THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

AMA 2233

Data Analytics and Visualization

Yancheng Yuan

Department of Data Science and Artificial Intelligence

Recap of last lecture

Simple Random Walk

Simple Random Walk: A walker stands at the position S_0 at time t = 0. Each time step, the walker will randomly choose to walk up (with probability p) or walk down for one unit. Mathematically, assume X_t be independent identical Bernoulli random variable with mass function

$$P(X_t = 1) = p, \quad P(X_t = -1) = 1 - p,$$

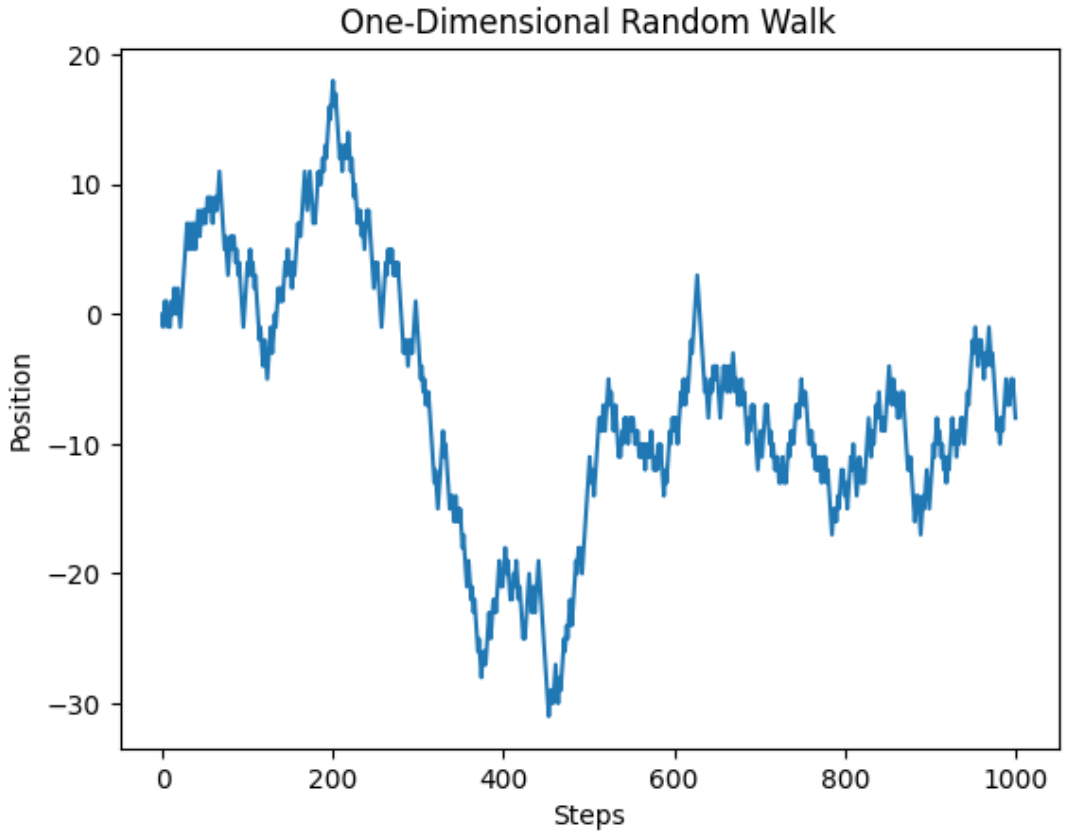
then

$$S_n = S_0 + \sum_{t=1}^{n} X_t$$

is the position of the walker located after n steps of random walks starting from the initial position S_0.

Remark: We can regard S_n as a time series.

Visualization of 1D Random Walk



An application

Gambler's ruin Theorem: To buy a new watch at price N dollars, the probability that he is ultimately bankrupted starting with k dollars is

$$1 - \frac{k}{N}$$

Implication: Without a good strategy, "to use a small matter to achieve great things" is almost impossible.

