

# Interim Report

Radek Chramosil

WorldQuant University

E-mail: radek@keemail.me

**This document is the Interim Report for the Capstone Project. It shows the basic structure for the database that holds data the data from crypto-currency exchange Binance and some initial data analysis.**

## Introduction

The structure of the database is described below. Additionally, I provide more details about the data analyses so far carried out as well as some suggestion for trading possible trading strategies. Some plots for second data set are provided. The asset I worked with is **TUSDBTC** as quoted on *Binance Exchange*. The asset is Bitcoin quoted in USD. The access to the full exchange is free and level 2 data is available for every tick. The data was collected on second basis for one hour. The collection was done with 1 second frequency as opposed to per trade basis (I am still unsure how to handle data collection per trade basis). The data set is therefore some 3440 full market book hits; some 100 quotes for bid prices and some 100 quotes for ask prices.

## Database Structure

I have included the below tables in this submission. The tables can be created by SQL commands in the “create\_db.sql” file. The schema for the database is shown in Figure 1.

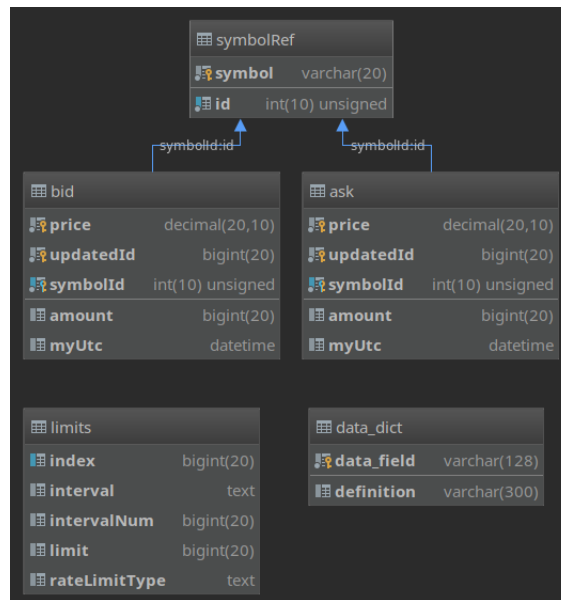


Figure 1: Schema

The tables can be dropped using the “rollback\_db.sql” file. This file provides a logical inverse to the create operation. Additionally, I built functionality that can create table on fly using Pandas and SQL Alchemy however the default behaviours are not very good as it fails to create foreign keys and so on.

## Tables

**symbolRef.** This is just a reference table for symbols. The symbols are strings, however it is more efficient to use integers for foreign keys and it is better to store integers than strings.

**bid and ask.** These two tables are identical from a data type point of view however they just store the bid and ask prices as well as my UTC time stamp and the server 'updatedId' flag. This is a large integer that gets changed each time there is a trade.

**limits.** . Basically, the exchange give users request or order limits in time intervals. We can make 1200 data request in a minute or 10 orders in a second and 100 000 orders in a day. If we are creating a real trading application we would have to consider these closely.

**data.dict.** This is manually created table and holds descriptions for the data fields in the database.

## Bid and Ask

Bid and ask prices are split into two tables. The tables are identical. The tables from Figure 2 and Figure 3 carry the data for the symbol and specific Exchange ID as given by *updateId* column. My time stamps are in the *myUtc* column representing the time on my computer.

price	amount	updatedId	myUtc	symbol
0.00020497	1806	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020503	147	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020504	6	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020511	9	48208950	2018-11-19 21:03:04	TUSDBTC
0.0002052	4453	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020538	4453	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020546	18528	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020556	4453	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020558	1194	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020576	34	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020577	15	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020586	5	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020587	5	48208950	2018-11-19 21:03:04	TUSDBTC

Figure 2: Ask Price

## Exchange API

It is important to handle the communication to the Exchange well. The class *DataEndPoint* hold all methods that can collect the data from the exchange as opposed to submit trades. I leave it

price	amount	updatedId	myUtc	symbol
0.00020471	504	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020468	183	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020466	281728	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020465	10	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020463	49	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020461	20	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020456	20	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020455	10	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020454	5	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020453	10	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020452	10	48208950	2018-11-19 21:03:04	TUSDBTC
0.0002045	6	48208950	2018-11-19 21:03:04	TUSDBTC
0.00020449	30	48208950	2018-11-19 21:03:04	TUSDBTC

Figure 3: Bid Price

to the reader to read over the comments for these methods to see what can be easily pulled out from the exchange.

## Data Analysis - First Look

I collected 1 hour of trading data on per second basis. Figure 4 shows cumulative bid and ask prices with quantity. The way how I think about this is it is supply and demand curves. We see that there is a lot of volatility over 1 minute interval. This is good as it means that we will be able to create strategies.

