WHAT CAN GOOGLE TELL US ABOUT BITCOIN TRADING VOLUME IN CROATIA? EVIDENCE FROM THE ONLINE MARKETPLACE LOCALBITCOINS*

Tea Livaić** and Ana Perišić

Polytechnic of Šibenik
Šibenik, Croatia
DOI: 10.7906/indecs.17.4.2
Regular article

ABSTRACT

Timely economic statistics is crucial for effective decision making. However, most of them are released with a lag. Thus, nowcasting has become widely popular in economics, and web search volume histories are already used to make predictions in various fields including IT, communications, medicine, health, business and economics. This article seeks to explore the potential of incorporating internet search data, in particular Google Trends data, in autoregressive models used to predict the volume of Bitcoin trading. Toda and Yamamoto procedure was applied in order to examine causality between Google search data and Bitcoin trading volume on the online marketplace LocalBitcoins, for the area of the Republic of Croatia. The results showed that internet search data can be useful for forecasting Bitcoin trading volume, since Google searches for the term “bitcoin” Granger causes Bitcoin trading volume in the online marketplace LocalBitcoins.

KEY WORDS

Bitcoin, Google Trends, Granger causality, Toda and Yamamoto approach

CLASSIFICATION

JEL: C1, E47

*This article is substantially developed part of the graduate thesis of T. Livaić.
**Corresponding author, *livaic.tea@gmail.com; +385 22 311 088;
The Polytechnic of Šibenik, Trg Andrije Hebranga 11, HR – 22 000 Šibenik, Croatia
INTRODUCTION

In the era of modern technology, the growing impact of digitization moves the classic trading model into the virtual world. Consequently, the need for alternative means of exchange arise. This is one of the cause of development, among others phenomena, of the cryptocurrencies, i.e. systems that use cryptography to allow the secure transfer and exchange of digital tokens in a distributed and decentralized manner. These tokens can be traded at market rates for fiat currencies. With the increasing informatization, massive new data sources resulting from human interaction with the internet offer a new perspective on the behavior of market participants in periods of large market movements [1]. These new data sources, such as Google Trends, gained substantial attention due to the ability to capture real-time online signals about consumer interest in a aspecific topic. A number of studies have examined how internet search data can be used to monitor various phenomena as they happen. This process is mostly described as “nowcasting”. Nowcasting is defined as the prediction of the present, the very near future and the very recent past [2]. The term is a contraction for now and forecasting and has been used for a long-time in meteorology and recently also in economics [3]. Nowcasting has become widely popular in economics, and web search volume histories are already used to make predictions in various fields including IT, communications, medicine, health, business and economics. Nowcasting is particularly relevant for those key macro economic variables which are collected at low frequency, typically on a quarterly basis, and released with a substantial lag [2]. Research using Google Trends has increased dramatically in the last decade, and in the process, the focus of research has shifted to forecasting changes, whereas in the past the focus had been on merely describing and diagnosing research trends, such as surveillance and monitoring [4]. It has been confirmed that Google search data constitute a reliable alternative when official data are lacking [5].

Although Google search data have been applied in wide range of areas and a number of articles have demonstrated the usefulness of Google trends, studies on the level of Croatia are lacking. Therefore, this article contributes to analyses of the utility of the Google search data on economic data, especially cryptocurrencies. The focus is placed on the analysis of the relationship of search data with the keyword “bitcoin” provided from Google Trends for the area of the Republic of Croatia, with the Bitcoin trading volume in Croatia in the LocalBitcoins marketplace. Causality test is based on the approach proposed by Toda and Yamamoto [6].

CRYPTOCURRENCY

Cryptocurrency is the name given to a system that uses cryptography to allow a secure transfer and exchange of digital tokens in a distributed and decentralized manner. These tokens can be traded at market rates for fiat currencies [7]. Cryptocurrencies are not real money, they represent an internet protocol through which is data transferred from one web-location to another. Every protocol has a purpose and a specified type of data which it transfers. The purpose of a protocol can be e.g. exchange for a fiat currency, the smart contract, exchange of data, etc. Unlike fiat currencies, cryptocurrencies are decentralized, meaning that there is no central authority which manages or issues them and their value is determined by their users.

