Sample Solution for HW1

This is the sample solutions of each question collected from and credited to these students. Reader: Siqi Wang (siw119@berkeley.edu)

Student: Haoyang Wang

Problem 1

(a)

LIMIT BUY ORDERS		LIMIT SELL ORDERS	
PRICE	Volume	Price	Volume
$49.75	500	$51.50	100
$49.50	800	$54.75	300
$49.25	500	$58.25	100
$49.00	200		
$48.50	600		

(b)

LIMIT BUY ORDERS		LIMIT SELL ORDERS	
PRICE	Volume	Price	Volume
$49.75	500	$51.50	100
$49.50	800	$54.75	300
$49.25	550	$58.25	100
$49.00	200		
$48.50	600		

(c)

LIMIT BUY ORDERS		LIMIT SELL ORDERS	
PRICE	Volume	Price	Volume
$50.00	200	$51.50	100
$49.75	500	$54.75	300
$49.50	800	$58.25	100
$49.25	550		
$49.00	200		
$48.50	600		

(d)

LIMIT BUY ORDERS		LIMIT SELL ORDERS	
PRICE	Volume	Price	Volume
$50.00	100	$51.50	100
$49.75	500	$54.75	300
$49.50	800	$58.25	100
$49.25	550		
$49.00	200		
$48.50	600		

(e)

LIMIT BUY ORDERS		LIMIT SELL ORDERS	
PRICE	Volume	Price	Volume
$50.00	100	$51.50	100
$49.75	500	$54.75	300
$49.50	800	$55.00	100
$49.25	550	$58.25	100
$49.00	200		
$48.50	600		

Student: Yutong Yang

Problem II

Suppose I play a game where I toss a fair coin over and over and record the values in a list ('H' for head, and 'T' for tail). I am interested in finding specific sequences in the list, such as 'HT', which would occur on the nth roll if the nth roll is 'T', and the $(n-1)$st roll is 'H'. Find the expected number of coin tosses until the first occurrence of sequence $'HHH'$.

Denote the event X as "the occurrence of sequence $'HHH'$".
Denote random variable F as the result of the first toss, we have $F = H$ or $F = T$.
Similarly, denote random variable S as the result of the second toss, we have $S = H$ or $S = T$.
Denote random variable R as the result of the third toss, we have $R = H$ or $R = T$.
The conditional expectation of X given F is denoted as $E(X|F)$.
By the law of conditional expectation, we have

$$\begin{aligned} &E[X] \\ &\quad = E[E[X|F]] \\ &\quad = p(F = H)E[X|F = H] + p(F = T)E[X|F = T] \\ &\quad = \frac{1}{2}E[X|F = H] + \frac{1}{2}(1 + E[X]) \end{aligned} \tag{1}$$

$E[X|F = H]$ is conditional on the result of the second toss, i.e. the random variable S.
Thus also by the law of conditional expectation, we have

$$\begin{aligned} &E[X|F = H] \\ &\quad = p(S = H)E[(X|F = H)|S = H] + p(S = T)E[(X|F = T)|S = T] \\ &\quad = \frac{1}{2}E[(X|F = H)|S = H] + \frac{1}{2}(2 + E[X]) \\ &\quad = \frac{1}{2} * 3 + \frac{1}{2} * (2 + E[X]) \end{aligned} \tag{2}$$

Combining equation (1) and (2), we may derive the expression of

$$\begin{aligned} &E[X] \\ &\quad = \frac{1}{2}(1 + E[X]) + \frac{1}{2}[\frac{1}{2}(2 + E[X]) + \frac{1}{2}[\frac{3}{2} + \frac{1}{2}(3 + E[X])]] \\ &\quad = 1 + \frac{3}{4} + (\frac{1}{2} + \frac{1}{4} + \frac{1}{8})E[X] \end{aligned} \tag{3}$$

