**TABLES**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

1. ExpectedVwap_ReturnVwap_20090803_20090831.csv is the file that is merging expected vwap calculated from lob table with the vwap returns calculated from VWAP table. The lag is created for the 15 minutes.

2. ExpectedVwap_combination_20090803_20090831.csv is the file that contains the lob data and corresponding estimated maximum buy vwap, minimum buy vwap, maximum sell vwap, minimum sell vwap, and the combination records that come up with all the four vwaps.

3. ReturnVwap_20090803_20090831 is the file that includes the vwap data and 15 minutes return for the vwap table.

**ABOUT ExpectedVwap_ReturnVwap_20090803_20090831.csv**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

1. Expected AA buy and sell vwap is merged with all A, AA, GE, GOOG, MSFT data since there are no explicit requirement to isolate AA from others in merge step.

2. Since the 15 minutes return is required, the return at 9:30:00 am every day for every company is set to NA, because there are not trading information before that in the same date.

3. buy_max_vwap, buy_min_vwap, sell_max_vwap, sell_min_vwap are the columns for the estimated maximum buy vwap, minimum buy vwap, maximum sell vwap, minimum sell vwap that calculated from lob.

**ABOUT ExpectedVwap_combination_20090803_20090831.csv**

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

1. The definitions for buy_max_vwap, buy_min_vwap, sell_max_vwap, sell_min_vwap are the same as above, and buy_max_com, buy_min_com, sell_max_com, sell_min_com are the corresponding combinations of rows in lob table that come up with buy_max_vwap, buy_min_vwap, sell_max_vwap, sell_min_vwap. For example, in the following records:

| symbol | date | interval | buy_max_vwap | buy_min_vwap | buy_max_com | buy_min_com | sell_max_vwap | sell_min_vwap | sell_max_com | sell_min_com |
|--------|------|----------|--------------|--------------|-------------|-------------|---------------|---------------|--------------|--------------|
| AA | 8/3/2009 | 9:29:59 | 11.6 | 9.686 5 | 196& 236& 40 | 126&45& 236&11& 40 | 13.07 | 12.03 13 | 394& 384& 352 | 284& 273& 270 |

11.6 in column buy_max_vwap is calculated by vwap made up from price and size records in rows 196,236,40 in lob.sas7bdat which is recorded as 196&236&40 in buy_max_com column. Since the first rows in lob in csv is occupied by the names of columns, it should be calculated by price and size records in rows 196+1, 236+1, 40+1 in lob in csv format. And in the lob.csv, it means that:

$$\text{Size: } H(196 + 1) + H(236 + 1) + H(40 + 1) = 10000$$

$$\text{Vwap: } \frac{H(196+1)*E(196+1)+H(236+1)*E(236+1)+H(40+1)*E(40+1)}{H(196+1)+H(236+1)+H(40+1)} = 11.6$$

Note: In lob.csv, H is the column for size, and E is the column for price.

2. The calculation of buy and sell is achieved in following steps:

  1) Randomly select 20 records in the buy/sell for given date and time

  2) Randomly select 5 of the 20 records to traverse to find all possible combinations whose sizes are sum up to 10000 shares

  3) If the number of combinations that can be summed up to 10000 shares is less than 3, then go back to 1) to try again; else, find the minimum vwap in buy/sell as buy/sell_min_vwap, and the maximum vwap in buy/sell as buy/sell_min_vwap, and the corresponding combinations of the records are recorded as buy/sell_min_com and buy/sell_max_com

The details of the codes is in assignment.r and subset_sum.r

We assume weak efficient market hypothesis here, in which the behavior of the security price is random walk in the small time interval.