Properties of the simple random walk

Theorem: The simple random walk is spatially homogeneous, that is

$$P(S_n = j \mid S_0 = a) = P(S_n = j + b \mid S_0 = a + b).$$

Theorem: The simple random walk is temporally homogeneous, that is

$$P(S_n = j \mid S_0 = a) = P(S_{n+m} = j \mid S_m = a).$$

Theorem: The simple random walk has Markov property, that is

$$P(S_{n+m} = j \mid S_0, S_1, \dots, S_m) = P(S_{n+m} = j \mid S_m).$$

Answer to Question 1

Theorem: For a simple random walk,

$$P(S_n = b \mid S_0 = a) = \binom{n}{\frac{1}{2}(n+b-a)} p^{\frac{1}{2}(n+b-a)}(1-p)^{\frac{1}{2}(n+b-a)}.$$

Count the paths

Motivation: To answer Question 2, we need to count the number of random walk paths $N_n(a,b)$ that $S_0 = a$ and $S_n = b$. Also, we need to count the number of random paths among them that revisits the x-axis.

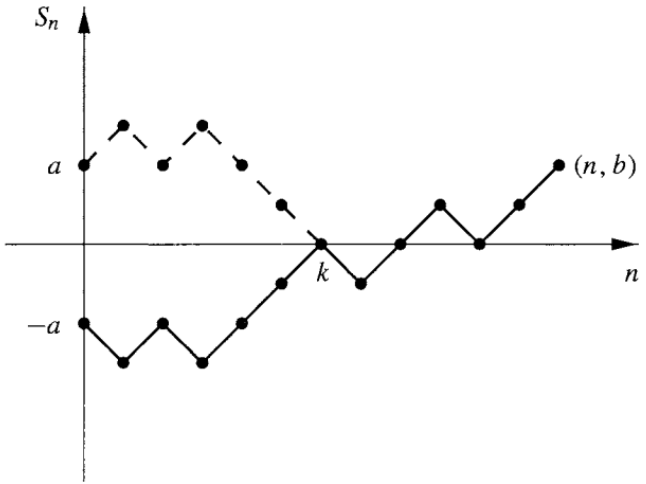
Theorem: For a simple random walk,

$$N_n(a,b) = \binom{n}{\frac{1}{2}(n+b-a)}.$$

The Reflection Principle

Theorem (The Reflection Principle): For a simple random walk and some $a, b > 0$, the number of random paths satisfying $S_0 = a$ and $S_n = b$ and intersecting with the x-axis somewhere is given by

$$N_n^0(a, b) = N_n(-a, b).$$



Ballot Theorem

Ballot Theorem: For a simple random walk and some $b > 0$, the number of random paths satisfying $S_0 = 0$ and $S_n = b$ and does not revisit the x-axis is

$$\frac{\boldsymbol{b}}{\boldsymbol{n}} * \boldsymbol{N_n(0, b)}.$$

Recap the portfolio management problem

Portfolio management with Maximum Drawdown Constraint: If you are a portfolio manager with initial funding M dollars from the investors. The investors put a constraint that the maximum drawdown during the investment period should not exceed 10% of the funding M. If the profit of your portfolio is just a random walk (increase absolute 1% or drop absolute 1% daily with equal probability), What is the probability is your portfolio ends with at least 10% profit after 364 trading days?

Question: What is the probability value in theory?

Answer: **0.25439**

Agenda

1. Introduction to Financial Time Series and Visualization
2. Moving Average
3. Two Simple Trading Strategies

Introduction to Financial Time Series and Visualization

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

Traditional Financial Data for Listed Companies

Market Data: Market data refers to the live streaming of trade-related data. It encompasses a range of information such as **price (Open/Close)**, **bid/ask quotes** and **market volume**.

Other types of Traditional Financial data: financial statements, press releases, and SEC filings, and so on.

Alternative Data

Alternative data: Alternative data includes external data from non-traditional sources such as financial transactions, web data, social apps, geolocation data, and so on.

Who generates alternative data? Alternative data is generated by individuals, business processes, sensors, and so on.

Where to find alternative data? Web scraping, raw data acquisition, and third-party licensing.

We will focus on market data for this lecture.

Market Data Source

Market Data Source: There are a lot of market data provider nowadays. In this lecture, we will use **Yahoo financial data**.

