Course of Financial Markets Analytics - Professor G. Forte - Academic year 2023/2024

On Portfolios tilting based on different metrics

Carlo Arpini 918543 - Chiara Mariani 918354
Università degli Studi di Milano-Bicocca, CdLM Data Science

Abstract

Past returns are no guarantees of future returns. This saying holds true as any stock market portfolio can be characterised broadly by some random variable in the very short term; but we also know that risk and return are two characteristics of any portfolio much intertwined. By studying how different portfolios behaved in the past we can get a look at their historical returns but also their risk, and most importantly we can try and draw a relation between their risk and their return - relation that will help us understanding how to better protect assets and try and chase future returns without exposing too much to risk. To do so, we decided to study 8 different portfolios tilted following different variables and compare them with the S&P500.

Contents

1 Introduction

As stated in the abstract, we want to study the risk return relationship and how tilting portfolios affect returns and risks, compared to the S&P500 [1].

To do so, we reconsidered the Security Market Line in form of a regression equation as in the Market Model. We implemented a rolling regression over a period of 180 days to obtain the parameters that characterized the regression equation, alpha and beta, their significance and goodness of the fit. All this data will be used to build tilted portfolios to be then compared in terms of returns and volatility with that of the S&P500.

We decided to choose the S&P500 because of several reasons; mainly because it has become the main index for comparison in the last 20 years, it is considered to be a good representation of the American economy, which arguably is the most important one in the world, and is in a sense thought of as the barometer of the market; then also it is a very easy index to download data of since we have price history of the index itself dating back to the 1970s; and lastly it is vast enought to contain a significant number of companies, which allows us to perform a rolling regression and rebalance portfolios easily. Thus the S&P500 was a natural choice and we can expect to extract some useful empirical information about the relationship between risk and returns.

In this regard, we actually expect there to be some component of risk in the S&P500 that might not be justified in light of its returns; that is, we can expect that lowering the risk component could lead to a better risk/return ratio, even at the expense of some returns. Our expectations are clearly dependent on the tilting of the different portfolios involved, but in general we can expect that portfolios with high risk will yield higher returns but will not be optimal in the risk/return ratio context; lower risk portfolios will maybe yield lower returns but have a higher risk/return ratio, and at last tilting based on alpha will probably give us a significant advantage over the index in terms of returns, but will come at a higher volatility price.

To start our analysis the first step is to acquire all the needed data; then we'll clean it up a bit and at that point we will be ready to compute all the coefficients of the Market Model. Lastly, we'll build the portfolio based on those coefficients and analyse the various characteristics with price chart comparisons and table of returns/volatility. We start with the data acquisition and cleaning part, present in the "DataDownload_Processing.ipynb" file.

2 Data acquisition

The first important thing to do to achieve our goal is to correctly acquire all the data necessary to execute the rolling regression and build tilted portfolios.

We collect data starting from 1st of June 2017 to 1st of July 2024, having a total of 1782 market days.

The reference market we select is the American one and, to be more precise, we choose the S&P500 as market index. Standard&Poor 500, known in abbreviated form as S&P500, is arguably the most famous U.S. stock index. It was created by Standard&Poor's in 1957 and tracks the performance of an equity basket consisting of the 500 largest-capitalization U.S. companies. We are able to download the data for this index's prices through FactSet [2]. The downloaded data did need some cleaning as the file was an xslx and had many columns; we converted the file from xslx to csv to have an easier time processing the data on the notebook. We then had to drop some columns and rows to actually have the index's data loaded in a dataframe the way we wanted, and lastly we converted all dates in datetime-like values.

Instead, to obtain the historical price series for a set of stocks of our interest the process is a bit more complex. We want to keep only all those tickers that were present throughout the whole timeframe considered (June 2017 - July 2024), because otherwise we could have problems with the rolling regression we'll later apply; this can be arguably fixed with rebalancing, but we will rebalance on a weekly basis and not a daily one; hence we can safely say that in general we could encounters issues (think of a stock getting removed on Tuesday from the S&P500 and our rebalance day is Monday; then we would still track that said stock in our portfolios for 4 more days before kicking it out). First of all, we download from Wikipedia [3] the list of all stocks' tickers at the 1st of July 2024 present in our index and from this we remove any ticker that got in or out in the S&P500; this way it's ensured that all stocks present in our list were there in 2017 as well as in 2024. At last, we download data about stock prices from Yahoo Finance through yfinance [4] and do some cleaning of unwanted columns and join everything together. Yahoo Finance will suffice as the source since we're only interested in stock prices, and they are accurate enough.