The problem in the Figure 4 is that we do not see much details. So, I plotted the same time interval but with close look at what was happening around the spread in Figure 5. Again we see the volatility in the price within one minute.

We also have have look the maximum bid and minimum ask prices and plot them over the 1 hour interval. This is plotted in Figure 6. This shows that we do have trends and turning points happing over time again. We should use the trading strategies that we learning for this and we can starting simulating the trade at bid and ask prices.

Finally, we check the spread for the two currencies as in Figure 7. We immediately see that that we could create mean reversion strategy for the spread. It may not be viable on profit versus fee basis however we could tests this.

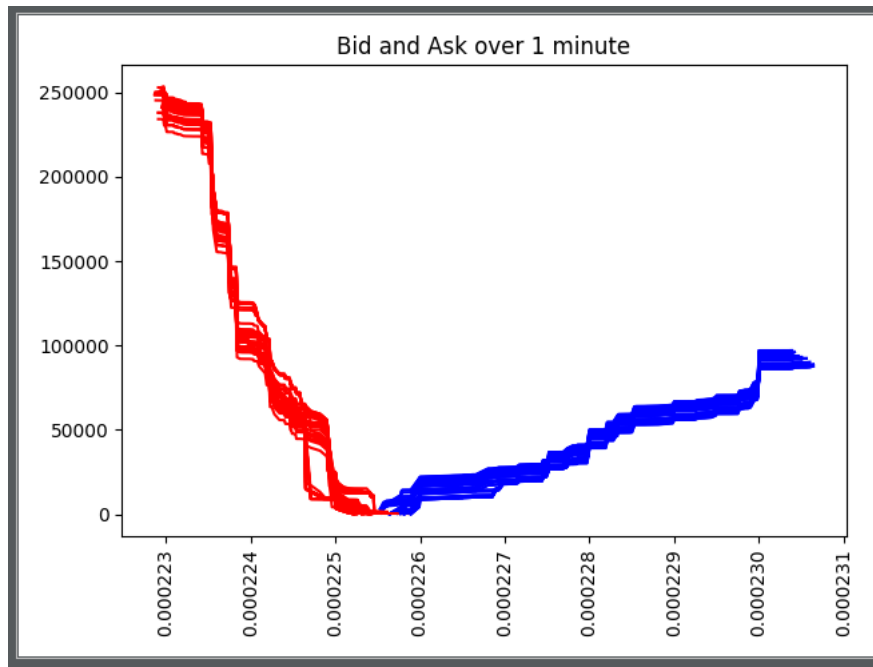


Figure 4: All bid and ask prices and quantities in 1 minute

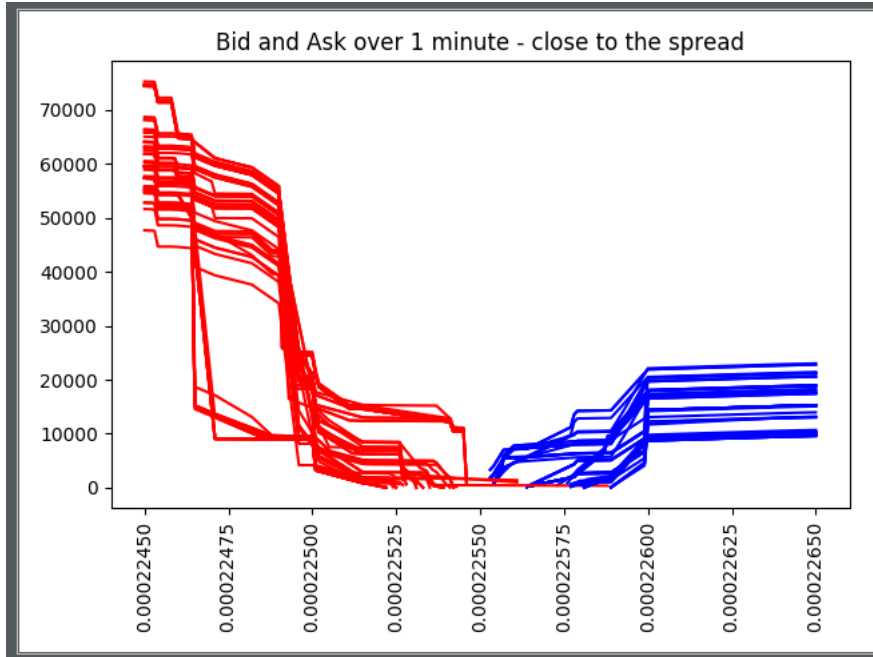


Figure 5: Detailed Price and quantity moments around the spread in 1 minute

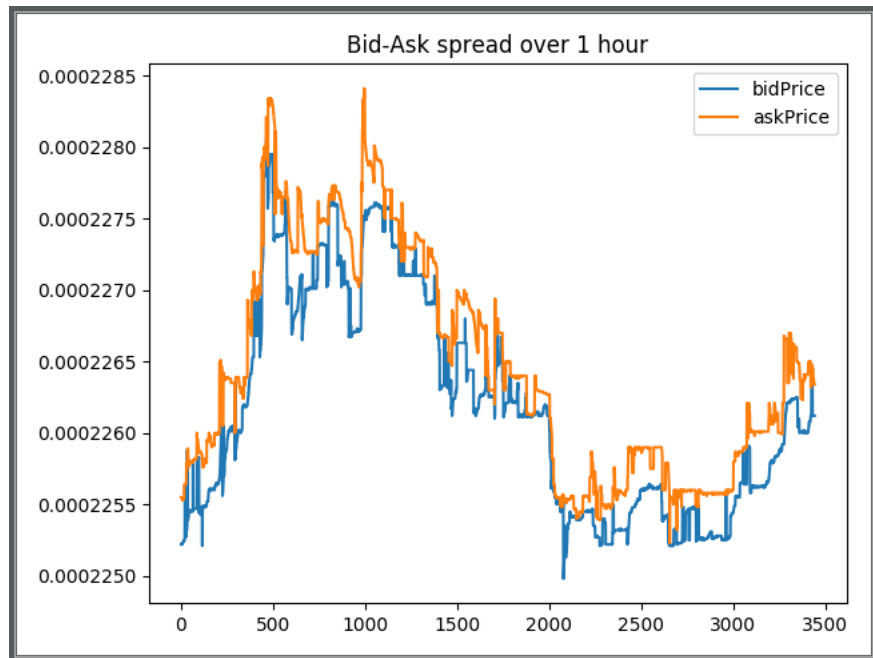


Figure 6: Max bid and min ask over 1 hour

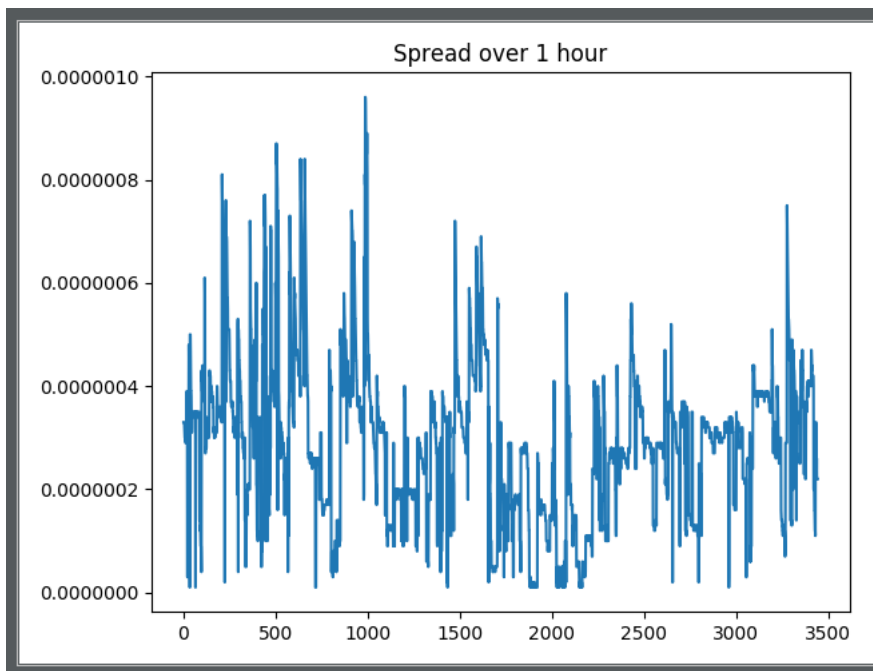


Figure 7: Max bid and min ask spread over 1 hour

## **Some Research - to be done**

Please advise how important this is. I focus on coding over last two weeks.

### **0.1 Electronic Trading Chain**

#### **0.1.1 Exchanges**

#### **0.1.2 Brokers**

#### **0.1.3 Algo. Trading engines**

#### **0.1.4 DMA Services**

### **0.2 Benefits and Pitfalls of DMA technologies**

#### **0.2.1 Automated Order Routing systems - AORs**

#### **0.2.2 Sponsored Access - SA**

#### **0.2.3 Direct access by non-intermediary market-members**

### **0.3 Risks in DMA**

### **0.4 Latency Types**

#### **0.4.1 Transmission Latency**

#### **0.4.2 Propagation Latency**

#### **0.4.3 Processing Latency**

### **0.5 Typical Order Messaging System**

## Summary

Looking at the data, it is easy to spot the same patterns as in the time frequencies however connecting to the exchange directly gives full details as to what is going in the market. We are not able to do this for equity or standard financial markets for free. Having a look at the data, we will be able to exactly model what is going in the market rather than second guess from some summary data such as OHLC from Yahoo. Going forwards, I will focus on creating and testing at least one trading strategy in the data.