Financial Markets Analytics project

Lorenzo Lobosco 851289, Giulia Beccaria 851286

Abstract

The goal of this project is to develop screening models using as input the data contained in the "euro.xls" file. The file includes several key indicators for companies: PE Ratio, Five Yr Avg Price Earnings, T12M Dil PE Cont Ops, 10 Year Moving Average PE, PX to Tang BV Per SH, Current EV to 12M Sales, Current EV to T12M EBITDA, Five Year Avg EV to T12 EBITDA, T12M Dil EPS Cont Ops, Trail 12M EBITDA per Share, Trail 12M Sales per SH, Net Debt per Share, Tang Book Val per SH, Normalized Accruals CF Method, EBITDA Margin, EBITDA Margin 3YR Avg, RSI 14D, PX Last, Mov Avg 50D, Mov Avg 20D, Mov Avg 10D, Mov Avg 5D, Mov Avg 40D, Mov Avg 30D, 3Mth Impvol 90.0 MNY DF, Volatility 90D, Volatility 30D, Volatility 180D, Cap Expend to Sales, PX Volume, T12M DVD Payout Ratio, EQY DPS Net 5YR Growth, EQY REC CONS, Best EPS, WACC Cost Equity, Normalized ROE, 5YR Avg Return on Equity, Cur Mkt Cap, Normalized Accruals BS Method, PX to Book Ratio. The data refer to a time period of 111 months, while the number of securities is 797. Based on the comparison of the benchmark portfolio, univariate screening, and multivariate screening results, it is evident that the portfolio generated through multivariate sequential screening exhibits superior performance, as expected.

Contents

1. Introduction

Stock screening models play a vital role in the investment and trading process, enabling investors and traders to identify potential investment opportunities based on specific criteria. These models employ thorough analysis of financial data to narrow down the selection of stocks, highlighting those that meet desired parameters. The aim of this project is to implement diverse screening models using Python and utilize them to construct portfolios. Subsequently, the portfolios will be assessed based on their returns, and appropriate absolute and relative performance metrics will be calculated by comparing them to a benchmark. To initiate the project, we loaded the "euro.xls" data and conducted a concise exploratory analysis to gain a better understanding of its structure and characteristics. After making preliminary adjustments, we computed the log return dataframe using the "PX_LAST" variable, which represents the closing price of a financial instrument, on a given trading day. By calculating the log return, we aimed to measure the rate of change in stock prices over time. In order to establish a benchmark for comparison, we considered a portfolio containing all the securities. Subsequently, we calculated the mean log return for each period and determined the total equity of the benchmark portfolio. To illustrate these findings, we presented corresponding graphs.

2. Univariate Screening and Transaction Costs

Let us examine the procedures involved in calculating the benchmark and conducting univariate screening in greater detail.

A. Benchmark calculation

We have computed the log returns for the stocks in our dataset as part of our project. To calculate the log returns, we divided the "Price Last" of the current period by the "Price Last" of the previous period and took the logarithm of that value. This process resulted in a dataframe with 798 rows (representing the number of stocks) and 111 columns (representing the number of periods).

We will utilize this dataframe throughout our project to assess the performance of our screening model. To compare our models effectively, we need to consider a benchmark. As mentioned earlier, our benchmark is a portfolio composed of all the stocks available in the dataset, with equal weighting. We computed the mean of the log returns for each period of the benchmark and also calculated the total equity of the portfolio. Subsequently, we plotted corresponding graphs to visualize these results.

The important property is that log-returns are symmetrical around 0 with respect to addition. This property makes it possible to speak of an average return. Log-returns are easier to handle computationally, especially in bulk, but non-log returns are easier to understand/imagine as a number of their own.

