**Questions Regarding the First Interview Task**

**Ziwei Jiang**

**Explain what you think causes the price discrepancy.**

The first thing I notices was the difference in trading quantity. The trading quantity of okex is significantly less than Bitmex. Thus, I assume that the price of okex would be more volatile, which was proved wrong in later data explorations. However, I think lower trading quantity will still affect the price in another way. Lower trading quantity could lead to higher commission fees, which would lead to a higher price level when the price is increasing.

**Describe the strategy or your method to arbitrage the opportunity.**

The general idea was to be maker on one side of the market and taker on the other side.

First, we need a ticker to determine which index to bid on and ask for. I simply used average bid price and average ask price here. When the average ask price of one index is significantly lower than the average bid price of the other index, an arbitrage opportunity appears.

Second, we need to determine at what time should we make the transaction. I think the important idea I introduced there is that I treat the time interval between the arbitrage opportunities appear is also a time series that could be described by ARCH model, which means the arbitrage opportunities happens in a mean reversal nature with changing volatility. The initial time interval series was generated by randomly sampling from the first 10000 time point which high frequency price data was recorded. You can also set the limit for time interval and minimum per-transaction profit to take limit of equipment and condition of transaction commission into consideration.

The philosophy here is that we don’t need to pursue the biggest possible profit per transaction, but we want to make sure that every transaction we made was profitable if the arbitrage opportunities happens as described in our model.

Last but not least, we would limit the scale of every transaction to 10% of our budget, since there could be some overlapping time intervals. Profit per transaction is accumulated on a daily basis.

**Describe the back test and the assumption you made on the back test.**

Given the scale of dataset, the back test was conducted on the first daily high frequency data file. The assumption made was described in previous question: the arbitrage opportunities happens in a mean reversal nature with changing volatility.

The gross profit was 2903143.70, third times as much as our initial investment, which is unrealistic. There should be some overfitting problem because I disregard the nature of price data itself. I didn’t make assumption about the price trend. Further back test with more data is needed.

**Describe how you have processed the data in order to conduct your back test.**

I used Python for all the data cleaning and data preprocessing. I converted the datetime column into timestamps to make it possible to compare two time point. No missing value encountered in my data preprocessing process since a statistical sampling method is applied to avoid missing values at a specified time point.