At the end of this phase we have two different csv files: one containing all the prices for the selected set of stocks, named *prices_stocks_list*, and the other containing all the prices for the S&P500 index, named *SP500_prices*.

3 Regression analysis

The analysis of acquired data, computing coefficients and portfolio building is present in the "Analysis and regressions.ipynb" file.

The objective when we work with a portfolio is returns and not prices: the graph of prices presents a lot of variation and uncontrolled movements, while the graph of returns is stationary, and of much simpler interpretation. Since we have downloaded the prices relative to the index and to the set of stocks, we can calculate the log-returns in a step of time as

$$r_t = \ln P_t - \ln P_{t-1} = \ln \frac{P_t}{P_{t-1}}$$

Once the log-returns are obtained we have all the elements to apply a rolling regression and estimate the parameters of interest on which to base the construction of our portfolios. All our reasoning starts from the fundamental law of asset management, proposed by Grinold and Kahn in 1997 [5]

$$IR = \frac{\alpha^B}{\omega} = IC * \sqrt{BR}$$

where the information coefficient IC gives a number of value added in terms of new information explaining returns and breadth BR represents how many factors we consider in our model. If our goal is to explain returns with different factors, we have to find good factors with high information coefficient and relevant high breadth to obtain at the end a good financial investment product. From a practical point of view, we implement a multi-factor regression model

$$r_{i,t} = \alpha + \beta_1 f_{1,t-1} + ... + \beta_K f_{K,t-1} + \epsilon$$

In this project we consider the basic model with only one explanatory factor called Market Model. In this case, the right expression is

$$r_{i,t} = \alpha + \beta_i R_{M,t} + \epsilon_{i,t}$$

where α represents the extra performance, β_i represents the estimated coefficients and $R_{M,t}$ corresponds to the market portfolio return which in this case coincides with the S&P500 index's return.

This Market Model is a simple rewriting of the Security Market Line in terms of regression model: the return of an asset depends on the market portfolio where the link is based on beta and there is an error which explains the approximation.

In this setting, we implement a 180-days rolling regression which progressively gets shifted one day ahead each time, ensuring smoothness of change among various variables of the model.

So, doing the regression with a sample of 180 days, we are able to save the estimated parameters for each stocks, which are:

- *alpha*: this represents the extra performance in those 180 days for our stock; it is the intercept of the model and its interpretation is that if it's positive it means that the stock has higher returns than the S&P500 index; if it's negative or zero the stock has lower or comparable returns to S&P500. Note that we will not have a single value for alpha for each stock but rather a time series of 1782-180=1602 values

- *alpha significance*: indicates the statistical significance of the corresponding estimated alpha

- *beta*: it is a coefficient measuring the sensitivity of the stock to market movements; it relates to systematic risk of a stock; this coefficient too is not a single value but rather a time series of values, as per alpha

- *beta significance*: indicates the statistical significance of the corresponding estimated beta
- *R squared*: it is a key statistic informing about the goodness of the fit of the model; if it is high it means that the model is able to explain a huge part of the variability of stocks' returns, and it too is expressed in a time series of values
- *systematic risk*: it is related to the relationship between the stock and movements of the market; it is the risk of being inside a specific market and it is calculated as $\beta_i^2 \sigma_M^2$ and as such it's totally dependent from the values of beta
- *specific risk*: it is the intrinsic risk of a single stock and it is calculated as the variance of the residuals where residuals are the difference between the returns of the stock and the predicted returns. Note that this way we do not have a time series of specific risk values, but a single value tied to each stock, as the only time series like value we can construct are the residuals, but they all get blended together in the variance
- *total risk*: it is the overall risk of a stock and it is given by the sum of systematic risk and specific risk

By proceeding in this way, we obtained several parameters with which we are able to sort and group stocks to create tilted portfolios.

4 Portfolios building

After implementing the rolling regression, we decide to build our tilted portfolio based on R^2, α, systematic, specific and total risk, as highlighted below.

- Portfolio 1 - based on extra performance α

In order to create the portfolio based on the value of the alpha coefficient, we can't simply average the alpha values calculated through the regression over the time period considered. But rather, in order to give greater validity to our work and results, we decided to go for a weighted average of the alpha values considering 1 over the value of alpha significance as weights. This way the greater the significance (and lower the p-value present in alpha significance) the greater the weight it has in the average. Operating in this way, we obtain a dataframe in which the first column contains the tickers of each stock analysed and the second column contains the value of the associated alpha coefficient. It is now possible to sort the tickers according to decreasing values of alpha and select the upper 10% quantile to obtain the stocks that will define the portfolio. This way, for portfolio 1, we select 38 stocks that have the highest extra performance value correctly calculated with a weighted average on alpha significance.

