Aalto University
School of Science

MS-E2177 - Seminar on Case Studies in Operations Research

Applying Advanced Analytics in Asset Allocation

Project Plan

Client:
Varma

Team:
Leevi Olander (Project Manager)
Julius Lind
Samuli Mäkinen
Rasmus Heikkilä

March 14, 2019

1 Background

The pension system provides individuals with income after they have retired from working life and no longer earn regular income from employment. In Finland, there are three statutory pensions: the national pension, the guarantee pension and earnings-related pension. The first two provide individuals with basic income security and are administered by the Social Insurance Institution of Finland (Kansaneläkelaitos). The earnings-related pensions are administered by many organizations. The national pension and earnings-related pension complement each other: as the amount of the earnings-related pension that an individual is entitled to increases, the amount of national pension decreases. [1]

The national pension and the guarantee pension are funded from the state budget. In contrast, the earnings-related pensions are funded by insurance contributions from employers and employees. Funding of the earnings-related pensions is based on two principles. The pay-as-you-go principle means that the contributions from currently employed individuals are used to finance the pensions of current retirees. The partial funding principle stands for investing a fraction of the insurance contributions. This decreases the amount of insurance contributions required to finance the pensions, as investment returns can be used to cover a part of the cost. [2]

Recently the pension expenditure has exceeded the total insurance contributions in both private and public sectors [3]. For instance, in the private sector in 2017 insurance contributions totalled 16.5 billion EUR, whereas the sum of pensions paid was 17.2 billion EUR [3]. As the number of pensioners in Finland increases and the working age population shrinks, the deficit will only increase. This emphasizes the importance of pension funds in the Finnish pension system.

Our client is the Varma Mutual Pension Insurance Company, which manages earnings-related pensions for private sector employees and self-employed individuals. Varma is responsible for the pensions of almost 900 000 Finns, making it one of the largest pension insurance companies in Finland. It is a mutual company owned by its clients: private companies, insured employees, self-employed persons and owners of guarantee capital. [4]

The 45 billion EUR investment portfolio of Varma consists mainly of equities (46% of the portfolio value in 2017), fixed-income investments (28%), real estate investments (8%) and hedge funds [4]. Investment returns are realized not only as changes in the value of the investments, but also as cash income in the form of dividends, interest payments and rents. Achieving good return on investments is of great importance to Varma, as the company uses the returns to cover the deficit between insurance contributions received and pensions paid [4].

2 Objectives

The key decision in managing the investment portfolio is the allocation of funds between equities and fixed-income investments, as these are the largest asset categories. The objective of our team is to use machine learning methods to make allocation decisions based on a large number of macroeconomic and financial market indicators. Our goal is to build two models: one to make allocation decisions between global stocks and global bonds, and another to make allocations between short and long duration global bonds. Time horizons of 3-6 months will be used in the allocation decisions.

Index	**Bloomberg ticker**
Global stock index (MSCI AC TR USD)	NDUEACWF Index
Global bond index (Bloomberg Barclays Global Agg 1-3 Year TR Unhedged USD)	LG13TRUU Index
Global bond index (Bloomberg Barclays Global Agg 7-10 Year TR Unhedged USD)	LG71TRUU Index

Table 1: The target indexes of the project and their Bloomberg tickers.

The allocation decisions will be based on a dataset of 61 time series of various market indicators. The dataset consists of daily, weekly and monthly time series from the last 8 to 15 years. Three of these time series can be considered our target variables, which depict the performances of global stock markets, global short- and long duration bonds. The target variables are shown in Table 1. The performance of the developed models will be evaluated against a static 50/50 allocation, using expected returns and volatility as performance measures.

3 Tasks

3.1 Supporting Tasks

To efficiently make use of conflicting personal schedules between team members, the team has to make decisions regarding tools. First, we agreed to use Python 3 as our shared coding language and chose to follow the PEP 8 coding convention [5]. Secondly, we opted to use Anaconda 3 [6], which is a widely used data science platform. Finally, we selected Aalto Version Control System [7] as our version control system.

3.2 Client Interaction & Reporting

In the initial meeting with our client, Varma, we confirmed the goals and objectives of this project. The client's previous research in this topic provided us valuable insights and NDAs were given for us to sign. After the NDAs were signed, we received the data that is used in this project. Subsequent meetings will be agreed upon on an as-needed basis, however at least so that a meeting will be organized around each deliverable deadline.

3.3 Data

Before method development, we must confirm and verify the validity of the data. The data may contain inaccuracies, for example abnormally high spikes in the data. Visual verification will be used in this task, whereafter data smoothing and frequency transformation methods will be applied due to all data not having the same frequency.

