Module Title: High Frequency Trading

Module Code: 7FNCE025W

Course: MSc Fintech and Business Analytics, Semester 2, 2023/2024

ASSIGNMENT 1

Table of Contents

Question 1

The dataset in Microsoft Access format includes BuyPrice, SellPrice, BuyVolume, and SellVolume for Facebook, offering insights into bid/ask prices and available stock quantities. It's crucial for analyzing market dynamics, spread, volume imbalance, and price discovery.

a. Midprice

The midprice of a stock is calculated as the average of its highest bid and lowest ask prices. It serves as an indicator of the stock's current market value, offering a midpoint reference for traders and analysts (O'Hara, 1995).

$$M = \frac{p^b + p^a}{2}$$

p^b - best bid price

p^a - is the best ask price

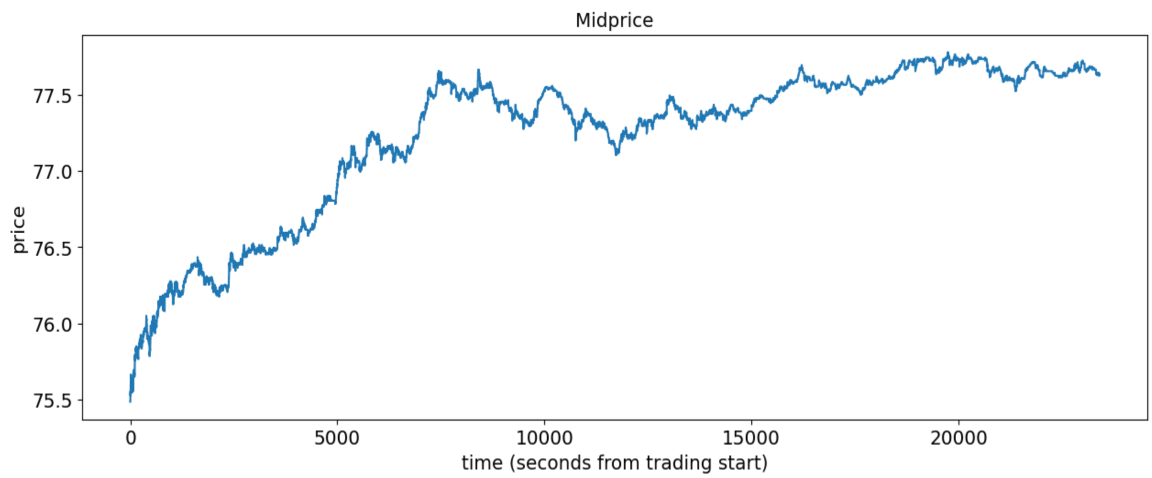


Figure 1. Midprice of Facebook over time

Delving into the trading data of Facebook, it was noted that 234,000 shares exchanged hands on a particular day. The statistical analysis unveiled a mean midprice of $77.24 for the shares during that trading session. Notably, the midprice peaked at $77.78 at the session's onset and dipped to its lowest point at $75.49.

count	234000
mean	77.23742
std	0.49785
min	75.48500
25%	77.10500
50%	77.40500
75%	77.59500
max	77.78000

Table 1. Summary statistics of Midprice

b. **Microprice** :

The microprice of a stock refines its valuation by weighting the bid and ask prices according to order volumes, providing a more precise market value indicator, especially in high-frequency trading (Álvaro Cartea, Jaimungal and José Penalva, 2015).

The graph for the data given is given below.

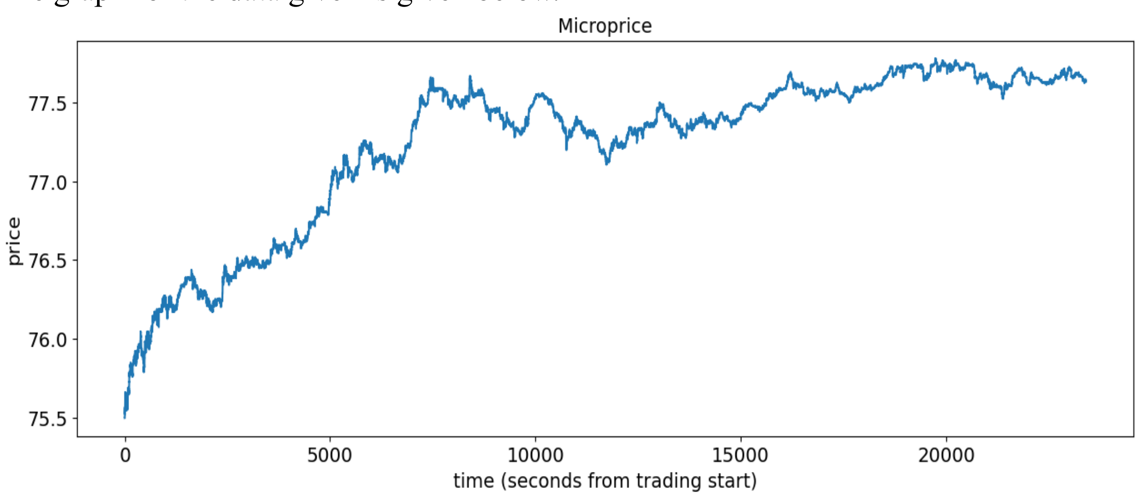


Figure 2. Microprice of Facebook over time

The dataset comprises 234,000 microprice values, each representing the microprice of a facebook at a specific time during the trading day. The dataset reveals that the microprice peaked at $77.78, approximately 250 minutes into the trading session, with a standard deviation of approximately 0.4978. The lowest microprice recorded was $75.498, observed roughly 20 minutes after the start of trading.