Therefore, we may solve the equation (3) and get

$$E[X] = 14$$

Student: Charles Wang
Problem 3

Let I_k denotes the result of the k^{th} roll
$E(X) = E(X|I_1 = 2)P(I_1 = 2) + E(X|I_1 \neq 2)P(I_1 \neq 2)$
In which $E(X|I_1 \neq 2) = 1 + E(X)$
$E(X|I_1 = 2) = E(X|I_1 = 2, I_2 = 3)P(I_2 = 3) + E(X|I_1 = 2, I_2 = 2)P(I_2 = 2) + E(X|I_1 = 2, I_2 \neq 2, 3)P(I_2 \neq 2, 3)$
In which, $E(X|I_1 = 2, I_2 = 3) = 2$, $E(X|I_1 = 2, I_2 = 2) = 1 + E(X|I_1 = 2)$, and $E(X|I_1 = 2, I_2 \neq 2, 3) = 2 + E(X)$
Plugging back into the $E(X|I_1 = 2)$.

$$\begin{aligned}
E(X|I_1 = 2) &= 2 * \frac{1}{6} + (1 + E(X|I_1 = 2))\frac{1}{6} + (2 + E(X))\frac{2}{3} \\
&= \frac{2}{6} + \frac{1}{6} + \frac{1}{6} * E(X|I_1 = 2) + \frac{8}{6} + \frac{4}{6} * E(X) \\
6 * E(X|I_1 = 2) &= 11 + E(X|I_1 = 2) + 4E(X) \\
E(X|I_1 = 2) &= \frac{11 + 4E(X)}{5}
\end{aligned}$$

Plugging back into the original equation.

$$\begin{aligned}
E(X) &= \frac{11 + 4E(X)}{5}\frac{1}{6} + (1 + E(X))\frac{5}{6} \\
30E(X) &= 11 + 4E(X) + 25 + 25E(X) \\
E(X) &= 36
\end{aligned}$$

(b) Find the expected value E(Y) (8 pts).
Ans:
$E(Y) = E(Y|I_1 = 2)P(I_1 = 2) + E(Y|I_1 \neq 2)P(I_1 \neq 2)$,
In which $E(Y|I_1 \neq 2) = 1 + E(Y)$ since we already roll one die and need to start over; and
$E(Y|I_1 = 2) = E(Y|I_2 = 2, I_1 = 2)P(I_2 = 2) + E(Y|I_2 \neq 2, I_1 = 2)P(I_2 \neq 2)$
In which $E(Y|I_2 = 2, I_1 = 2) = 2$ and $E(Y|I_2 \neq 2, I_1 = 2) = 2 + E(Y)$
Plugging all back into the original equation.

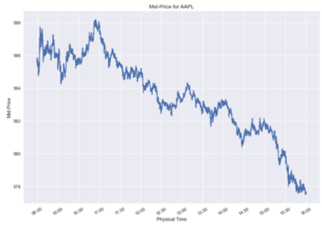
$$\begin{aligned}
E(Y) &= (2 * P(I_2 = 2) + (2 + E(Y))P(I_2 \neq 2))P(I_1 = 2) + (1 + E(Y))P(I_1 \neq 2) \\
&= (2 * \frac{1}{6} + (2 + E(Y))\frac{5}{6})\frac{1}{6} + (1 + E(Y))\frac{5}{6} \\
&= (\frac{1}{3} + \frac{5}{3} + \frac{5}{6}E(Y))\frac{1}{6} + \frac{5}{6} + \frac{5}{6}E(Y) \\
&= (2 + \frac{5}{6}E(Y))\frac{1}{6} + \frac{5}{6} + \frac{5}{6}E(Y) \\
&= \frac{12}{36} + \frac{5}{36}E(Y) + \frac{30}{36} + \frac{30}{36}E(Y) \\
E(Y) &= \frac{42}{36} + \frac{35}{36}E(Y) \\
E(Y) &= 42
\end{aligned}$$

(c) Compare you answers in (a) and (b) and explain the result (4 pts).
Ans:
Comparing the result between (a) and (b), we can see that E(Y) is greater than E(X). This can be explained by the probability of getting the sequence ”22” and ”23” is the same. However, after getting a ”2”, there emerges a difference between X and Y. In the case of not getting the sequence ”23”, there is still a chance that we rolled 22, in which we are only 1 additional steps from getting ”23” sequence. On the other hand, regarding Y, if we didn’t get the sequence ”22”, we have to at least have two additional steps to get the sequence ”22”. Therefore, E(Y) is greater than E(X).