Bitcoin is a Peer-to-Peer system based on complex cryptographic algorithms. Peer-to-Peer is a network where there is no central authority issuing new money or tracking transactions. The advantages of such a system are that it is possible to point out a simple transfer of money over the internet, without the intermediary, whereby a third party cannot prevent or manage
the user’s transactions. There is no central bank in the Bitcoin system that issues money and keeps and processes transactions, nor does it have a unique owner of the Bitcoin network. The key difference of Bitcoin in relation to centralized systems comes from the fact that every user has access to their transactions and transactions of the other parties. Each transaction contains a digital signature of the user who started it [8]. The precision of this value limits the extent to which units of the currency can be subdivided; the smallest unit is called Satoshi. By convention, $10^8$ Satoshi is considered the primary unit of currency, called one “Bitcoin” and denoted BTC or XBT [9].

Bitcoin is the first implementation of a concept called “cryptocurrency”, which was first described in 1998 by Wei Dai on the Cypherpunks mailing list, suggesting the idea of a new form of money that uses cryptography to control its creation and transactions, rather than a central authority. The first Bitcoin specification and proof of concept was published in 2009 in a cryptography mailing list by Satoshi Nakamoto [10]. Bitcoin’s genesis block was mined around January 3, 2009. The first use of Bitcoin as a currency is thought to be a transaction in May 2010, where one user ordered pizza delivery for another in exchange for 10,000 bitcoins [9]. Trading took off in 2011, when one bitcoin was worth about $0.05. In early 2013, bitcoin peaked above $200, only to drop back in value later on again. During the final months of 2013, the value increased to over $1100 and dropped in the following months. During the early months of 2015, the value of bitcoin has been relatively stable between $200 and $300 and after rising since the end of 2015, the value rose above $900 again [11]. The value began to grow steadily again in the year 2016. At the end of 2017, at one point, it reached the value of $20,000.

The Bitcoin’s success has ignited an exposition of new alternative crypto-currencies (altcoins); however, none of these have been able to jeopardize the Bitcoin’s dominant role in the field [12]. Most of altcoins rely on the same or similar blockchain technology as Bitcoin, and aim to either complement or improve certain Bitcoin characteristics [13].

GOOGLE TRENDS

Google Trends is an online tool that provides their users to explore how frequently specific terms, phrases and topics are entered into Google’s search engine relative to the site’s total search volume over a specific time period and in a specific place (country or region). Varian and Choi (2009) emphasize that Google Trends data do not report the raw level of queries for a given search term, but a query index which starts with the query share: the total query volume for search term in a given geographic region divided by the total number of queries in that region at a point in time. The query share numbers are then normalized so that they start at 0 in January 1, 2004. Numbers at later dates indicated the percentage deviation from the query share on January 1, 2004 [14].

Search results are proportionate to the time and location of a query by the following process [14]:

- each data point is divided by the total searches of the geography and time range it represents to compare relative popularity. Otherwise, places with the most search volume would always be ranked highest,
- the resulting numbers are then scaled on a range of 0 to 100 based on a topic’s proportion to all searches on all topics,
- different regions that show the same search interest for a term don’t always have the same total search volumes.

The index value is based only on the share of search query volume. The total aggregated volume for a particular search query is obtained from a particular geographical area and scale ranges from 0 to 100, where 100 represents the top of the search or the highest possible
frequency and intensity of the search for a specific term that is being searched. The first step in the creating of the index is to calculate the ratio of new search queries and total search volume to get relative values. The values are then divided for each period with the highest relative value. The highest query number is assigned a value of 100, while the rest is divided proportionally. If the search query number is insufficient, the index value is zero.

Data that is excluded comprises [14]:

- searches made by very few people: Trends only shows data for popular terms, so search terms with low volume appear as “0”,
- duplicate searches; trends eliminates repeated searches from the same person over a short period of time,
- special characters; trends filters out queries with apostrophes and other special characters.

A number of studies have utilized search data to make predictions in various fields. Jun et al. [4] assert that Google Trends has become such a popular source for big data research and applications since it provides an excellent platform for observing consumers’ information seeking activities and offers instant reflection of the needs, wants, demands and interests of its users. Moreover, they emphasize that Google Trends is easy to use because Google not only collects data but also provides a variety of options for comparison. Varian and Choi [15] were one of the first authors who have demonstrated the potential of including internet search history data in different predictive models. They showed that Google Trends data can be helpful in improving forecasts of the current level of activity for a number of different economic time series, including automobile sales, home sales, retail sales, and travel behavior. Moreover, Google search data have been used as measure of investor attention. For instance, Da et al. [16] proposed a direct measure of investor attention using search frequency from Google that captures investor attention in a more timely fashion and can be helpful in predicting stock prices. A comprehensive analysis of the trends in research studies in the past decade which have utilized Google Trends, as a new source of big data, together with an overview of the studies that have used Google Trends can be found in [4].