How?

```
1 # Import packages
2 import yfinance as yf
```

Available Data: The **yfinance** Api provides market data for publicly listed companies with different frequencies ranging from **1 minute** to **3 months.**

A first glance

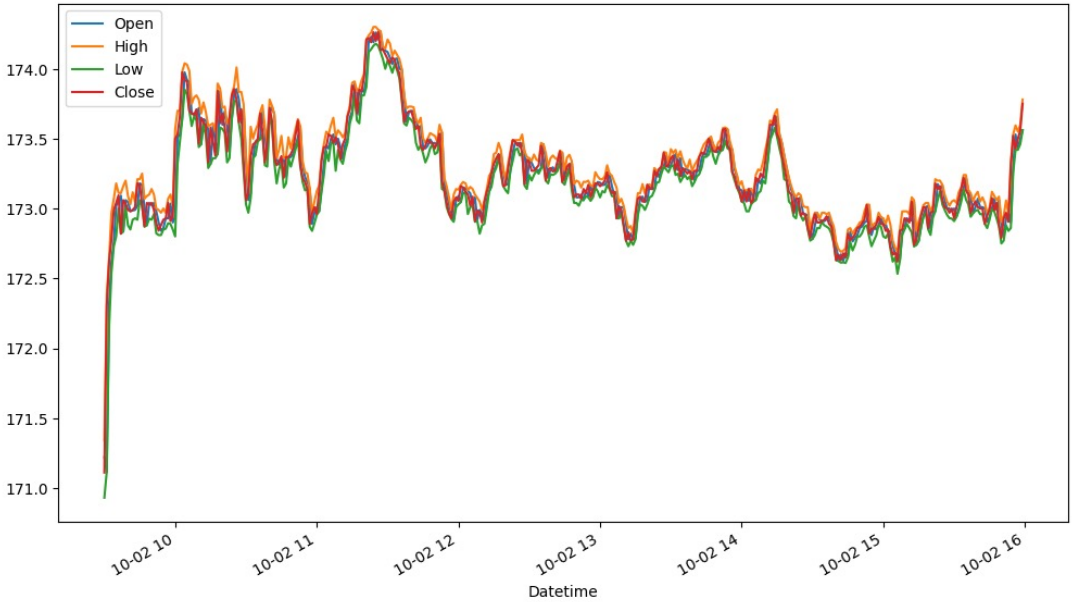
Date	Open	High	Low	Close	Volume	Dividends	Stock Splits
2000-01-03 00:00:00-05:00	0.794799	0.852586	0.770643	0.848323	535796800	0.0	0.0
2000-01-04 00:00:00-05:00	0.820377	0.838376	0.766854	0.776801	512377600	0.0	0.0
2000-01-05 00:00:00-05:00	0.786273	0.837902	0.780590	0.788168	778321600	0.0	0.0
2000-01-06 00:00:00-05:00	0.804273	0.810904	0.719961	0.719961	767972800	0.0	0.0
2000-01-07 00:00:00-05:00	0.731329	0.765433	0.723751	0.754065	460734400	0.0	0.0
...	...	...	...	...	...	...	...
2023-09-26 00:00:00-04:00	174.820007	175.199997	171.660004	171.960007	64588900	0.0	0.0
2023-09-27 00:00:00-04:00	172.619995	173.039993	169.050003	170.429993	66921800	0.0	0.0
2023-09-28 00:00:00-04:00	169.339996	172.029999	167.619995	170.690002	56294400	0.0	0.0
2023-09-29 00:00:00-04:00	172.020004	173.070007	170.339996	171.210007	51814200	0.0	0.0
2023-10-02 00:00:00-04:00	171.220001	174.300003	170.929993	173.750000	52115400	0.0	0.0

5975 rows × 7 columns

Visualize the Stock Price Data

Commonly Used Stock Price Data: Open Price, Close Price, Highest Price, Lowest Price.

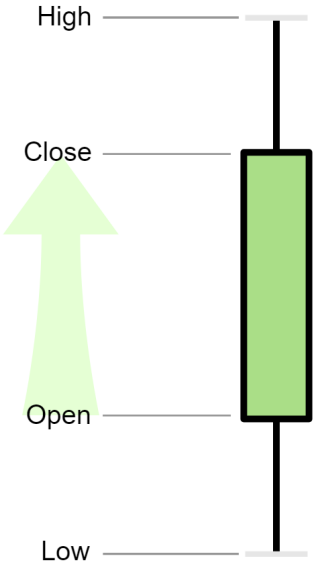
A first visualization idea: Plot a curve for each of them.



