Data-driven portfolio management

Programming for Data Science Final Project

Authors:
Elena María Gómez Orihuel
María Lara Trullenque
Ignacio López Carboneras
Jaime Guerrero Carrasco
Javier Rocamora García

UNIVERSIDAD
POLITÉCNICA
DE MADRID

April 2023

Data-driven portfolio management

Back-testing and analysis of investing strategies. Practical assignment for the Programming for Data Science subject at the Polytechnic University of Madrid

Authors

Elena María Gómez Orihuel

María Lara Trullenque

Ignacio López Carboneras

Jaime Guerrero Carrasco

Javier Rocamora García

EIT Health Master & EIT Digital Data Science Master

MADRID, April 2023

Contents

List of Figures

1 Introduction

This **data analysis report** is intended to **analyze the portfolio information generated** in previous tasks, which consists in a dataset containing **different strategies** in the form of **portfolios composed of 5 different assets**: stocks, corporate bonds, public bonds, gold, and cash. The purpose of this analysis is to **gain insights into the performance** of various investment strategies and **understand the potential risks and rewards associated with each asset**, in order to proceed with the development of a model for automatic financial advising addressing the customers of Smallville Asset Management company. By examining the historical data, we can **identify trends** and **patterns** that may **help inform investment decisions** in the future.

The analysis will be **focused on the two questions proposed**:

- Is it more probable to obtain a positive or negative return?
- Is it ALWAYS true that the higher the risk, the higher the obtained return is?

The implementation can be accessed via *Github*[1]

1.1 Domain

Prior to every analysis is important to have a domain of the **different concepts that are going to be assessed**. Regarding the **structure of the portfolios**, as we know, they are composed by the **combination of five assets** which are stocks, corporate bonds, public bonds, gold and cash. Each of these assets has its **unique characteristics and performance drivers**, which makes them **suitable for different investment goals** and risk profiles. Overall, the performance of each asset class can be influenced by a range of economic, political, and global events, that's why **diversifying** across different asset classes can **help to manage risk and potentially improve long-term returns**. The **context characteristics of each of the assets** is as follows:

- **Stocks**: Stocks are ownership stakes in companies and represent a share of the company's profits and losses. They are considered to be a high-risk, high-reward asset due to their volatility and potential for large returns over the long term. The performance of stocks is influenced by a range of factors, including the health of the economy, industry trends, company earnings, and political developments.
- **Corporate bonds**: Corporate bonds are issued by companies to raise capital and offer investors a fixed rate of return over a set period. They are generally considered to be less

[1]Source: `https://github.com/Elena-Gomez-Orihuel/programmingDS`

risky than stocks but offer lower potential returns. The performance of corporate bonds is tied to the creditworthiness of the issuer, interest rates, and economic conditions.

- **Public bonds**: Public bonds, also known as government bonds, are issued by national governments to finance public spending. They are considered to be the safest asset in this dataset, as they are backed by the government's ability to tax and print money. Public bond yields are primarily driven by interest rates and inflation expectations.
- **Gold**: Gold is a precious metal that has long been used as a store of value and a hedge against inflation and currency fluctuations. It is considered to be a safe-haven asset in times of economic uncertainty and can provide diversification benefits to a portfolio. The price of gold is influenced by a range of factors, including global macroeconomic conditions, geopolitical tensions, and currency movements.
- **Cash**: Cash is the most liquid asset in this dataset and provides a low-risk, low-return option for investors. It can be used as a temporary store of value or to take advantage of short-term investment opportunities. The value of cash is not subject to market volatility but is affected by inflation and interest rates.

-

2 Pre-analysis

And now, prior to answering the questions regarding the analysis report assignment, it seems of **utmost importance** to **perform a pre-analysis** of the dataset containing the different generated portfolios, with the aim of **checking the different insights** that these data throw at us. This analysis can be seen in depth on the *jupyter notebook*[1] made for the same.