count	234000
mean	77.23717
std	0.49773
min	75.49815
25%	77.10817
50%	77.40333
75%	77.59019
max	77.77684

Table 2. Summary statistics of Microprice

A key finding from our analysis is the notable difference between the midprice and microprice of the facebook, particularly at the start of the trading day. This difference, however, stabilizes to fluctuate between -0.0 and +0.02 as the day progresses. This observation underscores the dynamic nature of stock prices and highlights the importance of considering both microprice and midprice in market analysis.

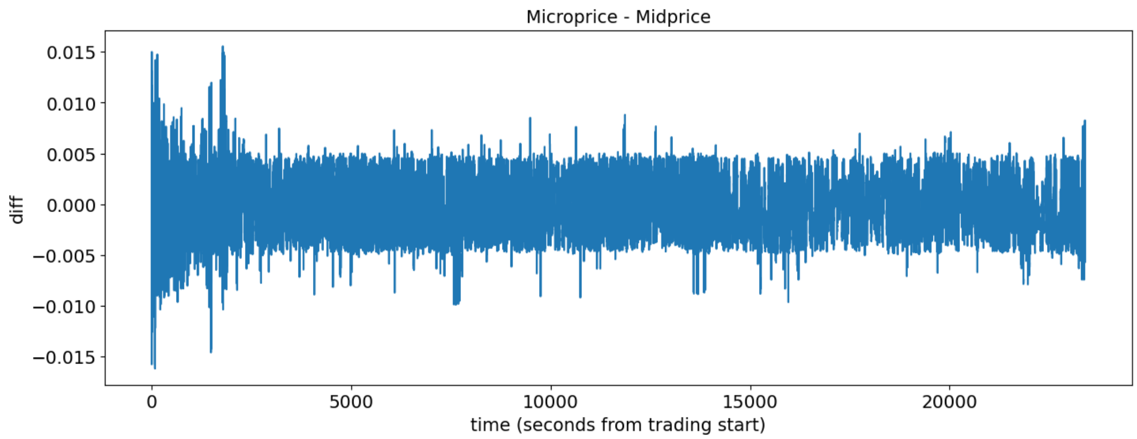


Figure 3. difference between Midprice and Microprice of Facebook over time

The detailed statistical summary includes the mean, standard deviation, and the minimum and maximum values of both midprice and microprice, providing a comprehensive overview of the stock's price behavior throughout the trading day.

	Midprice	Microprice
count	234000	234000
mean	77.23742	77.23717
std	0.49785	0.49773
min	75.48500	75.49815
25%	77.10500	77.10817
50%	77.40500	77.40333
75%	77.59500	77.59019
max	77.78000	77.77684

Table 3. Summary statistics of MidPrice and Microprice

c. **Spread**:

The spread of a stock is the difference between its bid (buy) and ask (sell) prices. It reflects the stock's liquidity, with narrower spreads indicating higher liquidity. Spreads are crucial for understanding trading costs and market efficiency (Harris, 2003).