The Equity Line is a graph showing the curve of realized profits and losses over time, considering an investment in our

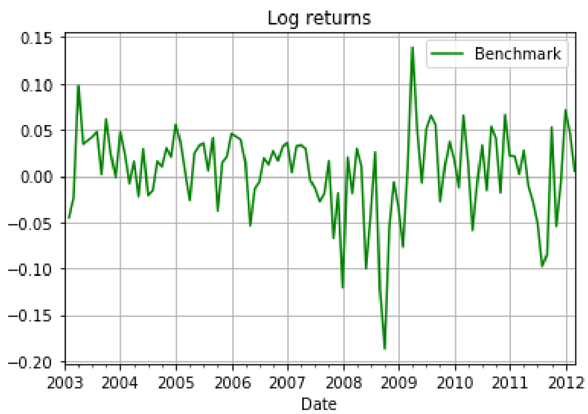

Figure 1 Benchmark Log Returns

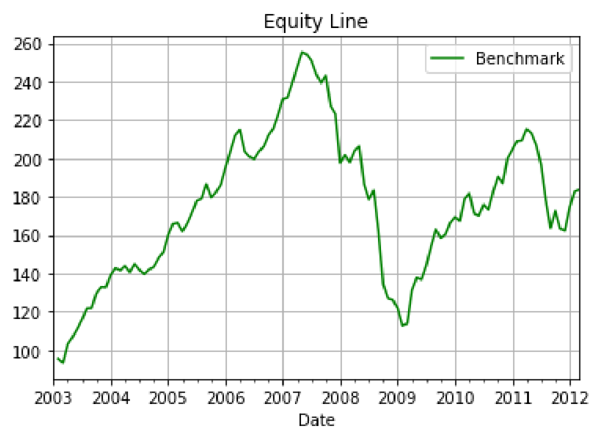

Figure 2 Benchmark Equity Line

portfolio of 100$. It is a graph that will show us the trend of gains and allow us to understand in a simple and intuitive way the constancy and reliability of the trading system.

B. Univariate Screening

In this chapter, we present an exclusive Univariate screening strategy that focuses on a single factor for selecting the optimal securities to include in our portfolio during each time period. Specifically, our attention is directed towards the EBITDA Margin factor, which serves as a profitability indicator by measuring the percentage of gross operating profit (EBITDA) relative to sales revenue.

The EBITDA margin is also a financial position index, as it also has the meaning of 'self-financing of operations'. In this capacity, the index signals the % measure of internal financial means generation from sales revenue.

A univariate screening with the EBITDA Margin would allow you to analyze the variability of the EBITDA margin between different companies or over time for the same company. You can use this index to identify companies with a high EBITDA margin, which could indicate greater operational efficiency and potential profitability.

Initially, we sort each security for each time period based on the value of the EBITDA margin factor. Subsequently, we select the top 50 securities for each column.

We have opted for a portfolio comprising 50 stocks in our project, as we prioritize maintaining a balanced and secure portfolio. We aim to strike a balance between concentration, avoiding a portfolio with too few securities, and excessive diversification. This is the rationale behind selecting a portfolio composed of 50 different stocks.

After calculating the portfolio for each period, following the factor value for each stock, we proceed to calculate the Turning Rate. This step is essential for assessing transaction costs, which impact the overall return of our portfolio. The Turning Rate for each period represents the percentage of the portfolio that changes from the previous period. Essentially, it measures the extent to which stocks enter and exit the portfolio in each period. Unlike the benchmark, where all stocks remain in the portfolio each period, here we must consider the Turning Rate. This is because transaction costs are incurred whenever there are changes in the stocks within the portfolio.

Initially, we set the turning rate to 1 for the first column, as all securities in period 1 will be new additions to the portfolio. For the subsequent columns, the turning rate is calculated by dividing the number of new top 50 securities in the current column, which were not previously present in the portfolio, by the total number of securities in the portfolio. To gain a deeper insight into the asset allocation of our stocks using the screening strategy, we also visualize the turnover rate through a plot. This plot allows us to observe the timing and extent of turnover within our portfolio. By examining the image, we can observe that our Univariate screening strategy results in a minimal turnover of stocks within our portfolio.

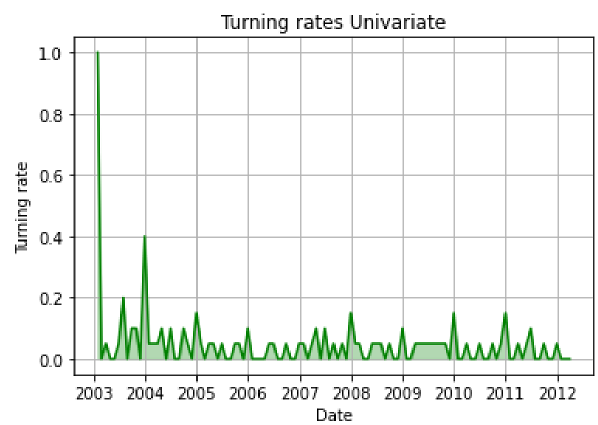

Figure 3 Univariate strategy Turning rates

The final aspect to consider is Transaction Costs. In our project, we have factored in fees of 20 basis points. We calculate the transaction costs by multiplying the turning rate by the basis points, resulting in the costs that need to be subtracted from the returns of our portfolio.