Looking directly at the dataset, with a **heatmap** in *figure 2.1* we can already have a **first idea** of **how the different assets are correlated with the attributes** of return and volatility:

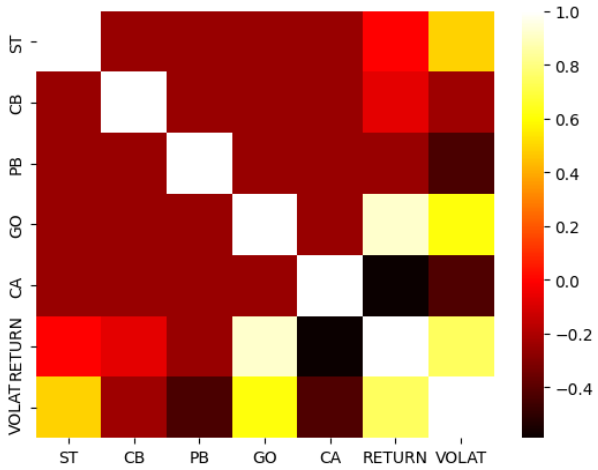


Figure 2.1: Heatmap of the dataset

In addition, looking in *figure 2.2* at the **individual contribution of each asset** on its totality to the portfolio gives us an idea of **how each asset marginally affects the portfolio** (marginal contribution of the asset to the portfolio):

Moving on to the different attributes of the portfolio, which are the main drivers of the later questions, first of all we have the '**return**' attribute in *figure 2.3*, which refers to the **anticipated amount of returns that an investment in a portfolio will generate over a period of time**, and it is calculated by taking the weighted average of the expected returns of each asset in the portfolio.

Hence, we can see a brief **summary of its statistics**. This summary already throws us

[1]Direct link to the notebook: `https://github.com/Elena-Gomez-Orihuel/programmingDS/blob/main/part3_analysis.ipynb`

	ST	CB	PB	GO	CA	RETURN	VOLAT
0	0	0	0	0	100	-6.691566	3.270975
5	0	0	0	100	0	23.904133	7.911984
20	0	0	100	0	0	0.221099	1.577834
55	0	100	0	0	0	3.918536	2.232727
125	100	0	0	0	0	5.133871	7.909930

Figure 2.2: Asset individual contribution to the dataset

```
Return Statistics:
------------------
count    126.000000
mean       5.297215
std        5.897931
min       -6.691566
25%        1.227764
50%        4.238132
75%        8.584090
max       23.904133
```

Figure 2.3: Return attribute statistics

interesting insights from the different portfolios generated. As we can observe, in this summary, the **mean return on investment** from the portfolios is **5,3%**, which means that **on average**, a **portfolio generates a return** of 5.3% over the period of time covered by the dataset.

Also it can be seen regarding the **maximum** and **minimum values** of the summary, that neither of the portfolios will, at least in the short term of a year covered, make us lose all the money and neither duplicate it, which is interesting to know also for those **people investing for their first time**. As it can be visualized, on the one hand, the **portfolio with the least return** has a **-6,69% return**, corresponding to the portfolio made out only by the cash asset. This negative result means that if we invested in a portfolio composed entirely only of the cash asset, we would end up losing money (specifically a 6,69% of the money we invested in the portfolio). This makes sense since cash is generally considered a low-risk investment due to its low expected return compared to other asset classes such as stocks and bonds (since it does not generate any income like stocks or bonds do), and also can have a negative return over time due to inflation (can erode the purchasing power of cash over time).

On the other hand, it can be seen that the **portfolio with the highest return** has a **23,9%** return, corresponding to a portfolio made up only by the gold asset. This means that if we invested in a portfolio composed entirely only of the gold asset, we would end up earning money (specifically a 23,9% more of what we invested). This also makes sense since gold has a history of maintaining its value, making gold a useful hedge against inflation and commonly known for serving as a value storage vehicle.

The **different percentiles** can be understood better in the next **box plot** in *figure 2.4*

regarding the 'Portfolio Returns'. There we can see how there are **only a couple portfolios which have atypical returns** thus giving us more return than the rest of the portfolios. Also it can be seen that the intermediate percentiles from 25 to 50 are closer to the first quartile than to the last quartile so that the majority of data are more prone to unpretentious results. The first quartile (25%) is 1.23%, which means that 25% of the returns are less than 1.23%, the median (50%) return is 4.24%, which means that 50% of the returns are less than 4.24%, and 50% of the returns are greater than 4.24%, finally the third quartile (75%) is 8.58%, which means that 75% of the returns are less than 8.58%.

