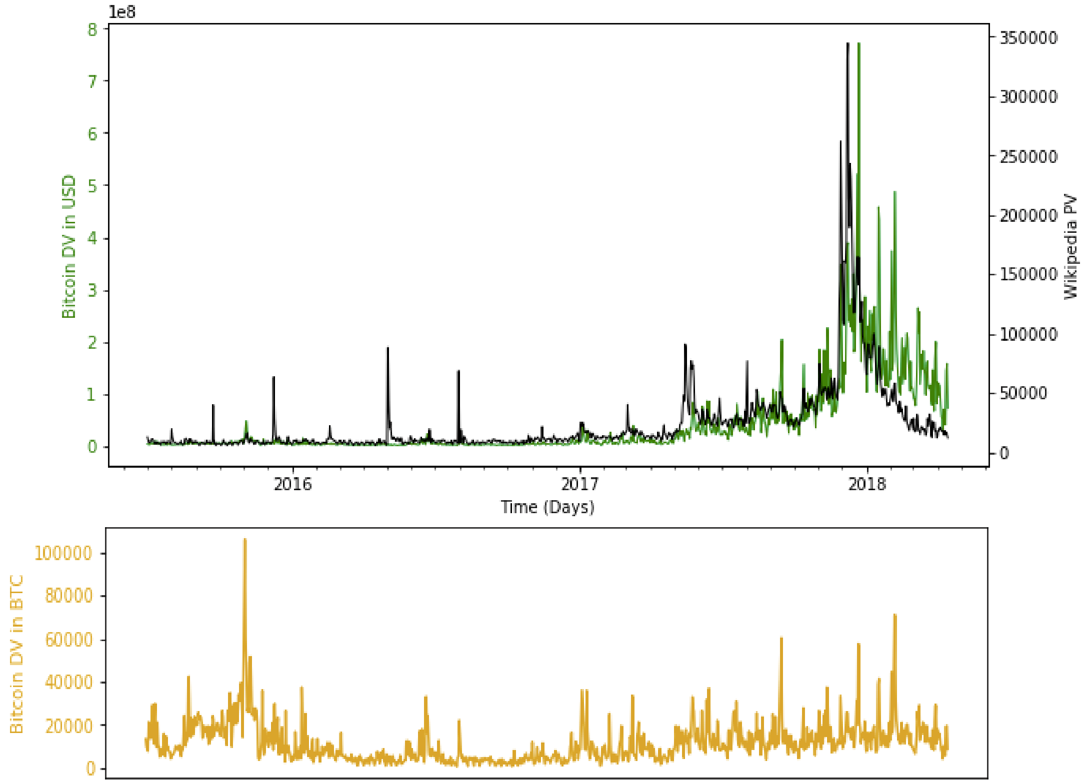



Figure 4 – Historical development of Wikipedia PV and Bitcoin DV

## 4.2 Correlation and cointegration analyses

### 4.2.1 Bitcoin price

| Variable/Transformation | r | p - value |
|---|---|---|
| **Google SV** | 0,89 | < 0,01 |
| 1st log differences | 0,2 | < 0,01 |
| **Wikipedia PV** | 0,75 | < 0,01 |
| 1st log differences | 0,04 | 0,16 |
| *Control Variable* | | |
| **Google SV (Worldwide)** | 0,92 | < 0,01 |
| 1st log differences | 0,15 | 0,016 |

Table 3 – Results of the correlation analysis concerning the Bitcoin price

| Variable | t-stat | p - value |
|---|---|---|
| **Google SV** | -4,08 | < 0,01 |
| **Wikipedia PV** | -3,9 | < 0,01 |
| *Control Variable* | | |
| **Google SV (Worldwide)** | -4,19 | < 0,01 |

Table 4 – Results of the cointegration test concerning the Bitcoin price

### 4.2.2 Bitcoin trading volume

| Variable/Transformation | r | p - value |
|---|---|---|
| **Google SV** | 0,93 | < 0,01 |
| 1st log differences | 0,5 | < 0,01 |
| **Wikipedia PV** | 0,75 | < 0,01 |
| 1st log differences | 0,27 | < 0,01 |
| *Control Variable* | | |
| **Google SV (Worldwide)** | 0,95 | < 0,01 |
| 1st log differences | 0,47 | 0,016 |

Table 5 – Results of the correlation test concerning the Bitcoin trading volume in USD

| Variable | t-stat | p - value |
|---|---|---|
| **Google SV** | -4,18 | < 0,01 |
| **Wikipedia PV** | -4,69 | < 0,01 |
| *Control Variable* | | |
| **Google SV (Worldwide)** | -3,05 | 0,09 |

Table 6 – Results of the cointegration test concerning the Bitcoin trading volume in USD

## 4.3 Granger causality and impulse response functions

For the following impulse response analyses, orthogonalized impulse response functions are used. The variables of the VAR model are ordered such that $y_1$ represents the Bitcoin price or trading volume and $y_2$ represents the measure of retail investor attention (see *3. Methodology* for further explanation). The impulse response functions with the reversed order of the variables are depicted in Appendix D.

### 4.3.1 Bitcoin price

In the case of the Google SV and Bitcoin WAP variable pair the Bayesian information criterion proposes the minimum lag and leads to a VAR model of order one. Based on the test statistics of the applied Granger causality test (see Appendix D for a summary of the test statistics), the null hypothesis can be rejected at a 5% significance level in both cases. This implies that impulses in Google SV are (Granger-) causal for responses in Bitcoin WAP (p-value = 0,01) and impulses in Bitcoin WAP are (Granger-) causal for responses in Google SV (p-value < 0,001). The VAR model's impulse response functions are shown in Figure 5 below.