By implementing this strategy, we have achieved returns for our portfolio while taking into account transaction costs. We compare the performance of our portfolio, obtained through univariate screening, with that of the benchmark. Specifically, we plot the average log returns for each period as well as the total equity line. To ensure comparability, we have adjusted the equity line by considering the size difference between the benchmark (composed of 798 stocks) and our portfolio (composed of just 50 stocks). It would be nonsensical to compare the total equity line directly between these two scenarios, so we have proportionally adjusted our portfolio to enable a meaningful comparison. The results are depicted below, Univariate strategy compare to benchmark (log returns and equity line):

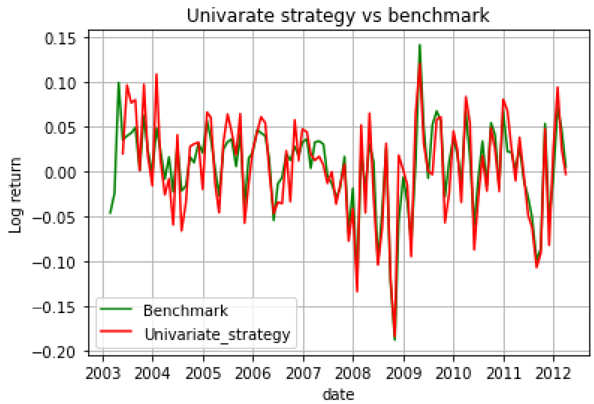

Figure 4 Univariate strategy vs benchmark (log returns)

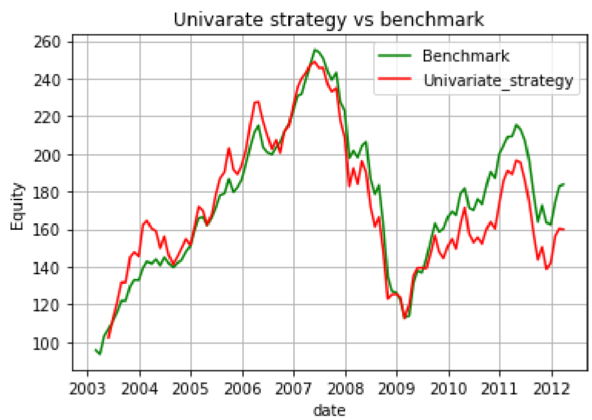

Figure 5 Univariate strategy vs benchmark (equity line)

C. Transaction Costs

One crucial aspect to consider in our project is the inclusion of transaction costs. We have established a default transaction fee of 20 basis points in our codes. Whenever we construct a portfolio using screening strategies, we calculate the turnover rate for each period. This turnover rate represents the percentage of the portfolio that changes from one period to the previous. To account for transaction costs, we subtract the product of the turnover rate and the 20 basis points transaction fee from the portfolio's return.

D. Metrics

For the valuation we use the absolute metrics Sharpe ratio, Traynor ratio and Sorbino ratio. As a relative metrics: the most used is the Information Ratio, it is a measure related to the benchmark, this is why it is called relative.