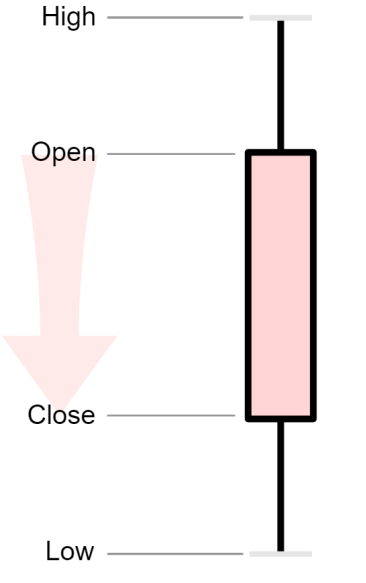
Candlestick Chart

Candlestick Chart: Candlestick represents four important pieces of information for that day: **open** and **close** in the thick body, and **high** and **low** in the "**candle wick**".

Increasing :
Bullish Candle Stick



Decreasing :
Bearish Candle Stick



An example


200
190
180
170
160
150
140
130
Price
5.0
2.5
0.0
−2.5
−5.0
150
100
50
Volume 10^6
2022-Oct-03
2022-Dec-13
2023-Feb-27
2023-May-09
2023-Jul-21
2023-Oct-02

Moving Average

Moving Average

Moving Average (MA): a moving average (rolling average or running average) is a calculation to analyze data points by creating a series of averages of different seA moving average is commonly used with time series data to smooth out short-term fluctuations and highlight longer-term trends or cycles. lections of the full data set.

Reason for MA: A moving average is commonly used with time series data to **smooth out short-term fluctuations and highlight longer-term trends or cycles**.

Low Pass Filter: Let $x = \{\dots, x_{t-s}, x_{t-s+1}, \dots, x_{t-1}, x_t, x_{t+1}, \cdots\}$ be a given time series and $w = \{\dots, w_{t-s}, w_{t-s+1}, \dots, w_{t-1}, w_t, w_{t+1}, \cdots\}$ be a filter (some specified weights), then the low frequency signal of the time series x under the filter F is

$$y_t = \sum_{i=-\infty}^{+\infty} w_i x_{t-i}$$

Delay Effect of Simple Moving Average

Moving Average in the time window T: Since we will not have infinite number of data points in practice, we usually consider the average in a given time window T:

$$y_t = \sum_{i=-(T-1)/2}^{(T-1)/2} w_i x_{t-i}$$

No Future Data: Since we do not have future data after time t, the above formula is not practical in practice. Instead, we use the following formula:

$$\hat{y}_t = \sum_{i=0}^{T-1} w_i x_{t-i}$$

Delay Effect of Simple Moving Average

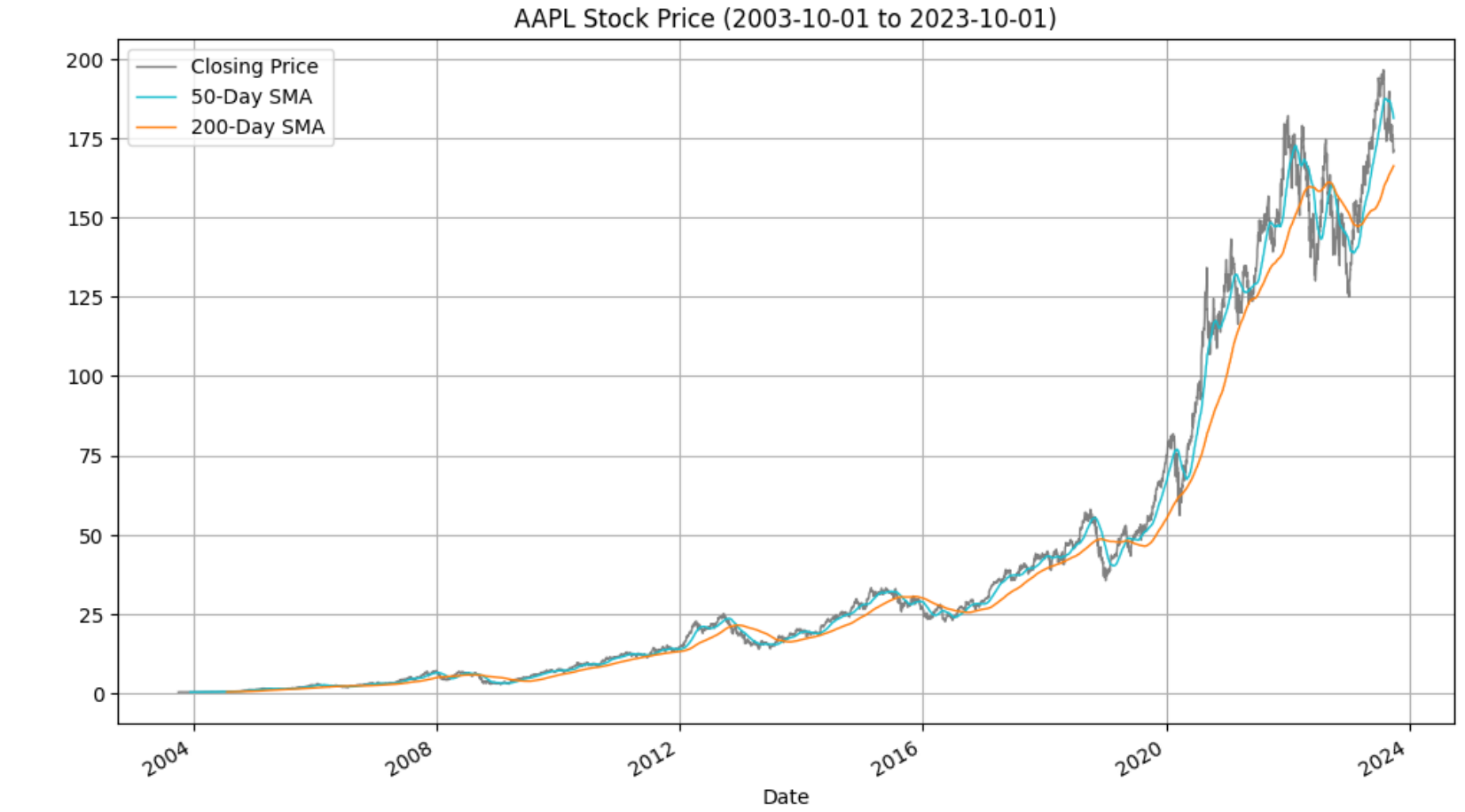
Delay effect of Simple Moving Average: We call it a simple moving average if we compute $\hat{y}_t$ as

$$\hat{y}_t = \frac{\sum_{i=0}^{T-1} x_{t-i}}{T}$$

Time Delay:

$$\hat{y}_t = \frac{\sum_{i=0}^{T-1} x_{t-i}}{T} = \frac{\sum_{j=-(T-1)/2}^{(T-1)/2} x_{t-(T-1)/2-j}}{T} = y_{t-(T-1)/2}$$

A simple visualization



Alleviate Time Delay Effect

Strategy to Alleviate Time Delay Effect: Instead of using uniform weights, we can put higher weights for the more recent data

$$\hat{y}_t = \sum_{i=0}^{T-1} w_i x_{t-i}$$

Linearly Decay Weights:

$$\hat{y}_t = (x_{t-T+1} + 2x_{t-T+2} + \cdots + Tx_t)/T(T+1)/2$$

Exponentially Decay Weights:

$$\hat{y}_0 = \alpha x_0, \hat{y}_t = \alpha x_t + (1 - \alpha)\hat{y}_{t-1}.$$

A typical choice for α is $\alpha = \frac{2}{T+1}$ if we use a T time window.

Two Simple Trading Strategies

Trend Trading with SMA

Trend Trading Strategy: Trend trading is a strategy that involves using technical indicators to identify the direction of **market momentum**.

How to identify a trend?

