# University of Cantabria

Science School

# Quantitative and Machine Learning Strategies For The Standard and Poor's 500 Index

Master's Degree dissertation to access to

**Master's in Data Science**

Author: Pablo Calatayud Pelayo
Director: Carlos López Gutiérrez

September 2022

This page is intentionally left blank

# Contents

**Abstract**

Lorem ipsum dolor..
**Key words:** A | B

# Introduction

The motivation of this Master's Degree Dissertation is the opportunity to analyse a matter that links data science with trading investing methods for the financial markets, designing strategies based on quantitative algorithms and machine learning to optimize the operational investing and take the human bias out of the equation with quantitative algorithmic trading.

Several decades ago, technology development in terms of hardware and software achieved the digitalization of the financial markets and the transformation in their relationship with the financial operators. The first IBM computers reached the New York Stock Exchange (NYSE) starting the digitalization path which completed its journey in 2005 when the NYSE Hybrid Marked was launched, creating a unique blend of floor-based auction and electronic trading [6].

On the one hand, this development has increased the range of algorithmic trading not only to the top financial organizations but also to the retail investors who have the chance to participate in this kind of trading through different digital brokers which are more and more competitive as its commissions have continuously decreased.

On the other hand, the development of complex technical algorithms supported by statistics models and machine learning techniques has increased the appeal of the market operators in these methods. Market actors want to utilize all these new techniques to search for market inefficiencies that can just be found through powerful computational algorithms, optimizing decision-making when it comes to investing in the markets.

This document is structured in six chapters

- In the first chapter, different technical analysis methods for trading are shown. Chartism trading is explored briefly, and the history and evolution of quantitative trading are explained in detail. Lastly, the concept of quantitative algorithmic trading and its particularities in comparison with discretionary trading is delved into.

- In the second chapter, financial markets and financial assets are described. Furthermore, primary and secondary markets along with the different trends and their categories are illustrated.

- In the third chapter, extraction, load, and data transformation processes are explained as well the different data sets to be used are described.

- In the fourth chapter, TBC
- In the fifth chapter, TBC
- In the sixth chapter, TBC

Scripts have been coded in Python, and it can be downloaded from Git-Hub repository link in the annex at the end of the work.

# Chapter 1

# Trading Systems

There are several types of trading strategies and systems for approaching the markets, but the main classification divide trading analysis into two opposite categories, fundamental analysis, and technical analysis. On the one hand, fundamental analysis tries to settle a theoretical price base on the company metrics like earnings, expenses, assets, and liabilities, considering the microeconomic and macroeconomic context of the business. Investing related to fundamental analysis is external to this work. On the other hand, technical analysis focuses on the balance between supply and demand analysing the price movement and the transaction volume, considering that the current price shows all the information available for the market players. Technical analysis is divided into two different branches, chartist analysis, and quantitative analysis.

## 1.1 Chartist analysis and Quantitative analysis.

Firstly, chartist analysis consists of studying charts and graphs generated by historical prices and looking for price patterns to determine trends. For instance, one famous pattern is the cup with a handle that theoretically predicts a price breakout.

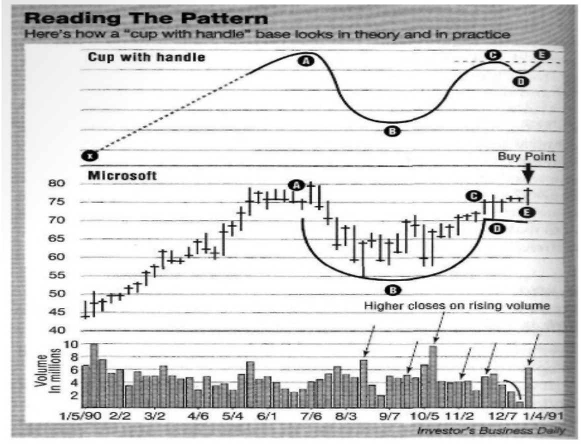


Figure 1.1: Example of the popular pattern cup with handle for Microsoft Corporation. [14].

Secondly, quantitative analysis is based on mathematical computation tools like indicators and oscillators that are built from historical prices and transactions. These techniques not only discard chartist partiality, but they enable the creation of a set of logical rules which can be combined to create quantitative algorithmic trading. This kind of trading tries to reduce the manual risk and psychological risk, taking the human bias out of the decisions.

Figure 1.2: Technical indicators like channels and moving averages at 200, 100, and 50 periods for Nasdaq-100 index. Graph built with TradingView

## 1.2 History of the quantitative analysis for the financial markets

It is difficult to establish who was the first quantitative investor, but it is clear that the first pioneer in this field was Harry Markowitz with his Modern Portfolio Theory (MPT) explained in the paper "Portfolio Selection" which was published in the Journal of Finance in 1952. Markowitz developed a mathematical model where an investor could achieve their best results by choosing an optimal mix of risk and return through diversification. Years after this great achievement, in 1973, Fischer Black and Myron Scholes published the Black-Scholes equation for the value of an option. That year later, Robert Merton establish a deeper argument behind their equation to estimate the theoretical value of options. Robert Merton formalism came to supplant theirs, and became the standard. As a consequence, the model is called by some authors as the Black-Scholes-Merton model [5]. These two milestones are considered the foundation of quantitative finance, and they were rewarded with the Nobel Prize in 1990 and 1997 respectively [12].