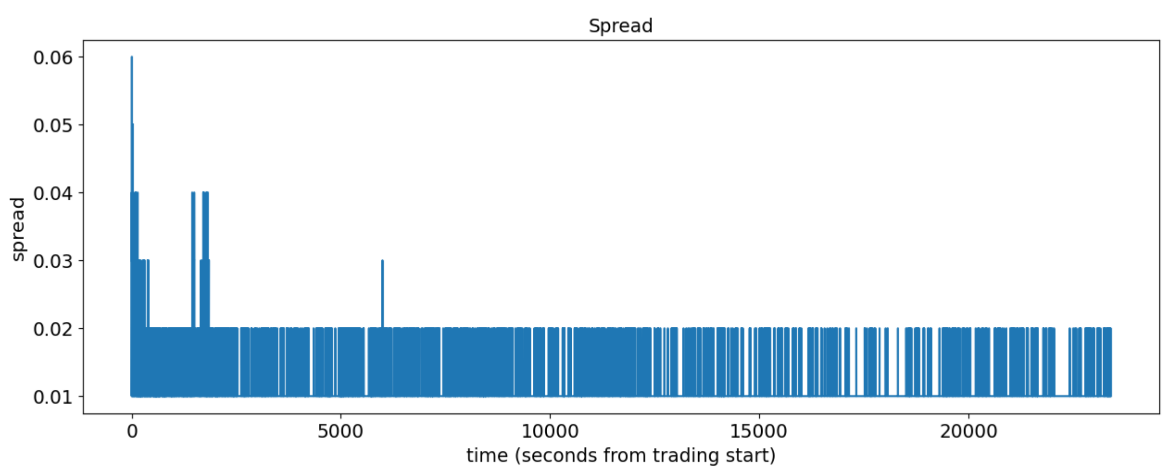


Figure 4. Spread of Facebook over time

The graph demonstrates the stock spread, with the minimum observed spread being 0.01 cents. In instances where bid and ask prices align, the spread can reduce to zero, a phenomenon increasingly observed in highly liquid assets, though such market conditions are usually transient. Daily trading data reveals a consistent minimum spread of 0.01, with the maximum spread of 0.06 occurring early in the trading day, reflecting typical early-day stock volatility.

count	234000
mean	0.01073
std	0.00285
min	0.01
25%	0.01
50%	0.01
75%	0.01
max	0.06000

Table 4. Summary statistics of spread

d. Volume imbalance

Volume imbalance in stock trading refers to the disparity between buy and sell orders, impacting price movements and market efficiency. It's particularly evident in the Chinese stock market, where short-horizon market efficiency and speculative trading are influenced by order imbalances, contributing to momentum patterns and trading strategies based on high-frequency data (Hu, 2018; Yang, 2021).

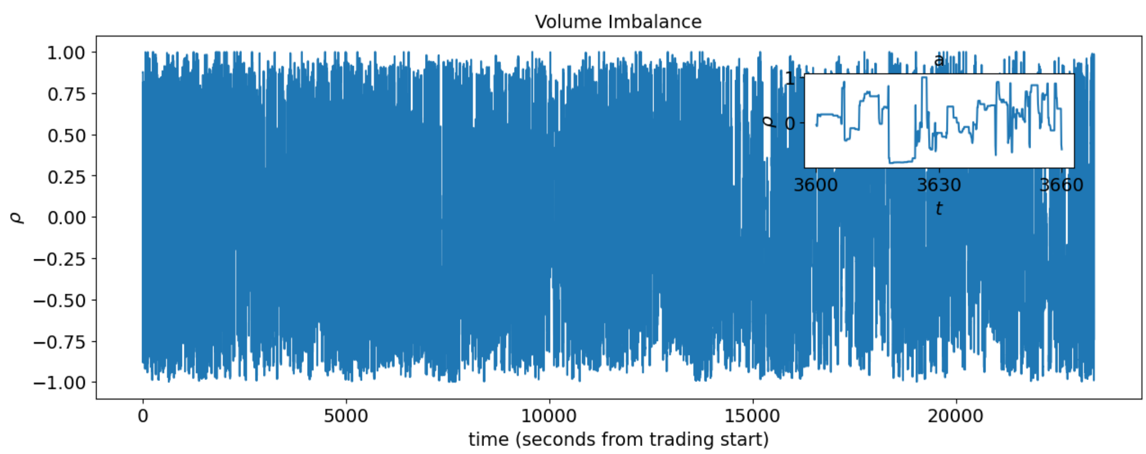


Figure 5. Volume Imbalance of Facebook over time

The graph illustrates the volume imbalance throughout the trading day, with imbalances oscillating between -1 and +1. A positive imbalance indicates a higher volume on the ask side, while negative signifies more on the bid side. The inset graph shows a one-minute snapshot with a narrower range of -0.5 to +0.5. Data for 234,000 shares traded shows imbalances ranging from -0.99 to +0.99, with an average daily imbalance of -0.049 and a standard deviation of 0.489.

count	234000
mean	-0.04936
std	0.48932
min	-0.99932
25%	-0.43677
50%	-0.06250
75%	0.33155
max	0.99949

Table 5. Summary statistics of volume imbalance

Question 2

The file 'FTSE_sample.mat' provides data on market orders (MO) on a minute-by-minute basis, including the highest bid and lowest ask prices at each minute's end, for a company listed on the FTSE during the 2008 financial crisis.

a. The data set comprises 62,190 observations of minutely returns for FTSE. The mean return is close to zero, indicating a relatively stable performance on average. The standard deviation suggests moderate variability in returns around the mean. The returns range from a minimum of -7.428571e-02 to a maximum of 5.734521e-02, showcasing the diversity in return values within the dataset. The quartile values provide insights into the central tendency and spread of the returns distribution.