RELATED RESEARCH

Given the current media attention focused on cryptocurrencies and the ability of the Google Trends online tool to explore the search for a particular key word, it is of interest to explore search terms related to cryptocurrencies, and especially Bitcoin. Such researches are not a novelty and authors have advocated the use of web search volume data to build prediction models. Liu and Tsyvinski [17] showed that cryptocurrencies have no exposure to most common stock markets and macroeconomic factors but in contrast, cryptocurrencies returns can be predicted by factors that are specific to cryptocurrency markets. They determined a strong time-series momentum effect and that proxies for investor attention strongly forecast cryptocurrency return. They constructed the deviation from Google searches for the word “Bitcoin”, “Ripple” and “Ethereum” in a given week compared to the average of those for the preceding weeks and showed that Google search volume can predict the future price movements. Kristoufek [18] studied the relationship between Bitcoin and search queries on Google Trends and Wikipedia, showing that search queries and prices are related with a pronounced asymmetry between the effect of an increased interest in the currency while being above or below its trend value. Matta et al. [19] compared trends of Bitcoin price and search queries on Google Trends, volume of tweets and particularly with those that express a positive sentiment. They found significant cross correlation values, especially between Bitcoin price and Google Trends data. The same group of authors [20] studied the existing relationship between Bitcoin’s trading volumes and the query volumes of Google search engine. They
achieved significant cross correlation values, demonstrating search volumes power to anticipate trading volumes of Bitcoin currency. Urquhart [21] employed Google trends as a proxy for investor attention and applied Granger causality tests to the data from period 2010 to 2017 restricted to the area of USA. The author concluded that previous day volume and volatility are significant drivers of attention of Bitcoin only in a subsample of data and that investors are attracted to Bitcoin after large increases in volatility and trading volume.

METHODS AND PROCEDURES

The analysis has been performed within a timeframe of 200 weeks. Weekly data have been collected in the period from the beginning of 2014 till the end of October 2017. The analysis was performed on log-transformed data, rather than the original series. Toda-Yamamoto [6] procedure was performed. According to Toda and Yamamoto [6], if one of the time series used in analysis is non-stationary, the model can be estimated with variables in levels, but an extra lag of integration must be added. That extra lag is later ignored by conducting the Wald test, where the test statistics follows the usual asymptotic χ² distribution under the null hypothesis.

As already stated, the Toda and Yamamoto approach refers to causality testing in the presence of nonstationary variables. They propose to estimate an augmented VAR model to correct for the observed unit roots. In the bivariate case considered in this article, the VAR setup has the following form [6, 22]:

\[
x_t = a_1 + \sum_{i=1}^{p+1} \beta_{1,i} x_{t-i} + \sum_{j=1}^{p+1} \gamma_{1,j} y_{t-j} + \epsilon_{1,t},
\]

\[
y_t = a_2 + \sum_{i=1}^{p+1} \beta_{2,i} y_{t-i} + \sum_{j=1}^{p+1} \gamma_{2,j} x_{t-j} + \epsilon_{2,t},
\]

where \( p \) is the optimal lag order chosen by information criteria (e.g. AIC, SC, HQ), \( a_1 \) and \( a_2 \) are constant terms, \( \beta_{1,i} \) and \( \beta_{2,j} \) are autoregressive parameters, while \( \epsilon_1 \) and \( \epsilon_2 \) are white-noise (mutually uncorrelated) error terms. Furthermore, \( x_t \) represents the logarithm of search volume of keyword “bitcoin”, while \( y_t \) represents the logarithm of trading volume of bitcoin (BTC) (in HRK equivalent) on the online market LocalBitcoins. Generally, if \( d \) is the maximum order of integration of the observed time series, the VAR setup is of the following form:

\[
x_t = a_1 + \sum_{i=1}^{p+d} \beta_{1,i} x_{t-i} + \sum_{j=1}^{p+d} \gamma_{1,j} y_{t-j} + \epsilon_{1,t},
\]

\[
y_t = a_2 + \sum_{i=1}^{p+d} \beta_{2,i} y_{t-i} + \sum_{j=1}^{p+d} \gamma_{2,j} x_{t-j} + \epsilon_{2,t},
\]

The causality testing procedure within the Toda and Yamamoto approach comes down to testing the following null hypotheses:

\( H_0: y \) does not Granger cause \( x \) (\( y_{1,1} = y_{1,2} = \cdots = y_{1,p} = 0 \)),

\( H_0: x \) does not Granger cause \( y \) (\( y_{2,1} = y_{2,2} = \cdots = y_{2,p} = 0 \)).