Students: Mehdi Badri

4 Problem 4

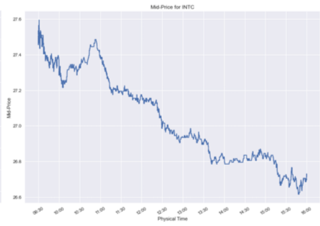
4.1 a



(a) AAPL

(b) AMZN

(c) GOOG



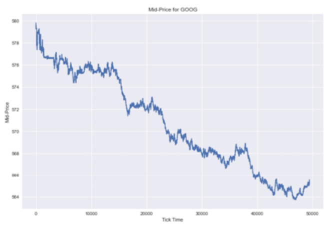
(d) INTC

(e) MSFT

Figure 1: Mid-prices of the stocks in physical time

(a) AAPL

(b) AMZN



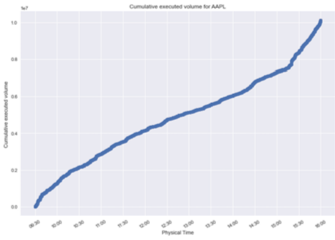
(c) GOOG

(d) INTC

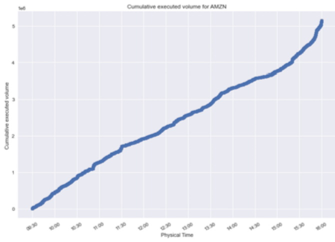
(e) MSFT

Figure 2: Mid-prices of the stocks in tick time

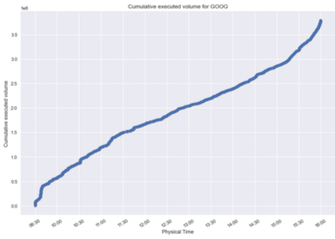
4.2 b



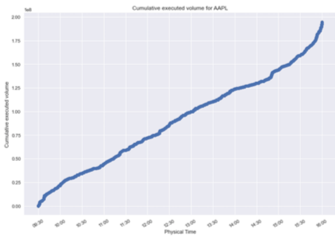
(a) AAPL



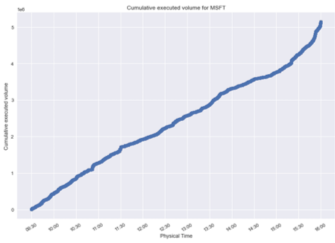
(b) AMZN



(c) GOOG

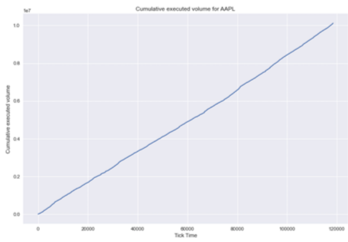


(d) INTC

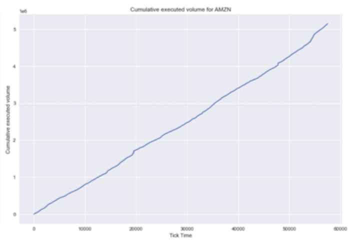


(e) MSFT

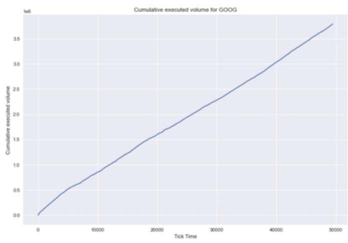
Figure 3: Total executed volume of the stocks in physical time