- **Sharpe ratio** = $\frac{r_p - r_f}{\sigma_p}$ where r_p is the portfolio performance, r_f is refereed to the returns of the strategy and the risk free rate is settled to 3% and σ_p is the standard deviation of the portfolio's return. The numerator represents the excess return earned by the portfolio above the risk-free rate, capturing the reward aspect. The denominator represents the total risk of the portfolio, as measured by its standard deviation, capturing the risk aspect. Therefore, a higher Sharpe ratio indicates that the portfolio is generating a greater return per unit of total risk, which is generally considered desirable. It suggests that the portfolio is providing better risk-adjusted performance compared to alternatives with lower Sharpe ratios.
- **Traynor ratio** = $\frac{r_p - r_f}{\beta_p}$ where r_p is the portfolio performance, r_f = returns of the strategy and the risk free rate is setted to 3% and β_p is the beta of the portfolio. The beta of the portfolio represents a proxy of the systematic risk considering only this component of the total risk.
- **Sortino Ratio** = $\frac{r_p - r_f}{SemiSD}$ where r_p is the portfolio performance, r_f = returns of the strategy and the risk free rate = 3% and SEMI standard deviation is equal to $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_i - \bar{r})^2}$ consider only cases when $r_i \le \bar{r}$
- **Information Ratio** = $\frac{TE}{TEV}$, where TE is the tracking error, also called α, it's the excess return of the fund over the the benchmark $R_p - R_{\text{benchmark}}$. Instead TEV=$\omega = \sigma(r_P - r_B)$ is the tracking error volatility, it computes the standard deviation

EBITDA Margin	Sharpe Ratio	Trainor Ratio	Sortino Ratio	IR
Univariate Screening	0.119	0.021	0.169	-0.15

Table 1 Metrics result

As observed in the table above displaying the metric results, the portfolio generated through univariate screening exhibits poor performance. Additionally, the comparative plots indicate that the benchmark outperforms our portfolio. These findings are further confirmed by the metrics. Therefore, we proceed with a multivariate analysis in order to achieve improved performance.

3. Multivariate Strategy

Due to the unsatisfactory results obtained from the univariate screening, we have decided to construct our portfolio using a multivariate screening strategy. To implement this approach, we began by considering four interest factors. We calculated the information ratios (IR) for each factor using the univariate screening method for each period until the current time. These IR values were then used as weights and multiplied by the standardized values of the respective factors. This process resulted in a final Z score for each stock during each period. We utilized these Z scores to select the top 20 stocks to include in our portfolio for each period.
From the provided list, here are the 4 factors that we have considered in our multivariate screening:
PE_RATIO (Price/Earnings Ratio): This factor indicates the ratio between the stock price and earnings per share, providing an indication of the relative value of the company compared to its earnings.
CURRENT_EV_TO_12M_SALES (Current Enterprise Value to Trailing 12-Month Sales Ratio): This factor measures the company's valuation relative to its recent sales, providing an indica-

tion of sales profitability.
FIVE_YEAR_AVG_EV_TO_T12_EBITDA (Five-Year Average Enterprise Value to Trailing 12-Month EBITDA Ratio): This factor represents the ratio between the enterprise value and EBITDA, a measure of the company's operating profitability over the past five years.
VOLATILITY_90D (90-Day Volatility): This factor measures the variation in stock prices over a 90-day period, providing an indication of stock stability.
After identifying the relevant factors, we proceeded to calculate their respective information ratios (IR). These IR values play a crucial role in our multivariate analysis as they enable us to assign weights to the different factors. To determine the weights for each period, we computed the proportion of each factor's IR divided by the total IR, which was obtained by summing the IR values of all factors. The resulting IR values obviously changing over time, for better undertand how these weighs change we depict a graphic representation of them:

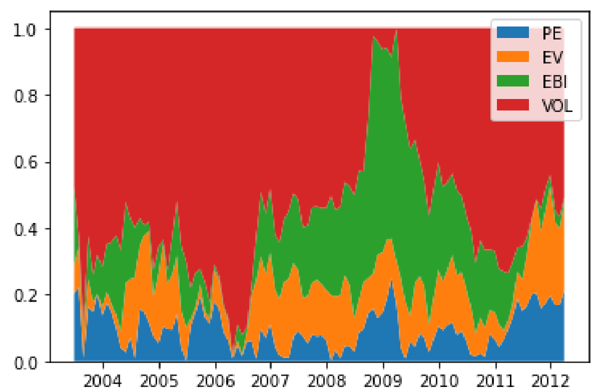

Figure 6 Factors weights over time

Subsequently, we standardized the values of each factor across all columns. This standardization process was conducted to enhance interpretability and comparability among the factor values within the data frame. As a result, we obtained separate standardized data frames for each factor. These standardized data frames, in conjunction with the weights derived from the information ratios (IR), were utilized to calculate the Z scores. The formula for computing the Z scores is illustrated below:
$Z_{score} = w_1 Z_{PE} + w_2 Z_{EV12m} + w_3 Z_{5vEBTIDA} + w_4 Z_{Volatility}$
where the values 'w' are the weights calculate through the IR values as descripted before.
In summary, the outcome of this procedure is a dataframe that comprises the Z-score values for each stock. Similar to the approach used in univariate screening, we have conducted the selection of the top 50 securities to include in our portfolio for each period. However, instead of solely considering the value of a single factor, we now take into account the Z scores derived from the multivariate analysis. Subsequently, we have constructed portfolios for each period and computed their respective returns, along with the associated transaction costs, using the same formulas and assumptions as in the univariate screening. As the Univariate screening strategy we compute the turning rate in order to consider the transaction costs in our portfolio returns.

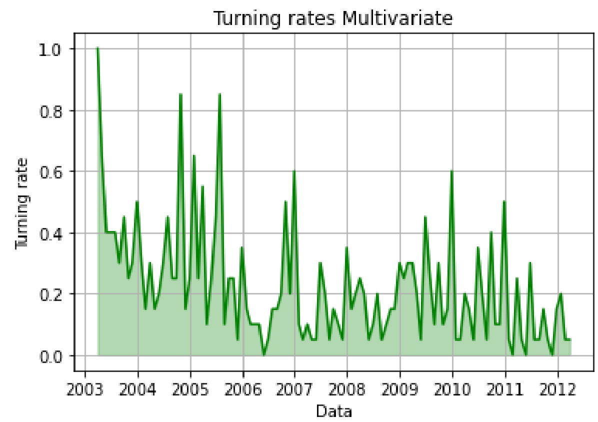

Figure 7 Multivariate strategy Turning rates

We can see from the graph that there is an higher turnover in the multivariate strategy than the univariate one. The resulting plots are provided below, allowing for a comparison of these values with the benchmark:

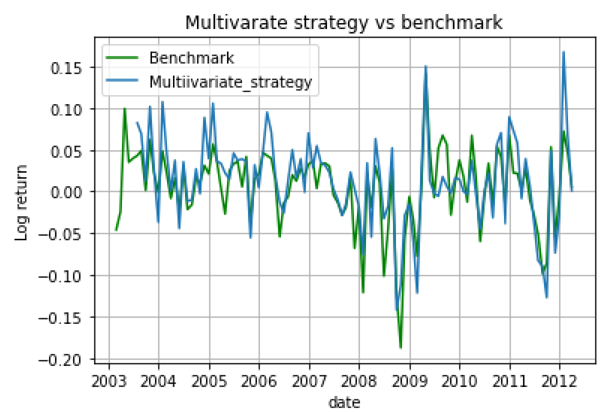

Figure 8 Multivariate strategy vs benchmark (log returns)

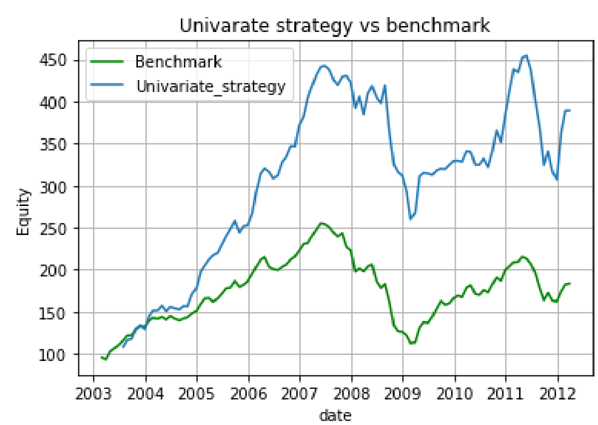

Figure 9 Multivariate strategy vs benchmark (equity line)

The graphs provide information on the multivariate strategy compared to the benchmark, for the log returns and the equity line. Below are also represented the metrics to evaluate the multivariate screening model:

	Sharpe Ratio	Traynor Ratio	Sortino Ratio	IR
Multivariate Screening	0.682	0.115	0.952	0.991