- Portfolio 2 - based on goodness of the model fit R^2

In order to create the portfolio based on the value of the R^2, we can simply do the average of $R-$squared values calculated through the regression over the time period considered. In this case there is no need to make additional considerations as the average of the values makes sense: R^2 only measures the goodness of the model. So, we sort the tickers according to decreasing values of R^2 and select the upper 10% quantile to obtain the stocks that define the portfolio. After this, we make another selection consider only the stocks with a $R^2 > 0.5$; this choice, albeit arbitrary, ensures we are selecting stocks where our model explains at least 50% of the variance; this is important and gives validity to the model itself. We can conclude that, for portfolio 2, we select 27 stocks that have the highest goodness of model fit over the time period considered.

- Portfolio 3 and 4 - based on systematic risk $\beta^2\sigma_M^2$

  In order to create the portfolio based on the value of the systematic risk, it is firstly necessary to calculate the systematic risk for each ticker as $\beta_i^2\sigma_M^2$. To do this, we firstly need to calculate a value of beta for each stock. As in the case of alpha coefficient, we do not simply do the average of beta values calculated through the regression over the time period considered, but rather, in order to give greater validity to our work and results, we do a weighted average of the beta squared values considering 1 over the value of beta significance as weights. The situation here is more complicated as we are dealing with beta values equal to zero because the significance value (recall it's expressed as p-values) is so small that the computer just approximates it to zero (orders of 10^{-74} and even smaller!). In this case, which occurs for just a handful of tickers, the weighted average is replaced by the arithmetic average. The reasoning is as follows: as the weight (inverse of the pvalue) tends to infinity, all other weights both at the numerator and at the denominator contribute to our weighted average less and less. As such the best approximation we have is to just consider all the other weights to zero and keep the ones where the pvalue tends to zero, which become of equal importance; hence the arithmetic average. Once we have calculated the beta coefficient value for each ticker, we multiply this value by the market variance, calculated from the log returns of the S&P500 index. We thus obtained the systematic risk value for each stock and we can sort the tickers according to decreasing values of $\beta_i^2\sigma_M^2$ and select the upper and lower 10% quantile to obtain the stocks that define the portfolios. We can conclude that, for portfolio 3 we select 38 stocks that have the highest systematic risk and for portfolio 4 we select 38 stocks that have the lowest systematic risk.

- Portfolio 5 and 6 - based on specific risk ϵ_{ei}^2

  In order to create the portfolio based on the value of the specific risk, we can simply take the specific risk calculated through the regression; note here that no further considerations are needed since all the values are the same over the time period considered. So, we sort the tickers according to decreasing values of σ_{ei}^2 and select the upper and lower 10% quantile to obtain the stocks that define the portfolios. We can conclude that, for portfolio 5 we select 38 stocks that have the highest specific risk and for portfolio 6 we select 38 stocks that have the lowest specific risk.

- Portfolio 7 and 8 - based on total risk ϵ_i^2

  In order to create the portfolio based on the value of the total risk, we can simply sum the systematic risk and the specific risk calculated as described in the previous portfolio. Then we sort the tickers according to decreasing values of σ_i^2 and select the upper and lower 10% quantile to obtain the stocks that define the portfolios. We can conclude that, for portfolio 7 we select 38 stocks that have the highest total risk and for portfolio 8 we select 38 stocks that have the lowest total risk.

Note that for risk-based portfolios we have chosen the different stocks present sorting through the variance and not the standard deviation; albeit the standard deviation is the correct measure for risk, sorting through the variance does not matter as all standard deviation values are positive and squaring them is passing them through a monotonous increasing functions: it does not alter their order. After this process, we have created eight different portfolios and to get the returns of each single portfolio we simply calculate the equally weighted average of stocks returns involved.

