

# 25579 SPRING 2025

# Assignment

October 2025

Kristoffer Glover, [kristoffer.glover@uts.edu.au](mailto:kristoffer.glover@uts.edu.au)

*Faced with the heightened economic uncertainty created by the current US administration, your boss wants the team to investigate the past performance of several defensive investment factors, with the goal of developing a new product designed to appeal to potential investors during these turbulent times. Naturally, as the most junior analyst on the team, you have been assigned the bulk of the groundwork for this new project...*

**Deadline: 11:59pm on Sunday November 9, 2025**

**Score:** marked out of 100 (worth 40% of the final mark)

## 0. Introduction and background

At a recent company-wide strategy session, your investment firm gathered to generate fresh ideas for future portfolio approaches. The Chief Investment Officer expressed concern that the fast-changing economic climate (driven by the unpredictable decision-making of the current US administration) makes it impractical to rely on a single, traditional investment factor.

To address this, the firm has decided to examine how the performance of different factors has varied over time, and whether combining them into a single synthetic factor could deliver more consistent results during turbulent time.

As a rising star on the quant team, you are tasked to explore the possible profitability of numerous strategies based on the following three factors:

- **LOW VOLATILITY:** favours firms whose stock prices exhibit relatively low fluctuations, typically measured by metrics such as standard deviation of returns, idiosyncratic volatility, or sensitivity to market movements.
- **PROFIT:** favours firms with strong profitability metrics such as gross profits-to-assets, return on equity, return on assets, cash flow-to-assets, and gross margin.
- **SAFETY:** favours firms with lower risk characteristics, including low beta, low leverage, reduced bankruptcy risk, and stable earnings.

Your report will have the following components:

1. A **statistical analysis** of the predictive power of the three investment factors (based on a quantile analysis).
2. An analysis of how different **macroeconomic environments** have affected the predictive power of each strategy in the past.
3. A presentation of the **results of backtesting** a long-only investment strategy for each of the three factors over the 1990–2023 period.
4. Optimization of these strategies based on the **rebalancing frequency**.
5. A proposal on how to **optimally combine** the three factors into a single factor [**Black Belt**].
6. Exploration of a **Machine Learning** approach using these three factors.

The deliverables for this assignment are:

- A Jupyter Notebook with your code;
- A pdf file with a (short) written report.

The report should be **no longer than 30 pages**, but my suggestion is to aim at **around 10 pages** of “main text” including tables and pictures. You can also add an appendix if you want to add more tables, etc. Please be sure that all the necessary information is in the main body of the report.

## 1. Statistical Analysis

In the first part of your assignment, you should present a detailed statistical analysis of the **predictive power of each of the three factors** (Low Volatility, Profit, and Safety). Specifically, you should perform a Quantile Analysis for each factor (use 4 bins, unless you think another number is more appropriate, which you should justify).

In your analysis you should also show **how the predictive power (from the quantile analysis) changes over time**. [You do not need to do an IC analysis.] After reading this part the reader should understand:

- **How strong** the predictive power of the three factors is over the entire sample.
- Whether this predictive power has **changed over time**.

Of particular interest to your boss is the correlation between the three factors. Therefore, you should calculate and comment on the **correlations** between the active returns of the three factors over time. Showing how this correlations has **changed over time** would also add to your analysis if you can.

For the **length of the quantile analysis**, you should aim at **around one-two pages**. If you feel the need to add many more tables and graphs, consider the possibility to relegate some of them to the appendix, leaving only your “main narrative” in the body of the report.

## 2. Predictive Power vs. Inflation

We know that factors tend to react differently to different economic environments and so your boss is worried that your historical analysis is useless because “*in the last couple of years the macroeconomic environment has changed significantly.*” Therefore, your boss is interested to know if the predictive power of the three factors is affected by changes in the macroeconomic environment. Specifically, your boss is interested in impact of inflation on factor performance. Data on the US inflation rate (at a monthly frequency) can be found in the file **Inflation.csv**.

You should use this data to define “low”, “normal”, and “high” inflation rate regimes as follows:

- Low: rate < 1.5%
- Normal: 1.5% < rate < 3%
- High: rate > 3%

You should then use these regimes to compute the information ratio of the top quartile portfolio for each of the three factors in each of these regimes (i.e., 3×3 numbers in total). This information should be neatly summarized in a table or a bar chart and after reading your analysis your boss should understand when the factors have done well and when they have not done so well.<sup>1</sup>

The length of this part should be around **one and a half pages** including both your comments and graphs and tables. As with the other parts of the assignments we will mainly mark the PDF report, so all your results must be presented and commented there.

## 3. Backtesting

For this part of your report, you should provide a detailed description of the **performance of the a long-only strategy** that invests in stocks that rank well according to each of the three factors (i.e., three backtests in total). Specifically, each strategy should:

- Invest in the top 300 stocks based on the individual factor;
- Be tested on data between Jan 1990 and Dec 2023;
- Rebalance monthly;
- Assume 0.15% roundtrip transaction costs.