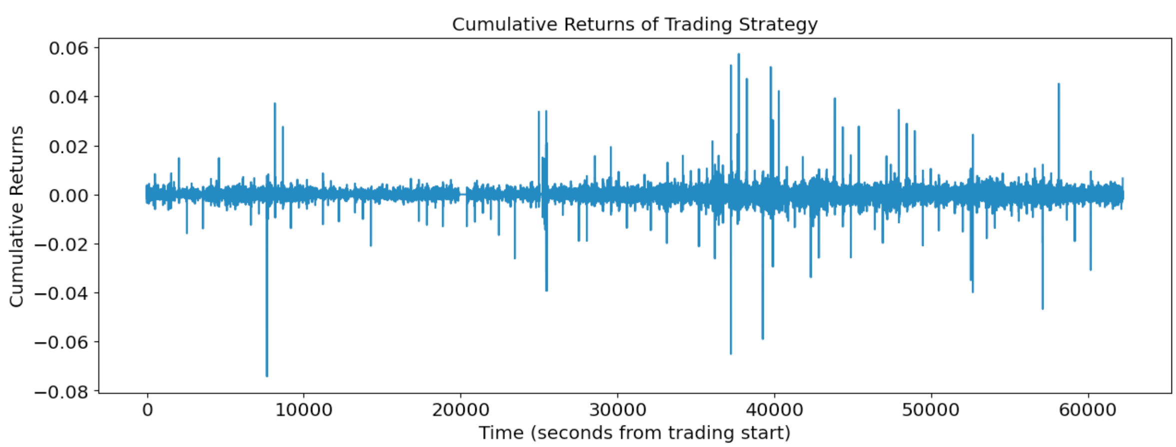


Figure 6. Minutely returns of FTSE

count	6.221900e+04
mean	-9.544055e-07
std	1.825031e-03
min	-7.428571e-02
25%	-5.875441e-04
50%	0.000000e+00
75%	5.805515e-04
max	5.734521e-02

Table 6. Summary statistics of minutely returns

b. The **Autocorrelation Function (ACF)** is a statistical tool used to analyse time series to quantify the correlation between observations at different time points within a series. It measures the relationship between a variable's current value and its past values at various lags. High autocorrelation values indicate a strong relationship between observations at specific lags, while low values suggest little to no relationship. The ACF is essential for identifying patterns, trends, and seasonality in time series data. (Alder and Wainwright, 1970)

   The plot features distinct peaks at lags 2 and -2, which implies a moderate positive autocorrelation of 0.4, suggesting a pattern that recurs every 2 minutes. The autocorrelation values at lags 19 and -19 are lower, indicating a reduced correlation at these intervals.

   The ACF plot indicates short-term periodicities in the minutely order flows, with a significant pattern occurring every 2 minutes. The long-term correlation decreases, as seen by the lower autocorrelation values at higher lags. This analysis is instrumental for traders and analysts in understanding the market dynamics at a minute level and could potentially inform trading strategies that capitalize on these short-term patterns.

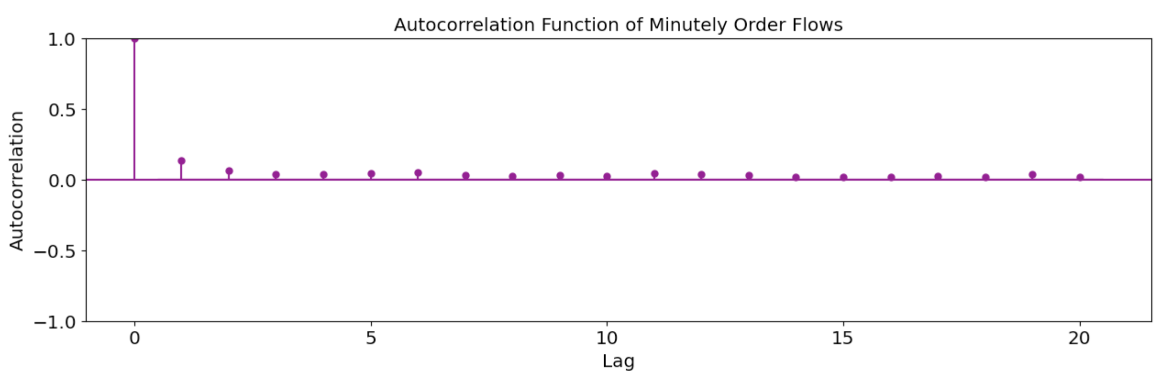


Figure 7. Autocorrelation of FTSE

c. Ordinary Least Squares (OLS) regression is a statistical technique that estimates the relationship between a dependent variable and one or more independent variables by minimizing the sum of the squares of the differences between observed and predicted values. (Maiti, 2018)

   The OLS Regression Results image displays a statistical analysis of minutely returns using the Ordinary Least Squares method. The model, with an R-squared of 0.123, explains 12.3% of the variance in returns. The lagged returns coefficient is -0.0379, indicating a negative impact on current returns, while order flow has a positive coefficient of 2.429e-05, both statistically significant with p-values of 0.000. The Durbin-Watson statistic of 2.001 suggests no autocorrelation in residuals. However, the Omnibus and Jarque-Bera tests indicate potential non-normality in residuals, with p-values of 0.000. The analysis was conducted with 62,218 observations.