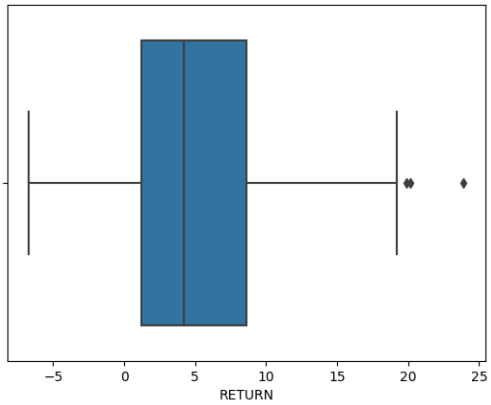
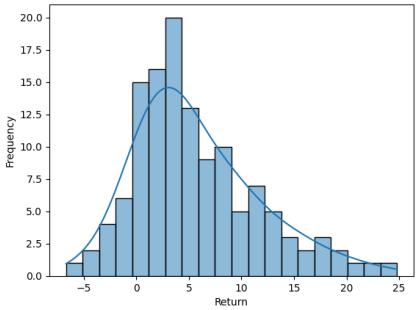
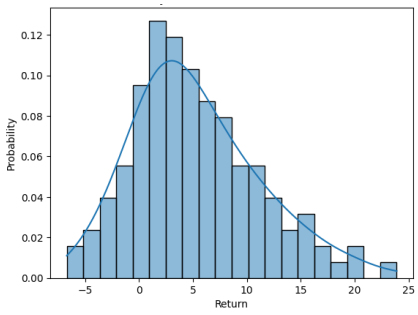


Figure 2.4: Box plot of portfolio returns

The next plot in *figure 2.5a* allow us to observe the **distribution of all the portfolios returns**. As it can be observed, as we get closer from both of the tails (begin and end tails) of the graphical plot to the mean portfolio return we commented before, more portfolio instances appear. So we could say the **tendency of the returns increases as the return rises**, and when it **reaches the mean portfolio return**, then the number of portfolio instances associated with each return **decreases** progressively. The graphic with the probability statistic in *figure 2.5b* is similar to the previous one, but in this case gives us the **percentage of portfolio instances associated with each specific return**.



(a) Distribution of Portfolio Returns



(b) Probability Distribution of Portfolio Returns

Figure 2.5: Portfolio Return Distributions

Also the next plot regarding the '**Cumulative Distribution Function of Portfolio Returns**' in *figures 2.6* allow us to see more clearly **how the major increase of portfolio returns is produced** approximately between the percentiles 25 and 75.

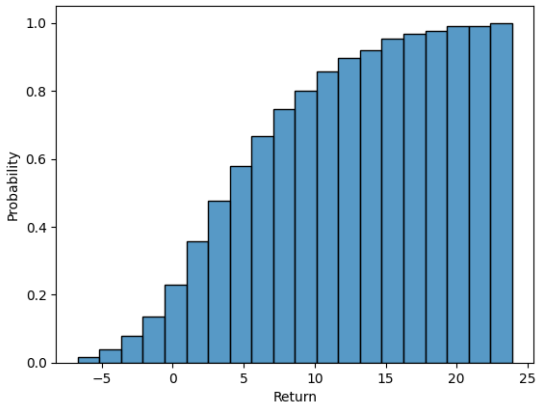


Figure 2.6: Cummulative Distribution Function of Portfolio Returns

Secondly, we have the 'volatility' attribute in *figure 2.7*, which is a measure of a portfolio risk, and refers to the tendency of a portfolio to deviate from its mean return and serves a measure of how wildly the total value of all the stocks in each portfolio appreciates or declines. Hence, we can see a brief summary of its statistics. As before, this summary also throws us interesting insights from the different portfolios generated. As we can observe, in this summary, the mean volatility of the portfolio returns is 3.204068, which is higher than the median. This indicates that there are some portfolios with high volatility levels that are pulling the mean upwards. The volatility statistics show that the standard deviation of the portfolio returns is 1.675337, meaning that the portfolio returns are quite volatile, with a significant amount of variation in the returns over time.

```
Volatility Statistics:
----------------------
count    126.000000
mean       3.204068
std        1.675337
min        1.223643
25%        1.755915
50%        2.757381
75%        4.334817
max        7.911984
```

Figure 2.7: Colatility attribute statistics

Also, it can be pointed out that there are no risk-free assets, which indicates that even filling all our capital on the least risked asset, still we will have risk in our portfolio. As it can be visualized, on the one hand, the portfolio with the least volatility has a 1,22% volatility, corresponding to the portfolio formed by the stocks and the public bonds assets (the first one on a 20% and the second one on an 80%). This could be due to a negative correlation between both of these assets. On the other side, it can be seen that the portfolio with the

highest volatility has a 7,91% return, corresponding to a portfolio composed only of the gold asset. This makes sense theoretically since, as it is seen in the analysis on the notebook, gold is the asset with the highest volatility.