Table 2 Metrics result

The table illustrates that the performance of the portfolio generated through multivariate screening is slightly superior to that of the univariate screening portfolio. The Sharpe Ratio, standing at 0.682, indicates that the strategy is delivering returns that exceed the risk taken, which is generally a positive sign for investors seeking a balance between risk and reward. Despite a relatively lower Traynor Ratio of 0.115 compared to the Sharpe Ratio, it still suggests that the strategy is yielding positive risk-adjusted returns, albeit with less emphasis on mitigating systematic risk. The Sortino Ratio, being notably high at 0.952, implies that the strategy excels in managing downside risk or volatility, potentially making it particularly attractive for risk-averse investors concerned about protecting their investments during market downturns. Lastly, the Information Ratio of 0.991 indicates that the strategy consistently outperforms its benchmark, suggesting a level of skill or access to superior information that enables the strategy to deliver excess returns with relatively low volatility compared to the benchmark. In summary, these results suggest that the multivariate screening investment strategy shows strong potential for delivering favorable risk-adjusted returns, effectively managing downside risk, and consistently outperforming its benchmark.

4. Correlation Problem

When considering the z-score as a composite measure for all chosen factors, a strong correlation between these factors may occur. This statistical feature can lead to redundancy of information and influence the screening results.

The collinearity occurs when explanatory variables are strongly linearly linked to each other leads to additional difficulty in separating the individual effect of the explanatory variables on the response.

Collinearity can be detected in several ways: examination of the correlation matrix of the predictors may reveal large pairwise collinearities. One measure to check if a certain explanatory variable may be predicted well using the others is through the Variance Inflation Factor, In general, a high VIF value (a general high value of VIF is considered when it is greater than 10) indicates that the explanatory variables are highly linearly associated.

To address the problem that correlations between univariate strategies are not accounted for in simple Z-scores, several theoretical solutions can be considered.

- **Weighted Z-scores**: when calculating a composite Z-scores, different weights can be assigned based on the correlation between strategies, instead of assigning the same weight to each univariate strategy. For example, highly correlated policies can be weighted to reduce the impact of information redundancy.
- **Smoothing factor**: The use of a leveling factor can help reduce the impact of correlations between strategies. This can be achieved by introducing a leveling term in the formulation of the compound Z score that takes correlations into account and promotes strategy diversification.
- **Principal component analysis** (PCA): principal component analysis can be used to reduce the dimensionality of univariate strategies and obtain their linear combinations, thus reducing correlations. This makes it possible to create new variables (principal components) that capture the most relevant information from the original policy, avoiding redundancy.

5. Sequential Screening

The sequential screening is a process of securities selection based on successive criteria, progressively narrowing down the final set of securities. The fundamental concept is to identify relevant factors based on the manager's assumptions and implement a classification of securities using "quantiles or fractals," thereby selecting portfolios with maximum or minimum exposure to the factor.

In our project, we have chosen Lakonishok's screening strategy. This strategy aims to buy stocks of troubled companies due to bad news or reputational issues to exploit their undervaluation.

Here are the steps we will follow to implement the screening strategy according to Lakonishok's model:

- **Step 1**: Market capitalization filtering. We select companies listed that fall within the top 30% based on market capitalization using the factor "CUR_MKT_CAP".
- **Step 2**: Undervaluation assessment. We calculate the industry average for the P/E (Price-to-Earnings), represented by "PE_RATIO" and P/B (Price-to-Book Ratio) ratios, "PX_TO_BOOK_RATIO". We select before the companies with P/E values lower than the industry average, and then we have done the same considering the P/B. This selection is in order to detect the stocks with a potential undervaluation.
- **Step 3**: strategy focuses on the "EQY_REC_CONS" factor in analyst consensus on earnings forecasts. We select companies with an increasing consensus over time, indicating a growing analyst agreement on earnings prospects. We calculate the difference compared to the previous period and choose stocks with a positive difference.
- **Step 4**: : Analizing RSI. We use the Relative Strength Index (RSI) indicator to evaluate the strength of a price trend. We look for companies where the RSI has reversed, transitioning from a low value to a high value, indicating a potential reversal of the downtrend. firstly we select stocks with RSI14D positive and then we select the stocks that have RSI14D higher that RSI30D. In this project the dataset shows the 14 and 30 days RSI, we use those factors instead of the 13 and 26-week of the Lakonishok strategy. At the end we select the 50 stock with a lowe level of RSI30D, in order to obtain a portfolio composed by 50 stocks also in this strategy.