Trading quantitative systems were put into practice in the 70s, and they had experienced a meteoric growth in the 80s that continues until today because of technological development, reduction of the computation costs, and the beginning of financial engineering subject. In 1982, the mathematician Jim Simons founded Renaissance Technologies, which is one of the first funds that started to use quantitative analysis, developing mathematical models to create investing strategies. Jim Simon was convinced that there were mathematical expressions not yet discovered to predict the market. Even though the inner workings of Renaissance Technologies remain a mystery, it is considered one of the best funds with annualized returns of 39.1% from 1988 to 2018 according to its reports.

**Returns Comparison**

| Investor | Key Fund/Vehicle | Period | Annualized Returns* |
|---|---|---|---|
| Jim Simons | Medallion Fund | 1988–2018 | 39.1% |
| George Soros | Quantum Fund | 1969–2000 | 32%* |
| Steven Cohen | SAC | 1992–2003 | 30% |
| Peter Lynch | Magellan Fund | 1977–1990 | 29% |
| Warren Buffett | Berkshire Hathaway | 1965–2018 | 20.5%* |
| Ray Dalio | Pure Alpha | 1991–2018 | 12% |

Figure 1.3: Historical returns comparison for some of the most important funds. [18].

The previous table shows the potential of quantitative trading.

## 1.3 Different types of quantitative algorithmic trading

There are several types of quantitative algorithmic trading methods, each with its costs and benefits, and not all of them are workable for the different market actors because of their characteristics and requirements.

### 1.3.1 High-Frequency Trading (HFT)

HFT is available for just a few financial institutions because it requires high-tech software development, intensive use of computational resources, and large amounts of capital. HFT is used by some of the best funds and the most important investment banking departments. It is estimated that just 2% of all financial institutions can access HFT, and 60% of NYSE transactions are made through this kind of trading [2]. HFT consists of transacting a large number of orders in fractions of a second, where speed is a key player. The difference between the acquisition and sell price is tiny, but because of the large number of transactions, there are high returns. This kind of trading is criticized by some market players, and some of its activities have been researched by the U.S. Securities and Exchange Commission (SEC). Defenders of HFT say that it apports high market liquidity, and does not affect long-term investors. The next article in the New York Times explains how HFT is used by financial institutions [16].

### 1.3.2 Arbitrage Trading (AT)

Arbitrage trading looks to profit from financial market opportunities. The opportunities occur from financial market inefficiencies and mispricing between different markets. Here, speed is key too, and opportunities don't usually last long because of computers' ability to scan these kinds of chances. Similar to HFT, this technique can only be implemented by some of the most capable financial institutions.

### 1.3.3 Medium and Low-Frequency Trading

Resources, strategies, and aims of this kind of trading are different in comparison with the previous ones. Market actors who use medium and low-frequency trading have worse market information and lower computational resources and slippage to be aware of. Slippage is the difference between the price at which the order should have been executed and the real price the order is done, these deviations are made in a fraction of a second, and it can have a high consequence on the operation outcome.

In consequence, a higher average profitability per trade (APPT) is required to compensate for slippage and commissions in comparison with HFT. A higher APPT can only be a result of spending more time in the market, which means accepting a higher risk. Strategies made in this dissertation are formulated on this kind of trading. There are several kinds of medium and low-frequency trading methods, but the most popular are day trading, scalping, and swing trading.

#### Day Trading

Day Trading tries to take advantage of small inefficiencies that are negligible for the funds. These operations do not take longer than one session and must be implemented in small timeframes of hours and minutes. Day-trader investors are supported by indicators of volume, volatility, price movement, and momentum among other techniques, and they have to consider some daily effects like First Hour Effect and Last Hour Effect where the session volatility is concentrated. The main disadvantage of this kind of trading is the high latency of operations, which derivates in a high number of commissions.

The advantage of day traders is that they are indifferent to closed market occurrences, so it removes the closed market risk.

#### Scalping

Scalping can be placed between day trading and HFT. Similar to day trading, scalping is geared towards profiting from minor price changes in a stock's price. Scalpers can place hundreds of trades in a day in the belief that small profits can be compounded into large gains. In this case, the timeframes are of seconds which means high latency trading and like day trading the main risk is to compensate for the commissions with the small earnings per trade. Scalping strategies are rarely profitable, and most of the scalpers end in bankruptcy, leaving the broker as a unique beneficiary.

#### Swing Trading

Swing trading can hold the trades open for different temporal windows. It tries to capture short-medium term price movement, and it can vary from one day to several weeks or months. In general, swing traders have more unprofitable trades than profitable because of the stop-loss protection, but they have a favourable risk/reward owing to the bigger size of the winning trades. APPT is higher than in the two previous methods, and the commission expenses are lower because of the fewer numbers of trades.
The strategies discussed in this dissertation are primarily swing trading and day trading.