3.4 Clustering of Indicators & Factor Analysis

In this task the team will cluster the data sets into smaller groups with similarities. The groups are not defined beforehand, but could refer to, for example, growth, inflation and risk. By letting the data speak for itself, the clustering may bring significant and interesting new information. The team has chosen to implement fuzzy clustering based on Cross Correlation, hierarchical clustering and some machine learning techniques may also be applied. We chose these methods because k-means, DBSCAN and other conventional clustering methods are not suitable for time series clustering. These methods require statistics like mean and variance to be calculated from the time series and then clustering based on these statistics. Generally these statistics are insufficient to represent time series comprehensively, thus clustering based on them is ineffective. Additionally, factor analysis will be applied via Principal Component Analysis (PCA) [8] and Independent Component Analysis (ICA) [9].

3.5 Predictive Models

The team is tasked to apply supervised learning and other models to identify relationships with predictive power. The client suggested the following techniques to be analyzed: Lasso Regression [10], K-nearest neighbors [11] and Random Forests [12]. However, we will strive to implement other methods as well, for example, Neural Networks [13] and use the features extracted by PCA and ICA. These predictive models may give signals, which are then employed to dynamically allocate assets.

3.6 Performance Benchmark

In the final phase of the project we assess the performance of the models to a static model. The comparable static model is a 50/50 allocation in stocks and bonds and the performance will be measured by standard return and risk metrics.

4 Schedule

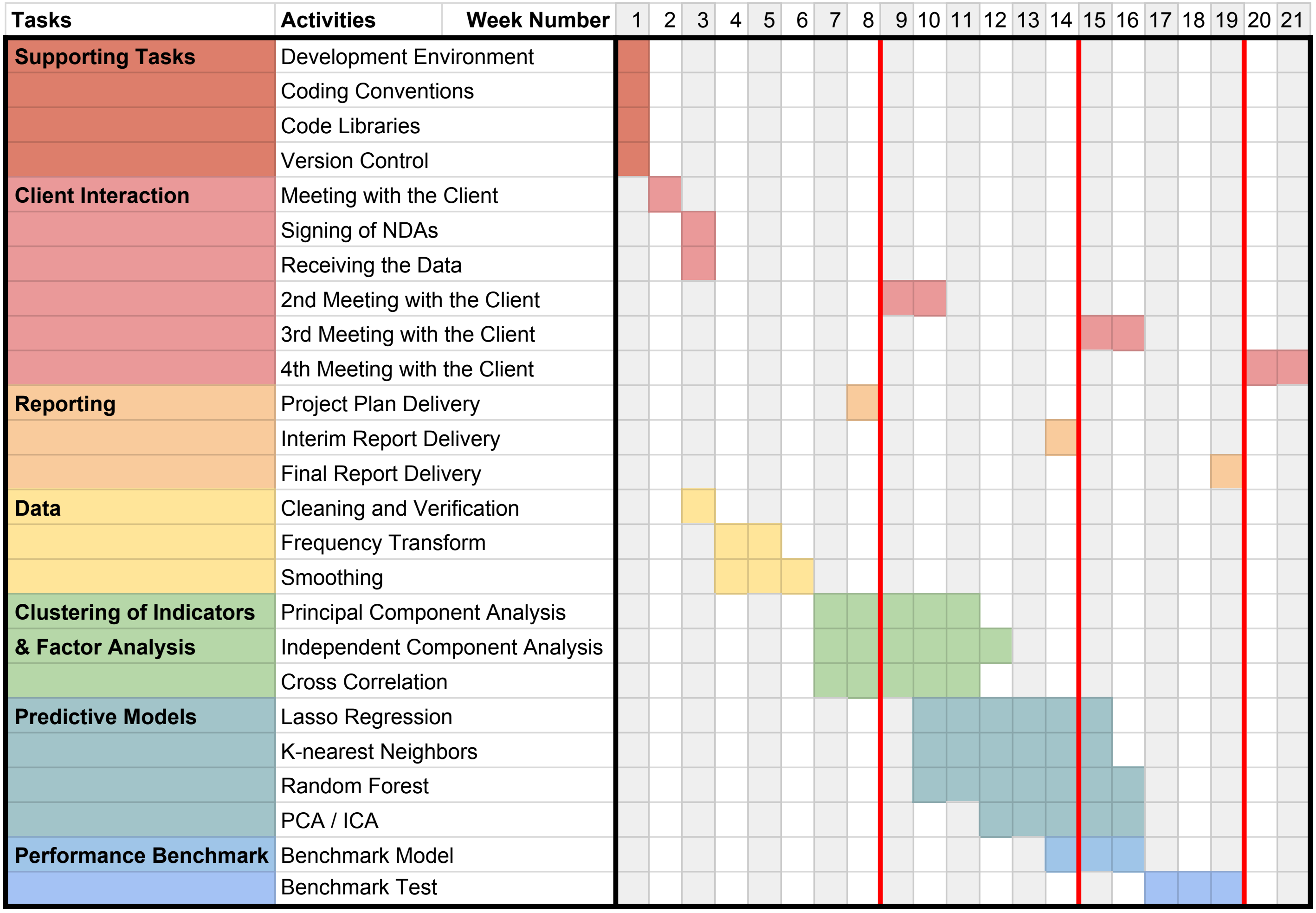

Figure 1: A Gantt chart of the project schedule. The week numbers corresponds to the weeks of the course MS-E2177, which started on 11.01.2019. The red vertical lines show the deadline for deliverables.

