

Report on Bachelor / Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University

Student:	Ondrej Karlicek
Advisor:	Jan Sila
Title of the thesis:	Application of Machine Learning in Portfolio Construction

OVERALL ASSESSMENT (provided in English, Czech, or Slovak):

Please provide a short summary of the thesis, your assessment of each of the four key categories, and an overall evaluation and suggested questions for the discussion. The minimum length of the report is 300 words.

Short summary

The thesis contributes to the portfolio selection literature. It combines several methods from machine learning literature. The author calculates the characteristics of companies and then clusters them to select a representative of the group. Each of those representatives is a constituent in a calculated portfolio. Since the author optimizes the portfolio with respect to the Sharpe ratio, the thesis also predicts returns split into direction and magnitude parts. The exercise is done on daily data with a particular focus on 2020.

Contribution

The work covers a wide selection of methods from econometrics (GARCH, VaR) to statistics (PCA, RandomForests) and finance (portfolio optimization). It contributes by combining them into a process that should help pre-select appropriate stocks for the portfolio problem, which is optimization in 2020.

The thesis sometimes lacks clarity in motivation as the research question is not clearly delivered. The thesis describes its motivation as „interesting combination of models, whereby clustering we want to obtain different group of stocks“. If one covers the whole universe of stocks, one would still be able to select the same companies.

Clustering would help pick them from somehow differently defined groups, opposite to typical industry-related grouping. But then I would assume one would like to work with the clusters as new „mini universes“, rather than selecting a single stock from it and then combine them across all the clusters. Another concern is with the dataset period. Focusing only on rebalancing in 2020 is limited. The portfolio is compared within a period of high market fluctuations and panic caused by the Covid pandemic in March 2020. Using daily data and monthly rebalancing, one's portfolio would hardly defend against such structural break. The robustness and significance of results are questionable with respect to the number of rebalances done in the sample.

Also, the thesis uses a benchmark as a 1/N portfolio of the selected stocks, whereas a typical stock index might seem more appropriate.

The author eliminates half of the stocks with the highest Value at Risk in the cluster. But would those not be the ones with the possibly highest gains? This is a quite arbitrary condition imposed on the preselection procedure. I would like to see a more detailed discussion on this and the data on the actual VaR of the cluster constituents. Judging from the overall market performance prior to the Covid crisis I would suspect those not to be too extreme for half of the stocks. Perhaps eliminate stocks having the largest drawdowns across clusters at some empirical quantile.

Also, the part of returns predictions is heavily influenced by the preceding periods of market growth. The random forest model most likely cannot learn to predict negative returns. That explains why the in-sample accuracy is quite high, but it completely fails in Period 3, when covid hits and everything „goes red“. The author proceeds to forecast the magnitude of the return with GARCH model. Yet, the combination of models suffers from the market crash and its limited flexibility given the time period selection. Thus the Sharpe ratio based optimization suffers from having poorly estimated future returns. The author comments on this, but perhaps a remedy would be to use Global minimum

Report on Bachelor / Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University

Student:	Ondrej Karlicek
Advisor:	Jan Sila
Title of the thesis:	Application of Machine Learning in Portfolio Construction

variance portfolio which is returns independent. It would at least diversify the volatility and might be a better performer than the 1/N portfolio.

Methods

The range of methods covered is wide, and the thesis is quite technical. The author shows a good understanding of those and offers relevant references. In light of my comment in the previous section, the methodology part is perhaps too wide and, therefore shallow. As the author combines many methods, the text needs to explain and introduce lots of parameters and details to connect it all together. This limits a bit an easy flow of the text.

Literature

The thesis offers relevant literature and references to all the subtopics. The thesis demonstrates author's understanding, yet it would be better if the research question were more linked through the whole work and particularly the literature review.

Manuscript form

The thesis is written in an appropriate language, and outputs quality attains a high academic standard. In terms of the text itself, the readability sometimes suffers from the necessity to cover a wide array of independent parts. So sometimes, the connection between those is not clear.

Overall evaluation and suggested questions for the discussion during the defense

The thesis is quite ambitious in terms of the number of models put together. Yet the motivation behind this particular process is not clearly presented. I believe that at this stage, the author shows a very good deal of skill and ability to produce this pipeline of calculations. This is actually one of the strongest aspects of the work. As a recommendation for the future, I suggest elaborating on the motivation and perhaps form a more narrow and detailed discussion.

I have included a majority of comments in the Methodology section on which I would like author's comments. In particular:

- The benchmark selection – why not include the whole market index? That is what an investor is trying to beat, as it is the „easiest“ investment.
- How did the methodological setup deal with a rapid decline in the market? What were the causes and how could it be remedied? If I were an investor, what should I include in this strategy to be better prepared for such a black swan event?
- When the returns prediction failed, is it still sensible to use a Sharpe ratio based portfolio?
- What is the key takeaway from this work? How would you behave as an investor?

In my view, the thesis fulfills the requirements for a bachelor thesis at IES, Faculty of Social Sciences, Charles University, I recommend it for the defense and suggest a grade D.

The results of the Urkund analysis do not indicate significant text similarity with other available sources.

Report on Bachelor / Master Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University

Student:	Ondrej Karlicek
Advisor:	Jan Sila
Title of the thesis:	Application of Machine Learning in Portfolio Construction

SUMMARY OF POINTS AWARDED (for details, see below):

CATEGORY	POINTS
<i>Contribution</i> (max. 30 points)	15
<i>Methods</i> (max. 30 points)	19
<i>Literature</i> (max. 20 points)	15
<i>Manuscript Form</i> (max. 20 points)	18
TOTAL POINTS (max. 100 points)	67
GRADE (A – B – C – D – E – F)	D

NAME OF THE REFEREE: Jan Šíla

DATE OF EVALUATION: 22.5.2021

Digitally signed (22.5.2021)
Jan Šíla

Referee Signature

EXPLANATION OF CATEGORIES AND SCALE:

CONTRIBUTION: *The author presents original ideas on the topic demonstrating critical thinking and ability to draw conclusions based on the knowledge of relevant theory and empirics. There is a distinct value added of the thesis.*

METHODS: *The tools used are relevant to the research question being investigated, and adequate to the author's level of studies. The thesis topic is comprehensively analyzed.*

LITERATURE REVIEW: *The thesis demonstrates author's full understanding and command of recent literature. The author quotes relevant literature in a proper way.*

MANUSCRIPT FORM: *The thesis is well structured. The student uses appropriate language and style, including academic format for graphs and tables. The text effectively refers to graphs and tables and disposes with a complete bibliography.*

Overall grading:

TOTAL	GRADE
91 – 100	A
81 - 90	B
71 - 80	C
61 – 70	D
51 – 60	E
0 – 50	F