RESULTS

The log-transformed time series are presented in Figure 1. The Bitcoin trading volume data are provided from LocalBitcoins trading site where people from different countries can exchange their local currency to bitcoins [23].

Stationarity and the maximum level of integration of the Bitcoin volume and Google Trend data was tested for each time series with Augmented Dickey-Fuller (ADF) test and Kwiatkowski-Phillips-Schmidt-Shin test (KPSS). The combination of these two tests is convenient since they have opposite null-hypotheses. The results suggest that Bitcoin volume is non-stationary (\( ADF = -2.19, p = 0.5; KPSS = 3.79, p < 0.01 \)), while the analysis of the first
Figure 1. Bitcoin trading volume (left) and search queries (right) evolution.

First differences suggests that the process of first differences is a stationary process (KPSS = 0.04675; \( p > 0.1 \); ADF = –7.1031; \( p < 0.01 \)), so we can conclude that Bitcoin contains the unit-root. Also, results suggest that Google Trend is non-stationary (ADF = –3.1439; \( p = 0.099 \); KPSS = 0.96, \( p > 0.01 \)), while the process of first differences is stationary (ADF = –7.6428; \( p = 0.01 \); KPSS = 0.0641; \( p > 0.1 \)). Consequently, Google Trend contains the unit-root.

Appropriate lag order of the VAR model was determined using lag-length selection criteria. Several measures were used to determine the appropriate lag order of the VAR model: Akaike Information Criterion (AIC), Schwarz Information Criterion (SC), Final Prediction Error (FPE) and Hannan-Quinn (HQ) Information Criteria. According to the SC and HQ criteria, the optimal lag order was \( k = 3 \), while the AIC, LR and FPE criteria suggested \( k = 5 \). VAR was estimated with lag order \( k = 3 \) and \( k = 5 \). Results of the autocorrelation error tests for the VAR model with \( k = 3 \) suggested the existence of autocorrelation errors, so the model with the lag order \( k = 5 \) was analysed. The results are summarized in Table 1.

The results of the autocorrelation analysis of the residuals indicated that there is no problem of autocorrelation of residuals. Also, the model satisfies the stability conditions.

Table 1. Lag-length selection criteria.

<table>
<thead>
<tr>
<th>Lag</th>
<th>Log L</th>
<th>LR</th>
<th>FPE</th>
<th>AIC</th>
<th>SC</th>
<th>HQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>–414,0024</td>
<td>NA</td>
<td>0.408968</td>
<td>4.781636</td>
<td>4.817947</td>
<td>4.796366</td>
</tr>
<tr>
<td>1</td>
<td>–290,1860</td>
<td>243,3632</td>
<td>0.103177</td>
<td>3.404437</td>
<td>3.513370</td>
<td>3.448627</td>
</tr>
<tr>
<td>2</td>
<td>–264,7448</td>
<td>49,42027</td>
<td>0.080642</td>
<td>3.157986</td>
<td>3.339541</td>
<td>3.231636</td>
</tr>
<tr>
<td>3</td>
<td>–250,3882</td>
<td>27,55808</td>
<td>0.071596</td>
<td>3.038945</td>
<td>3.293122*</td>
<td>3.142055*</td>
</tr>
<tr>
<td>4</td>
<td>–244,2613</td>
<td>11,62005</td>
<td>0.069873</td>
<td>3.014497</td>
<td>3.341296</td>
<td>3.147067</td>
</tr>
<tr>
<td>5</td>
<td>–238,4114</td>
<td>10,96001*</td>
<td>0.068414*</td>
<td>2.993235*</td>
<td>3.392656</td>
<td>3.155265</td>
</tr>
<tr>
<td>6</td>
<td>–238,1010</td>
<td>0.574545</td>
<td>0.071393</td>
<td>3.035643</td>
<td>3.507686</td>
<td>3.227133</td>
</tr>
<tr>
<td>7</td>
<td>–233,4411</td>
<td>8.516240</td>
<td>0.070875</td>
<td>3.028059</td>
<td>3.572724</td>
<td>3.249009</td>
</tr>
<tr>
<td>8</td>
<td>–229,9509</td>
<td>6.298439</td>
<td>0.071320</td>
<td>3.033919</td>
<td>3.651205</td>
<td>3.284328</td>
</tr>
</tbody>
</table>

*selected lag order by the criterion
The results for the causality testing procedure within the Toda and Yamamoto approach are summarised in Table 2.