The different percentiles can be understood better in the next graphic box plot in *figure 2.8* regarding the 'Portfolio Volatilities'. There we can see how, for this volatility attribute of the portfolios, there are no atypical values within the quartile range. The first quartile (25%) is 1.76%, which means that 25% of the volatilities are less than 1.76%, the median (50%) volatility is 2,76%, which means that 50% of the returns are less than 2,76%, and 50% of the returns are greater than 2,76%, finally, the third quartile (75%) is 4,33%, which means that 75% of the returns are less than 4,33%.

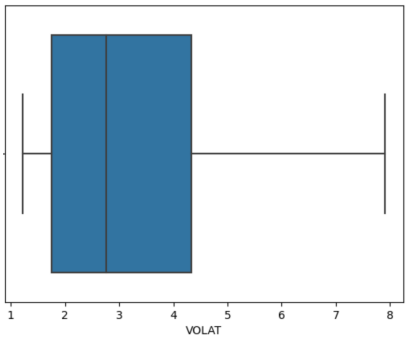
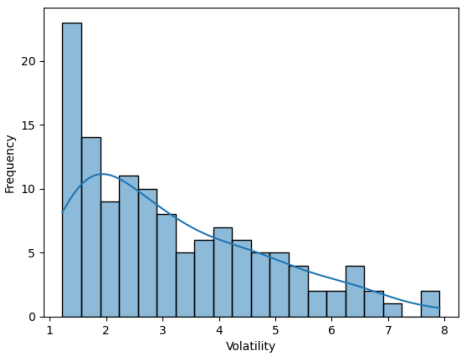
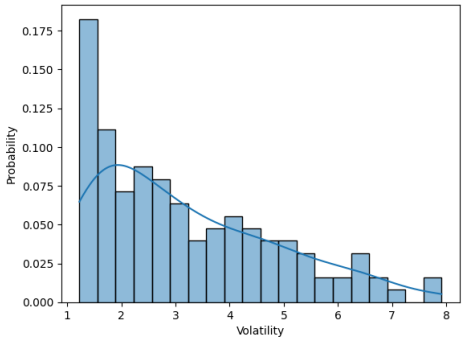


Figure 2.8: Box plot of portfolio volatilities

The next plots in figures *figures 2.9a and 2.9b* allow us to observe the distribution of all the portfolios' volatilities. As it can be observed, in this case the tendency of the number of portfolios decreases progressively as the volatility rises.



(a) Distribution of Portfolio Volatilities



(b) Probability Distribution of Portfolio Volatilities

Figure 2.9: Portfolio Volatility Distributions

Also the next graphic plot regarding the 'Cumulative Distribution Function of Portfolio Volatilities' in *figure 2.10* allows us to see how, contrary to the previous plot from the returns, in this case the accumulation is more progressive for each level of volatility.

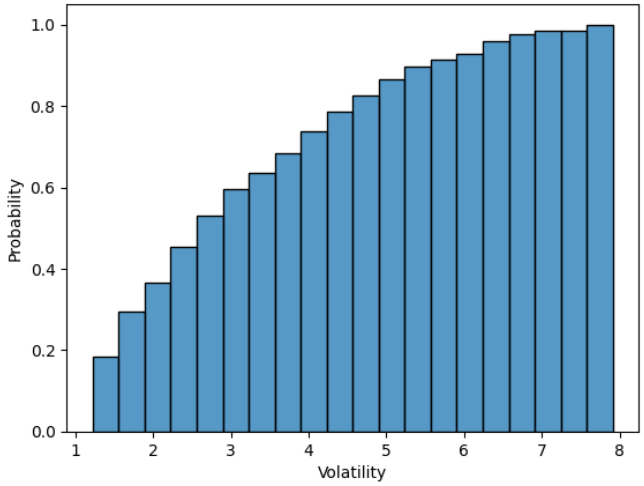


Figure 2.10: Cummulative Distribution Function of Portfolio Volatilities

In the end, as it can be seen in the scatter plot graphic below in *figure 2.11* regarding the Return and Risk attributes, for this portfolio-generated dataset the portfolio with the highest return is indeed the portfolio with the highest volatility. Also, there can be seen a positive correlation between return and risk, which is very interesting since it shows the ultimate purpose of the financing portfolios, and that is to choose the best portfolio that fits our needs, meaning to choose the portfolio with the highest return given the volatility, or in other words, to choose the one with lowest volatility given a return.

