Report on Bachelor Thesis

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

Student:	**Ondřej Karlíček**
Advisor:	**Jan Šíla**
Title of the thesis:	**Application of Machine Learning in Portfolio Construction**

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

Short summary

The thesis under the review attempts to improve portfolio construction using machine learning methods. Author finds that predictions from considered models are not satisfactory and does not improve naive portfolio selection results.

Contribution

Author described lots of details and lists number of advanced techniques and calculations. Unfortunately, it results in bit confusing technical exercise without solid results and contribution. The main finding of the author is that he could not outperform the naive portfolio selection using the considered methods which is fine and could potentially be an important result. The problem is that it needs to be shown robustly and I cannot see the reason why author obtained such results. Specifically, there is number of possible explanations. The issue could be the choice of stocks based on the 95% VaR as risk measure (I will elaborate this in more detail below) as well as mixing the VaR quantile-based optimization with mean-variance based optimization. The problem could also be technical where number of important steps need to be done in the training that are not discussed (as hyperparameter search, dropouts, ensembles, and other commonly used techniques) since we know that machine learning without these techniques easily overfits the data and produces bad out of sample results. While I realize this topic is advanced for a bc. level student, it unfortunate since author's findings may well stem from improper application of the methods.

Methods

Author uses sophisticated methods including decision trees, random forests from machine learning literature and volatility and Value-at-Risk models from the time series literature as well as portfolio optimization. Although it seems impressive, the use of the methods is rather confused (not surprisingly since these are mostly advanced methodologies). While it is rather hard to follow the analysis from the description, my main confusion is from mixing the classical portfolio optimization with Value-at-Risk which is quantile of returns. Even more confusion comes from the fact that author uses 95% quantiles as risk measure.

Literature

The thesis demonstrates author's good understanding of the literature.

Manuscript form

The structure of the thesis is logical, although the text is hard to follow, and it reads as unfinished draft notes at number of places. Otherwise the form of the thesis is on a good level.

Overall evaluation and suggested questions for the discussion during the defense

I have few questions for the defense that author should discuss:

1/ How are the results influenced by the early choice of cluster of 11 stocks? Would it be more practical to re-evaluate the choice of the stocks every month?

Report on Bachelor Thesis

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

Student:	**Ondřej Karlíček**
Advisor:	**Jan Šíla**
Title of the thesis:	**Application of Machine Learning in Portfolio Construction**

2/ How can these results be useful to a practitioner? Can one conclude that considered machine learning techniques are not useful? Is this in contrast to other literature?

3/ Why the risk / VaR 95% is chosen? Especially, what is the motivation of using 95% quantile as risk measure? How does it connect to mean-variance setting? Here the main question is that usually we use volatility as risk measure in mean variance setting, or in case we believe second moment does not describe well the risk preferences, we can use more flexible measure such as VaR. But 95% VaR is connected to returns that fall below some extreme right tail. If returns above this threshold are omitted, then we are not looking at most profitable stocks which does not make sense to me at all. I would expect 5% threshold to avoid left tail risk

4/ Do author see a potential in reducing errors in tuning the models used? Why standard tuning of the out-of-sample results has not been considered (hyperparameter optimization, shrinkage/dropouts, ensembles and model averaging, early stopping etc.).

In conclusion, the thesis attempts to use machine learning methods to improve portfolio construction. While focusing on number of problems including clustering and dimensionality reduction, prediction improvement from machine learning (recution of errors from prediction), author shows a hard work, but at the same time tries to solve demanding advanced problem with too many tools. This distracted focus then does not allow to look deeply enough to make sound empirical contribution either supporting the usefulness of machine learning methods or showing that these methods are useless for such problem. The results from Urkund do not indicate significant text similarity with other available sources. The thesis fulfills the requirements for a bachelor thesis at IES, FSV UK, and I recommend it for the defense and suggest a grade D.

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

CATEGORY		**POINTS**
Contribution	*(max. 30 points)*	20
Methods	*(max. 30 points)*	15
Literature	*(max. 20 points)*	20
Manuscript Form	*(max. 20 points)*	12
TOTAL POINTS	*(max. 100 points)*	**67**
GRADE	**(A – B – C – D – E – F)**	**D**

NAME OF THE REFEREE: Jozef Barunik

DATE OF EVALUATION: 10.5.2021

Jozef Baruník

Digitally signed by Jozef Baruník
Date: 2021.05.10 16:38:56 +02'00'

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**