DVM Final assignment: Bitcoin prices and search interest in Google

Albert Xavier Lopez Barrantes

19th December 2018

Abstract

Almost one year ago Bitcoin prices peaked at 20.000$ and from today's perspective it's easy to visualize the bubble around cryptocurrencies. However, from an economic point of view and specially for those who like behavioral finance, I found an interesting correlation between Bitcoin prices and the search interest on "Bitcoin" in Google around the world for the last 2 years. I got the Search interest data from Google Trends and it can be a very interesting relation to study when analyzing massive trends on financial assets or any other sentiment analysis.

1. Introduction

For this final assignment I created a dataset from scratch from a topic a year ago was very trendy inside and outside the financial sector, Bitcoin. This was the first cryptocurrency and the first design was published in the paper "Bitcoin: A peer-to-peer Electronic Cash System" by Satoshi Nakamoto:

> *"A purely peer-to-peer version of electronic cash would allow online payments to be sent directly from one party to another without going through a financial institution. Digital signatures provide part of the solution, but the main benefits are lost if a trusted third party is still required to prevent double-spending. We propose a solution to the double-spending problem using a peer-to-peer network. The network timestamps transactions by hashing them into an ongoing chain of hash-based proof-of-work, forming a record that cannot be changed without redoing the proof-of-work. The longest chain not only serves as proof of the sequence of events witnessed, but proof that it came from the largest pool of CPU power. As long as a majority of CPU power is controlled by nodes that are not cooperating to attack the network, they'll generate the longest chain and outpace attackers. The network itself requires minimal structure. Messages are broadcast on a best effort basis, and nodes can leave and rejoin the network at will, accepting the longest proof-of-work chain as proof of what happened while they were gone."*

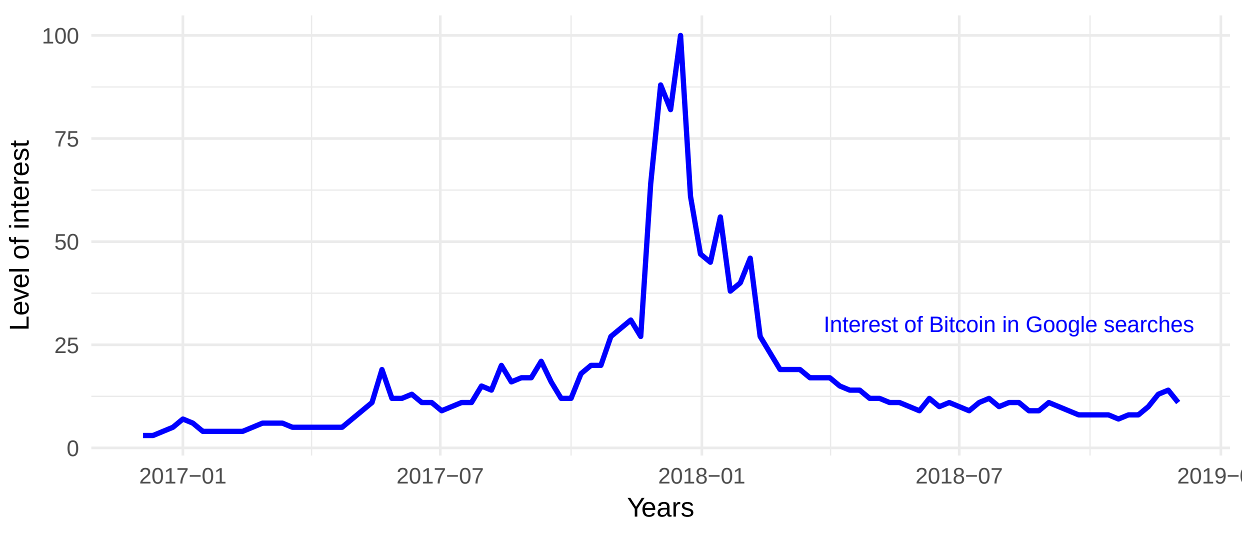
A currency to perform transactions between individuals without the need of a traditional financial institution involved in the process generated a huge expectation around the world rising Bitcoin prices to a peak of 20.000$ in January 2018. However, due to the low number of transactions per second cryptocurrencies are able to support and the non-recognition as a currency by most companies and governments, caused a big drop in the price to the current 3.000 dolars price. Coming from an economic background, this is what we call a bubble, and it's nothing more than a huge increase on expectations on a short period of time, often coming from the irrational decisions we do as human beings.