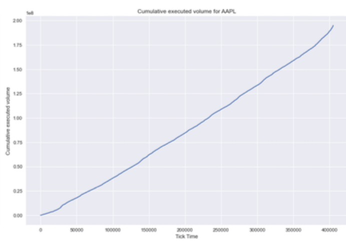
(a) AAPL



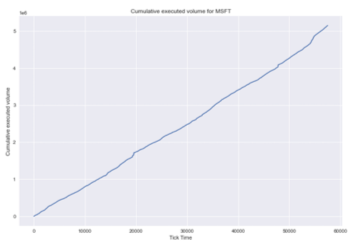
(b) AMZN



(c) GOOG



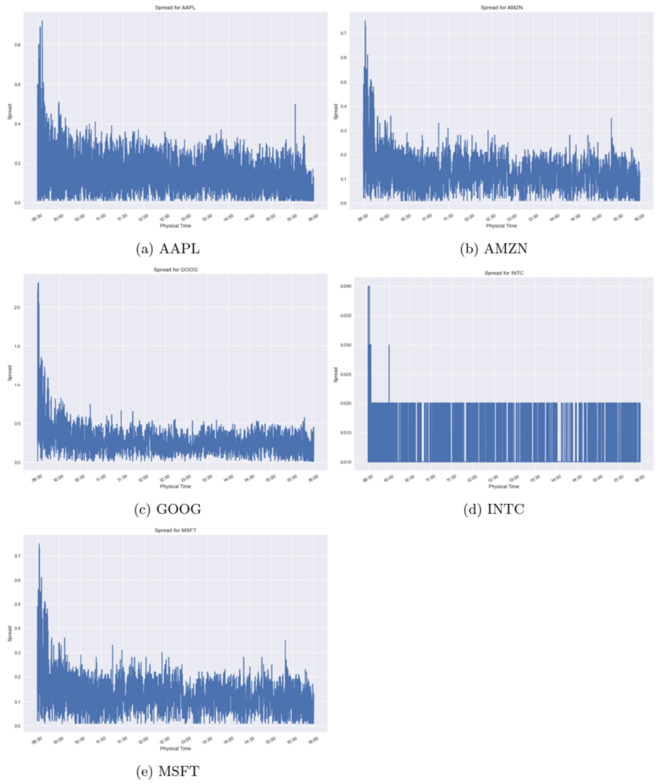
(d) INTC



(e) MSFT

Figure 4: Total executed volume of the stocks in tick time

4.3 c

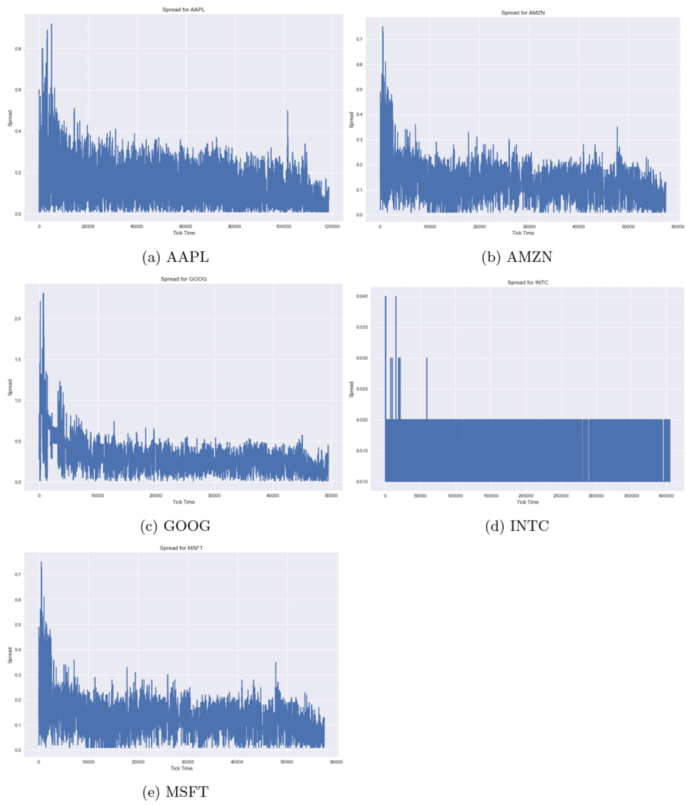


(a) AAPL (b) AMZN