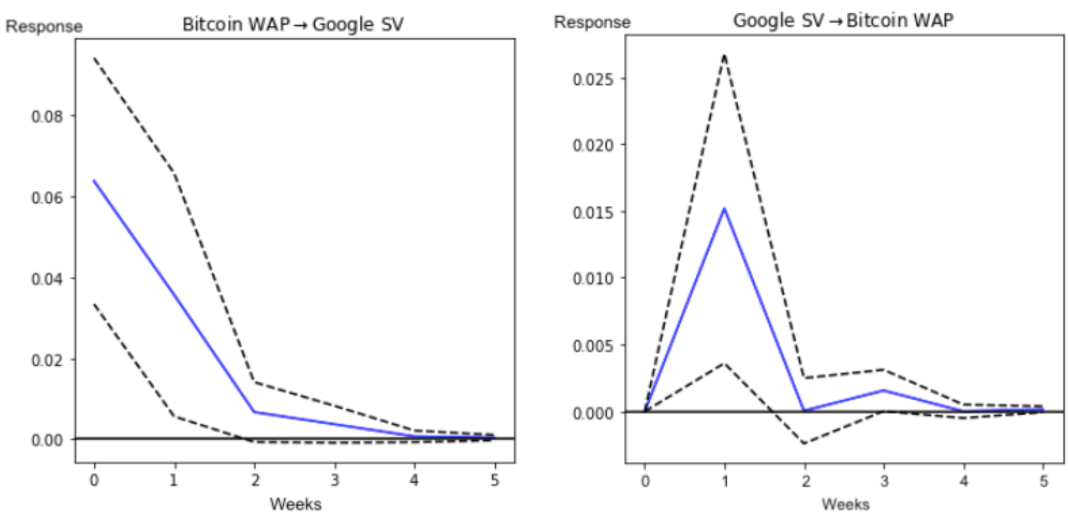



Figure 5 – Orthogonalized impulse response functions of the Google SV/Bitcoin WAP - VAR model. (Dashed lines show 95% significance bounds)

For Wikipedia PV and Bitcoin DP the impulse response functions are obtained through a VAR model of order six, following the Bayesian information criterion as well. The response of Bitcoin DP to a shock in

Wikipedia PV is negligible: The Granger causality test fails to reject the null hypothesis at every common significance level (p-value = 0,867), implying that a shock in Wikipedia PV does not (Granger-) cause responses in Bitcoin DP. The opposite is the case when impulse and response are reversed. Based on the Granger two-step test, Bitcoin DP is (Granger-) causal for Wikipedia PV at a significance level of 5% (p-value = 0,027). The corresponding impulse response function is depicted in Figure 6 below.

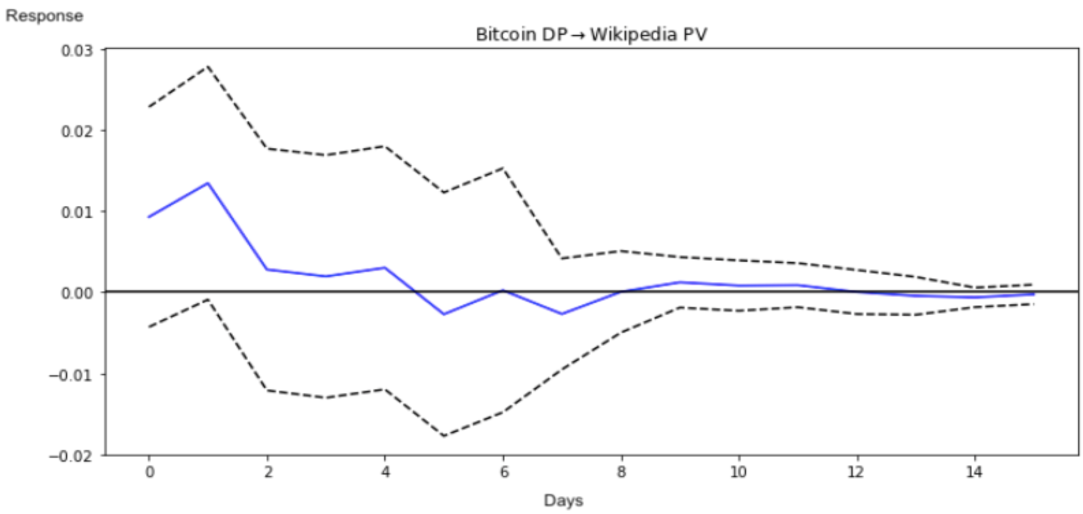



Figure 6 – Orthogonalized impulse response function illustrating the response in Wikipedia PV to an impulse in Bitcoin DP. (Dashed lines show 95% significance bounds)

### 4.3.2 Bitcoin trading volume

The following VAR models were estimated using the trading volume stated as the number of bitcoins traded. The results obtained for the trading volume in USD are depicted in Appendix E. Due to the application of first logarithmic differences both results are very similar since the transformation adjusts for the trend of the Bitcoin price in the USD representation of the trading volume. Nevertheless, I base my discussion on the findings about the trading volume in bitcoins traded since potential influences of the Bitcoin price are avoided in advance.

When investigating the Google SV and Bitcoin WAV variable pair, I uncover no statistically significant response in one of the variables to a shock in the respective other variable, based on the Granger causality test (p-value = 0,66 and 0,45). However, it bears mentioning that in this particular case the result obtained for the control variable of worldwide Google search volume is significantly different and therefore must be mentioned: The Granger causality tests applied to the VAR model with worldwide search queries declares that Google SV does (Granger-) cause Bitcoin WAV with a p-value of 0,063. This causality is represented by the impulse response function in Figure 7 below. However, a statistically significant (Granger-) causality for the opposite cause-effect relation, Bitcoin WAV → Google SV, can also not be proven for the VAR model with worldwide Google searches (p-value = 0,88).