Thinking about it, here we can see two things. First, we can observate a classical financial bubble on a specific asset which is observable on the Bitcoin price time series (I got the weekly closing prices from a cryptocurrency brokerage firm):

Second I started to think on a possible indicator to visualize the behaviour of society towards Bitcoin. Nowadays and thanks to Google Trends we can know the concerns of society about a specific topic or issue. Not only that, this tools gives us the interest on a topic from a specific part of the world and the time frame we desire. Since Bitcoin it's been a global topic I got the global interest on the word "Bitcoin" around the world for the last 3 years. Specifically, this time series I got is defined by google like this:

> *"Google Trends is a search trends feature that shows how frequently a given search term is entered into Google's search engine relative to the site's total search volume over a given period of time. A value of 100 indicates the maximum popularity of a term, meanwhile values around 50 mean half the interest related to the maximum value and 0 not enough data from that term."*

In other words, the data you get using this tool is a standrized parameter of google searches on the word "Bitcoin" and the visualization of this data is the following:



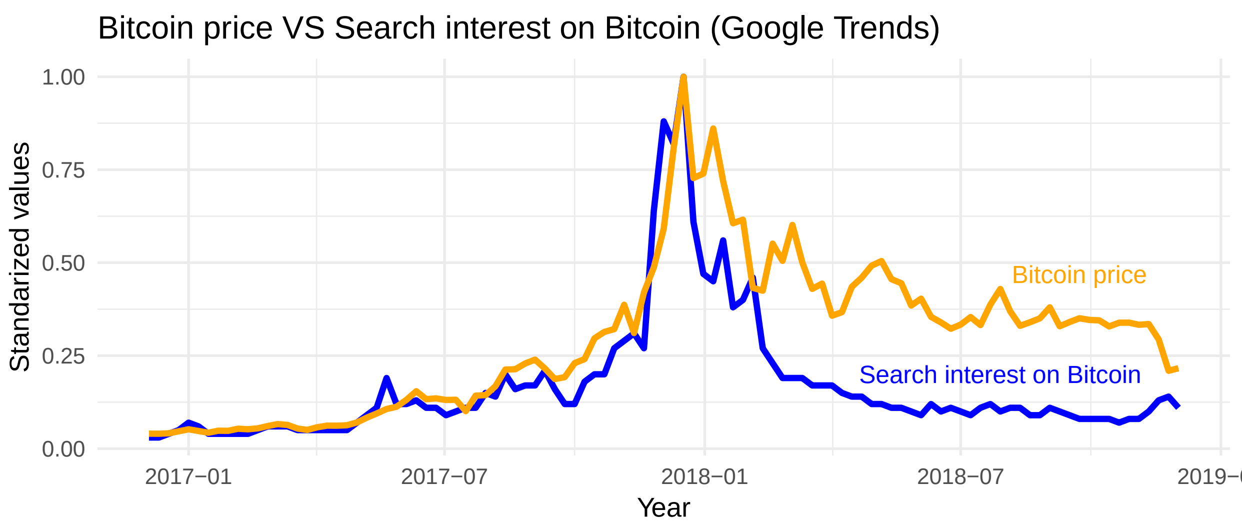
So once we visualize price and interest time series a significant set of questions arise. Are they correlated? Which significance levels do we get? If correlated, could we predict Bitcoin prices using Google Trends as a sentiment tool towards this digital currency? All of them are reasonable and interesting questions to prove from a satistical point of view, and I will try to cover most of them along this paper.

2. Statistic Tests

In this part I will go furher in the study of both time series studying correlations, autocorrelations, randomness and so on.

2.1 Correlation tests

At first sight, the most interesting thing to explore is to check if there is some kind of correlation between Bitcoin prices and the search interest on Bitcoin provided by Google Trends:



So let's run some correlation methods between both variables to check if there is correlation or not. To do so I will be using the cor.test() function from R changing between some of the possible methods. First I'm running a Pearson correlation test with a null hypothesis of no correlation, where I get the following results:

```
##
##  Pearson's product-moment correlation
##
## data:  bitcoin$price.USD. and bitcoin$interest
## t = 11.168, df = 103, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6390504 0.8159899
## sample estimates:
##       cor
## 0.7400692
```

I get a p-value of 2.2e-16 which is less than the significance level of alpha = 0.05. So we can say Bitcoin prices and the level of interest showed by Google Trends is significantly correlted with a correlation coefficient of 0.74 and a p-valye of 2.2e-16. Second, I will perform Kendall rank correlation test under the same null hypothesis. This test is used to estimate a rank-based measure of association.

```
##
##  Kendall's rank correlation tau
##
## data:  bitcoin$price.USD. and bitcoin$interest
## z = 7.9846, p-value = 1.41e-15
## alternative hypothesis: true tau is not equal to 0
## sample estimates:
```

```
##       tau
## 0.5392618
```

