University of St.Gallen

School of Management, Economics, Law, Social Schiences and International Affairs

The effect of Twitter activity on Bitcoin price

Documentation

Software Engineering for Economists
(7,610,1.00)

Dimitrios Koumnakes - 10-613-370
Severin Kranz - 13-606-355
Joël Sonderegger - 11-495-488
Alen Stepic - 11-475-258
Chi Xu - XX-XXX-XXX

Fall Term 2017

Supervisor
Prof. Dr. Philipp ZAHN
Department of Economics

December 27, 2017

Abstract

(Insert text)

Contents

List of Figures

1 Introduction

In the academic environment accountability and reproducibility is important. However, the publishing process of papers and journals seem to be outdated. New ways of data collection and data processing exist by using computational economics. The usage of algorithms can increase effectiveness and efficiency. Hence, much lager data sets can be proceeded. However this creates also new problems regarding to traceability and reproducibility. Academic paper, which were often cited exist, where the initial computation is not reproducible. Furthermore, some academic paper even contain errors. Replicating data or existing results do not provide any new knowledge at all. Nevertheless, the ability to reproduce increases trustworthiness and indicate the quality of the conducted work. This explains why reproduction is of great relevance.

1.1 Goal of the paper

The goal of this documentation is the provision of a description. This description should enable the reader to reproduce the results discussed in the separate paper. Thus, it contains an explanation how the input data have been gathered, stored, aggregated and analysed. In other words, the input data, the model core, the model parameters and the applied math program are explained.

1.2 Methodology

This documentation consists out of four chapters. The first chapter contains a short introduction and provides the reader with an overview about the topic. Furthermore it points out the relevance of documentation. The second chapter discusses the input data. This includes the process of gathering and storing twitter tweets as well as the gathering of the bitcoin price data. The third chapter discusses how the data is aggregated by pointing out the core model and its parameters. Finally the fifth chapter discusses how the analysis has been conducted.

1.3 Scope

The scope of the documentation is the provision of an overview about the different steps which have been conducted to obtain the results in the paper. It does not contain any discussions about the results of the separate paper. It is not a deep description of the code as the code itself as the code is documented separately. Nevertheless, important lines of code are discussed.

2   Data Collection

Here, we provide a detailed description of how the data for the sequential analysis is gathered and stored. This includes two subsections the (1) tweets data and the (2) bitcoin price data.

2.1   Tweets Data

To collect the needed twitter data a python script has been written. Whit the python script real-time twitter data are streamed and stored by the usage of a raspberry pi.

2.1.1   Python Script

Twitter offers different Application Programming Interfaces (API) for collecting data. Based on twitter, the biggest an probably most relevant API's are (1) Ads API, (2) Filter realtime Tweets, (3) Search Tweets and (3)Direct Message API. (How to add a reference???) However, the timeframe for gathering data for free is limited to 7 day's. As python offers different libaries tweepy a open-source libary is been used for streaming the data. By using tweepy twitter can be reached in two way's (??) or by the streaming API. The latter has been used in contex of the assignment.

Stream - push then pull, easier - authentic Twitter Authentification - printscreens and how to proceed - pip install??

Twitter Streaming API ....

2.1.2   Hardware Setup

(Severin's Part)

2.2   Bitcoin Price Data

We wrote a Python script which collects Bitcoin price data as there was no preexisting data collection that satisfied our needs. The Bitcoin price is best expressed by the Bitcoin Price Index. The Bitcoin price index (BPI) is an index of the exchange rate between the Bitcoin (BTC) and the US dollar (USD) (**kristoufek2015main**). The objective of the script was to gather hourly Bitcoin Price Index Data for the time period in which we gather the Tweets data.

2.2.1 Execution

By executing the python script `CollectCryptocurrencyData.py` hourly data for the Bitcoin Price Index is retrieved.

```
$ python CollectCryptocurrencyData.py
```

2.2.2 Output

After successfully running the python script `CollectCryptocurrencyData.py` the file `bpi.csv` is generated in the folder `/data`. It is important to note that every execution of the script overwrites any existing `bpi.csv` file.

The file `bpi.csv` contains historical Bitcoin Price Index data for one month on an hourly basis. Each data point consists of the following parameters:

- time: Timestamp on an hourly basis
- average: Average price (in USD)
- high: Highest price (in USD)
- low: Lowest price (in USD)
- open: Opening price (in USD)

2.2.3 API: Bitcoinaverage.com

Bitcoinaverage.com offers a free API that provides historical Bitcoin Price Index data. The following requests delivers data for an per hour monthly sliding window.

```
https://apiv2.bitcoinaverage.com/indices/global/history/BTCUSD?period=monthly&?format=json
```

The time timestamps is UTC time.

3 Data Aggregation

(Joel's Part)

4 Data Analysis

(Dimitri's Part)