Your report should include:

1. A single (combined) table with the main diagnostics for each strategy;
2. A single (combined) graph with the performance of the three strategies and the benchmark;
3. A single (combined) bar graph with the total active return of the three strategies each year.

---

<sup>1</sup> This [blog post](#) might help inspire you for the type of analysis that can be made.

The results should **be properly commented in your report**. After reading this part, the reader should know:

- Whether the strategies performed well or not over this period;
- Whether the ability of the strategies to “beat the market” was constant across this period or has changed in time.

The length of this part should be around **two pages**, including comments and **graphs/tables**.

## 4. Optimization

One of the key decisions when implementing any active investment strategy is the trade-off between rebalancing frequency and the signal decay. Therefore, you should find the optimal **rebalancing frequency** for all three strategies. When determining what is ‘optimal’ use the information ratio of the backtested performance as the thing you are trying maximize.

If your optimal rebalancing frequency is different for the three strategies, then you should also comment on the reason why you think this might be the case.

### Notes:

1. When performing the backtests for strategy optimization, you should use the same time-period and transaction cost assumption as in Part 3;
2. In the interest of time, you don’t need to consider rebalancing frequencies longer than six months;
3. Do not repeat the backtesting function multiple times in your code, instead use for-loops.

For the length of this part, you should aim at around **one page**. If you feel the need to add many more tables and graphs, consider the possibility to relegate some of them to the appendix, leaving only your “main narrative” in the body of the report.

## 5. Combining the Factors

As noted by your boss during the strategy meeting, rather than choosing which of the three factors to implement for your investment strategy, it may be beneficial to combine all three factors into a single factor and then implement the strategy based on this combined factor. Therefore, in this part of your report, you should **propose a “clever” way to combine** the three defensive factor values (low volatility, profit, and safety) into a single factor. Obviously, the aim is to combine them in a way that will improve the **out-of-sample** performance of your strategy. You should also avoid using decision trees, since this is bases of the final part of your analysis.

In your report should include:

- A **description of the method** used to combine the factors and the economic/statistical rationale behind it;
- A **brief summary** of the performance of a backtest of the new combined factor with the same assumptions as in Part 3 above. Specifically, it should include:
  - A table with the main diagnostics;
  - A graph with the performance of your combined factor, each individual factor, and the benchmark.

#### ATTENTION

This part is the “**black-belt**” component of the assignment in which you will need to research methods of combining factors and how to implement this in python.

This section is worth 20% of the total Assignment mark. If you do not want to attempt this part, then **simply take the average value of the three factors each month** as the combined factor to perform your backtest. A maximum of 5% is then available for this section (for the backtest results of the simple average strategy).

The length of this part should be around **one-two pages** including both your comments and graphs and tables. As with the other parts of the assignments we will mainly mark the PDF report, so all your results must be presented and commented there.

## 6. Machine Learning Approach

You have been experimenting with decision trees lately, and so to finish up your report, you decide to investigate whether these trees can be used to improve the performance of your strategy. Specifically, you are wondering whether a decision tree model can exploit nonlinearities in the three factors (and their relationship) and find a ‘better’ way of picking good and bad stocks (rather than simply ranking based on the signal).

To do so, you should build and train a decision tree with the following characteristics:

- **Target Variable:** the tree should try to predict if the next month’s return will be above or below the median return of all stocks that month;
- **Sample:** You should train the model on data from the beginning of 1990 to the end of 2014 (we will use the rest for out-of-sample testing);
- **Max Depth:** 3;
- **Factors:** the tree should use only the three factors (Low Volatility, Profit, and Safety).

Once the decision tree is trained, you should plot the tree itself and describe **in words** how the decision tree uses the three factors to decide which stocks are labelled as “good” and which are “bad.” How much does each variable contribute to reducing the Gini Impurity of the model (i.e., the **feature importance**).

To assess the quality of the model you should use the out-of-sample predictions from the model for Jan 2015 to Dec 2023 in the `ml_analysis()` function and report the main diagnostics. Comment on the performance you see.

Finally, after presenting the results of your decision tree to your boss, she notes that time-series variables can also be used to as features in a decision trees, and so she recommends that you re-train your model with the addition of the inflation variable used in Part 2. In other words, include the variables in the file “Inflation.csv” along with the original three factors, and retrain the decision tree with the same characteristics as described above.

Again, you should plot the tree itself and describe how the decision tree uses the three factors **AND** the inflation time-series to decide which stocks are “good” and which are “bad.” How does the **feature importance** change between the two decision trees, and how does the out-of-sample performance (using `ml_analysis()` for Jan 2015 to Dec 2023) vary from the simple 3-feature tree described earlier?