OLS Regression Results

Dep. Variable:	returns	R-squared:	0.123
Model:	OLS	Adj. R-squared:	0.123
Method:	Least Squares	F-statistic:	4361.
Date:	Tue, 12 Mar 2024	Prob (F-statistic):	0.00
Time:	00:59:15	Log-Likelihood:	3.0804e+05
No. Observations:	62218	AIC:	-6.161e+05
Df Residuals:	62215	BIC:	-6.161e+05
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	1.003e-06	6.86e-06	0.146	0.884	-1.25e-05	1.45e-05
lagged_returns	-0.0379	0.004	-10.088	0.000	-0.045	-0.031
order_flow	2.429e-05	2.61e-07	93.129	0.000	2.38e-05	2.48e-05

Omnibus:	44198.801	Durbin-Watson:	2.001
Prob(Omnibus):	0.000	Jarque-Bera (JB):	144438701.075
Skew:	-1.791	Prob(JB):	0.00
Kurtosis:	239.015	Cond. No.	1.44e+04

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.44e+04. This might indicate that there are strong multicollinearity or other numerical problems.

d. A trading strategy based on cumulative order flows and cumulative returns over the previous 30 minutes was designed which illustrates a financial trading strategy based on past returns. The strategy calculates a rolling sum of returns over a 30-period window and then determines a trading position based on the sign of these cumulative returns. The position is taken to be the opposite of the sign of the cumulative returns (short if the cumulative returns are positive and long if they are negative), which is a form of mean-reversion strategy. The strategy's returns are calculated by multiplying the previous period's position by the current returns.

The strategy yielded a return of 2.24% CAGR whereas the benchmark FTSE here is performing at a return of -0.159%.

returns   -0.159546
strategy   2.244781

e. The trading strategy, based on cumulative order flows over the previous 30 minutes, yielded a return of **2.24% CAGR** whereas the benchmark FTSE here is performing at a return of -0.159%. This difference in the performance indicates that the trading strategy **outperformed** FTSE during the observed period. The graph below shows the cumulative sum of both the actual returns and the strategy returns, applying the exponential function to these sums to show the compounded growth of returns over time.

   The graph demonstrates that the strategy starts to significantly outperform the actual returns after a certain point in time, suggesting that the strategy might be capturing some predictive patterns in the returns data.

   The graph is a visual representation of the effectiveness of the strategy over time. When evaluating the performance of a trading strategy against a benchmark or the raw returns of an asset. The graph indicates that the strategy has periods of both underperformance and outperformance relative to the actual returns, with the latter part showing a notable divergence where the strategy outperforms. Investors can consider conducting a further analysis of the strategy, its potential risks, and market conditions to identify the areas where there could be an improvement and achieve more favourable outcome.

   We made money in this trade because mean-reversion strategy is designed to capitalize on the assumption that prices will revert to their historical average levels after periods of deviation.

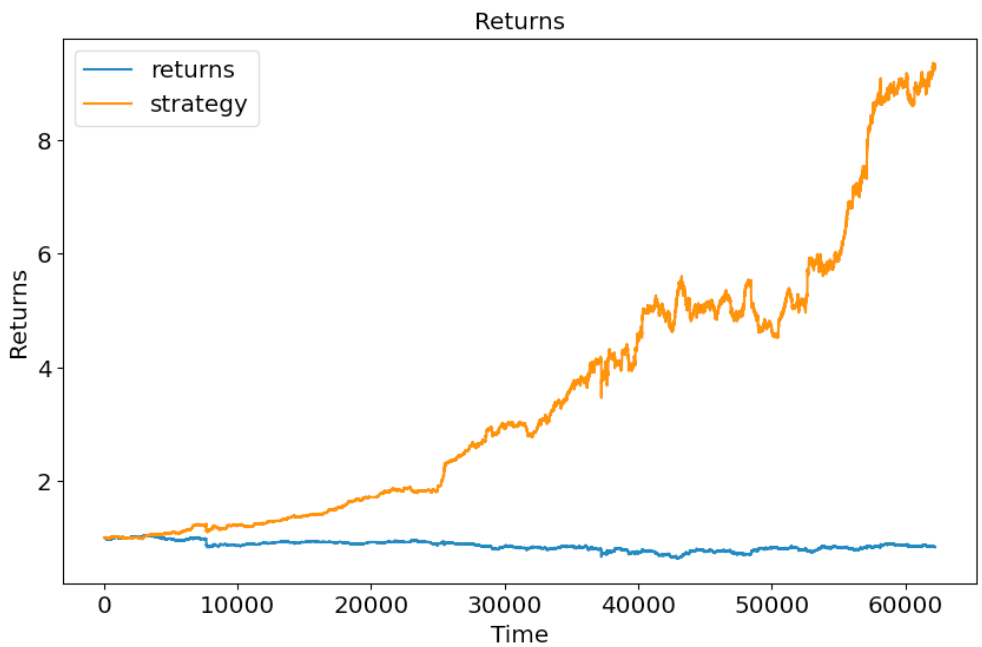
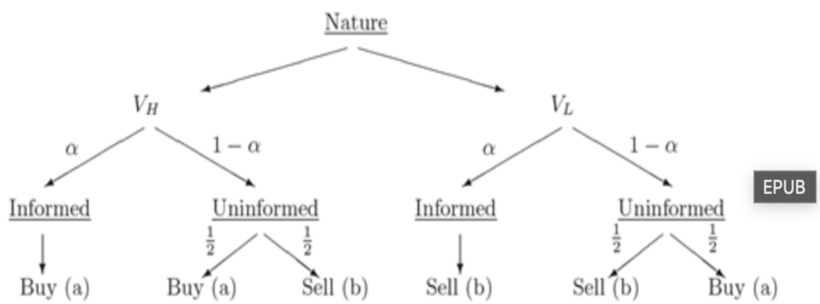