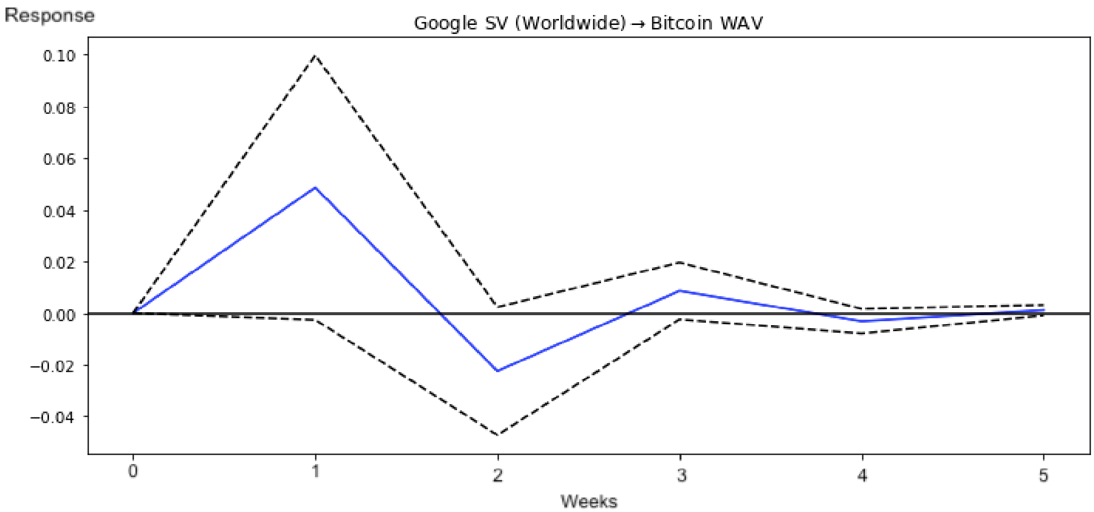



Figure 7 – Orthogonalized impulse response function illustrating the response in Bitcoin WAV to an impulse in Google SV (Worldwide). (Dashed lines show 95% significance bounds)

The Granger-causality test furthermore claims that Wikipedia PV is (Granger-) causal for Bitcoin DV at a significance level of 5% (p-value = 0,002). The corresponding impulse response function is depicted in Figure 8 below. When shock-variable and response-variable are reverted and a response in Wikipedia PV to a shock from Bitcoin DV is investigated, the Granger-causality test fails to reject the null hypothesis even at a 10% significance level (p-value=0,18). No statistically significant causality is present.

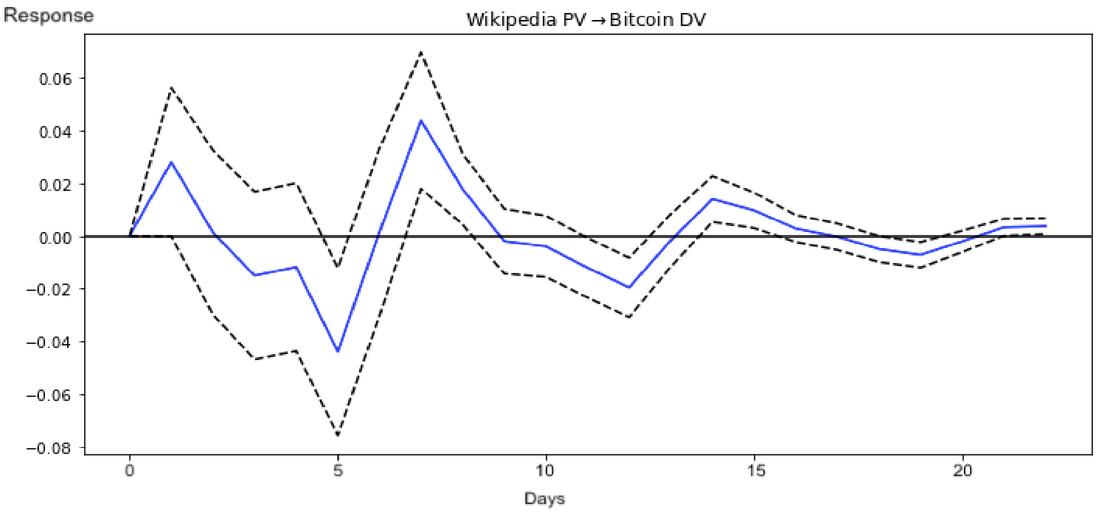



Figure 8 – Orthogonalized impulse response function illustrating the response in Bitcoin DV to an impulse in Wikipedia PV. (Dashed lines show 95% significance bounds)

# 5 Discussion

The maxima of Google search volume, weekly average Bitcoin price and weekly average Bitcoin trading volume in USD occur on two consecutive weeks in December 2017. When applying Wikipedia pageviews as the measure of retail investor attention, the maxima of the three variables on a daily level are a bit more dispersed but also appear in only three consecutive weeks in December. This distribution and the historical development of the variables, visible in Figures 1-4, already suggest a co-movement of the investigated time series which is proven statistically: The correlation between retail investor attention and the Bitcoin price as well as the Bitcoin trading volume is strongly positive and significant in the case of the non-transformed, non-stationary time series. After adjusting for stationarity these positive correlations become a lot weaker but remain significant, except for the Wikipedia pageviews and the daily price of Bitcoin. Apart from the correlation analyses, the cointegration test illustrates that there is a long-run relationship between the two measures of investor attention and the Bitcoin prices/trading volume. Nonetheless, all of these findings are only descriptive in nature and can therefore not verify a cause-and-effect relationship.

Having said that, also dynamic relationships between the variables are found. I prove a significant bi-directional causal relationship between Google searches and the Bitcoin weekly average price (p-value = 0,01 / <0,001). The impulse response function (Figure 5) shows that an increase in attention (i.e. in Google searches) causes an increase in the Bitcoin price. With no fundamental value behind bitcoins due to the absence of cash flows and consequently lacking evaluation techniques, it seems that attention entails meaningful information about the Bitcoin price. On the other hand, an increase in the Bitcoin price also causes an increase in Google searches. The corresponding impulse response functions (Figure 5) indicate that the cumulative response in Google searches is a lot stronger than the cumulative response in the Bitcoin price. These results correspond to the findings of Kristoufek (2013) who discovered the same mutual relationship between Google searches and the Bitcoin price, as well as the difference in magnitude, for an earlier time period, from May 2011 to June 2014.