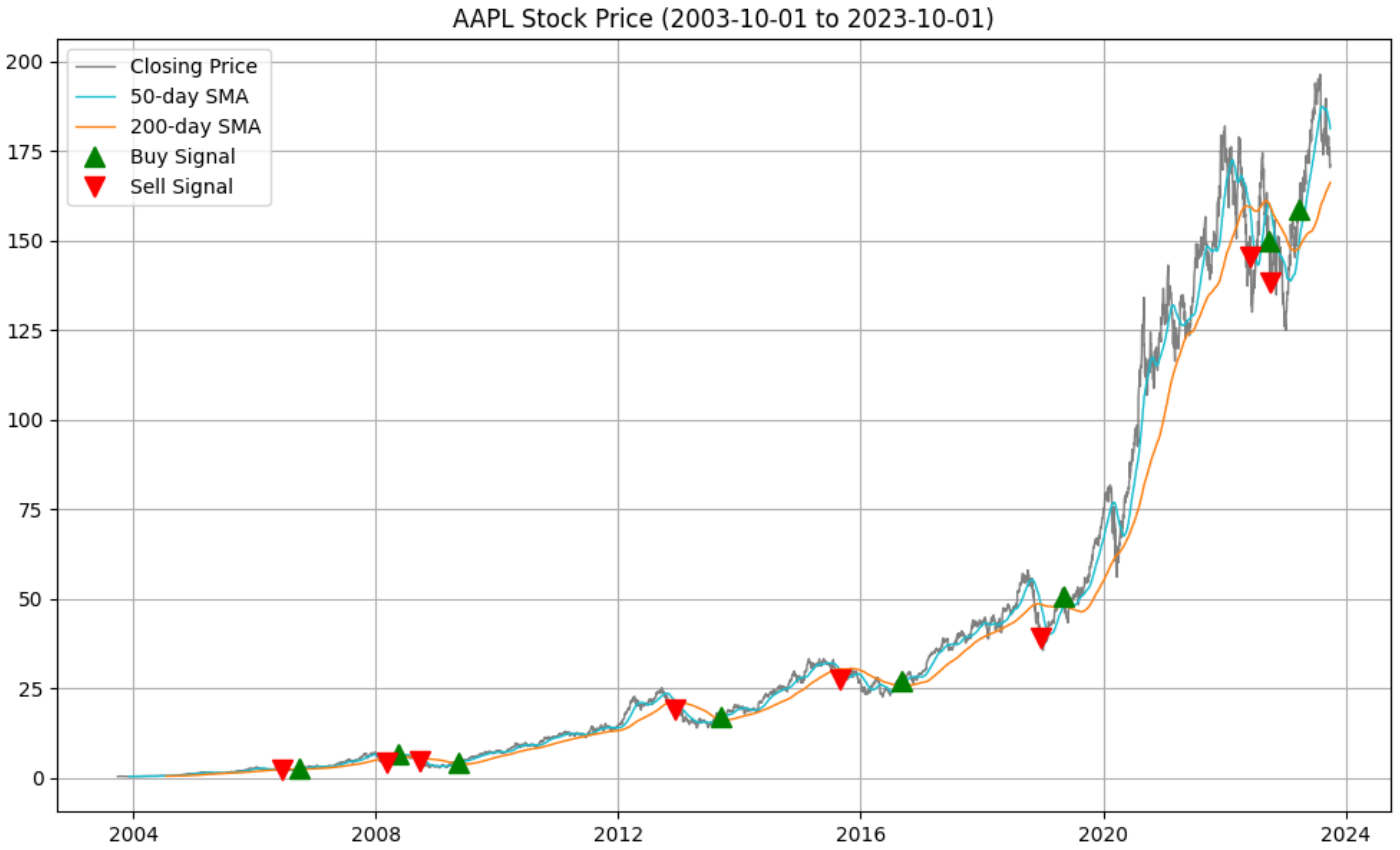
Two useful indicators

Principle idea: We can compare a pair of simple moving averages with each covering different time frames. If a shorter-term simple moving average is above a longer-term average, an uptrend is expected. On the other hand, if the long-term average is above a shorter-term average then a downtrend might be the expected outcome.

Death Cross: A death cross occurs when the 50-day SMA crosses below the 200-day SMA. This is considered a bearish signal, indicating that further losses are in store.

Golden Cross: The golden cross occurs when the 50-day SMA breaks above the 200-day SMA. This is considered a bullish signal, which implies further gains are in store.

Death Cross and Golden Cross



SMA Based Trend Trading

Trading Strategy:

Buy stocks with all cash when a **golden cross occurs**.

Sell all stocks when a **death cross occurs**.

