25579 AUTUMN 2024

Assignment

April, 2024
Marco Navone, marco.navone@uts.edu.au

Worried about the persistence of high interest rates, your company is exploring possible new investment strategy focused on investing in companies with high degrees of "financial health" as measured by their ability to service their debt. Your task is to assess the feasibility of such a strategy.

Deadline: 11.59pm Monday May 27, 2024.
Score: marked out of 100 (worth 40% of the final mark)

0. Introduction and background

Last week your investment company held an all-hands-on-deck meeting to brainstorm new ideas for future investment strategies. Your Chief Investment Officer is particularly concerned that all players in the industry, including your own company, offer very similar products based on a mixture of "classic" (i.e. old school) investment factors such as value, momentum, and profitability…

Another interesting consideration is that the recent raising of interest levels has put on the spot the ability of leveraged companies to generate enough cash flows to be able to service current and future debt. This would be similar, in sprit, to the "safety" component of the quality investment factor, where one would invest in firms with low levels of debt.

As a result, a new idea has emerged: would it be profitable to systematically invest in companies with high levels of:

- Interest Coverage Ratio (INTCOV).
- Cash flow to current liabilities Ratio (OCF_LCT).
- Cash flow to debt Ratio (CASH_DEBT).

As a rising star in the quant team, you are tasked to **explore the possible profitability of a strategy** based on this idea.

Your report will have the following components:

1. **Statistical analysis** of the predictive power (based on the Information Coefficient Analysis) of the three ratios, as well as of a synthetic factor calculated as the average of the three.

2. A presentation of the **results of backtesting** of a 130/30 investment strategy based on the "average" synthetic factor over the 2007–2023 period.

3. Optimization of this strategy based on the **number of shares** over/underweighted and the **rebalancing frequency**.

4. Address some of the concerns of the CIO of your company [***Black Belt***]**.**

5. Exploration of a **Machine Learning** approach to this strategy.

The deliverables for this assignment are:

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

There is **no formal length requirement** for the report. My suggestion is to aim at anywhere **between 8-10 pages** of "main text" including tables and pictures. You may 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 these three financial ratios** (Interest Coverage Ratio, Cash flow to current liabilities Ratio, and Cash flow to debt Ratio) as well as a syntenic "Average" factor built as the average of these three.. The analysis should be based on the Information Coefficient (you should NOT use Quantile Analysis).

In doing this analysis you should use all the available data (1980-2023) and you should show **how the predictive power changes in time**. After reading this part the reader should understand:

- **How strong** is the predictive power of these factors over the entire sample?
- Has this predictive power **changed over time**?

For the **length of the statistical 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. Backtesting

For this part of the report, you should create and backtest a long-short 130/30 strategy based on the "Average" factor (you can forget about the three original ratios). Your strategy should:

- Overinvest in the top 300 stocks and underinvest in the bottom 300.
- Have an active percentage of 30% (should, in fact be a 130/30 strategy).
- Be tested on data between 2007 and 2023.
- Rebalance monthly.
- Assume 0.2% roundtrip transaction costs.

After reading this part of the report your boss should have a clear picture of **how the strategy performs** over the entire period and whether the performance is focused in a **particular period of time** (for example the earlier part of the sample, or the covid years…).

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

3. Optimization

You should find the optimal parameters of the strategy in order to maximize the performance. Here you should focus on the **number of shares** to over-under-weight and on the **rebalancing frequency**.

For each parameter you should try **three** different values. You should use a **reasonable range** of numbers to capture meaningful differences in the strategy (for example you should not test 301, 302 and 303 shares…).

After reading this part the reader should know what combination of parameters produces the best performance.

Notes:

1. When performing the backtests for strategy optimization, you should use the same time period and transaction cost assumption as in part three.
2. If you can, try not to repeat the backtesting function nine times, 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.

4. Addressing your boss's concerns

Your boss, the Chief Investment Officer of the company, has two main concerns that would like you to explore.