**Table 2. Causality test results.**

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>$\chi^2$</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y$</td>
<td>4,4245</td>
<td>0.4900</td>
</tr>
<tr>
<td>$x$</td>
<td>18,8567</td>
<td>0.0020</td>
</tr>
</tbody>
</table>

Based on the causality test results it can be concluded that $x$ Granger causes $y$, but not vice-versa.

**CONCLUSION AND DISCUSSION**

Cryptocurrencies, such as Bitcoin, Ethereum and Ripple, can be regarded as a new asset class, a fully digital, sui-generis financial instruments but allocating capital into cryptocurrencies remains in the domain of pure speculation due to their strong volatility [24]. Therefore, due to their rapidly increasing and very volatile exchange rate, cryptocurrencies have been a lightning rod of interest for millions of people. Today bitcoin is one of the most trending topics on search engines and social media. In recent years, the availability and the timeliness of internet search data have encouraged researchers from various fields to employ these kind of big data sources in order to build different prediction models.

Research have revealed that Google search data can be used as a proxy for investor attention in different markets. Moreover, search data are able to capture investor attention in a more timely fashion and thus can be helpful in building prediction models. This is also especially beneficial for cryptocurrency market which is highly influenced by news, technological development, various social and government factors and factors that are specific to cryptocurrency markets.

The purpose of this study was to examine whether the information extracted by web search media could be helpful and used by investment professionals in Bitcoin limited to the area of Republic of Croatia. We analyzed query volume search of “bitcoin” keyword on the online platform Google Trends and the volume of Bitcoin trading in the online marketplace LocalBitcoins, restricted to the area of the Republic of Croatia within a timeframe of 200 weeks. Due to non-stationary variables included in the analysis, causality was tested by applying Toda and Yamamoto approach. The results confirmed that query volumes of Google search engine have a significant causal effect on Bitcoin trading volume. Similar results were reported by various research, using data not restricted to a specific area.

Moreover, information available on social media satisfies one of the basic principles of nowcasting – promptly exploitation of information. This kind of information can be relevant for building prediction models in cryptocurrency market and already has been used as a strong measure of investor attention. Different forms of social media, such as discussion forums, apps and web-sites, can provide a goldmine of information. For instance, user comments and replies in online cryptocurrency communities proved to affect the number of transactions among users [25] while it is demonstrated that the number of tweets from Twitter is a significant driver of next day trading volume and realized volatility [26]. Due to the fact that research have already shown that social media signals can be used as a stronger measure of investor attention, we believe that incorporating data collected from social media could be helpful in building prediction models. Thus, today, challenges grow from not only using internet search data for prediction purposes but also combining them with other sources of big data. Being accurate, able to collect big data and releasing reliable information for free, we believe that Google Trends will continue to have a growing impact in the future.
REFERENCES

Scientific Reports 3, No. 1684, 2013, 
http://dx.doi.org/10.1038/srep01684.

CEPR Discussion Paper No. DP7883. 2010, 
https://ssrn.com/abstract=1640405, accessed 15th October 2019,

Working Paper No. 1564, ECB, 2013, 
http://dx.doi.org/10.1016/B978-0-444-53683-9.00004-9,

Technological Forecasting and Social Change 130, 69-87, 2018, 
http://dx.doi.org/10.1016/j.techfore.2017.11.009,

Working Paper No. 717, Banque de France, 2019, 
http://dx.doi.org/10.2139/ssrn.3370917,

Journal of Econometrics 66(1-2), 225-250, 1995, 
http://dx.doi.org/10.1016/0304-4076(94)01616-8,

Zbornik sveučilišta Libertas 1-2(1-2), 191-210, 2017,

Zbornik Veleučilišta u Rijeci 3(1), 145-158, 2015,

IEEE Symposium on Security and Privacy. IEEE, San Jose, pp.104-121, 2015, 
http://dx.doi.org/10.1109/SP.2015.14,

https://bitcoin.org, accessed 12th February 2018,

Electron Markets 28(3), 321-338, 2018, 
http://dx.doi.org/10.1007/s12525-018-0308-3,

PLoS ONE 10(4), No. 0123923, 2015, 
http://dx.doi.org/10.1371/journal.pone.0123923,

http://dx.doi.org/10.1016/j.intfin.2017.11.001,

https://support.google.com/trends, accessed 18th February 2018,

Google Research Blog, 2009, 
http://dx.doi.org/10.2139/ssrn.1659302,
What can Google tell us about Bitcoin trading volume in Croatia? Evidence from the online …