## 1.4 Pros and Cons of Quantitative Algorithmic Trading

Some main advantages have been mentioned previously. Firstly, it allows the creation of a set of logical rules, casting aside the partiality of chartism, news, or human emotions. Secondly, quantitative strategies can be back-tested and optimized to the historical data and even though favourable past results do not guarantee a similar future outcome, it helps to have a better understanding of the strategy. Lastly, once a strategy is designed, it is effortless to implement and analyse the same strategy in different assets.

The disadvantages of algorithmic trading are that it requires a continual follow-up of the system. It is necessary to track that the system is working properly, the orders are being executed at the right timing and price, and although it was a good strategy for historical data, the market may have changed and under the new market conditions, the strategy is not profitable anymore.

# Chapter 2

# Financial Markets and Financial Assets

## 2.1 Financial Markets

In order to understand the magnitude of the financial markets, it is of note that around 160.95 trillion USD were traded during 2021 in 46 billion transactions according to the World Federation of Exchanges [7], These figures can be given some context by considering that this amount of money corresponds to 190% of the world GDP in 2020 which was estimated at 84.75 trillion USD [3]. However, other data resources increase this last number to 1000% of the global GDP.

From the perspective of external market actors, the determination of an asset price is facile, but nothing could be further from the truth. Obtaining the right price is a challenging task, and financial markets play a key role in this objective. Buyers and sellers meet their interests in the financial markets, observing the market from opposite positions in order to complete a transaction. This fact reflects the subjectivity and bias of human interactions, where one side wants to buy, and another would like to sell. Financial markets trade in all types of securities and are crucial to the smooth operation of a capitalist society.

In contemplation of the financial markets, generating an index to reflect market behaviour is a common practice. The aim of these indexes is to have dynamic and precise information on market trends. For example, in Spain, there is the Iberian Index 35 (IBEX 35) as a benchmark index. This index is compounded by the 35 most important companies which fulfil some liquid, capitalisation, and volume traded requirements. Determining the weight of each component of the IBEX 35 begins with adding up the total market cap for the index by adding together the market cap of every company in the index. The weight of each company in the index is calculated by taking the company's market cap and dividing it by the total market cap of the index. As consequence, the companies with larger market capitalisation are going to have more importance in the index movements [15]. The most-traded index is the Standard and Poor's 500 (S&P 500) which is compounded by the largest 500 U.S. companies which achieve some liquidity and floating stock requirements among other characteristics. S&P 500 is a weighted index too [17].


Figure 2.1: Historical price of S&P500 from 2000 to 2022. Logarithmic scale. Plot built with TradingView.

Lastly, it should be noted that it is not possible to invest directly in a financial market index, but it can be done through several financial products. The most popular option is to invest in an Exchange Trade Fund (ETF) that replicates the S&P 500 index. ETFs will be explained ahead.

### 2.1.1 Primary and Secondary Market

Financial Markets can be classed in different ways. There are many kinds of financial markets, including but not limited to forex, commodity, stock, and bond markets. Another classification corresponds to separating the capital markets between the primary market and the secondary market [9].

#### Primary Market

Companies resort to the primary market when they sell new stocks and bonds to the public for the first time. Therefore, the market is also called the new issues market. The most common cooperative operation is the initial public offering (IPO), in this case, the company offers the securities and to complete the process it hires an underwriting financial firm to review it and create a fair price for the securities. In addition, all issues on the primary market are subject to strict regulation and companies must complete a bureaucratic policy monitoring by the SEC. A recent popular IPO arrangement is called the special purpose acquisition company (SPAC) which is based on raising money through an IPO with the advantage that an SPAC has fewer regulatory requirements, and can go public faster. Other issues included in the primary market are private placement, a rights issue, and a preferred allotment [4].

**Secondary Market**

The secondary market is the place where investors purchase and sell securities or assets from other investors. Trading that occurs on the secondary market gets its name because the share has been traded at least once from the original transaction that created the security [13].

### 2.1.2 Efficient Market

Continuing, market-making prices, it is extensive to understand the efficient market concept. There is an efficient market when prices show all information available to market players at any given time. In conjunction with the previous statement, market efficiency is linked to price speed assimilation of new information. According to this, prices move randomly until new information is revealed - Random Walk Theory. Once new information is spread, the price is going to get adjusted to the new data. The efficient market hypothesis (EMH) has sparked a controversial debate among investors about the market efficiency levels.

- Weak EMH claims that past price movements and volume do not affect a stock's price, and they cannot be used to predict future prices.
- Semi-strong EMH implies that a stock price reflects all public information relative to the company and to the macroeconomic context.
- Strong EMH is the strictest version of the EMH investment theory. It states that all public and private information is accounted for in a stock's price.

In accordance with these hypotheses, some conclusions are extracted. If weak EMH is fulfilled, the market cannot outperform by technical analysis. If semi-strong EMH is affected, fundamental and technical analyses are useless in predicting a stock's future trend. As a consequence, only material non-public information (MNPI) can be advantageous for trading. Lastly, strong EMH considers that the market cannot be surpassed [11].

In any case, consensus claims that it is not possible to predict the market in the long and medium-term. Therefore, the aim is to model the value of the stock in order to get acceptable entrance and exit signals. Some trades will be profitable and others not, but the goal is to have a consistently profitable model.