Question 1: Is this strategy **simply a re-skinned version** of some other well-known strategy such as value, momentum and/or profitability? In other words, are these "new" investment factors or just a different way of measuring older factors?

Question 2: These new ratios measure the ability of the company to generate enough cash flows to payoff interest on existing debt. It stands to reason that the predictive power of these factors should be **stronger when interest rates are higher**. Is this true?

It is **up to you to decide** how to perform your analysis but after reading your report your boss should have clear answers to these two questions.

ATTENTION

This is the **"black-belt"** part of the assignment in the sense that you will have to develop some original code and come up with **your own way** to perform the calculation and present the results in the most effective way. Among the data provided you will find some useful indicators (interest rates and some "classic" investment factors).

This part is worth 25% of the Assignment mark. If you do not want to attempt this analysis, you can simply try to **answer the two questions theoretically** based on sound economic reasoning. A maximum of 10% can be awarded for this non-numerical analysis.

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 have to be presented and commented there.

5. Machine Learning Approach

Your company has been experimenting with decisions trees, so you decide to round-up your report with a simple analysis of the inclusion of these three signals (you can forget about the "Average" in this part) in a decision tree model.

Build and assess the predictive power of a decision tree with the following characteristics:

- **Target Variable**: the tree should try to predict if the next month return will be **positive or negative**.
- **Sample**: From the beginning of 2007 to the end of 2023.
- **Max Depth**: 3

- **Factors**: the tree should use data from
    - The three new factors (INTCOV, OCF_LCT, CASH_DEBT).
    - A value factor (EPQ).
    - A momentum factor (MOM12).
    - A profitability factor (GP).
    - A low-volatility factor (TVOL).
- **Accuracy** (out of sample) should be measured with a 10-fold Cross Validation model.

Your main goal in this section is to assess whether the three new factors:

- Significantly affect how the decision tree predicts future performance.
- Increase the (out of sample) forecasting accuracy of the model.

6. Data

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

- **INTCOV.zip** contains the monthly data of the Interest Coverage Ratio of US stocks from 1908 to 2023. The ratio is measured as *Net Income* divided by *Interest Expenses*.
- **OCF_LCT.zip** contains data on the Cash flow to Current Liabilities Ratio measured as *Operating Cash Flow* divided by *Current Liabilities*.
- **CASH_DEBT.zip** contains data on the Cash flow to Debt Ratio measured as *Operating Cash Flow* divided by *Total Debt*.
- **GP.zip**, **MOM12.zip**, **TVOL.zip** and **EPQ.zip** contain monthly data of traditional investment factors for profitability (Gross Profitability), momentum (12-Months Momentum), low volatility (Total Volatility) and value (Quarterly Earnings to Price).

  All the factors are already normalized and winsorized (i.e. they are "Ready to go" as the other ones used in class).

- **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.
- **names.zip** contains name and industry affiliation for all US companies.
- **INT_RATES.csv** contains the monthly time series for the following macro-economic indicators:
    - **10YTR:** yield to maturity of a 10-years US Government Bond.
    - **1YTR:** yield to maturity of a 1-year US Government Bond.

7. Timeline

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

Assignment Part	Week	Notebooks / Concepts
1. Statistical Analysis	6	07 - Information Coefficient
2. Backtesting	8 and 9	09 – Full Backtesting 11 – Long-Short Portfolios
3. Optimization	8 and 9	09 – Full Backtesting 11 – Long-Short Portfolios
4. Black-Belt		Your own coding/research skills
5. Machine Learning	12	13 – Decision Trees

8. Submission

You will submit:

1. A pdf file with your report (**A_XXXXX.pdf**) where XXXXX is your UTS student ID number. Files with wrong names or wrong extension **will attract 5% penalty**.

2. A Jupyter Notebook with all your code. This should be **the same** Jupyter Notebook that has been emailed to your UTS address from **apmfilesexchange@gmail.com.** If you cannot find this email, please **contact the subject coordinator** at marco.navone@uts.edu.au as soon as possible.

Your pdf report should contain a professional-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.