We get a Tau correlation coefficient of 0.539 with a p-value under 0.05 so we can say it's significant. Third, I'm running a Spearman test to estimate a rank-based measure of association and with the same null hypothesis as in the previous two tests:

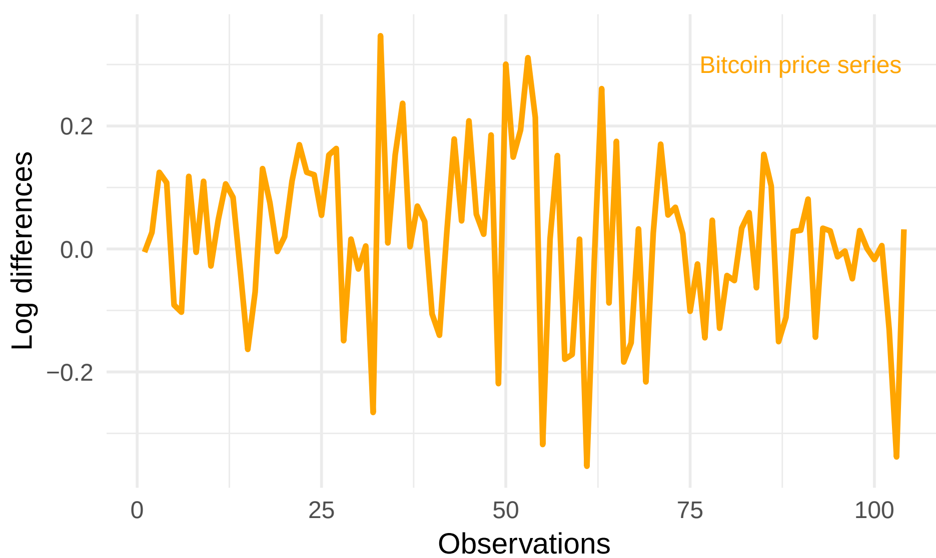
```
##
##  Spearman's rank correlation rho
##
## data:  bitcoin$price.USD. and bitcoin$interest
## S = 59257, p-value = 2.644e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.6928441
```

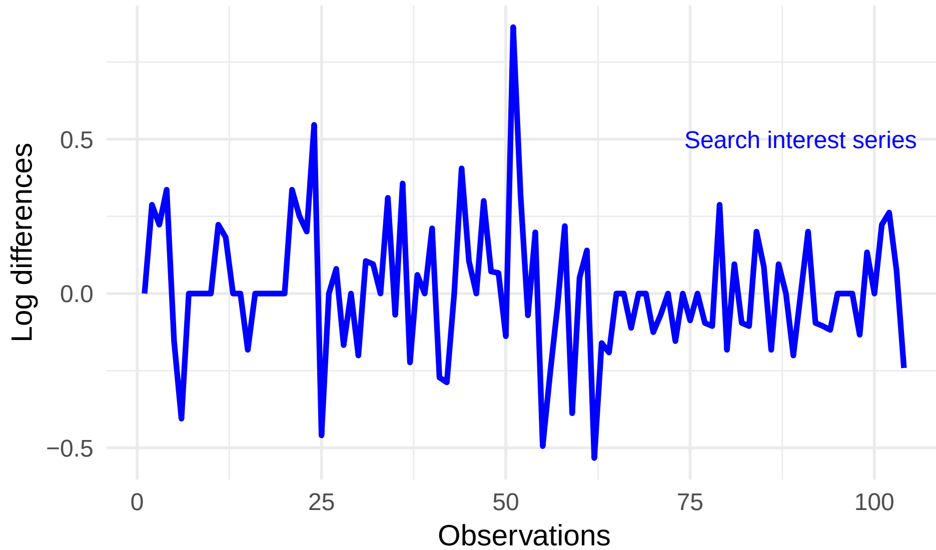
We get a rho coefficient of 0.6928 with a p-value = 2.644e-16 which is very significant, and again indicating a positive correlation between both time series. To sum up, results from three correlation methods gave us some evidence in favor of a positive correlation between Bitcoin price and the interest in google searches on Bitcoin.