### 2.1.3 Market Trends

Trending markets are of primary interest in technical analysis. Identifying a trend can be made through a price movement or a temporal window. On the one hand, when focusing on price movement, it is possible to classify trends as a bullish trend (an uptrend), a bearish trend (a downtrend), and a neutral trend (when the price goes up and down by small movements and so is neither bullish nor bearish)

On the other hand, concentrating on a temporal window, it is possible to divide the trends into primary trends, secondary trends, and tertiary trends. A primary trend goes from 1 year up to 3 years, and it indicates the global motion direction. Inside primary trends, it is possible to observe secondary trends,


Figure 2.2: Historical price for the S&P500 from August 2020 to April 2021. Primary uptrend in green, Secondary neutral trend in yellow and lastly, tertiary downtrend in red. Graph built with TradingView.

### 2.1.4 Dow Theory

Once the previous concepts have been introduced, the Dow theory must not be overlooked The theory was introduced by Charles H. Dow who along with Edward Jones and Charles Bergstresser co-founded the Dow Jones & Company Inc. They created the popular Dow Jones Industrial Average (DJIA) which currently is a stock market index formed by the top 30 American companies. Dow also co-founded The Wall Street Journal, which is one of the most famous financial publications in the world. Because of his early death, Dow could not publish his theory, but his followers carried on his work and published it with several contributions,

1. The market discounts everything. Dow's theory accepts the EMH.
2. There are three primary kinds of market trends. Primary, secondary and tertiary trends.
3. There are three primary kinds of market phases.
   - In the bullish market: Accumulation, public participation, and excess.
   - In the bearish market: Distribution, public participation, and panic.
4. Indexes must confirm each other. This means that the signals that happen on one index must match with the signals on the others.
5. Volume must confirm the trend. Volume should follow the primary trend and decrease on the pullbacks.
6. Trends persist until a clear reversal occurs. Sometimes, It is difficult to determine whether there is a change of trend or a pullback. Dow's theory advocates caution and search for confirmation.

In the bearish market: Distribution, public participation, and panic [1].

## 2.2 Financial Assets

An asset is a store of value, over which ownership rights are enforced and from which their owners may derive economic benefits by holding or using them over a period of time. Financial assets are a subset of economic assets that are financial instruments [8]. Financial assets can be described owing to liquidity, risk, and return.

- Liquidity is the capacity to convert an asset into ready cash without affecting its market price. The most liquid assets are in the forex market.
- In financial terms, risk is considered as the chance that an outcome or investment's gains will differ from an expected outcome.
- Return is the price movement made on an investment over some period of time. Normally, returns are shown as a percentage derived from the ratio of profit to the investment.

As discussed earlier in this report, there is a key role for the pair return/risk in financial investing, and several assets have been enumerated. However, this manuscript will be concentrated on the stock market and the S&P 500 Index. The most popular option is to invest in an Exchange Trade Fund (ETF) that replicates the SP 500 Index.

### 2.2.1 Stock Market and ETF

A stock is a security that represents the ownership of a fraction of a firm. This entitles the owner of the stock to a proportion of the corporation's assets and profits equal to the percentage of stock they own. Owning stock gives you political and economical rights. For example, the right to vote in shareholder meetings, receive dividends if and when they are distributed, and it gives you the option to sell your shares to somebody else following the company and SEC policy. There are two main types of shares: common and preferred, and they differ in their political and economic benefits.

Before introducing the ETF concept, It is relevant to understand what a fund is. A fund can be described as a pool of money allocated for a concrete function. In the investing sector, the most popular is the mutual funds, which are managed by professionals who allocate the funds into different possible financial products to produce capital gains for the fund's investors. There are several kinds of mutual funds according to the securities they invest in. Investing in a share of a mutual fund is different from investing in shares of stock. The price of a mutual fund share is referred to as the net asset value per share, and shares are not bought in the stock market. The most common mutual funds are stock funds, bond funds, index funds, balanced funds, and international funds [10].

In contrast, ETFs operate like mutual funds, but they can be purchased or sold on a stock exchange as regular shares. ETFs can contain any kind of investments, including stocks, indexes, sectors, commodities, etc. The first ETF was the SPDR S&P 500 ETF (SPY), which replicated the S&P 500 Index, which remains active today.

The core of this essay is to discuss different quantitative and machine learning techniques over the S&P 500 Index. Strategies will be compared with the buy-and-hold as a benchmark.

# Chapter 3

# Extract, Load and Transformation of Historical Stock Data

A time series is a series of data points indexed chronologically. In the case of historic pricing series, the sequence of data is equally spaced in time, referred to as the *time frame* of the data, this makes it a discrete-time series. As discussed previously, different TF are required, which can vary greatly depending on the type of trading.