One last thing before actually computing returns is the problem of rebalancing: we want to rebalance weekly and keep an equally weighted aspect for each of our portfolios. But this poses a problem: if we were to rebalance say on a specific day of the week, we could end up rebalancing "too early" or "too late", in the sense that in general if we choose Monday to rebalance and the next one is a holiday, two weeks will have passed with us not rebalancing; similarly if we choose Monday as our rebalancing day, but then Tuesday to Friday there is some holiday and market is closed we end up rebalancing for no reason at all the next Monday.
The easiest way to tackle this problem, and also the one we chose to follow, is to just realise that in a standard market week we have 5 market days; our prices/returns data is already by definition only stored in market days, so we'll rebalance every 5 days.

5 Final Statistics

Final statistics on every portfolio are almost all defined by the returns of the portfolios and their overall risk; so the first and most logical thing to do is to compare different portfolios with a price chart, as can be seen in Figure 1.

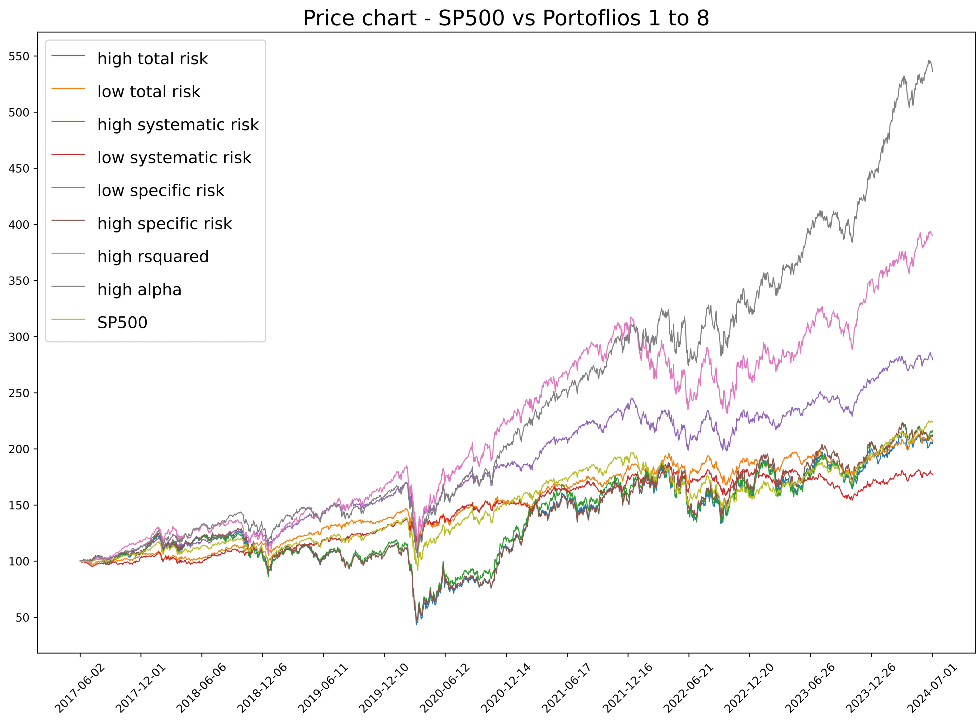

Figure 1: Price chart of the different portfolios VS the S&P500 index

From said figure clearly the first things that we can observe is that, as expected, the portfolio made by stocks which actually returned more than the index (had positive alpha) shoots up much more in value than the index; no surprise here. Furthermore it can also be seen that all those stocks with an high R^2 value actually outperform the index once again; and the same can be said of those stocks that bear a low specific risk. This probably is the most interesting of the three, but to view it more clearly and draw some conclusions we should eliminate the top two performing portfolios to better compare the risk-based ones.