The mutual causality is a very interesting finding as it demonstrates how the attention dedicated to Bitcoin and the price of Bitcoin whip each other up: Enlarged attention stimulates buying decisions by retail investors, which creates price pressure and rising prices that raise attention even more. This self-reinforcing cycle of cause and effect can serve as an explanation for the price explosions apparent in the Bitcoin market ever since its creation. However, based on the impulse response function in Figure 5, I cannot conclude that the sharp declines in price which followed the vigorous increases throughout the existence of Bitcoin (the most recent one in the beginning of 2018, see Figure 1) are a direct consequence of increased attention. I expected a different behavior of the impulse response function based on the results of Da et al. (2011) concerning the stock market. In their paper, the three authors prove the existence of a long-run reversal in share prices, following an abnormal increase in attention measured by the Google search volume. For Da et al. this follows the understanding that attention does not resemble fundamental information about the firm and can therefore only constitute temporary price pressure (Da et al., 2011, p. 1483). With the same thesis one could assume a similar behavior of the Bitcoin price. Nonetheless, my obtained impulse response function does not acknowledge a price reversal in the weeks following the impulse in attention. Methodically, Da et al.’s research does not resemble mine since the authors used the abnormal Google search volume, expressed as the deviation of the search

volume from its eight-week moving median, as a variable rather than the plain search volume. In order to truly falsify the thesis of Da et al. (2011) for Bitcoin, I am prompted to repeat the analyses using a measure of abnormal attention (i.e. whether the Google search volume is above or below its moving average). As for now, some external factor(s) other than attention must be regarded the trigger for the frequent collapses in the Bitcoin price.
Contrary to the discussed findings on the Google search volume, retail investor attention measured by Wikipedia pageviews does not cause a statistically significant response in the Bitcoin price (p-value = 0,867). Only for the opposite cause-effect relation, Bitcoin price → Wikipedia pageviews, a significant causality is proven in this research (p-value = 0,027). Again, this is in line with the findings of Kristoufek (2013). It seems like an increase in the price of Bitcoin encourages people to pay attention to the currency by looking it up on Wikipedia. However, this attention does not translate into buying/selling of bitcoins and no price reaction is provable. When taking the different results concerning Google searches and Wikipedia pageviews into account, my findings confirm Kristoufek's thesis that Wikipedia and Google are used for different purposes and the individuals using them *"[…] can have different motives"* (Kristoufek, 2013, p.4).
The findings comprising the trading volume of Bitcoin are informative as well. My results claim that the Bitcoin trading volume does, unlike the Bitcoin price, not affect attention, neither the Google search volume, nor the pageviews on Wikipedia (p-value = 0,45 / 0,18). However, I prove that the two measures of attention are conversely (Granger)-causal for the trading volume:
While for the US Google search volume, no statistically significant causality is measurable as well (p-value = 0,66), the result is significantly different when the control variable of worldwide Google search volume is applied. In this case the Granger causality test confirms that the search volume is causal for the Bitcoin trading volume with a p-value of 0,063. With this 10% significance level in mind, the result is still to be evaluated critically but can be further discussed. Whether the causal relationship can be considered valid or not depends on the assumption made about the internationality of the trading activity on bitstamp.net. Is it appropriate to use retail investor attention on a global scale even though trading data from an US-American

exchange is applied? Concerning the discussed results on the Bitcoin price, I would argue that it is appropriate based on, assuming an efficient Bitcoin market, the law of one price. Supporting this claim, causality is proven at the 1% significance level for both US and worldwide Google search volume. I find significantly differing results only for the Bitcoin trading volume and in this case, it is more difficult to argue that the attention of retail investors in, say, Germany is causal for the trading volume at an exchange in the US. An argument in favor of the results concerning the worldwide search volume is that Bitstamp is a very well-known Bitcoin exchange with one of the largest trading volumes in the world. It is therefore not unrealistic to assume that retail investors from all around the world trade there.

Next to the Google search volume, my results claim that Wikipedia pageviews are (Granger-) causal for the Bitcoin trading volume as well, in this case even at an unambiguous 1% significance level. Despite being an interesting finding, this can however not serve as an indicator for the validity of the same causality relating to the worldwide Google search volume since a similar issue occurs: The Wikipedia pageviews refer to the anglophone Wikipedia article for Bitcoin and the views are accumulated by people from all around the world, not just the US.

The impulse response functions of those two causal relationships (Google SV (Worldwide) / Wikipedia PV → Bitcoin trading volume) look very much alike (Figure 7 and 8). The only distinguishing feature is that the function depicting the response of an impulse in the Google search volume lasts longer and its profile is smoother. These observations can be explained by the application of weekly average, instead of daily, data points. Both functions have a peculiar look for which a definite explanation cannot be found simple-mindedly, which leads me to a necessary critique of the applied methodology: Other variables which also have an impact on the Bitcoin trading volume are omitted from the models. However, their influence is still included in the residuals. One of the, among potentially multiple, omitted variables with an influence on the trading volume might the Bitcoin price for which a causal relationship with the Google search volume and Wikipedia pageviews is proven and already discussed. A hypothesis explaining the responses in the Bitcoin trading volume emerges when the Bitcoin price is taken into account as well (compare Figures 7 and 8): At first, the initial shock in attention of retail investors leads to an increase in the traded volume of bitcoins. A simple explanation for this development is that new retail investors enter the Bitcoin market which leads to increased trading activity. Then the number of bitcoins traded decreases again, which I propose is due to the rise in the Bitcoin price which is proven to be another consequence of an impulse in attention. Keep in mind that this decrease in the number of traded bitcoins does not imply a simultaneously decreasing trading activity: The amount of money in US-dollars moving in the market can still rise due to the accompanied increase in the price of one bitcoin. In the last step, as has been proven, the increase in price causes attention to increase (further) and the process is reiterated. Even though there is only one shock in attention at the beginning of the impulse response functions, we can observe a repeated reaction based on the mutual causality of retail investor attention and the Bitcoin price. The cycle is initiated by the single impulse in the beginning, kept alive by repeated stimulus in attention but declining in power over time due to diminishing causal reactions. The process eventually ends after four weeks for Google searches and after about twenty days for Wikipedia pageviews. Following the frequent changes in sign of the response, the cumulative response in the trading volume is approximately zero.

My interpretations are substantiated by the historical development of the variables, visible in Figures 1-4: The development of the trading volume in the number of bitcoins traded (Figure 2 and 4, bottom) demonstrates the oscillating movements on the daily and weekly level that are suggested by the impulse response functions. Furthermore, this up and down motion of the trading volume does not follow a trend but moves around a constant mean, following the cumulative response of zero in both impulse response functions. This is also observable in the fact that the two time series, representing the trading volume in the number of bitcoins traded, are stationary before any transformation is applied (see Appendix A). The trading volume expressed in US-Dollars (Figure 2 and 4, top) does however follow trends, which resemble the trends in the Bitcoin price (Figure 1 and 3).