The length of this part should be around **two pages** including both your comments and graphs and tables.

## 7. Data

Together with the assignment you will find the following data files:

- **FACTORS.zip** contains the monthly data of three “low volatility” factors, the “profit” factor, and the “safety” factor, for a sample of US stocks. (Note: All factors have already been normalized.)
- **Prices.zip** contains the monthly (adjusted) prices for US companies from 1980.
- **Market\_cap.zip** contains the monthly market capitalization for US companies from 1980.
- **Inflation.csv** contains a monthly time series of the US inflation rate (per annum).

## 8. Timeline

The content of the assignment is covered in different weeks of the subject. Here is a rough estimation of when the material (and code) necessary for each session is covered:

Assignment Part	Week	Notebooks / Concepts
1. Low Vol and How to Measure It	7 and 8	None. Concepts of Investment Factor, Risk Premium and Market Anomaly
2. Quantile Analysis	8	08 – Quantile Analysis
3. Predictive Power vs. Economic Environment	8	08 – Quantile Analysis plus your own coding/research skills
4. Backtesting	9 and 10	09 – Full Backtesting 11 – Long-Short Portfolios
5. Optimization	9 and 10	09 – Full Backtesting 11 – Long-Short Portfolios
6. Combining the Factors		Your own coding/research skills
7. Machine Learning	12	13 – Decision Trees plus your own coding/research skills

## 9. Submission

You will submit:

1. A PDF file with your report (**A\_XXXXX.pdf**);
2. A Jupyter Notebook with all your code (**A\_XXXXX.ipynb**).

Where XXXXX is your UTS student ID number.

Your PDF report should contain a professionally looking **cover page** (no specific format required) with your name and Student ID.

Tables in the report should be edited and not simply cut and paste images from the Notebook. The graphs coming from the notebooks should be exported as image files and then inserted in the report and not “screenshotted and pasted”.

The files will be **submitted electronically** using the electronic drobox in Canvas.

## 9.1 Structure of the Jupyter Notebook

1. You should start with an empty notebook and add all the cells that you need;
2. Your python notebook should contain your **name and student ID** at the top in the initial markdown cell;
3. The first code cell should contain all the **import statements** (including the *apmodule*). No import statements should be located anywhere else in the file;
4. The second code cell should contain all the **user-generated functions** (if any). You should not report here the functions from our own *apmodule*, but use them from the module;
5. The third code cell should contain the initial loading of all the files;
6. The notebook can import **any of the libraries used** in the subject including the *apmodule* library. If you want to import a library that we have not used in class, please check with me beforehand;
7. Your code should be **well commented** using markdown cells and # comments in the code cells. Markdown cells are preferred where possible. Use # comments only when you need to add a comment in the middle of a code block;
8. You should use markdown cells to **help the reader navigate the file**, basically explaining what you are doing in the code cells (or at least the most relevant ones);
9. You should choose **which results to present** in the notebook in order to help the reader without overwhelming. For example, if you simply modify a column in a DataFrame it may not be necessary to show the result on screen;
10. I need to be able to **run your notebook in one go** from top to bottom, so before submitting make sure that all your code runs properly. Do not worry about the running time of your notebook. You do not need to optimize for speed;
11. You can **copy portions of the notebooks** created in class, but please **do not follow** their structure. The notebooks used in class were designed for teaching purposes and do not reflect the sequence of objectives of your notebook. Not everything we have done in class fits with this research project.

You should consider the notebook as the **natural complement to the PDF report**. Imagine that your reader will go through the report and then, from time to time have a look at the corresponding section of the notebook to understand some technical aspect of your model. This implies that:

- All the important information is in the report. You **will not receive marks** for required elements of the assignment that are in the code but **not** in the report;
- The notebook should not contain significant elements if the results of such analysis are not commented in the report.