## 3.1 Historical stock data structure

Stock price data has always the same structure. The next chart illustrates this,

| | Open | High | Low | Close | Adj Close | Volume |
|---|---|---|---|---|---|---|
| **Date** | | | | | | |
| **2020-12-31** | 134.08 | 134.74 | 131.72 | 132.69 | 131.52 | 99116600 |
| **2021-01-04** | 133.52 | 133.61 | 126.76 | 129.41 | 128.26 | 143301900 |
| **2021-01-05** | 128.89 | 131.74 | 128.43 | 131.01 | 129.85 | 97664900 |
| **2021-01-06** | 127.72 | 131.05 | 126.38 | 126.60 | 125.48 | 155088000 |
| **2021-01-07** | 128.36 | 131.63 | 127.86 | 130.92 | 129.76 | 109578200 |

Figure 3.1: Historical daily data price for Apple Inc. Data extracted from Yahoo Finance.

Historical price data has a data frame structure where the index corresponds with a particular time and the columns with the data attributes for that precise time. The attributes of the columns are open, high, low, close, adjusted close, and volume. Python makes it simple to read, write and execute time-series data through powerful packages such as Python Data Analysis Library (PANDAS).

- Open. Price at the beginning of the period.
- High. Maximum price reached in the period.
- Low. Minimum price reached in the period.
- Close. Price at the end of the period.
- Adjusted Close. Close price considering any corporate actions [1].
- Volume. Number of trades in the period.

On the one hand, it is important to highlight that the closing price of one day and the opening price of the day after may not match because of pre-market and after-hours trading. On the other hand, it is of note the massive number of indicators, oscillators, and mathematical expressions that can be created with only these 6 attributes.

## 3.2 Data Extraction and Load

There are several websites, applications programming interfaces (APIs), and stock data providers on the internet. According to the timeframe and quality required, it is possible to opt for free or paid resources. For this manuscript, the daily stock price is accurate enough and the free API selected is Yahoo Finance. The data range selected for analysis is from 01/01/1995 to 01/06/2022. The ticker for the Standard and Poor's 500 index in Yahoo Finance is *GSPC* and the data extraction is shown in the next frame,

[1] Indexes are not disrupted by corporate actions, in consequence Adjusted close and close always match for the S&P 500 Index.

In [1]:
```python
#Importing Required Libraries
import pandas as pd
import yfinance as yf
import matplotlib.pyplot as plt
```

In [2]:
```python
#Extracting S&P 500 data from Yahoo Finance
S_AND_P_500 = yf.download("^GSPC", start = "1995-01-01", end = "2022-06-01")
```

```
[*********************100%***********************]  1 of 1 completed
```

In [3]:
```python
S_AND_P_500
```

Out[3]:

| Date | Open | High | Low | Close | Adj Close | Volume |
|---|---|---|---|---|---|---|
| 1995-01-03 | 459.209991 | 459.269989 | 457.200012 | 459.109985 | 459.109985 | 262450000 |
| 1995-01-04 | 459.130005 | 460.720001 | 457.559998 | 460.709991 | 460.709991 | 319510000 |
| 1995-01-05 | 460.730011 | 461.299988 | 459.750000 | 460.339996 | 460.339996 | 309050000 |
| 1995-01-06 | 460.380005 | 462.489990 | 459.470001 | 460.679993 | 460.679993 | 308070000 |
| 1995-01-09 | 460.670013 | 461.769989 | 459.739990 | 460.829987 | 460.829987 | 278790000 |
| ... | ... | ... | ... | ... | ... | ... |
| 2022-06-13 | 3838.149902 | 3838.149902 | 3734.300049 | 3749.629883 | 3749.629883 | 4572820000 |
| 2022-06-14 | 3763.520020 | 3778.179932 | 3705.679932 | 3735.479980 | 3735.479980 | 4126400000 |
| 2022-06-15 | 3764.050049 | 3837.560059 | 3722.300049 | 3789.989990 | 3789.989990 | 4474610000 |
| 2022-06-16 | 3728.179932 | 3728.179932 | 3639.770020 | 3666.770020 | 3666.770020 | 4511200000 |
| 2022-06-17 | 3665.899902 | 3707.709961 | 3636.870117 | 3674.840088 | 3674.840088 | 6954110000 |

6915 rows × 6 columns

Figure 3.2: Historical daily data price for S&P500 index. Data extracted from Yahoo Finance.

## 3.3 Data Transformation

There are two important transformations with respect to the historical price. Firstly, it is required to treat possible missing values. In the case of historic pricing series, there is a consensus on the treatment of null values. The consensus is to get the previous row to fill the missing value, and the perturbation generated is that the performance for the null value is 0 because there is no price change in that row.

In [6]:

```
#Backfill Missing Values
S_AND_P_500.bfill()
```

Out[6]:

| Date | Open | High | Low | Close | Adj Close | Volume |
|---|---|---|---|---|---|---|
| 1995-01-03 | 459.209991 | 459.269989 | 457.200012 | 459.109985 | 459.109985 | 262450000 |
| 1995-01-04 | 459.130005 | 460.720001 | 457.559998 | 460.709991 | 460.709991 | 319510000 |
| 1995-01-05 | 460.730011 | 461.299988 | 459.750000 | 460.339996 | 460.339996 | 309050000 |
| 1995-01-06 | 460.380005 | 462.489990 | 459.470001 | 460.679993 | 460.679993 | 308070000 |
| 1995-01-09 | 460.670013 | 461.769989 | 459.739990 | 460.829987 | 460.829987 | 278790000 |
| ... | ... | ... | ... | ... | ... | ... |
| 2022-06-13 | 3838.149902 | 3838.149902 | 3734.300049 | 3749.629883 | 3749.629883 | 4572820000 |
| 2022-06-14 | 3763.520020 | 3778.179932 | 3705.679932 | 3735.479980 | 3735.479980 | 4126400000 |
| 2022-06-15 | 3764.050049 | 3837.560059 | 3722.300049 | 3789.989990 | 3789.989990 | 4474610000 |
| 2022-06-16 | 3728.179932 | 3728.179932 | 3639.770020 | 3666.770020 | 3666.770020 | 4511200000 |
| 2022-06-17 | 3665.899902 | 3707.709961 | 3636.870117 | 3674.840088 | 3674.840088 | 6954110000 |

6915 rows × 6 columns

In [5]:

```
#Checking NAN
S_AND_P_500.isna().sum().sum()
```

Out[5]: 0

Figure 3.3: Treating the missing values for S&P500 index. Data extracted from Yahoo Finance.

Secondly, updating the corporate actions. It is necessary to adjust the open, high, and close if those attributes are to be used. In the case of the S&P500 index, there are not any corporate actions, so this step is not necessary, but it must be done as follows,

In [6]:

```
#Adjusting Historical price
S_AND_P_500_adjusted = S_AND_P_500.copy()
factor = S_AND_P_500_adjusted["Adj Close"]/S_AND_P_500_adjusted["Close"]
S_AND_P_500_adjusted["Open"]=factor*S_AND_P_500_adjusted["Open"]
S_AND_P_500_adjusted["High"]=factor*S_AND_P_500_adjusted["High"]
S_AND_P_500_adjusted["Low"]=factor*S_AND_P_500_adjusted["Low"]
S_AND_P_500_adjusted["Close"]=factor*S_AND_P_500_adjusted["Close"]
S_AND_P_500_adjusted.drop("Adj Close", axis=1, inplace=True)
S_AND_P_500_adjusted
```

Out[6]:

| Date | Open | High | Low | Close | Volume |
|---|---|---|---|---|---|
| 1995-01-03 | 459.209991 | 459.269989 | 457.200012 | 459.109985 | 262450000 |
| 1995-01-04 | 459.130005 | 460.720001 | 457.559998 | 460.709991 | 319510000 |
| 1995-01-05 | 460.730011 | 461.299988 | 459.750000 | 460.339996 | 309050000 |
| 1995-01-06 | 460.380005 | 462.489990 | 459.470001 | 460.679993 | 308070000 |
| 1995-01-09 | 460.670013 | 461.769989 | 459.739990 | 460.829987 | 278790000 |
| ... | ... | ... | ... | ... | ... |
| 2022-05-24 | 3942.939941 | 3955.679932 | 3875.129883 | 3941.479980 | 3901640000 |
| 2022-05-25 | 3929.590088 | 3999.330078 | 3925.030029 | 3978.729980 | 4322190000 |
| 2022-05-26 | 3984.600098 | 4075.139893 | 3984.600098 | 4057.840088 | 3961940000 |
| 2022-05-27 | 4077.429932 | 4158.490234 | 4077.429932 | 4158.240234 | 3560560000 |
| 2022-05-31 | 4151.089844 | 4168.339844 | 4104.879883 | 4132.149902 | 5192220000 |

6902 rows × 5 columns

Figure 3.4: Adjusting prices for any corporate actions for S&P500 index. Data extracted from Yahoo Finance.

Following the previous steps, it is possible to work with any daily historical stock data found in Yahoo Finance.

## 3.4 Dividing Historical Price

As mentioned above, the time range for analysis goes from 01/01/1995 to 01/06/2022. The next plot illustrates the price movement for that time and how the data set has been sliced to obtain training, validation, and test sets,


Figure 3.5: Historical price from 01/01/1995 to 01/06/2022 for S&P500 index. Data extracted from TradingView.

In this data division, it is possible to observe that each set has more or less pronounced ups and downs. This is an important aspect to generate a general model that works on both trends.

As the previous data shows, the data has been split as follows,

- Training. From 01/01/1995 to 31/12/2004.
- Validation. From 01/01/2005 to 31/12/2015.
- Test. From 01/01/2016 to 01/06/2022.

# Chapter 4

# Quantitative Methods

Firstly, a quantitative strategy based on indicators and oscillators will be designed. After that, the strategy will be optimized in the training and validation subset. Lastly, it will be checked in the test subset. In the next chapter, the aim is to enhance the quantitative strategy through machine learning techniques.

## 4.1 Indicators and oscillators

In this section, the indicators and oscillators used are explained.

### 4.1.1 Moving Average (MA)

There are different types of moving averages, and this kind of indicator is a common tool for any algorithmic strategy. It is used to determine the direction of a trend, and it sums up the data of a time series over a specific time period. The TF can vary from short MAs, such as 9 periods, to long MAs, such as 200 or 300 periods. In addition, it can be used with the price action or another MA to signal an action. Like a filter, moving averages smooth the noise of price action created by the temporal series, and it allows seeing a clearer trend. In contrast, there is a lag in the data observation for the current time. There are different kinds of MAs, but the most popular are the simple moving average and the exponential moving average.

