Aalto University
School of Science

AALTO UNIVERSITY

SEMINAR ON CASE STUDIES IN OPERATIONS RESEARCH

MS-E2177

Data-driven categorization of stocks in asset portfolio management: Project Interim report

Tommi Kantala (Project manager)

Hannu Pantsar

Ilkka Särkiö

Client: Elo

May 16, 2018

1 Project status and accomplished tasks

The project properly started with some initial analysis of the data and finding suitable R packages for time-series modeling. A bit surprisingly, the GARCH packages in R were quite limited and only rugarch and rmgarch packages seem to be feasible options. We spent quite some time trying and testing the univariate and multivariate GARCH models. However, the GARCH models cause a lot of problems. We ran several times out of memory, the model becomes singular and the computation took very long, at least on Linux computers. The computational servers Brute and Force had older version of R which caused even more problems. The times we managed to have an proper result, the prediction was not very good. Eventually, the GARCH modeling was set aside. Looking backwards, we spent too much time with the GARCH modeling and did not simultaneously proceed with other parts as rapidly as we could have. The GARCH modeling will most likely not be used later.

The basic stock clustering framework is now complete. All the code done for is aimed to be clear and easy to use, with plenty of comments. This means that we do not have to spend time later fixing much of it. We still have some features to implement to the model, but this should now be quite doable and easy, since the basic structure is well documented and already some what tested. We still need to run more test to see how stable the clusters are, how they compare to GICS and how different external variables change this. Some preliminary results are reviewed during the Finnair excursion.

The core part of the literature review is now done, with some more text to write. We also naturally keep expanding it as the project continues, since new ideas and problems may arise. Some parts of the literature review must be still reformatted to the final report.

We are behind schedule due to the unforeseen difficulties with the GARCH modeling and this has also caused some changes in the scope of the project compared to the initial plan. Our initial project schedule was very optimistic and we have additional two weeks to finish the project, which were not taken into account when doing the initial schedule. The project schedule for remainder of the project can be seen in Table 1. The physical absence of Hannu will create some problems, but we can solve

these by dividing task with care. More of future risks and realized risk are tabulated in Tables 1 and 2

2   Changes to objectives and the scope of project

We reviewed and tested correlation forecast models but computation was slow and imposed many challenges e.g. inaccuracies, computation time and memory problems. The correlation forecast model proved to be far more difficult to implement than expected. Thus the approach of using multivariate time series forecast modeling for the stock return data has been abandoned for now. This has been done in mutual agreement with the client. Return to the forecasting approach has been deemed possible only in case we succeed on the current tasks at hand relating to clustering the assets.

Since the forecasting approach has been abandoned, the main objective focus is now on clustering of the assets. The clusters are compared to the GICS-classification of the assets and the results are analyzed. As better and better clustering solutions are formed, more variables can be added to the clustering. Between- and within-cluster correlation analysis is performed between data based clustering and GICS. These include macroeconomic variables and key performance indicators of company performances. The additional objective of evaluating the cluster stability over time has been added. This means that the assets are clustered based on data and then compared against subsequent clusters. If the clusters stay the same over time, the clusters are said to be stable.

The objective of wrapping the project code to a usable script that can be used with different data sets remains the same and is going well at the moment. Script usage improvements will be performed until the end of the project to ensure the desirable quality.

All in all the scope of the project has been narrowed down after the group understood the excessive workload that the original objectives would have caused. Even though a major objective was removed, the workload is still challenging enough and the result of the project will be beneficial to the client.

3 Project plan for the remainder of the project

The project plan for the project parts left is in Table 1

	Week start date	9.4	16.4	23.4	30.4	7.5	14.5
Topic	Activity Week	13	14	15	16	17	18
Client interaction	Receive data						
	Steering meeting with client, model evaluation						
Reporting	Interim report						
	Final report						
Modelling	Literature review						
	Improving the clustering model						
	Clustering result evaluation						
	Script usage improvements						

Figure 1: Updated project schedule.

4 Updated risk management plan

We have gathered some realized risks to 1. The risk matrix for rest of the project is in Table 2.

Table 1: Realized risks so far, their impact and proposed actions.

Risk	Realization	Impact	Proposed actions
Team member inactivity	Low	Project difficulties have been found later than optimal	Project priority must be increased over work and other school projects for the remainder of the project.
Too great workload	Medium	Project scope has been narrowed.	Further work must be fitted to a narrower project scope.
Forecasting inaccurate	Realized	Forecast models only reviewed, not implemented	Forecast left out from project model implementation.
Data delivery delays	Low	Low impact as of now. If there are wishes from the client to include more data sets into the project, there might be project delays.	Active communication with client
Computational resource insufficiency	Medium	One of the reasons to drop forecast models from the project scope	

Table 2: Updated risks, impacts and mitigation plans

Risk	Probability	Impact level	Impact	Precautions and mitigation
Team member inactivity	Low	Medium to High	Project delays	Project priority must be increased over work and other school projects for the remainder of the project.
Too great workload	Low	High	If workload increases in the end, project might be delayed	Focus on remaining main tasks and completing ongoing tasks before initiating new approaches.
Clustering analysis doesn't generalize	High	Low	Expected project results might be different.	Agree with client that this might be viable result
Data delivery delays	Low	Low	Some features might be left out from the project.	Active communication with client
Insufficient computational resource	Low	Medium	Slows down project progress or shrinking of project scope/model complexity	Ensure online computational resources early on (e.g. Aalto University servers) and pay attention to program efficiency during project implementation
Team member absence in late May	Certain	Medium	If project is late in May, smaller team size may result in delays and excessive workload	Schedule must be observed to ensure fast enough progress