Method	Corr_coefficient	p_value
Pearson	0.7400692	2.2e-16
Kendall	0.5392618	1.41e-15
Spearman	0.6928441	2.644e-16

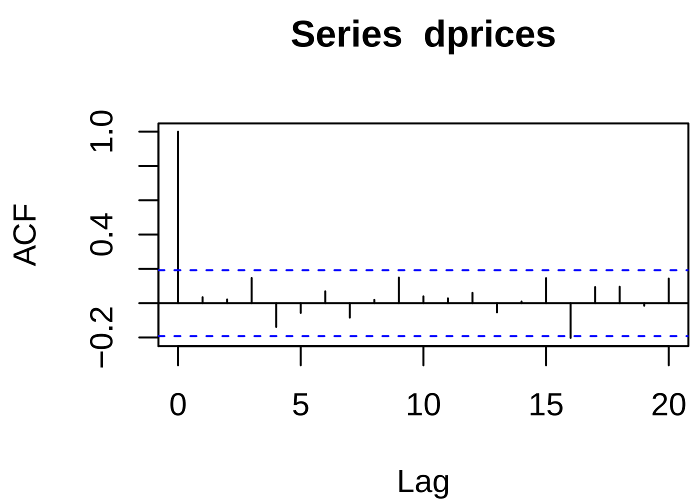
2.2 Time series analysis

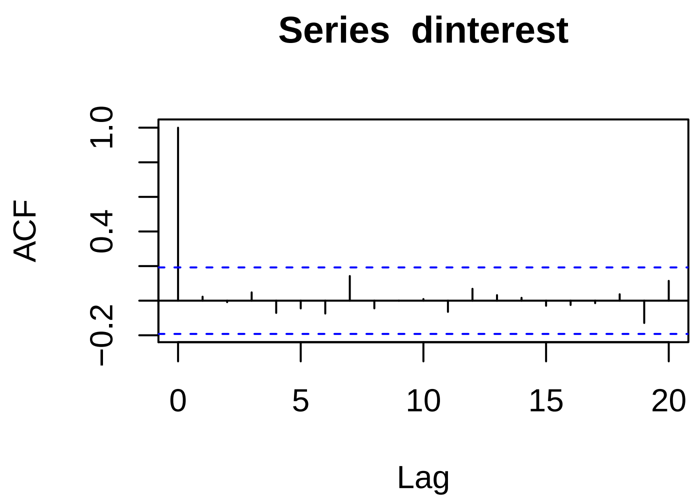
Almost all time series related to market prices have common characteristics that have to be treated before working on them. On of the most important is **heteroscedasticity**, which means a non constant variance. So the first step here is to apply differences or log differences in order to get **homoscedastic** properties in our time series:





Next step is to plot the ACF to check the existence of autocorrelations inside time series and detect possible autocorrelations with past observations:





In the first ACF we can see an interesting thing in Bitcoin price series. In the ACF we can see what seems a significant correlation with the 16th lag. I performed an permutation test to prove the significance level of this autocorrelation in the 16th lag:

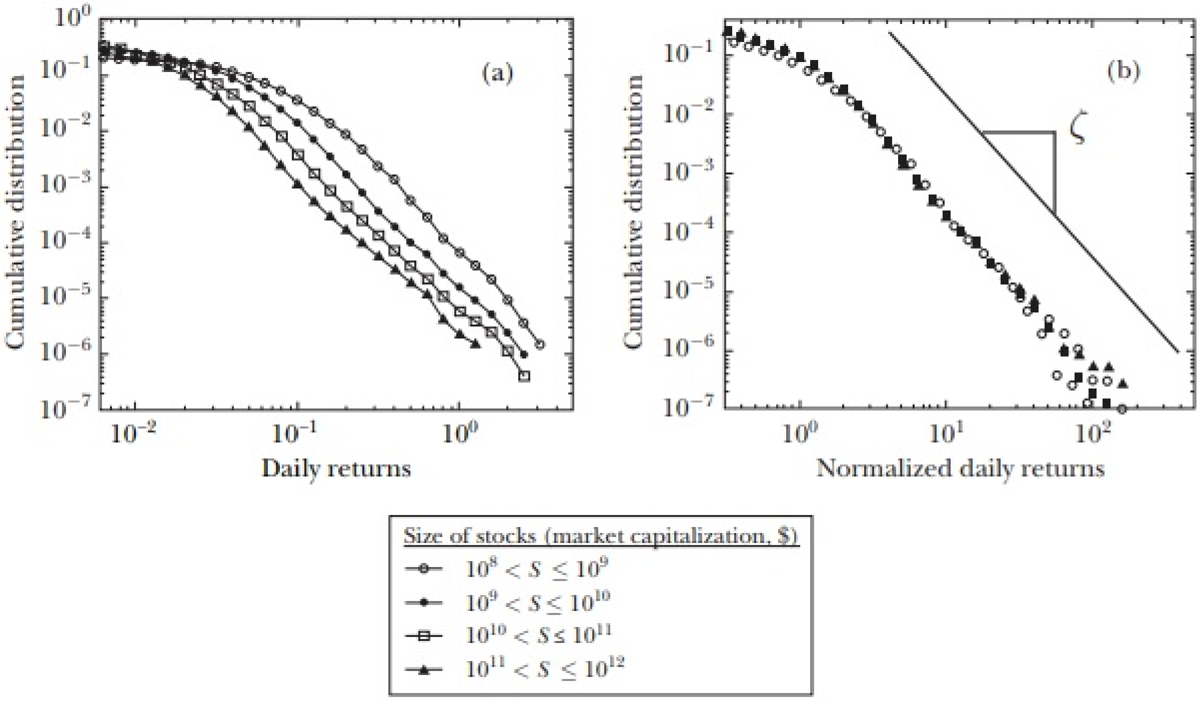
```
nr=10000 #number of rearrangements to be examined
st=numeric(nr)
sttrue=acf(dprices,plot=F)$acf[17]
n=length(dprices)
for (i in 1:nr){
  d= sample (dprices,n)
  st[i]<-acf(d,plot=F)$acf[17]}
length(st[st >=sttrue])/nr
```

```
## [1] 0.9849
```