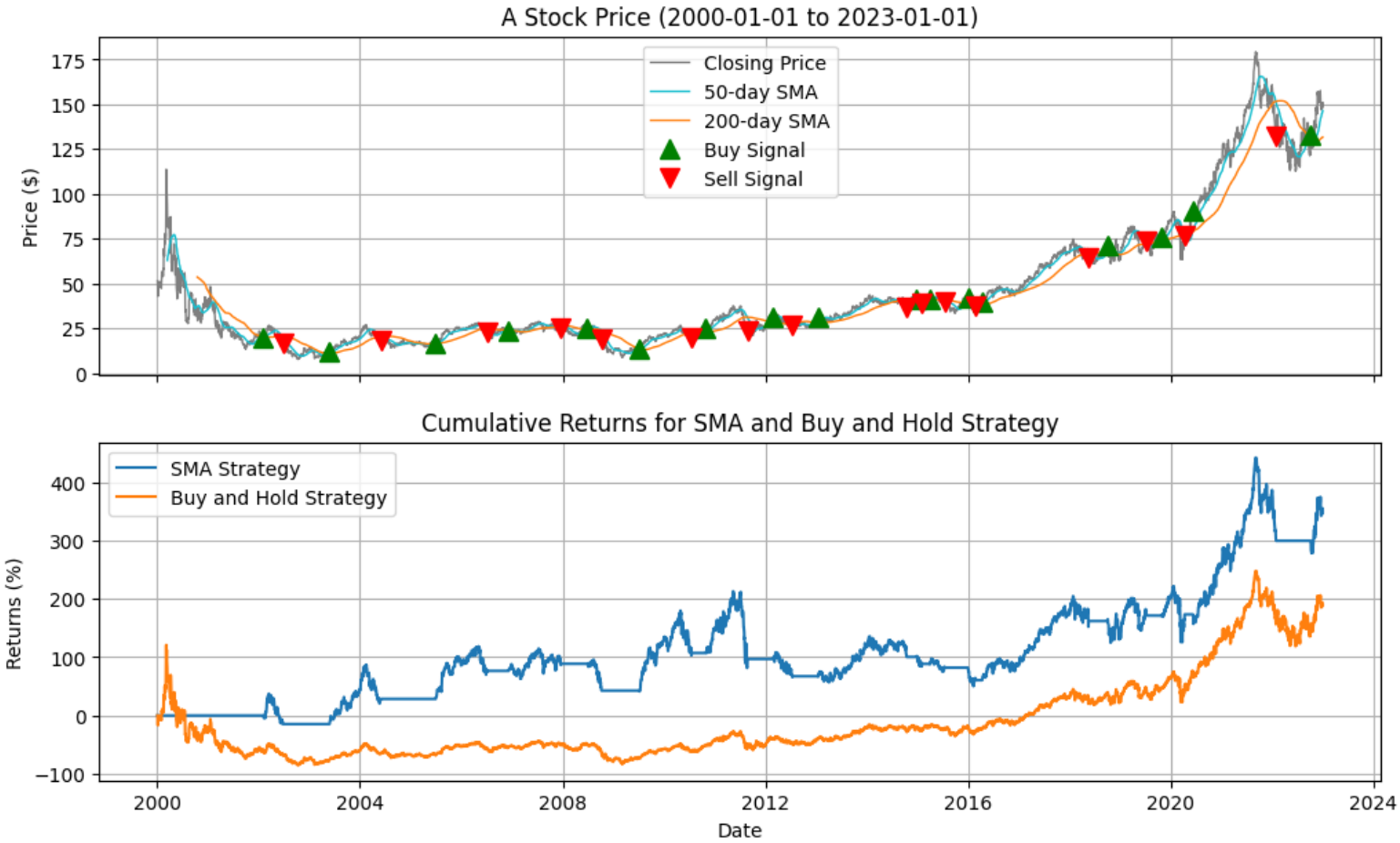
Evaluate the Trading Strategy: Backtesting

Backtesting: Backtesting is to evaluate how a trading strategy would have performed in the past. It aims to determine whether a trading strategy is worth implementing in the live market.

How? A backtest utilises historical data to qualify a strategy. A long-term investor may have to consider many years back, whereas a short-term trader can use data of several weeks or a couple of months. Backtesting will generally follow the steps below:

1. Collect historical stock data.
2. Define the trading strategy and parameters.
3. Look for trades that meet the conditions of the strategy.
4. Determine the net return based on the strategy.