In the end, including an influence of the Bitcoin price into my considerations about the relationship between retail investor attention and the Bitcoin trading volume also validates the application and explains the results of the worldwide Google search volume and the anglophone Wikipedia article: The attention of retail investors from all around the world causes the Bitcoin price at any given bitcoin exchange based on efficient markets which consequently, together with attention-based influences once more, defines the traded volume.

# 6 Conclusion

All in all, the research objective to develop a better understanding of the Bitcoin market and its past behavior by analyzing its relationship with retail investor attention has been achieved.

Firstly, I document a strong and significant positive correlation of retail investor attention and the Bitcoin price and trading volume. This is proven for searches on google.com as well as for Wikipedia pageviews and thereby confirms my personal experience outlined in the introduction.

Moreover, these relationships are also found to be causal: An increase in the Bitcoin price causes an increase in the Google search volume and in Wikipedia pageviews. Conversely, based on Google search queries, investor attention also exerts an effect on the Bitcoin price. The volume of Google searches is therefore identified as a factor with an influence on the value of Bitcoin. Furthermore, the positive feedback loop, resulting from the mutual causality of Google searches and the Bitcoin price, creates a steady increase in both variables and serves as a potential explanation for Bitcoin’s past price explosions.

In order to examine this phenomenon further, I currently apply the causality of Google searches and the value of Bitcoin to the estimation of statistical models aimed at forecasting the Bitcoin price. Once the backtesting is complete, new insights concerning this relationship and its practical use for determining optimal buying/selling decisions for Bitcoin will be available.

Finally, Wikipedia pageviews and Google searches are identified to be causal for the trading volume in the Bitcoin market. The causal entanglement of attention, the trading volume and the Bitcoin price offers an explanation for the historically volatile but stationary volume of bitcoins traded and the trend-following trading volume in US-Dollar.

## Bibliography

Anon., 2018, Search engine market share worldwide, Statista.
www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/

Anon., 2018, Search Engine Market Share Worldwide, StatCounter Global Stats.
www.gs.statcounter.com/search-engine-market-share

Da, Zhi, Engelberg, Joseph, and Pengjie Gao, 2011, In Search of Attention, *Journal of Finance* 66, 1461–1499.

Dickey, David A., Jansen, Dennis W., and Daniel L. Thornton, 1994, A Primer on Cointegration with an Application to Money and Income, In: Rao B.B. (eds) *Cointegration*, (Palgrave Macmillan, London), 9-45

Dimpfl, Thomas, and Stephan Jank, 2016, Can Internet Search Queries Help to Predict Stock Market Volatility?, *European Financial Management* 22, 171-192.

Engle, Robert F., and Clive W. J. Granger, 1987, Co-Integration and Error Correction: Representation, Estimation, and Testing, *Econometrica* 55, 251-276.

Kristoufek, Ladislav, 2013, BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era, *Scientific Reports* 3, 3415.

Lütkepohl, Helmut, 2005, *New Introduction to Multiple Time Series Analysis*, (Springer Science & Business Media, Berlin).

Phillips, Peter C.B., 1986, Understanding spurious regressions in econometrics, *Journal of Econometrics* 33, 311-340.

Sims, Christopher A., 1980, Macroeconomics and Reality, *Econometrica* 48, 1-48.

Yule, G. Udny, 1926, Why do we Sometimes get Nonsense-Correlations between Time-Series?--A Study in Sampling and the Nature of Time-Series, *Journal of the Royal Statistical Society* 89, 1-63.

# Appendix

**Appendix A:** Results of the unit-root tests (ADF and KPSS)

| Variable/Transf | ADF | p-value | KPSS | p- value |
|---|---|---|---|---|
| **Bitcoin WAP** | 3,0609 | > 0,05 | 0,7582 | < 0,05 |
| log | -0,2618 | > 0,05 | 1,1135 | < 0,05 |
| 1st differences | -2,9134 | < 0,05 | 0,1083 | > 0,05 |
| 1st log differences | -11,9824 | < 0,05 | 0,1161 | > 0,05 |
| **Bitcoin DP** | -1,4307 | > 0,05 | 2,7805 | < 0,05 |
| log | -0,0746 | > 0,05 | 4,1742 | < 0,05 |
| 1st difference | -5,6322 | < 0,05 | 0,0817 | > 0,05 |
| 1st log difference | -31,7937 | < 0,05 | 0,1174 | > 0,05 |
| **Bitcoin WAV USD** | -2,7215 | > 0,05 | 0,6559 | < 0,05 |
| log | -1,6863 | > 0,05 | 0,7604 | < 0,05 |
| 1st differences | -4,517 | < 0,05 | 0,0526 | > 0,05 |
| 1st log differences | -10,316 | < 0,05 | 0,0882 | > 0,05 |
| **Bitcoin DV USD** | -1,6516 | > 0,05 | 2,482 | < 0,05 |
| log | -8,1912 | < 0,05 | 0,0426 | > 0,05 |
| 1st differences | -7,6843 | < 0,05 | 0,0268 | > 0,05 |
| 1st log differences | -8,1912 | < 0,05 | 0,0426 | > 0,05 |
| **Bitcoin WAV BTC** | -5,4693 | < 0,05 | 0,2338 | > 0,05 |
| log | -2,5171 | > 0,05 | 0,2401 | > 0,05 |
| 1st differences | -9,0727 | < 0,05 | 0,0377 | > 0,05 |
| 1st log differences | -10,0733 | < 0,05 | 0,0522 | > 0,05 |
| **Bitcoin DV BTC** | -3,2975 | < 0,05 | 0,6475 | < 0,05 |
| log | -2,6849 | > 0,05 | 0,9497 | < 0,05 |
| 1st differences | -10,8907 | < 0,05 | 0,0161 | > 0,05 |
| 1st log differences | -8,7341 | < 0,05 | 0,024 | > 0,05 |
| **Google SV** | -2,7929 | > 0,05 | 0,5576 | < 0,05 |
| log | -1,6393 | > 0,05 | 0,6916 | < 0,05 |
| 1st difference | -6,0175 | < 0,05 | 0,042 | > 0,05 |
| 1st log difference | -6,0344 | < 0,05 | 0,1293 | > 0,05 |
| **Wikipedia PV** | -3,5367 | < 0,05 | 1,6456 | > 0,05 |
| log | -1,8077 | > 0,05 | 3,238 | > 0,05 |
| 1st difference | -8,0801 | < 0,05 | 0,0277 | > 0,05 |
| 1st log difference | -7,9689 | < 0,05 | 0,0624 | > 0,05 |
| *Control Variable* | | | | |
| **Google SV (Worldwide)** | -2,5841 | > 0,05 | 0,636 | < 0,05 |
| log | -1,8941 | > 0,05 | 0,8359 | < 0,05 |
| 1st difference | -6,1779 | < 0,05 | 0,0457 | > 0,05 |
| 1st log difference | -11,1784 | < 0,05 | 0,1386 | > 0,05 |