Figure 8. Returns of FTSE over the time vs the returns of strategy

Question 3



A. Probabilities that MM receives a coming buy order.
Probability that MM receives a coming buy order if $V=V_H$

- Informed traders will always buy because they know the value is high.
- Liquidity traders will buy with a probability of 0.5 because they trade randomly.

Since $P(buy|V_H)$ is different for informed traders and uninformed traders
For Informed traders:
$P(buy|V_H) = \alpha$
For Uninformed traders:
$P(buy|V_H) = 0.5*(1-\alpha)$
Therefore, for MM the value of $P(buy|V_H)$ is
$P(buy|V_H) = 0.5*(1-\alpha) + \alpha$
Therefore,
$P(buy|V_H) = 0.5*(1-\alpha) + \alpha = 0.5 + 0.5\alpha$

Probability that MM receives a coming buy order if $V=V_L$

- Informed traders will never buy because they know the value is low.
- Liquidity traders will buy with a probability of 0.5 because they trade randomly.

Since $P(buy|V_L)$ is different for informed traders and uninformed traders
For Informed traders:
$P(buy|V_L) = \alpha *0=0$
For Uninformed traders:
$P(buy|V_L) = 0.5*(1-\alpha)$
And since $P(V_L) = 1-p$
Therefore, for MM the value of $P(buy|V_L)$ is
$P(buy|V_L) = 0.5*(1-\alpha)$
Therefore,
$P(buy|V_L) = 0.5*(1-\alpha) = 0.5 - 0.5\alpha$

B. Probabilities that MM receives a coming sell order.

The probability of receiving a buy order when $\mathbf{P(buy|V_H) = 0.5+0.5\alpha}$, as derived from the previous question. Therefore, the probability of receiving a sell order when $V=V_H$ is:
$\mathbf{P(sell|V_H) = 1 - P(buy|V_H) = 1 - (0.5+0.5\alpha) = 0.5-0.5\alpha}$
For $V=V_L$:

The probability of receiving a buy order when $V=V_L$ is $\mathbf{P(buy|V_L) = 0.5-0.5\alpha}$, as derived from the previous question. Therefore, the probability of receiving a sell order when $V=V_L$ is:

$\mathbf{P(sell|V_L) = 1 - P(buy|V_L) = 1 - (0.5 - 0.5\alpha) = 0.5 + 0.5\alpha}$

C. According to Bayes Theorem, what is the probability MM receives a buy order?

Baye's Theorem

$\mathbf{P(A|B) = (P(B|A) *P(A))/P(B)}$
Where:
P(A|B) is the posterior probability of A occurring given that B has occurred.
P(B|A) is the likelihood of observing B given that A is true.
P(A) is the prior probability of A occurring.
P(B) is the marginal probability of observing B.

Given:

$\mathbf{P(V=V_H) = p}$,
$\mathbf{P(V=V_L) =1-p}$,

$\mathbf{P(buy|V_H) = 0.5+0.5\alpha}$.
$\mathbf{P(buy|V_L) = 0.5-0.5\alpha}$.

The marginal probability of receiving a buy order,
P(buy), can be calculated as:

$P(buy) = P(buy|V_H) *P(V_H)+P(buy|V_L) *P(V_L)$

Substituting the given probabilities:
$\mathbf{P(buy) = (0.5+0.5\alpha) *p+(0.5-0.5\alpha) *(1-p)}$

To find the probability for $V=V_H$ given a buy order,
$P(V_H|buy) = (P(buy|V_H) *P(V_H))/P(buy)$

Substituting the known values:
$\mathbf{P(V_H|buy) = ((0.5+0.5\alpha) *p)/(0.5+0.5\alpha) *p+(0.5-0.5\alpha) *(1-p)}$

$= \mathbf{(p+\alpha p)/(1-\alpha+2\alpha p)}$

Similarly, to find the probability that $V=V_L$ given a buy order, $P(V_L|buy)$, we apply Bayes' Theorem:
$P(V_L|buy) = (P(buy|V_L) *P(V_L))/P(buy)$

Substituting the known values:
$\mathbf{P(V_L|buy) = ((0.5-0.5\alpha) *(1-p))/(0.5+0.5\alpha) *p+(0.5-0.5\alpha) *(1-p)}$
$\mathbf{= (1-\alpha-p+\alpha p)/(1-\alpha+2\alpha p)}$