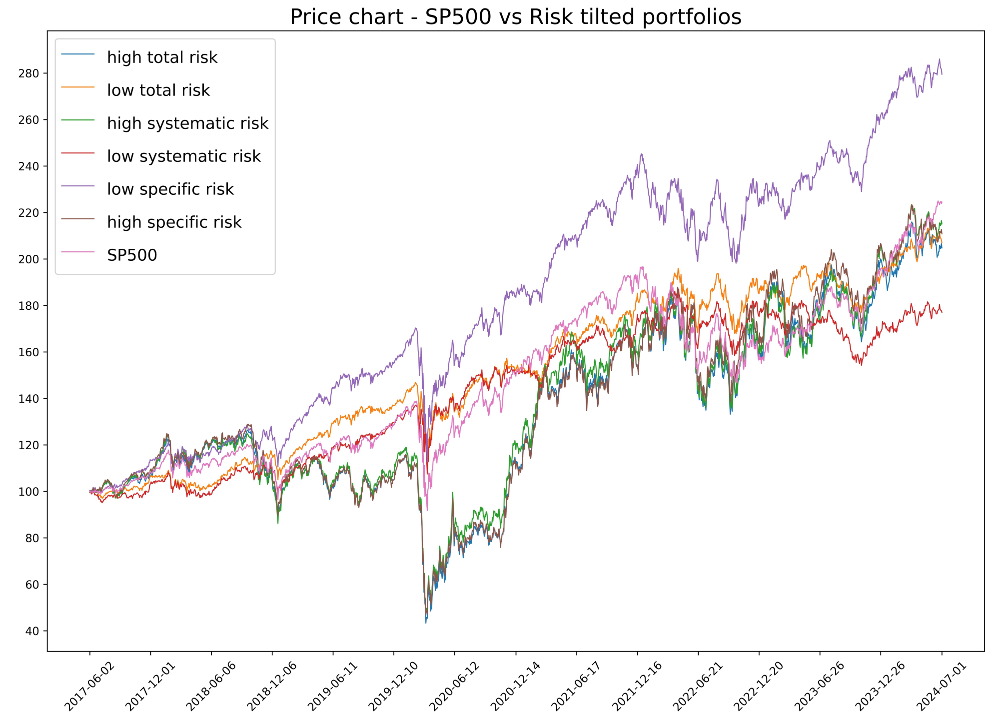

Figure 2: Price chart of the different risk tilting based portfolios VS the S&P500 index

One peculiar tract that stems from the analysis of the risk based portfolios in Figure 2 is that the risk-reward relation does not seem to "always" hold true, or better we can say that it looks like some portfolios carry an unnecessary amount of risk. The "worst performing" portfolios seem to be the ones who carry too much risk, the high-risk ones, even though their returns over the considered period are not too bad after all. But they can be considered worst performing in the sense that they fall too much lower values than the others portfolios. Instead, it looks like low risk portfolios do better, but it's keeping a low specific risk that does the trick - probably because keeping a low systematic risk signifies having a low beta, which translates in missing a lot of "good market" days, where prices all move upwards together. Instead, a low specific risk lowers the possibility of black swan events while keeping a fair value for beta, meaning that it "catches" good market days whilst not having particularly negative days on its own because of news about the stock itself.

It is clear at this point that we need to account for risk and not only for returns. As such, we can construct a table with not only returns of the portfolios considered but also risks and some ratios to better help us understand the risk-return relationship, such as the Sharpe index or the Sortino one.

The Sharpe ratio is defined as:

$$S_i = \frac{r_i - r_f}{\sigma_i} \tag{1}$$

with at the numerator the difference between the return of the portfolio and risk free return and at the denominator the standard deviation of the portfolio, while the Sortino one is instead:

$$\hat{S}_i = \frac{r_i - r_f}{\sqrt{\frac{\sum_{j=1}^{N}(\min(0,R_j)^2}{N}}} \tag{2}$$

with the same numerator but at the denominator we only are considering the downside standard deviation: that is, each time the R-jth daily return of our portfolio is negative we consider it, but if it is positive we set it to zero. Usually the minimum function in the downside standard deviation is applied to zero and the difference between the R-jth daily return and the minimum acceptable return (usually the risk free daily return), but for our case we consider the minimum acceptable return to be zero, as we will consider the risk free rate.

Portfolio	Total return(%)	Yearly average return(%)	Volatility	Sharpe	Max Downside	Sortino
highest_alpha	434.781705	27.064789	21.047313	1.285902	14.901871	1.816201
highest_rsquared	288.456361	21.392470	23.724671	0.901697	16.542363	1.293193
low_specific_risk	178.485683	15.755935	17.801558	0.885087	12.529358	1.257521
low_total_risk	106.236610	10.894344	15.160281	0.718611	10.593524	1.028397
SP500_index	123.875494	12.202087	19.324872	0.631419	13.883671	0.878880
low_systematic_risk	76.508481	8.455661	14.462752	0.584651	9.988151	0.846569
high_systematic_risk	115.849395	11.618410	31.584752	0.367849	22.569641	0.514780
high_specific_risk	111.779250	11.315277	31.374767	0.360649	22.317296	0.507018
high_total_risk	105.441891	10.833196	32.248686	0.335927	22.965530	0.471715

Figure 3: Risk-return table with different metrics

To do this we summarised results in the table shown at Figure 3; the table itself is already sorted in descending order based on Sharpe ratio. This table not only highlights better the difference between various portoflio returns, but leads us to some interesting cocnlusions:

- First of all we see that the higher the return, the higher the volatility of the portfolio, as expected; but this holds true only in a weak form, since we can see that S&P500 portfolio had a higher volatility than the low specific risk portfolio but it also had a lower return
- Interestingly, volatility and maximum downside are somewhat independent, in the sense that higher volatility did not imply an higher maximum downside; rather, a higher risk implied a higher maximum downside
- In a somewhat naive way, we can confirm stocks are indeed a risky asset: even the less volatile and with lower maximum downside portfolio, which returned less than all others, still had a high volatility and downside, and a corresponding higher return
- Another peculiar fact is that albeit returning about 2 basis points less than the S&P500 on average every year, the low total risk portfolio has a higher Sortino and Sharpe indexes. This highlights that the S&P500 actually has indeed a higher return but does not have a better risk-return ratio; hence an investor in the S&P500 would assume more risk for less return than what would be “fair”
- At last, considering the difference between Sortino and Sharpe ratios, we see that both ratios actually were identical in order; but it is interesting to notice that if we excluded the portfolio with high alpha stocks the values for the Sharpe and Sortino indexes for high R^2 and low specific risk portfolios are very similar and much higher than that of the S&P500 values.

6 Conclusion

The results obtained clearly show that in general an increase of risk generates higher returns, but those returns are not always coherent with the increase in risk. Instead, it has been shown that over the timeframe considered a strategy that chased stocks with low specific risks would have done very well, and with the minimum downside possible. Strategies that tried to follow stocks with a high alpha value would've even probably done better, although it's clear that predicting which stocks will have and will maintain a high alpha over a period of time is all but simple, especially when compared with just identifying stocks with low specific risk.

In general risk based strategies look very promising: it looks more important than ever to avoid high risk, both total, specific and systematic, given all those portfolios performed even less than the S&P500 and had worse volatility. Avoiding systematic risk too seems like a mistake, most probably because in the period considered the S&P500 used to increase and hence avoiding systematic risk meant having a low correlation coefficient with the index, which meant a stock decreasing. In a sense, it is as if avoiding systematic risk means avoiding those days were market is green for everyone. All this reflects in the low total risk portfolio, which scores still better than the S&P500 by Sharpe and Sortino ratios, albeit having lower returns than the S&P; this portfolio is probably both dragged upwards by the absence of specific risk and a bit downwards by the absence of systematic risk, but the absence of risk combined with its returns makes it a portfolio with a more optimized risk/return approach than the S&P500.

At last, the high R^2 portfolio scored much higher than the S&P500, and not too much less than the high alpha stock portfolio. This is probably due a series of factors:

- first and foremost, it is the least populated portfolio. This means it has a higher volatility and hence a higher risk and correspondingly an higher return over a longer period of time; this is highlighted by the fact that it has a higher volatility than the high alpha stock portfolio and is only behind those portfolios specifically built with high risk. It could have gone down instead of upwards too... but in a period of time where the market overall granted great returns it is normal its returns were even greater;
- having a high value for the R^2 coefficient means that we actually ended up selecting stocks whose returns over 180 days were a good fit for a linear model. This automatically means selecting stocks which will not have significant sudden changes in prices not coherent with what the market was doing; this automatically means excluding stocks with a high specific risk or a low systematic risk, while keeping ones that had a good fit for their alpha and beta values;
- having a good fit for alpha and beta values over a period of time where generally speaking the market moved upwards also implies somewhat of an high alpha.

All in all the best strategy seems to seek all those stocks which minimize specific risk, while trying to still stay correlated to the market. Some further developments that might be inherently good for this type of analysis are:

1. Considering a non zero risk free rate; this could change the order of Sharpe and Sortino values computed for the different portfolios and be more indicative in our current situation;
2. Selecting different types of portfolios tilting based on both values of R^2 and high values of alpha or beta, so basically combining different tilting together, this could prove more interesting results and higher returns;
3. Considering a longer timeframe or different percentiles to analyse what are the differences with respect to the table of Figure 3

References

[1] URL: 'https://www.spglobal.com/spdji/en/indices/equity/sp-500/#overview'.

[2] URL: 'http://www.factset.com/'.

[3] URL: 'https://en.wikipedia.org/wiki/List_of_S%26P_500_companies'.

[4] URL: 'https://github.com/ranaroussi/yfinance?tab=readme-ov-file'.

[5] Richard C Grinold and Ronald N Kahn. "Active portfolio management". In: (2000).