Figure 1 shows the scheduling of the tasks related to this project. The tasks between sections have some overlappings, which can partly be explained by a decentralized working schedule, where team members independently work on tasks.

5 Resources

Our team consists of four Systems and Operations Research students. We all have some background in machine learning and neural networks and we have a good knowledge of multiple mathematical and statistical methods that could be useful in this project. However, our knowledge in stock and bond markets and financial modelling is not extensive and we will need to allocate some time to study the fundamentals to be able to apply the mathematical methods. Leevi has been selected to be the project manager as he has more time available to put into the project. For other team members the workload is distributed evenly in the way that their strengths and knowledge has been taken in consideration.

Our client has a Cross Assets and Allocation Team of four investment professionals with a lot of experience from financial markets. They are willing to help us during the project by, for example, discussing about methods used, providing data and other materials and helping us to find clear direction for the project. They have already done some research on the project topic so they have a good knowledge of the possible problems we might face and the indicators we use in the models. The client team has already provided us data that consists of multiple different indicators from last 8 to 15 years depending on the indicator.

In addition to data and material provided by the client, we will study literature on the subject and investigate what kind of methods have been tried and how they have performed. We try to use academic research as our main source but we understand that there might be limitations to it regarding to this subject, at least in well performing models. To investigate the literature, we have access to scientific journals in the university's library. We also have the support of professors Ahti Salo and Antti Punkka who are the staff of the project course together with our project manager Leevi.

6 Risks

Risk	Likelihood	Impact	Effect	Mitigation measures
Model fails in performance benchmark	High	High	Model is not useful to the client.	Testing multiple models and careful study of relevant literature.
Data quality issues	Medium	Medium	Misleading, incorrect or inaccurate results.	Understanding the limitations both in data and methods used.
Too large workload	Medium	Medium	Lower quality and delays in project schedule.	Start with a small number of models, expand the scope only if time permits. Maintain active discussion with the client about task prioritization.
Team member absence	Low	High	High workload for other team members.	Good communication between the project manager and the rest of the team. Clear schedule and fast reaction to delays.
Communication issues with the client	Low	Medium	Result is not what the client wanted.	Good communication with client by email and frequent meetings with the client.
Issues with computational resources	Very low	Low	Delays in project schedule.	Use algorithms with low enough computational burden.

Table 2: Risks related to the project.

We have identified the risks in Table 2 that could affect the outcome of the project. The most significant risk is that the model fails in performance benchmarking and therefore is not useful to the client. The efficient market hypothesis states that markets use all available information in determining the asset prices. This means that developing a predictive model that beats the market in forecasting the asset price development is impossible assuming the hypothesis is true. Although real markets are not necessarily always efficient, our data is in daily, weekly and monthly frequencies. At this timescale we expect the global financial markets to be efficient.

References

[1] Finnish Centre of Pensions. *Pensions part of social security.* Jul. 3, 2018. [Online]. https://www.etk.fi/en/the-pension-system/pension-security/pension-as-social-security/. Accessed Feb. 23, 2019.

[2] Finnish Centre of Pensions. *Principles of pension financing.* Jan, 22, 2019. [Online]. https://www.etk.fi/en/the-pension-system/pension-financing-and-investments/financing-principals/. Accessed Feb. 23, 2019.

[3] The Finnish Pension Alliance. *Circulation of pension money.* [Online]. https://www.tela.fi/en/circulation_of_pension_money. Accessed Feb. 23, 2019.

[4] Varma Mutual Pension Insurance Company. *Varma annual and CSR report 2017.* 2018. [Online]. https://www.varma.fi/globalassets/vuosikertomus/varmas-annual-and-csr-report-2017.pdf. Accessed Feb. 23, 2019.

[5] Python.org. *PEP 8 – Style Guide for Python Code* [Online]. https://www.python.org/dev/peps/pep-0008/. Accessed Feb. 25, 2019.

[6] Anaconda.com. *Anaconda Distribution* [Online]. https://www.anaconda.com/distribution/#download-section. Accessed Feb. 25, 2019.

[7] Aalto Version Control System. *Aalto University Version Control System* [Online]. https://version.aalto.fi. Accessed Feb. 25, 2019.

[8] Jolliffe, Ian. *Principal component analysis.* Springer Berlin Heidelberg, 2011.

[9] Hyvärinen, Aapo, Juha Karhunen, and Erkki Oja. *Independent component analysis.* Vol. 46. John Wiley & Sons, 2004.

[10] Tibshirani, Robert. *Regression shrinkage and selection via the lasso.* Journal of the Royal Statistical Society: Series B (Methodological) 58.1 (1996): 267-288.

[11] Altman, Naomi S. *An introduction to kernel and nearest-neighbor nonparametric regression.* The American Statistician 46.3 (1992): 175-185.

[12] Breiman, Leo. *Random forests.* Machine Learning 45.1 (2001): 5-32.

[13] Haykin, Simon. *Neural networks* New York: Prentice hall, 1994.