(c) GOOG (d) INTC

(e) MSFT

Figure 5: Spread of the stocks in physical time



(a) AAPL (b) AMZN

(c) GOOG (d) INTC

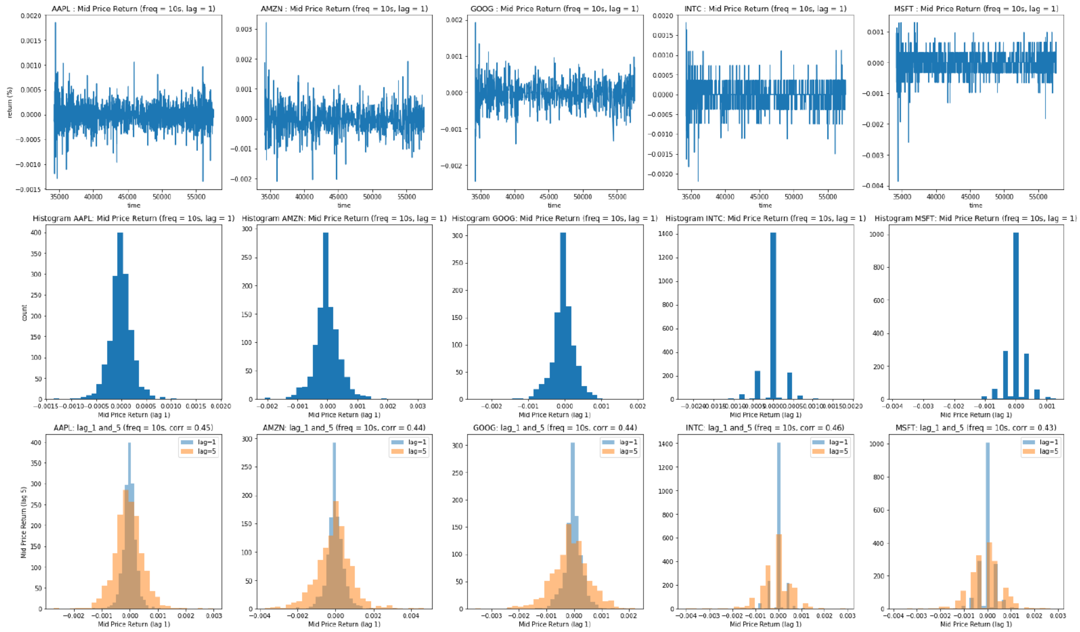
(e) MSFT

Figure 6: Spread of the stocks in tick time

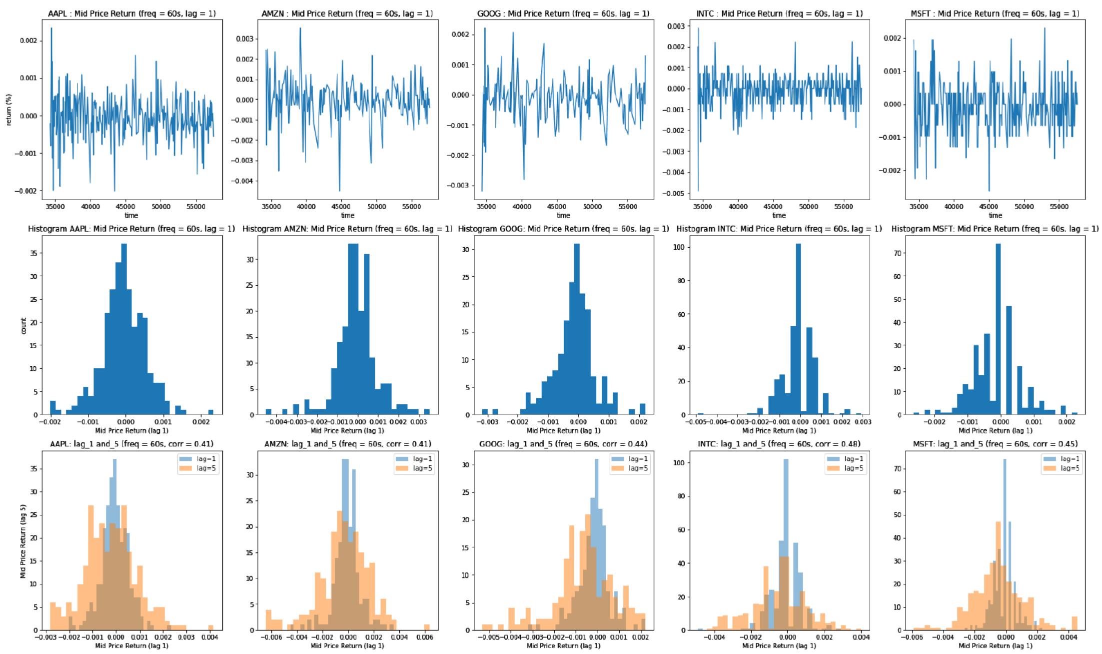
Students: Stanislas Bucaille

Problem 4(d)