8.1 Structure of the Jupyter Notebook

1. You should start with the notebook you **received by email**.
2. Your python notebook already contains your **name and student ID** at the top in the initial markdown cell. Do not modify this cell.
3. The first code cell should contain all the **import statements** (including the apmodule).
4. The second code cell should contain all the **user-generated functions**. 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. The file can only load and use the **datafiles provided with the assignment** or other datafiles used in the subject. No external data can be used.
8. Your code should be **well commented** using either markdown cells or # 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.
9. 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).
10. 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.
11. 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.
12. 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 (for example the quarterly IC analysis) if the results of such analysis are not commented in the report.

9. Marking Guide

Component	Value	Below Expectations 0 – 50%	Meets Expectations 50 – 80%	Exceeds Expectations 80 – 100&
Statistical Analysis:	15	• There are serious methodological errors in the analysis. • The analysis is incomplete. • The results are not well presented.	• There are no serious methodological errors, and the analysis is complete. • The presentation of the results is sufficient.	• The analysis is rich, and the problem is analysed from multiple points of view (complementary techniques, multiple time horizons, …) presenting a comprehensive picture. • The results are presented in a professional manner providing enough information without overwhelming the reader.
Backtesting	10	• There are serious methodological errors in the analysis. • The analysis is incomplete. • The results are not well presented.	• There are no serious methodological errors, and the analysis is complete. • The presentation of the results is sufficient.	• The analysis is rich, and the problem is well analysed presenting a comprehensive picture. • The results are presented in a professional manner providing enough information without overwhelming the reader.
Optimization	10	• There are serious methodological errors in the analysis. • The results are not well presented. • The results are not well commented.	• There are no serious methodological errors, and the analysis is complete. • The presentation of the results is sufficient.	• The analysis is rich, and the problem is well analysed presenting a comprehensive picture. • The results are presented in a professional manner providing enough information without overwhelming the reader.
Black Belt:	25	• There are serious methodological errors in the analysis. • The analysis is missing or incomplete. • The results are not well presented. • The analysis is purely theoretical or based on external evidence.	• The analysis uses data from the assignment and is methodologically correct. • The presentation of the results is sufficient.	• The analysis is rich, and the problem is analysed from multiple points of view (complementary techniques, multiple measures, …) presenting a comprehensive picture. • The results are presented in a professional manner providing enough information without overwhelming the reader.
Machine Learning	10	• There are serious methodological errors in the analysis. • The results are not well presented. • The results are not well commented.	• There are no serious methodological errors, and the analysis is complete. • The presentation of the results is sufficient.	• The analysis is rich, and the problem is well analysed presenting a comprehensive picture. • The results are presented in a professional manner providing enough information without overwhelming the reader.

[Continues in next page]

Component	Value	Below Expectations 0 – 50%	Meets Expectations 50 – 80%	Exceeds Expectations 80 – 100&
Quality of the Notebook	15	• The notebook does not run properly. • The structure of the notebook is confusing and unhelpful. • The notebook is not commented, or the comments are insufficient and/or unclear.	• The notebook runs properly. • The notebook is well structured. • The comments are sufficient to allow the reader to follow the process.	• The notebook is very well designed. • The coding is elegant and makes good use of functions and loops to avoid repetitions. • The variables are properly named and easy to follow. • The choice of which intermediate results to present is effective to help the reader. • The comments are rich and well-made, allowing the reader to follow both the financial logic and the code structure.
Quality of the Report	15	• There are frequent spelling and punctuation errors. • The language is often non correct and does not convey financial concepts in an effective way. • There is not a common graphic style and colour palette. • Graphs and tables are badly formatted or simply copy-pasted from the notebook.	• There are only minor spelling and punctuation errors. • The language is correct, and the financial terminology properly used. • There is evidence of editing to make the feel professional.	• The document has a truly professional tone and feel. • The narrative structure is easy to follow. • The language is very effective, and the key ideas presented in a succinct but clear way. • Tables and graphs are effective in communicating the key results.