**Appendix B:** Distribution of the original data and all transformations

| Variable/Transf | Mean | Skewness | Kurtosis |
|---|---|---|---|
| **Bitcoin WAP** | 1671,45 | 2,92 | 8,29 |
| log | 6,46 | 0,89 | 0,28 |
| 1st differences | 27,87 | 1,7 | 25,04 |
| 1st log differences | 0,02 | -0,06 | 8,33 |
| **Bitcoin DP** | 2703,4 | 1,99 | 3,33 |
| log | 7,06 | 0,65 | -0,85 |
| 1st differences | 7,61 | 0,84 | 24,94 |
| 1st log differences | 0 | -0,16 | 4,37 |
| **Bitcoin WAV USD** | 24611510,51 | 3,5 | 14,05 |
| log | 15,71 | 0,89 | 0 |
| 1st differences | 298807,64 | -0,99 | 26,54 |
| 1st log differences | 0,01 | 0,3 | 0,18 |
| **Bitcoin DV USD** | 39578451,36 | 3,4 | 16,94 |
| log | 16,12 | 0,51 | -0,9 |
| 1st differences | 67021,72 | 0,32 | 46,6 |
| 1st log differences | 0 | 0,31 | 0,7 |
| **Bitcoin WAV BTC** | 12781,7 | 1,93 | 5,09 |
| log | 9,26 | -0,11 | -0,14 |
| 1st differences | -36,86 | 0,03 | 5,41 |
| 1st log differences | 0 | 0,27 | 0,29 |
| **Bitcoin DV BTC** | 11567,95 | 2,7 | 14,85 |
| log | 9,06 | -0,25 | -0,22 |
| 1st differences | -4,38 | 0,07 | 6,81 |
| 1st log differences | 0 | 0,33 | 0,78 |
| **Google SV** | 6,7 | 4,71 | 25,91 |
| log | 1,27 | 1,4 | 1,48 |
| 1st differences | 0,03 | -0,28 | 43,94 |
| 1st log differences | 0 | 0,71 | 3,49 |
| **Wikipedia PV** | 23853,22 | 4,64 | 29,15 |
| log | 9,67 | 1,15 | 0,85 |
| 1st differences | -0,84 | 2,49 | 72,06 |
| 1st log differences | -6,69 | 1,55 | 29,38 |
| *Control Variable* | | | |
| **Google SV Worldwide** | 7,68 | 4,12 | 19,74 |
| log | 1,39 | 1,24 | 0,89 |
| 1st difference | 0,04 | 0,35 | 37,45 |
| 1st log difference | 0 | 0,44 | 2,15 |

## **Appendix C:** Results of the Granger causality f-test

For the Bitcoin price:

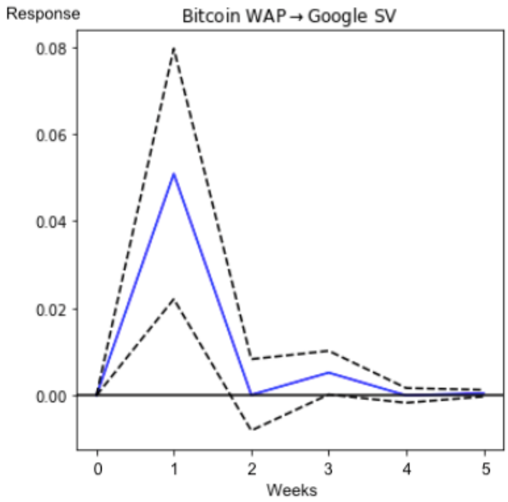
| Impulse -> Response | Test Statistic | Critical Value (p-value 0.05) | p-value |
|---|---|---|---|
| Google SV --> Bitcoin WAP | 6,6912 | 3,8597 | 0,01 |
| Bitcoin WAP --> Google SV | 12,3286 | 3,8597 | < 0,001 |
| Wikipedia PV --> Bitcoin DP | 0,4192 | 2,1031 | 0,867 |
| Bitcoin DP --> Wikipedia PV | 2,3849 | 2,1031 | 0,027 |
| *Control Variable* | | | |
| Google SV Worldwide --> Bitcoin WAP | 10,0994 | 3,8597 | 0,002 |
| Bitcoin WAP --> Google SV Worldwide | 14,081 | 3,8597 | < 0,001 |

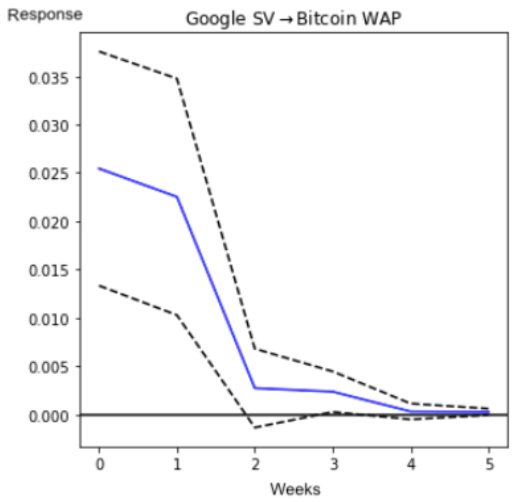
## For the Bitcoin trading volume:

