General Description:

You are working on the market tick data for two of A shares' symbol (we picked the two symbols randomly).

The format of the data is hdf5. This type of data can be opened and manipulated in python via h5py package. The columns of interest are mostly the bid and ask price, volume, and the total turnover etc.

Using the historical data, please make statistical study and discovery among one or more of the below questions, and write an analysis report with your results and observations. **We will be evaluating the process of the analysis rather than the absolute result of your model**, so please make your logical reasoning process as clear and solid as possible.

Note:

1. The market data may consists of the auction period, the regular trading period, the lunch break period. So please treat the data carefully.
2. **You do not need to finish or even start to work on all of the questions**, the questions are simply to give you some directions or ideas to think of, you only need to **choose one of the directions or sub directions** to work on, only work on more directions if you are truly interested or have more than enough time.
3. The transaction cost is generally 2bps for buying stock and 12bps for selling the stock. Bid offer spread can be found in market data.
4. Feel free to explore only one symbol, or both symbols, or the relationship between those symbols.

Questions:

1. Given the historical data, could you make any observations for market statistics? You can think along these lines but feel free to be more creative and go beyond these examples and directions:
   a) Are there any patterns in volume/volatility/number of trades/number of book orders?
   b) Are there any seasonality effect on the market? For example, any effect grouped by weekday, by month, or day of the month, etc.?
   c) How about some intra-day seasonality patterns? For example, it is well known that traded volumes in most intraday market follow an U shaped pattern, with market open being the most liquid time of the day, and gradually decaying afterwards
   d) How about some periods with relatively high return or low return on the stock, or periods with relatively low or high volatility, can you make any interesting observations about these periods?

2. Could you try to find features or models to forecast the direction of the stock?

a) You can work to choose any kind of prediction horizon, for example, 1 Min/5min/10min/1hour return

b) You can also only select and work on a specific period of the day. For example, you can try to make a model only to predict the expected return after auction period, or the expected return after lunch break, etc.

c) You can choose to work on any features that you think might be of interest. For example, is past return useful to predict future return? If so or if not, you can write your analysis process and result.

d) How do you evaluate the fitness of your features or models? Please explain your methodology and analysis carefully. You can choose your own method of designing the in sample and testing data during the 4-month historical data.

e) If you manage to find something that can be of predictive power, either a single feature or a more complex model, how would you construct a trade around it? Would your trade be sensitive to transaction cost or market bid offer spread?