### Simple Moving Average (SMA)

The simple moving average (SMA) is a straightforward technical indicator that is obtained by summing the $n$ recent data points in a given set and dividing the total by $n$. It applies an equal weight to all observations in the period,

$$SMA(n) = \frac{\sum_{i=1}^{n} P_i}{n}$$

Where $P_i$ is the price for the $i$ period and $SMA(n)$ is the price for the period $n$.

### Exponential Moving Average (EMA)

In contrast with the SMA, an EMA places a greater significance on the most recent data points. As a result of this, the EMA reacts faster to recent price changes, and it goes closer to the price than the SMA,

$$EMA(t) = P_t \cdot k + EMA(y) \cdot (1 - k)$$

Where, $t$ = last period, $y$ = next-to-last period, $N$ = number of days in EMA, $k = \frac{2}{N+1}$.

The next plot shows an example of a 50-period SMA and a 50-period EMA,


Figure 4.1: Historical daily price for the S&P 500 Index from 2018 to 2020 included. SMA and EMA for 50 days. Graph built with python.

### 4.1.2 Oscillators

**Moving Average Convergence Divergence (MACD)**

The MACD is a trend follower indicator based on two EMAs and their relationship. For the standard parameters, there is a short EMA of 12 periods and another long EMA of 26 periods. The MACD line is calculated by subtracting both EMAs. An EMA of 9 periods is then derived from the MACD line, called the “signal line”. Finally, the MACD triggers a signal when it crosses above (buying signal) or below (selling signal) its signal line.

$$MACD(c, l) = EMA(c) - EMA(l)$$


Figure 4.2: Historical daily price for the S&P 500 Index from 2018 to 2020 included. MACD with the standard parameters. Graph built in TradingView.

It is usual to show the MACD with a histogram which reflects the distance between the MACD and the signal. When the MACD is above the signal line, there is an uptrend (in green) and vice versa for the opposite case (in red). The MACD is considered an oscillator because it oscillates around zero, but it is not bounded to it.

### Relative Strength Index (RSI)

The RSI is one of the most famous indicators utilized in technical analysis. Because of the RSI mathematical expression, it varies from 0 to 100. It is considered overbought (selling signal) when the RSI values are higher than 70, and it is considered oversold (buying signal) when the RSI is lower than 30, The standard is to use the 14 periods to calculate the RSI, but for n periods the expression is as follows,

$$RSI = 100 \cdot (1 - \frac{1}{(1 + RS)})$$

where RS is the average ratio of average gain divided by the average loss

$$RS = \frac{\sum_{i=1}^{n} U_i}{\sum_{i=1}^{n} D_i}$$

and for each trading period, an upward change U or downward change D is calculated. For up periods,

$$U = close_{now} - close_{previous}$$
$$D = 0$$

for down periods,

$$U = 0$$
$$D = close_{previous} - close_{now}$$


Figure 4.3: Historical daily price for the S&P 500 Index from 2018 to 2020 included. RSI with the standard parameters. Graph built in TradingView.

In the figure 4.3, it is possible to observer how the RSI moves from 0 to 100 and how most of the time the indicator is between the range 30-70. When it crosses the overbought or oversold threshold, it triggers the signal.

### Stochastic

The stochastic oscillator is a momentum indicator, comparing a particular closing price of a security to a range of its price over a certain period of time. Similarly to the RSI, the stochastic moves from 0 to 100 using overbought and oversold signals. The stochastic formula is commonly used ofr 14 periods but for the standard configuration of n periods is the next,

$$\%K = \left(\frac{C-L_n}{H_n-L_n}\right) \cdot 100$$

Where,
$C$ is the most recent closing price
$L_n$ the lowest price traded of the n previous periods
$H_n$ the highest price traded of the n previous periods
$\%K$ is the current value of the stochastic indicator

The $\%K$ is known as the fast stochastic indicator, while the slow stochastic is built as $\%D = 3$-periods MA of $\%K$ for the standard parameters.

This stochastic indicator is referred to as the Fast Stochastic Indicator, and the Slow Stochastic Indicator incorporates a $\%K$ smoothing period of 3 for the standard set up. Changing that smoothing period to one convert the Slow Stochastic Indicator into the Fast Stochastic Indicator. The buying signal is when $\%K$ crosses above the $\%D$ line and vice versa for going short.

## 4.2 Backtesting parameters

In order to evaluate the strategies, it is important to define the parameters that are going to be analysed. As it was mentioned previously, the risk/return pair are two factors that must be taken into account. Firstly, in order to analyse the strategy profitability, the profit factor is selected. Secondly, as a deep measure of the risk strategy, the maximum drawdown is going to be chosen.

### Profit Factor

This concept allows traders to review the profitability of trading systems quickly. The profit factor is the gross profit ratio to the gross loss. A profit factor higher than one indicates a profitable strategy. Unanimity said that a convenient strategy should have a profit factor greater than 1.5. This means that you get 1.5-monetary units for each 1-monetary unit invested.

$$Profit\ Factor = \frac{GrossWinningTrades}{GrossLosingTrades}$$

As a result of this expression, it is possible to observe that the equation considers not just the return but the risk.