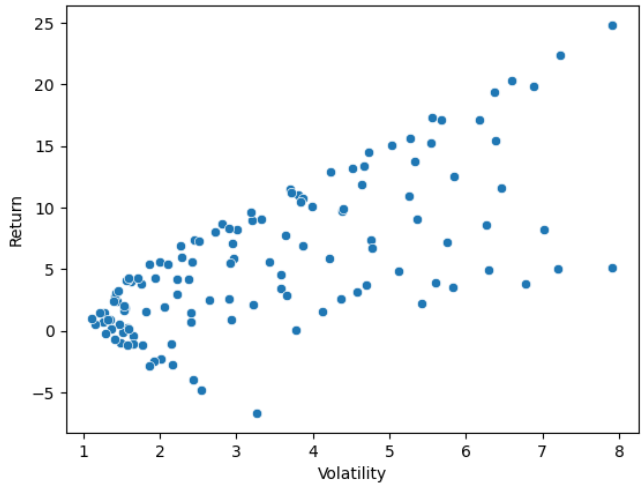


Figure 2.11: Return VS Risk

3 Results

Now that we finished already the addressing of a pre-analysis of all the data generated, we are able to answer the questions proposed for this assignment:

3.1 Is it more probable to obtain a positive or negative return?

Given the previous analysis we could say that yes, it's more probable to obtain a positive return. First of all, as we pointed out previously in the pre-analysis, the mean return is 5.30%, which means that on average, a portfolio generates a return of 5.3% over the period of time covered by the dataset.

Also both the distribution of Portfolio Returns and the probability Distribution of Portfolio Returns in *figures 2.5a and 2.5b* allow us to see, apart of the tendency of the portfolio returns, the range where more instances of a portfolio are accumulated, and indeed most of them are positive.

Also, we also observed from all the values that there is an 80.4% probability of obtaining a positive return from any of these investment strategies, in contrast to the 17.46% probability of obtaining a negative return. This is clearly shown in the previous analysis on the plot referred to as the Cumulative Distribution of Portfolio Returns in *figure 2.6*.

3.2 Is it ALWAYS true that the higher the risk, the higher the obtained return is?

No, it is not always true that the higher the risk, the higher the obtained return is. Though many investors believe they should take a high-risk approach to generate higher returns, academic research shows that's not necessarily true. There is no guarantee that taking greater risk results in a greater return. Rather, taking a greater risk may result in the loss of a larger amount of capital. A more correct statement may be that there is a positive correlation between the amount of risk and the potential for return.

For instance, for our specific case, this is clearly seen in the scatterplot graphic performed on the previous analysis regarding both the return and risk portfolio attributes in *figure 2.11*. In the same, it is evident that, although there seems to be a positive correlation between the volatility and the return, still we can find portfolios where for a specific volatility, there are different returns, or in other words, for a specific return, different volatilities. In finance that is common to see in portfolios with more than 2 assets (without taking into account a third asset free of risk), due to the fact that the amount of combinations is not linear anymore. And

indeed this fact is the final purpose of the portfolio, and that is to choose the best portfolio that fits our needs, meaning to choose the portfolio with the highest return given a volatility, or in other words, to choose the one with the lowest volatility given a return.

4 Conclusions

This portfolio analysis shows how the study of different investment strategies can help generate conclusions about issues such as the obtained return and volatility of different portfolio placements based on historical data.

The use of statistical summaries and different visual representations provides support to the analyst when making decisions as the one regarding the questions that were analyzed in this work. Firstly, we could observe how the return tends to be positive on average, which means that generally, an investment with a random portfolio will more likely have a positive reward for the client.

Secondly, we analyzed the hypothesis on whether higher volatility was always directly related to a higher return. Here, we could establish that the hypothesis was not true. While there exists a positive correlation, there is no guarantee that taking greater risk results in a greater return. This fact should be taken into account by clients, especially if they do not have a great margin of loss.

All in all, the process followed in this work can be taken as the first step toward the generation of an automatic portfolio-generating tool based on historical data.

References

colaboradores de Wikipedia. (2021, 10). *LaTeX*. Retrieved from https://es.wikipedia.org/wiki/LaTeX