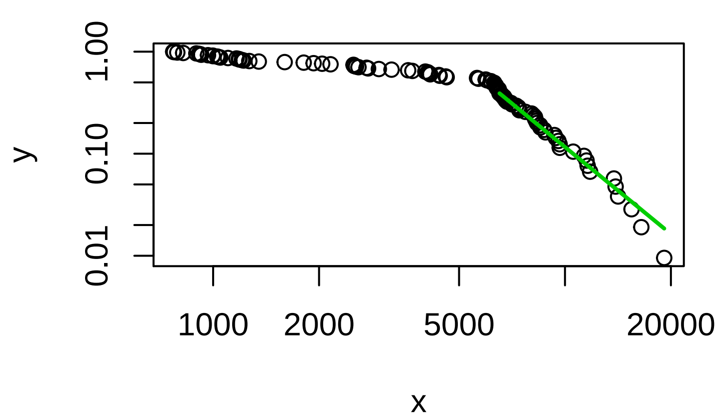
I get a p-value > 0.05 so we can conclude the 16th lag is not significant.

Finally, I wanted to check the Taylor Power Law that we saw in the last session, and check if Bitcoin prices follow the same path. In order to check that I made some reasearch on how it's been applyied in the past to other financial datasets and found a paper called "Power Laws in Economics" by Xavier Gabaix where he checks multiple financial time series checking if the power law also appears in there. Interestingly, he found that market returns time series are consistent with a "cubic" law. That means, the slope is close to 3 reflecting the "cubic law" when plotting the cumulative distribution against normalized daily returns. Here are the figures he got on his research:

Figure 4
Cumulative Distribution of Daily Stock Market Returns for Different Sizes of Stocks



Trying to apply the same process with the package poweRlaw() I get something similar. However, given the fact my database is small, results from Bitcoin prices have to be considered carefully.



I would like to have more time to work a bit more on this last part, since it seems to be following the same path as results from professor Xavier Gabaix.

Conclusions

Creating this dataset from scratch has helped me to understand a bit more some concepts I wanted to check from behavioral finance and some tools provided in this lectures have been very helpful.
First of all I worked on financial time series to check if there was any kind of correlation between Bitcoin prices and the search interest on them in Google and found a positive correlation between them. The intuition behind this positive correlation is quite stright forward. If supply is relatively fixed in the long-term then demand should be the largest single contributor to Bitcoin prices (given there are no actual quarterly earnings or interest rates associated with Bitcoin). In this case demand, or at least the interest from the demand side, can be observed from the serach interest in the biggest search engine of internet, Google. So finding a positive correlation between both time series seems pretty reasonable and the different correlation tests agree on the hypothesis with a high level of confidence. Second I made a typical time series analysis, this time trying to get any significant autocorrelation in each of the time series. To do so I performed some ACF functions and even in on one case it seem to be something, a deep analysis using a permutation test denied the hypothesis of autocorrelation with past information. Third and final, I tried to see if Bitcoin prices followed the power law, seen previously by other researchers in other asset prices time series and found some signs that they might follow as well this typical "Cubic Law". However, given the lag of knowledge on the field and time I couldn't prove my results.

References

Gabaix, Xavier. 2016. *"Power Laws in Economics: An Introduction."* Journal of Economic Perspectives, 30 (1): 185-206.

Nakamoto, Satoshi. 2008. *"Bitcoin: A peer-to-peer electronic cash system."* Consulted, 1:2012.