## 10. Marking Guide

Component	Value	Below Expectations 0 – 50%	Meets Expectations 50 – 80%	Exceeds Expectations 80 – 100%
<b>Quantile Analysis</b>	10	<ul style="list-style-type: none"> <li>There are serious methodological errors in the analysis.</li> <li>The analysis is incomplete.</li> <li>The results are not well presented.</li> <li>The results are not well commented.</li> </ul>	<ul style="list-style-type: none"> <li>There are no serious methodological errors, and the analysis is complete.</li> <li>The presentation of the results is sufficient.</li> </ul>	<ul style="list-style-type: none"> <li>The analysis is rich, and the problem is analysed from multiple points of view presenting a comprehensive picture.</li> <li>The results are presented in a professional manner providing enough information without overwhelming the reader.</li> </ul>
<b>Macroeconomic Analysis</b>	10	<ul style="list-style-type: none"> <li>There are serious methodological errors in the analysis.</li> <li>The analysis is missing or incomplete.</li> <li>The results are not well presented.</li> <li>The analysis is purely theoretical or based on external evidence.</li> </ul>	<ul style="list-style-type: none"> <li>The analysis uses data from the assignment and is methodologically correct.</li> <li>The presentation of the results is sufficient.</li> </ul>	<ul style="list-style-type: none"> <li>The analysis is rich, and the problem is analysed from multiple points of view (complementary techniques, multiple measures, ...) presenting a comprehensive picture.</li> <li>The results are presented in a professional manner providing enough information without overwhelming the reader.</li> </ul>
<b>Backtesting</b>	10	<ul style="list-style-type: none"> <li>There are serious methodological errors in the analysis.</li> <li>The analysis is incomplete.</li> <li>The results are not well presented.</li> </ul>	<ul style="list-style-type: none"> <li>There are no serious methodological errors, and the analysis is complete.</li> <li>The presentation of the results is sufficient.</li> </ul>	<ul style="list-style-type: none"> <li>The analysis is rich, and the problem is well analysed, presenting a comprehensive picture.</li> <li>The results are presented in a professional manner providing enough information without overwhelming the reader.</li> </ul>
<b>Optimization</b>	10	<ul style="list-style-type: none"> <li>There are serious methodological errors in the analysis.</li> <li>The results are not well presented.</li> <li>The results are not well commented.</li> </ul>	<ul style="list-style-type: none"> <li>There are no serious methodological errors, and the analysis is complete.</li> <li>The presentation of the results is sufficient.</li> </ul>	<ul style="list-style-type: none"> <li>The analysis is rich, and the problem is well analysed presenting a comprehensive picture.</li> <li>The results are presented in a professional manner providing enough information without overwhelming the reader.</li> </ul>

Component	Value	Below Expectations 0 – 50%	Meets Expectations 50 – 80%	Exceeds Expectations 80 – 100%
<b>Combining Factors</b>	20	<ul style="list-style-type: none"> <li>The simple maximum is used to combine factors.</li> <li>There are serious methodological errors in the analysis, or the analysis is incomplete.</li> <li>The results are not well presented and commented.</li> </ul>	<ul style="list-style-type: none"> <li>The presentation of the method used is complete and based on coherent reasoning.</li> <li>The presentation of the results is sufficient.</li> </ul>	<ul style="list-style-type: none"> <li>The presentation conveys a deep understanding of issues at hand and the pros and cons of different methods.</li> <li>The analysis is complete and functional.</li> <li>The results are presented in a professional manner providing enough information without overwhelming the reader.</li> </ul>
<b>Machine Learning</b>	20	<ul style="list-style-type: none"> <li>There are serious methodological errors in the analysis.</li> <li>The results are not well presented.</li> <li>The results are not well commented.</li> </ul>	<ul style="list-style-type: none"> <li>There are no serious methodological errors, and the analysis is complete.</li> <li>The presentation of the results is sufficient.</li> </ul>	<ul style="list-style-type: none"> <li>The analysis is rich, and the problem is well analysed presenting a comprehensive picture.</li> <li>The results are presented in a professional manner providing enough information without overwhelming the reader.</li> </ul>
<b>Quality of the Notebook</b>	10	<ul style="list-style-type: none"> <li>The notebook does not run properly.</li> <li>The structure of the notebook is confusing and unhelpful.</li> <li>The notebook is not commented, or the comments are insufficient and/or unclear.</li> </ul>	<ul style="list-style-type: none"> <li>The notebook runs properly.</li> <li>The notebook is well structured.</li> <li>The comments are sufficiently to allow the reader to properly follow the process.</li> </ul>	<ul style="list-style-type: none"> <li>The notebook is very well designed.</li> <li>The coding is elegant and makes good use of functions and loops to avoid repetitions.</li> <li>The variables are properly named and easy to follow.</li> <li>The choice of which intermediate results to present is effective to help the reader.</li> <li>The comments are rich and well-made, allowing the reader to follow both the financial logic and the code structure.</li> </ul>
<b>Quality of the Report</b>	10	<ul style="list-style-type: none"> <li>There are frequent spelling and punctuation errors.</li> <li>The language is often non correct and does not convey financial concepts in an effective way.</li> <li>There is not a common graphic style and colour palette.</li> <li>Graphs and tables are badly formatted or simply copy-pasted from the notebook.</li> </ul>	<ul style="list-style-type: none"> <li>There are only minor spelling and punctuation errors.</li> <li>The language is correct, and the financial terminology properly used.</li> <li>There is evidence of editing to make the feel professional.</li> </ul>	<ul style="list-style-type: none"> <li>The document has a truly professional tone and feel.</li> <li>The narrative structure is easy to follow.</li> <li>The language is very effective, and the key ideas presented in a succinct but clear way.</li> <li>Tables and graphs are effective in communicating the key results.</li> </ul>