Backtesting SMA Based Trend Trading on AAPL



Moving Average Convergence/Divergence (MACD)

MACD Line: MACD is a trend-following momentum indicator that shows the relationship between two exponential moving averages (EMAs). The MACD line is calculated by subtracting the 26-period EMA from the 12-period EMA.

Signal Line: 9-day EMA of the MACD line.

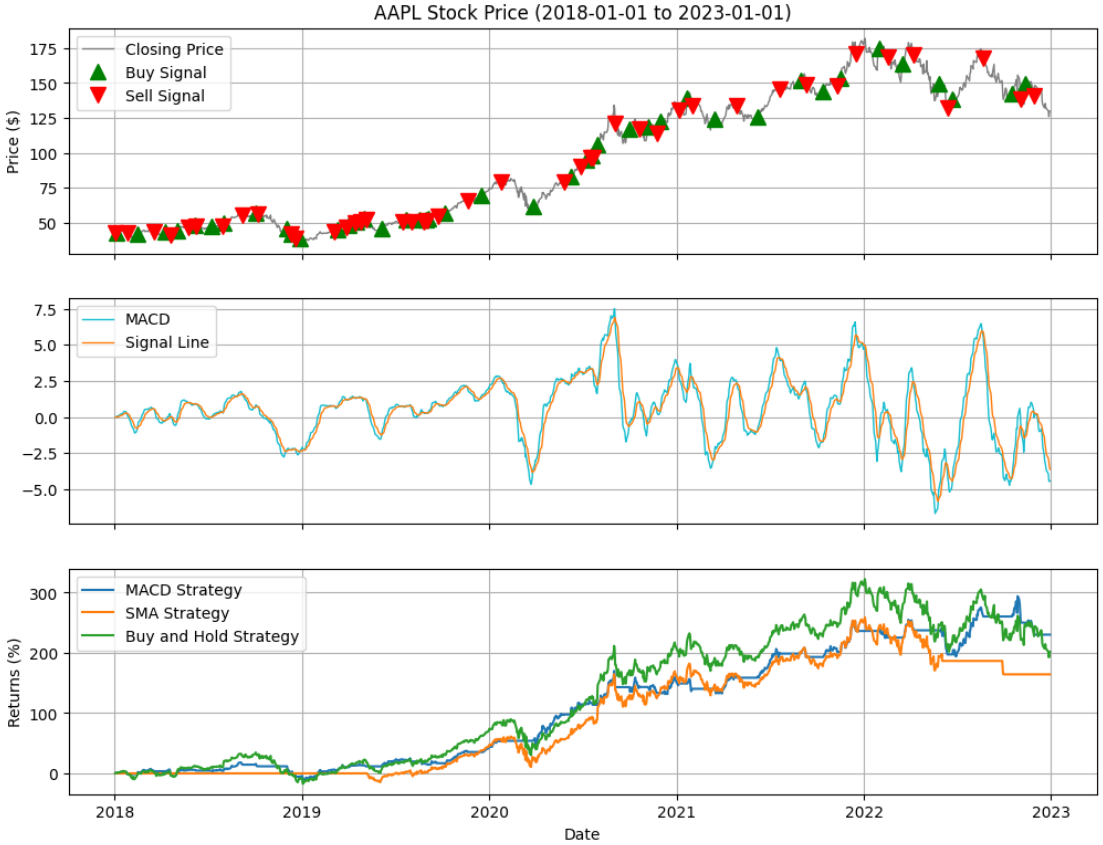
Trading Strategy:

Buy stocks with all cash when a **MACD Line crosses above the Signal Line**.

Sell all stocks when a **MACD Line crosses below the Signal Line**.


200
190
180
170
160
150
140
130
Price
5.0
2.5
0.0
−2.5
−5.0
150
100
50
Volume 10^6
2022-Oct-03
2022-Dec-13
2023-Feb-27
2023-May-09
2023-Jul-21
2023-Oct-02

Backtesting



Some observations: It seems that Trend Trading Strategy is mysterious, some time it works well, sometimes it pushes you to the darkness.

Why? Is it possible to say more about the trend trading strategy?


Thank You!