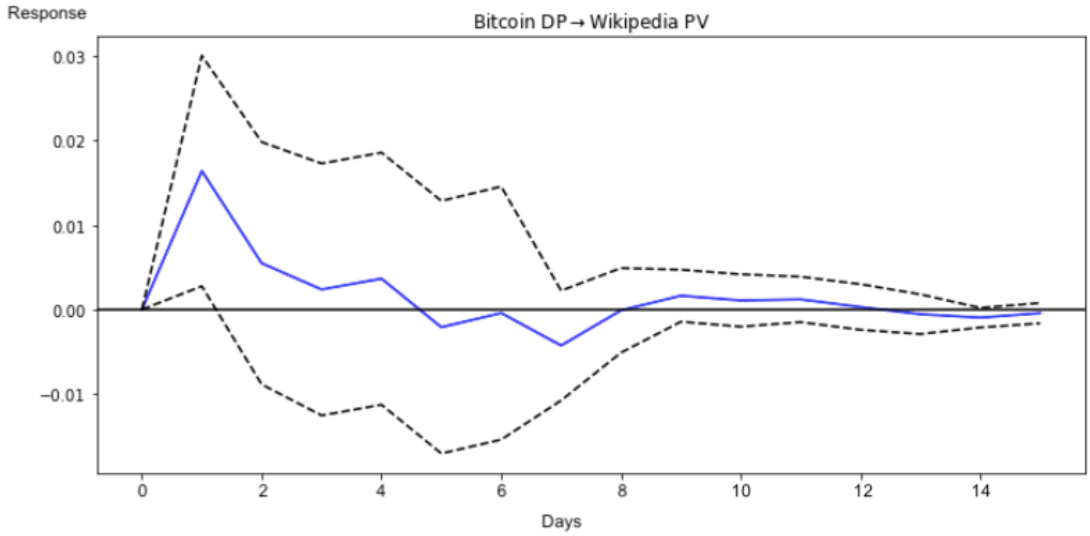
| Impulse -> Response | Test Statistic | Critical Value (p-value 0.05) | p-value |
|---|---|---|---|
| Google SV --> Bitcoin WAV USD | 0,0273 | 3,8597 | 0,87 |
| Bitcoin WAV USD --> Google SV | 0,1429 | 3,8597 | 0,71 |
| Google SV --> Bitcoin WAV BTC | 0,1968 | 3,8597 | 0,66 |
| Bitcoin WAV BTC --> Google SV | 0,5648 | 3,8597 | 0,45 |
| Wikipedia PV --> Bitcoin DV USD | 3,8465 | 2,1031 | 0,001 |
| Bitcoin DV USD --> Wikipedia PV | 1,1586 | 2,1031 | 0,33 |
| Wikipedia PV --> Bitcoin DV BTC | 3,5924 | 2,1031 | 0,002 |
| Bitcoin DV BTC --> Wikipedia PV | 1,4883 | 2,1031 | 0,18 |
| *Control Variable* | | | |
| Google SV Worldwide --> Bitcoin WAV USD | 5,8252 | 3,8597 | 0,016 |
| Bitcoin WAV USD --> Google SV Worldwide | 1,4724 | 3,8597 | 0,23 |
| Google SV Worldwide --> Bitcoin WAV BTC | 3,4823 | 3,8597 | 0,063 |
| Bitcoin WAV BTC --> Google SV Worldwide | 0,0214 | 3,8597 | 0,88 |

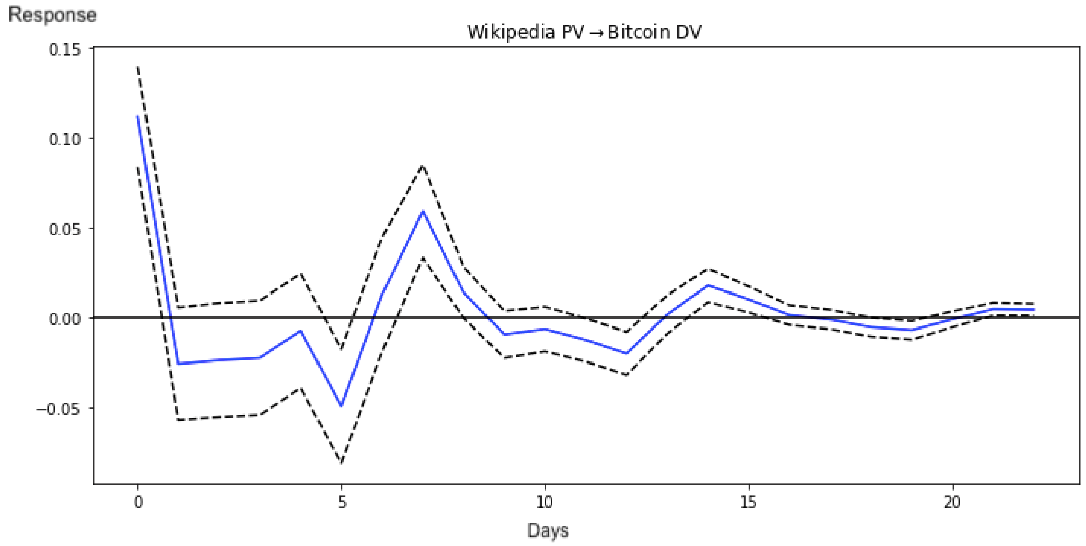
**<u>Appendix D:</u>** Orthogonalized impulse response functions