By implementing these steps, we will be able to identify securities that meet Lakonishok's selection criteria, allowing us to identify investment opportunities in undervalued companies or those experiencing a trend reversal. Of course also in this strategy we compute the turning rate in order to understand how change the asset allocation.

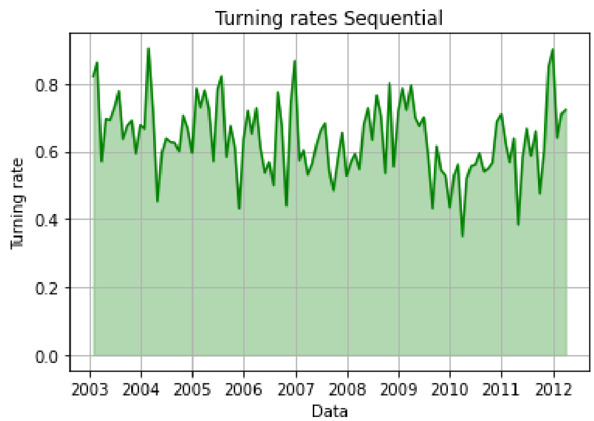

Figure 10 Sequential strategy Turning rates

In sequential screening we note that there is an high level of turnover in portfolio through time. Then there are the graphs in comparison with the benchmark of the sequential screening strategy.

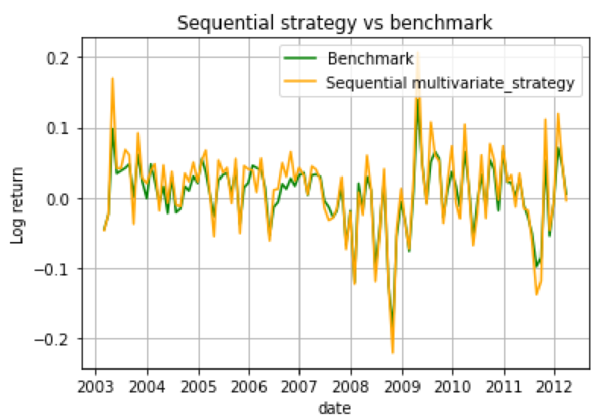

Figure 11 Sequential strategy vs benchmark (log returns)

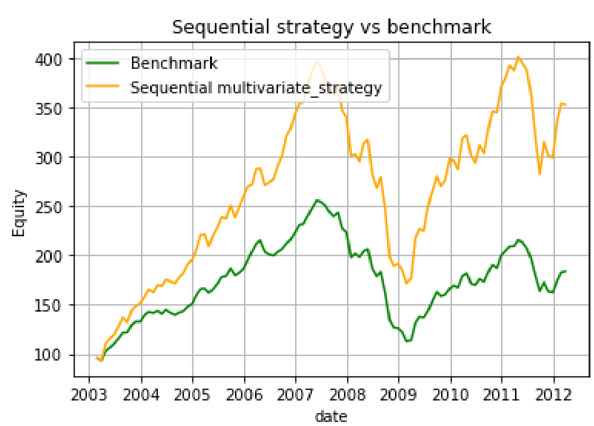

Figure 12 Sequential strategy vs benchmark (equity line)

	Sharpe Ratio	Traynor Ratio	Sortino Ratio	IR
Sequential Screening	0.563	0.092	0.737	1.024

Table 3 Metrics result

The evaluation metrics were also computed, and both the graphs and metrics clearly indicate that the portfolio has achieved excellent performance. Throughout the entire examination period, the equity line consistently outperformed the benchmark.

6. Conclusions and further developments

In conclusion, as anticipated, we have obtained the performance results for each screening strategy and the benchmark. The sequential screening strategy has proven to be the best-performing strategy. Below is the table comparing all the metrics:

	Sharpe Ratio	Traynor Ratio	Sortino Ratio	IR
Univariate Screening	0.119	0.021	0.169	-0.15
Multivariate Screening	0.682	0.115	0.952	0.991
Squential Screening	0.563	0.092	0.737	1.024

Table 4 Metrics result

We can enhance this project by implementing a solution to the collinearity problem that was previously analyzed just theoretically. Additionally, we could incorporate more advanced screening models, such as employing machine learning strategies.