Graphs: 10-sec data

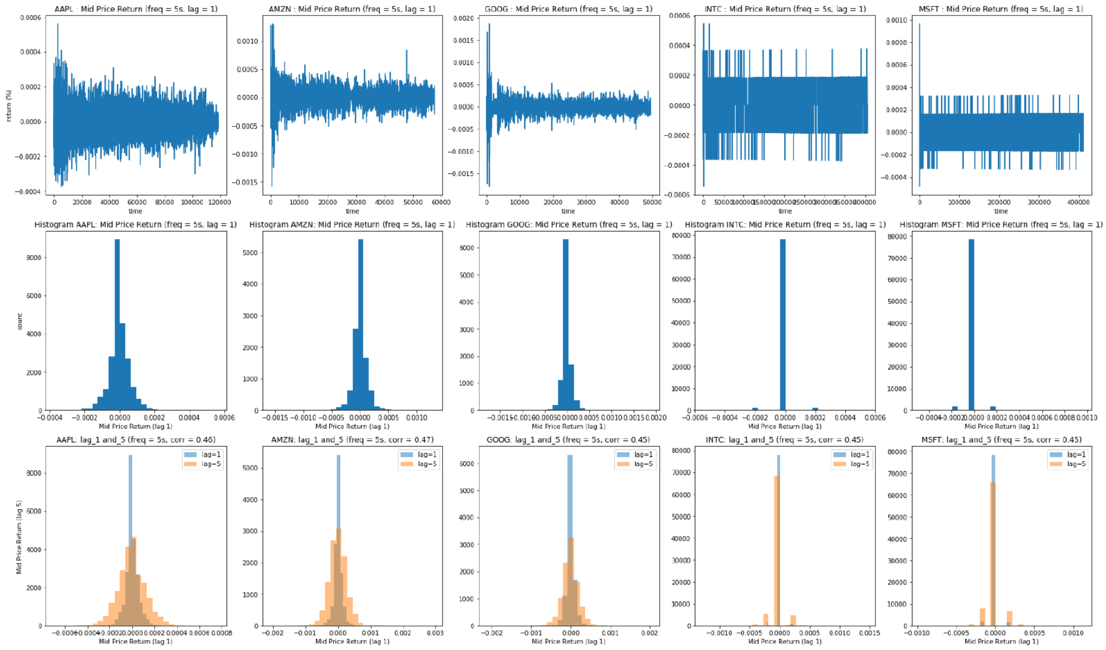


Graphs: 60-sec data

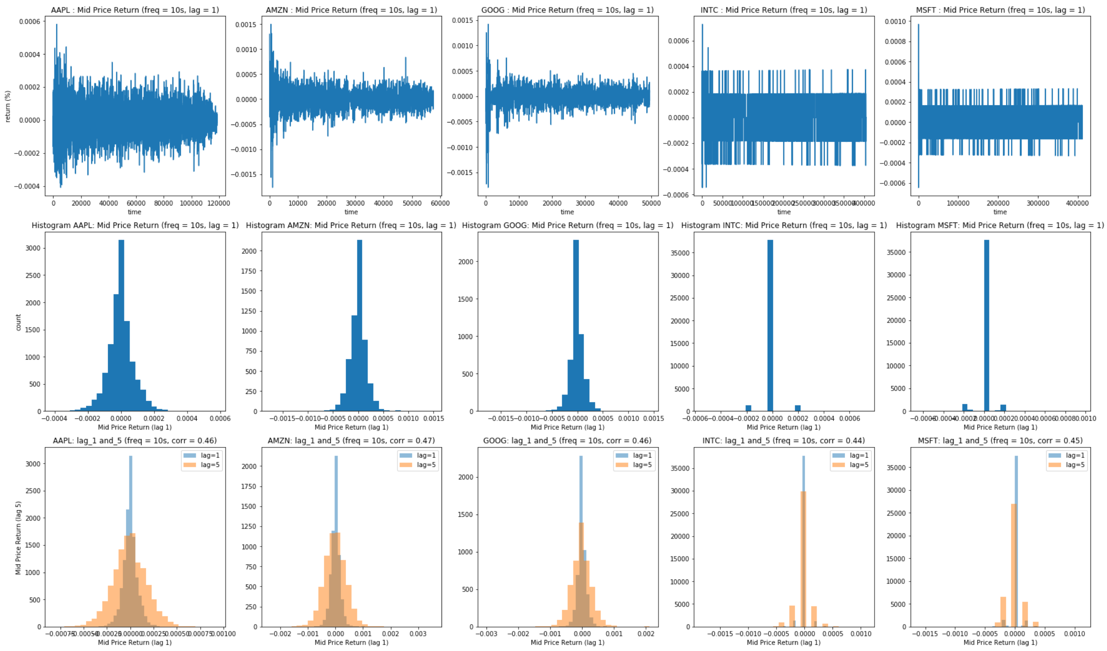


Problem 4(e)

Graphs: 5-tick data



Graphs: 10-tick data



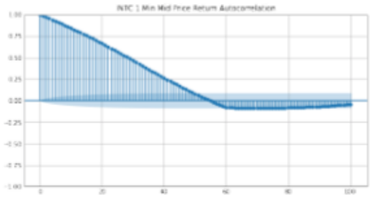
Student: Yutong Yang

Problem IV: (f)

Comment on the observed stylized facts.

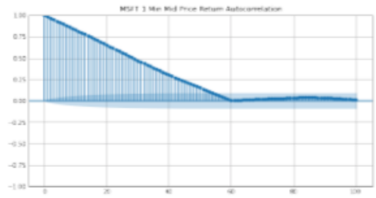
1. Asset return distributions

(a) AAPL



(b) INTC

(c) GOOG



(d) MSFT

Figure 23: Autocorrelation Function of 1 Min Mid Price Returns

- Absence of autocorrelations: We may observe from the autocorrelation functions of midprice returns that linear autocorrelations of asset returns over long periods diminishes. For the 5 tick midprice returns, the autocorrelation disappears after a lag of 5 at most.
- Long range dependence: If one looks at autocorrelation function of absolute returns as a function of time lag, it is empirically shown that it decays according to the power law distribution. We may observe from the autocorrelation functions of midprice returns, especially for INTC,GOOG and MSFT, that it decays according to the power law distribution.

- Heavy tails and aggregational normality: The distribution of asset price returns shows fat tails; as one increases the period of time, asset returns show lower tails. We may observe from the Mid Price returns of AAPL below.

- Intermittency: At any micro or macro time scale, asset price returns must display high degree of volatility. This is obvious according to Figure 26.

2. Volumes and order flow

- Order book volumes: From the cumulative executed volumes in part (b), we may observe that the increment of volume is more intense during the start and end of the day.

3. Non-stationary patterns

- From the cumulative executed volumes in part (b), we saw that the increment of volume is more intense during the start and end of the day, the volume of trade is not stationary during one day in physical time. While in contrast, if we look at the cumulative executed volume in tick time, we see that it nearly follows a linear shape, meaning that making a transformation from physical time to tick (or transaction) time may help adjusting for intraday non-stationarity.

Student: Charlotte Jin
Problem 4(g)

(7) According to the above discussion, I think that when we want to model the high frequency and high volatility of the stock market, it is better that we use tick time. However, if we are interested in the common trends of the stock market, it is fine to use physical time which can smooth out the high fluctuations. In addition, if we want to study the distribution of stock return, it is better to look at physical time because this can allow for more possible values, therefore fully capture the distribution patterns.