Variable order: $y_1$: Google SV/Wikipedia PV; $y_2$: Bitcoin variable

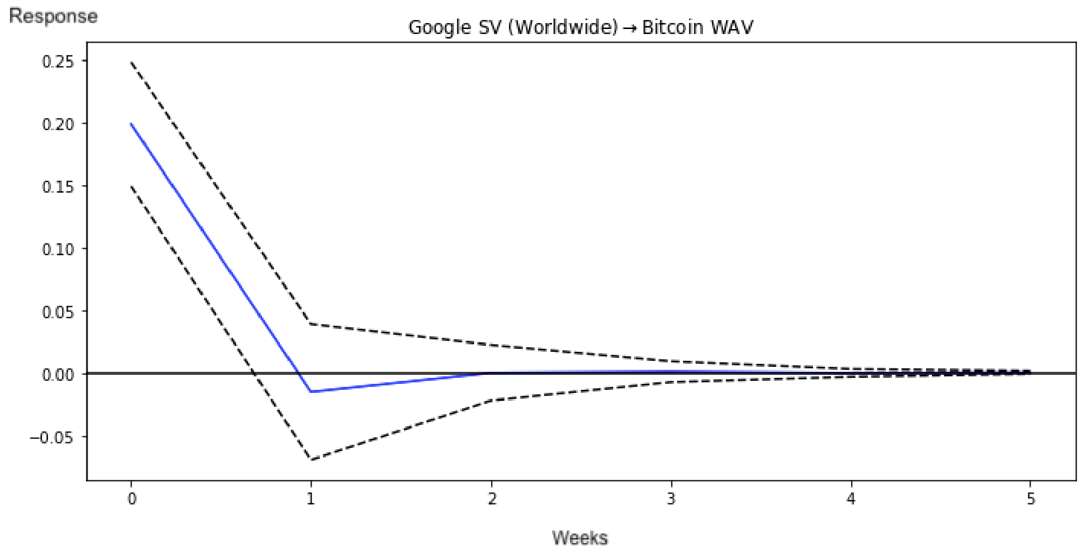


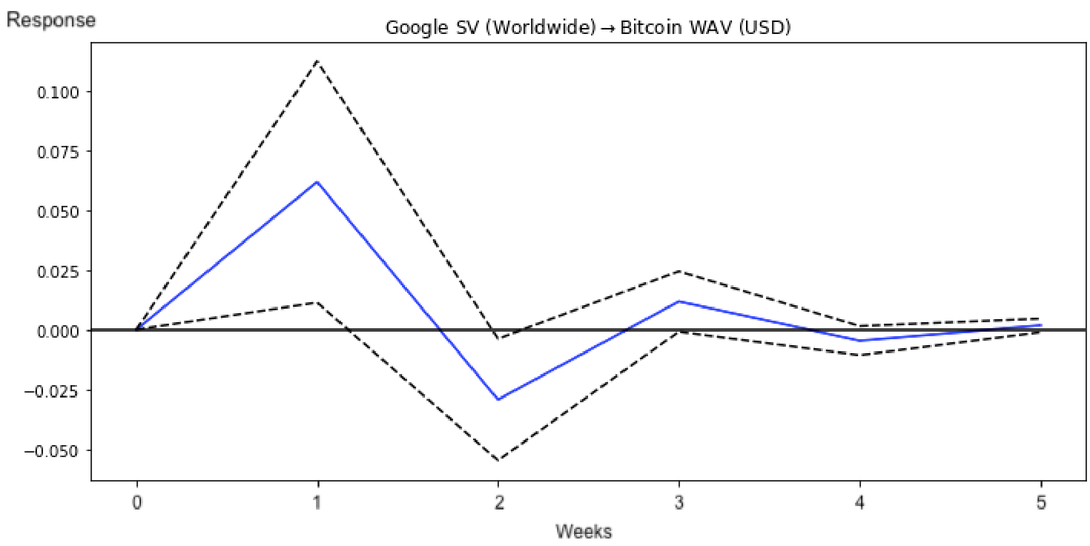



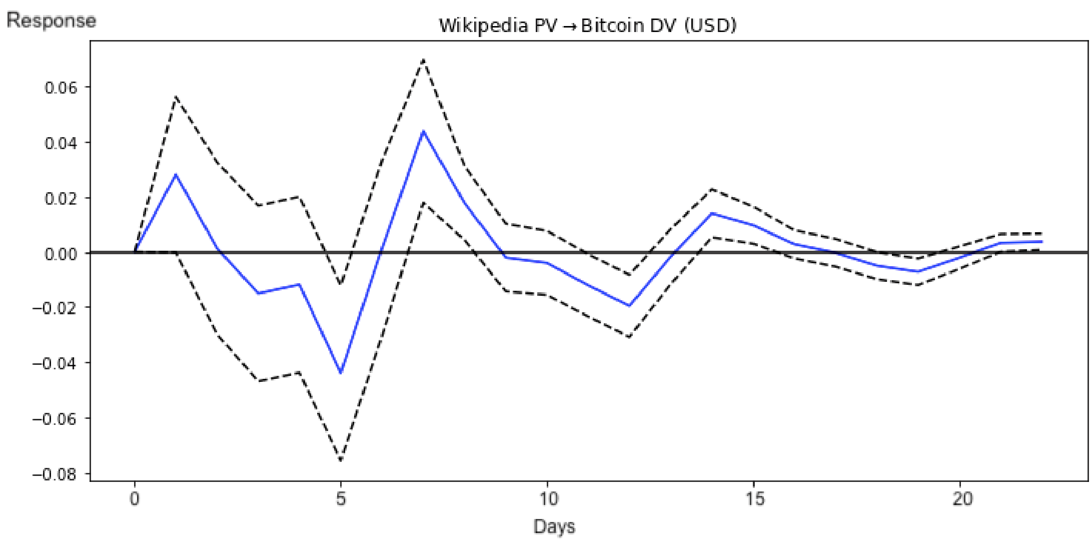



Control Variable (Bitcoin SV Worldwide):



**<u>Appendix E:</u>** Orthogonalized impulse response functions with the trading volume in USD



# Eidesstattliche Erklärung (Declaration of academic honesty)

„Hiermit versichere ich, dass diese Bachelorarbeit von mir persönlich verfasst ist und dass ich keinerlei fremde Hilfe in Anspruch genommen habe. Ebenso versichere ich, dass diese Arbeit oder Teile daraus weder von mir selbst noch von anderen als Leistungsnachweise andernorts eingereicht wurde. Wörtliche oder sinngemäße Übernahmen aus anderen Schriften und Veröffentlichungen in gedruckter oder elektronischer Form sind gekennzeichnet. Sämtliche Sekundärliteratur und sonstige Quellen sind nachgewiesen und in der Bibliographie aufgeführt. Das Gleiche gilt für graphische Darstellungen und Bilder sowie für alle Internet-Quellen. Ich bin ferner damit einverstanden, dass meine Arbeit zum Zwecke eines Plagiatsabgleichs in elektronischer Form anonymisiert versendet und gespeichert werden kann. Mir ist bekannt, dass von der Korrektur der Arbeit abgesehen werden kann, wenn die Erklärung nicht erteilt wird.“

Ort, Datum, eigenhändige Unterschrift