**Maximum Drawdown (MDD)**

Drawdowns (DD) are utilized for measuring the historical risk of an investment. A drawdown refers to how much an investment is down from the highest peak of the operation. Because of this, DD are commonly quoted as a percentage, this percentage means the decreasing percentage from the highest peak to the lower value until get that peak value again. The maximum drawdown is the higher DD over a temporal window.

$$MDD = \frac{Trough\ Value\ -\ Peak\ Value}{Peak\ Value}$$

In consequence, an MDD is an indicator of downside risk over a specified time period.

## 4.3 Quantitative Strategies

In this section

### 4.3.1 MACD

**MACD for standard parameters**

**MACD for optimized parameters**

### 4.3.2 RSI

**RSI for standard parameters**

**RSI for optimized parameters**

### 4.3.3 Stochastic

**Stochastic for standard parameters**

**Stochastic for optimized parameters**

# Chapter 5

# Machine Learning Methods

## 5.1 Data Aggregation

## 5.2 Some return Prices

## 5.3 Quantitative methods

## 5.4 S&P500 Data

# Chapter 6

# Conclusions

# Referencias


[1] Investopedia Adam Hayes. *Dow Theory*. url: https://www.investopedia.com/terms/d/dowtheory.asp.

[2] David Urraca Ordiz Andrés A. García Manzano Carlos Prieto Mota. *Estrategias y Mercados, Fundamentos de Trading Algorítmico. 6ª Ed. Vol I. Pag. 19.* 2020.

[3] The World Bank. *Current GDP in US$*. url: https://data.worldbank.org/indicator/NY.GDP.MKTP.CD.

[4] James Chen. *Primary Market*. url: https://www.investopedia.com/terms/p/primarymarket.asp.

[5] Emanuel Derman. *My life as a quant. Pag. 143-150*. John Wiley  Sons, Inc, 2004. isbn: 978-0-470-19273-3.

[6] New York Stock Exchange. *The history of NYSE*. url: https://www.nyse.com/history-of-nyse.

[7] World Federation Of Exchanges. “2021 Market Highlights”. In: ().

[8] INTERNATIONAL MONETARY FUND. *MONETARY AND FINANCIAL STATISTICS MANUAL AND COMPILATION GUIDE. Pag. 55*. International Monetary Fund, Publication Services, 2016. isbn: ISBN 9781513579191.

[9] Adam Hayes. *Financial Markets*. url: https://www.investopedia.com/terms/f/financial-market.asp.

[10] Adam Hayes. *Mutual Fund*. url: https://www.investopedia.com/terms/m/mutualfund.asp#:~:text=our%5C%20editorial%5C%20policies-,What%5C%20Is%5C%20a%5C%20Mutual%5C%20Fund%5C%3F,market%5C%20instruments%5C%2C%5C%20and%5C%20other%5C%20assets..

[11] Investopedia J.B. Maverick. *The Weak, Strong, and Semi-Strong Efficient Market Hypotheses*. url: https://www.investopedia.com/ask/answers/032615/what-are-differences-between-weak-strong-and-semistrong-versions-efficient-market-hypothesis.asp#:~:text=Though%5C%20the%5C%20efficient%5C%20market%5C%20hypothesis%5C%20theorizes%5C%20the%5C%20market%5C%20is%5C%20generally,technical%5C%20analysis%5C%20can%5C%20aid%5C%20investors..

[12] Investopedia by James Mcwhinney. *A Simple Overview of Quantitative Analysis*. url: https://www.investopedia.com/articles/investing/041114/simple-overview-quantitative-analysis.asp.

[13] Will Kenton. *Secondary Market*. url: https://www.investopedia.com/terms/s/secondarymarket.asp#:~:text=The%5C%20secondary%5C%20market%5C%20is%5C%20where,when%5C%20they%5C%20are%5C%20first%5C%20issued.

[14] William J. O'Neil. *24 Essential Lessons for Investment Success. Pag. 46*. McGraw-Hill, 2000. isbn: 0-07-136033-6.

[15] Banco Santander. *¿Qué es el Ibex 35?* url: https://www.bancosantander.es/glosario/ibex-35#:~:text=El%5C%20IBEX%5C%2035%5C%20es%5C%20el,Sistema%5C%20de%5C%20Interconexi%5C%C3%5C%B3n%5C%20Burs%5C%C3%5C%A1til%5C%20Electr%5C%C3%5C%B3nico%5C%20(.

[16] The New York Times. *The Thirty-Milliseconds Advantage*. url: https://archive.nytimes.com/www.nytimes.com/imagepages/2009/07/24/business/0724-webBIZ-trading.ready.html.

[17] Investopedia Will Kenton. *The SP 500 Index: Standard Poor's 500 Index*. url: https://www.investopedia.com/terms/s/sp500.asp.

[18] Gregory Zuckerman. *The Man Who Solved The Market. How Jim Simons Launched The Quant Revolution. Pag. 333*. Penguin Random House UK, 2019. isbn: 978-0-241-42215-1.