D. According to Bayes Theorem, what is the probability MM receives a sell order?

We are provided the probabilities of receiving a sell order given $V=V_H$ and $V=V_L$ as
$\mathbf{P(sell|V_H) = 0.5-0.5\alpha}$
$\mathbf{P(sell|V_L) = 0.5+0.5\alpha}$
We can calculate the marginal probability of receiving a sell order, P(sell), as:

$P(sell)=P(sell|V_H) *P(V_H)+P(sell|V_L) *P(V_L)$
Substituting the given probabilities:
$\mathbf{P(sell)=(0.5-0.5\alpha) *p+(0.5+0.5\alpha) *(1-p)}$

To find the probability that $V=V_H$ given a sell order, $P(V_H|sell)$, we apply Bayes' Theorem:
$P(V_H|sell) = P(sell|V_H) *P(V_H)/P(sell)$

Substituting the known values:
$\mathbf{P(V_H|sell) = (0.5-0.5\alpha) *p/(0.5-0.5\alpha) *p+(0.5+0.5\alpha) *(1-p)}$
$\mathbf{= (p - \alpha p)/ (1+\alpha-2\alpha p)}$

Similarly, to find the probability that $V=V_L$ given a sell order, $P(V_L|sell)$, we apply Bayes' Theorem:
$P(V_L|sell) = P(sell|V_L) *P(V_L)/P(sell)$
Substituting the known values:
$\mathbf{P(V_L|sell) = (0.5+0.5\alpha) *(1-p)/(0.5-0.5\alpha) *p+(0.5+0.5\alpha) *(1-p)}$
$\mathbf{= (1 - p + \alpha - \alpha p)/ (1+\alpha-2\alpha p)}$

E. Ask Price A

The expected value of V given a buy order, which is the ask price A, can be calculated using the conditional probabilities of VH and VL given a buy order. These can be derived using Bayes' theorem:
$P(V_H|buy) = P(buy|V_H) *P(V_H)/P(buy)$
$P(V_L|buy) = P(buy|V_L) *P(V_L)/P(buy)$

The expected value of V given a buy order is:
$E[V|buy] =P(V_H|buy) *V_H+P(V_L|buy) *V_L$

Given that $P(V_H|buy)$, $P(V_L|buy)$, and V_H, and V_L, this formula allows us to calculate A directly from the calculated values above.
Therefore,
$A = (p+\alpha p)/(1-\alpha+2\alpha p) *V_H + (1-\alpha-p+\alpha p)/(1-\alpha+2\alpha p) *V_L$

F. Bid Price B

The expected value of V given a sell order, which is the bid price B, can be calculated using the conditional probabilities of V_H and V_L given a sell order. These can be derived using Bayes' theorem:

$P(V_H|sell) = P(sell|V_H)\times P(V_H)/P(sell)$
$P(V_L|sell) = P(sell|V_L)\times P(V_L)/P(sell)$

The expected value of V given a sell order is:
$E[V|sell] = P(V_H|sell)\times V_H + P(V_L|sell)\times V_L$

Given that $P(V_H|sell)$, $P(V_L|sell)$, and V_H, and V_L, this formula allows us to calculate A directly from the calculated values above.

Therefore,
$B = (p - \alpha p)/ (1+\alpha-2\alpha p) *V_H + (1 - p + \alpha - \alpha p)/ (1+\alpha-2\alpha p)*V_L$

G. Updating beliefs and setting ask and bid prices.

Assuming that time is discrete, the trading happens sequentially at times t = 1,2,3,..., and at each trade the MM updates beliefs about V,

After k buy orders and l sell orders:

A(k +l) = E[V|k +1 BUY S, l SELLS],

B(k +l) = E[V|k BUY S, l +1 SELLS]

equivalent to:

$A(BUY|k+l) = V_H *P(V = V_H| (k+1) +l) + V_L *P(V = V_L| (k+1) +l)$,

$B(SELL|k+l) = V_H *P(V = V_H|k + (l+1)) + V_L *P(V = V_L|k + (l+1)$

where the four conditional probabilities can be calculated with the Bayes theorem as:

P(V|k BUY S,l SELLS) = (P(k BUY S,l SELLS|V)*P(V))/ (P(k BUY S,l SELLS))

Where P(k BUY S,l SELLS|V) is the binomial probability of having a price of V after k buys in k +l trades and P(k BUY S,l SELLS) is the binomial probability of having k Buys in k + l trades (hence l sells).

