# CISC5352: Financial Data Analytics Homework (2)

# High-frequency trading data Basics (100 points)

The TAQ_JNJ_1004_1015_2010_trading_.csv has all transactions of JNJ stock from Oct 4, 2010 to Oct 15, 2010 (Note: this csv is not a very standard csv because not all data are separated in different columns)

1. Trim all transactions not in the normal trading window: 9:30 am -4:00 pm EST and write the transactions in the normal trading window in to a csv file: TAQ_JNJ_1004_1015_2010_trading_normal_hours.csv. We assume we only consider the transactions during normal trading windows for the following problems.

2. Visualize the first and last 5000 transactions from file:

   (a) TAQ_JNJ_1004_1015_2010_trading_normal_hours.csv

3. Let $t_i$ be the trading time in which $i^{th}$ transaction took place and $P_{t_i}$ be the transaction price at $t_i$, The price change in the time interval $\triangle t_i = t_i - t_{i-1}$ is $y_i = \triangle P_{t_i} = P_{t_i} - P_{t_{i-1}}$

   (a) Count the number of transitions with price change, i.e., the cardinality of the set $\Theta = \{i | y_i \neq 0\}$.

   (b) Visualize the price change $y_i$, $i = 1, 2, ...$ such that the y-axis is the range of price changes (e.g. -\$0.5 to \$0.5). (Note: your x-axis is the index of $i$)

   (c) Draw a histogram of the frequency of all transactions with a price change

4. Partition the trading time sequence into 5/10/30/60-minute intervals by picking the *mean* transaction price and volumes in each interval and compute the log-return (aka 'U sequence') and write it into a corresponding csv file, in addition to visualizing them.

   (a) TAQ_JNJ_1004_1015_2010_5_min_trading_unit.csv

   (b) TAQ_JNJ_1004_1015_2010_10_min_trading_unit.csv

   (c) TAQ_JNJ_1004_1015_2010_30_min_trading_unit.csv

(d) TAQ_JNJ_1004_1015_2010_60_min_trading_unit.csv

5. Compute "section volatility": $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} u_i^2 - \frac{1}{n(n-1)}(\sum_{i=1}^{n} u_i)^2}$ for each case, where $n$ is the number of transaction units in each case and $u_i$ is the $i^{th}$ item in the log return sequence

6. Compute "volume section standard deviation":

$$s_v = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n} v_i^2 - \frac{1}{n(n-1)}(\sum_{i=1}^{n} v_i)^2}$$

for each case, where $n$ is the number of transaction units in each case and $v_i$ is the $i^{th}$ item in the volume sequence.

7. Compute section *skewness* and *kurtosis* values for each case and draw your conclusions based on the definitions of skewness and kurtosis.

$$shewness(X) = \frac{n}{(n-1)(n-2)}\frac{\sum_{i=1}^{n}(u_i - \bar{u})^3}{s^3}$$

$$Kurtosis(X) = \frac{1}{n}\frac{\sum_{i=1}^{n}(u_i - \bar{u})^4}{s^4} - 3$$

**Repeat 1-7 for TAQ_CAT_Feb_2010_trading_.csv, which includes all transactions for caterpillar (CAT) stock in Feb 2010**

# Bid-ask spread

1. TAQ_CAT_QUOTE_0104_2010.csv includes all quote information for caterpillar (CAT) stock transaction on Jan 04, 2010.

2. Compute and visualize its bid-ask spread: $\delta_p = P_{offer} - P_{bid}$, which is an important information for market liquidity, for each transaction.

3. Partition the trading time sequence into 5/10/30-minute intervals by picking the *mean* bid and offer prices and their corresponding in each interval and write it into a corresponding csv file, in addition to visualizing them.

   (a) TAQ_CAT_QUOTE_0104_2010_5_min_trading_unit.csv
   (b) TAQ_CAT_QUOTE_0104_2010_10_min_trading_unit.csv
   (c) TAQ_CAT_QUOTE_0104_2010_30_min_trading_unit.csv

4. Compute the following entropy values for offer and bid volume as follows

   (a) $H_{offer} = -\sum_{i=1}^{N} p_i log_2 p_i$,

(b) $p_i = \frac{V_i^{(offer)}}{\sum_{k=1}^{N} V_i^{(offer)}}$, where $V_i^{(offer)}$ is the mean offer volume in the $i^{th}$ interval.

(c) $H_{bid} = -\sum_{i=1}^{N} q_i log_2 q_i$,

(d) $q_i = \frac{V_i^{(bid)}}{\sum_{k=1}^{N} V_i^{(bid)}}$, where $V_i^{(bid)}$ is the mean bid volume in the $i^{th}$ interval.

(e) Relative entropy for offer and bid volume $H(P||Q) = \sum_{i=1}^{N} p_i log_2 \frac{p_i}{q_i}$

(f) Note: $log_2 t$ should be automatically set as zero if $t = 0$. Similarly $log_2 \frac{2}{0} = log_2 \frac{0}{0} = 0$

# What should you turn in?

- 1. A folder contains your source files and related output.
- 2. Please name your folder as first-name_last-name_CISC5352_homework_2. For example, John_Smith_CISC5352_homework_2 if your name is John Smith.
- 3. Send the zipped file (.zip instead of ,rar) of your folder to Blackboard before 11:59 pm Dec 16, 2016