After simplifying and applying binomial probability calculation to the equations we get,

$$A(k+l) = \frac{(\alpha+1)^{k+1}(1-\alpha)^{l} p V_H + (1-\alpha)^{k+1}(\alpha+1)^{l}(1-p)V_L}{(2\alpha p-\alpha+1)^{k+1}(-2\alpha p+\alpha+1)^{l}}$$

$$B(k+l) = \frac{(\alpha+1)^{k}(1-\alpha)^{l+1} p V_H + (1-\alpha)^{k}(\alpha+1)^{l+1}(1-p)V_L}{(2\alpha p-\alpha+1)^{k}(-2\alpha p+\alpha+1)^{l+1}}$$

H) If p=0.2, α=0.8, simulate the case (for at least 50 times) if V=V_H =2 (whileV_L =1). Plot A and B against time.

After simulation the results are as follows and the code for the simulation can be found at : https://shorturl.at/qtwZ1

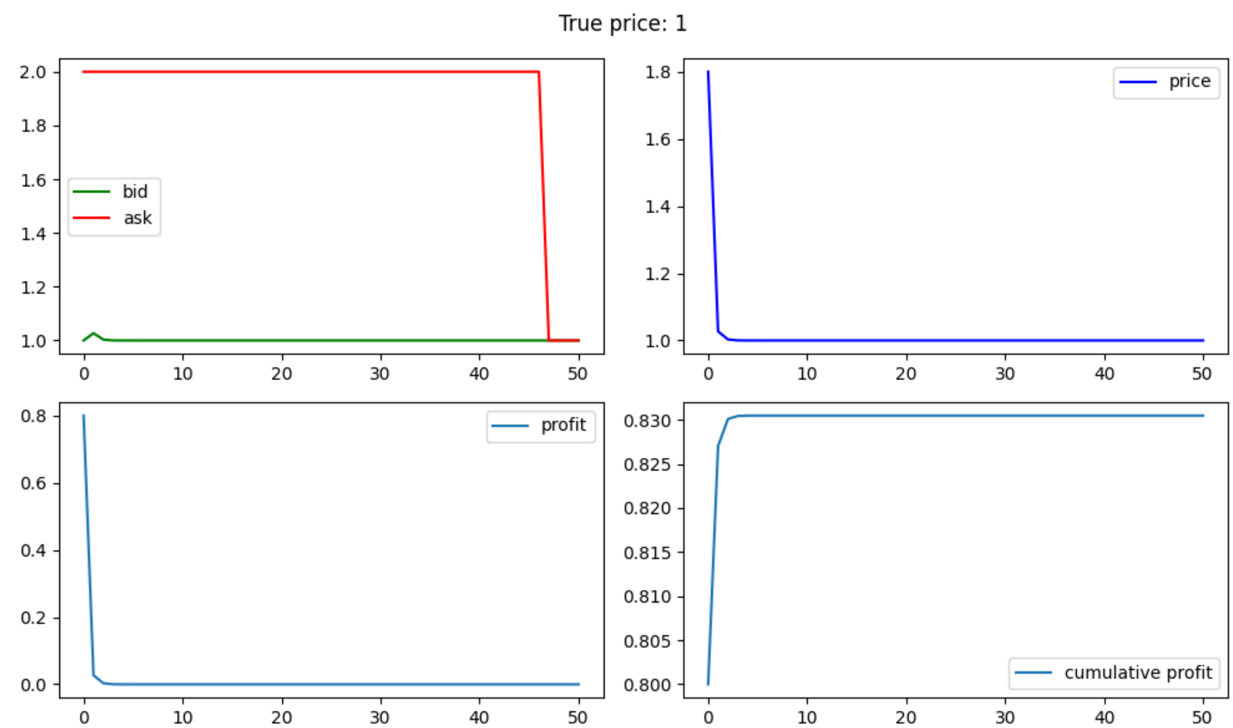


Figure 9. Simulation Results

References

Alder, B.J. and Wainwright, T.E. (1970). Decay of the Velocity Autocorrelation Function. *Physical Review A*, 1(1), pp.18–21. doi:https://doi.org/10.1103/physreva.1.18.

Álvaro Cartea, Jaimungal, S. and José Penalva (2015). *Algorithmic and high-frequency trading*. Cambridge: Cambridge University Press.

Harris, L. (2003). *Trading and exchanges : market microstructure for practitioners*. Oxford ; New York: Oxford University Press.

Hu, Y. (2018). Short-horizon market efficiency, order imbalance, and speculative trading: evidence from the Chinese stock market. *Annals of Operations Research*, 281(1-2), pp.253–274. doi:https://doi.org/10.1007/s10479-018-2849-4.

Maiti, M. (2018). OLS versus quantile regression in extreme distributions. *Contaduría y Administración*, 64(2), p.102. doi:https://doi.org/10.22201/fca.24488410e.2018.1702.

O'hara (1997). *Market microstructure theory ...* Blackwell Publishers.

Yang, L. (2021). Last hour momentum in the Chinese stock market. *China Finance Review International*, 12(1), pp.69–100. doi:https://doi.org/10.1108/cfri-06-2021-0106.

Appendix

GitHub link for the code:

https://github.com/conquerorpulkit/HFT/blob/73e6f0ca59bfe7b8c00c690168957428e2060c3c/PulkitGuptaHFTCW1.ipynb