Opponent's Report on Dissertation Thesis

Institute of Economic Studies, Faculty of Social Sciences, Charles University
Opletalova 26, 110 00 Praha 1, Czech Republic
Phone: +420 222 112 330, Fax: +420 222 112 304

Author:	**Martin Hronec**
Advisor:	Doc. PhDr. Jozef Baruník Ph.D. (IES)
Title of the Thesis:	Asset Pricing and Portfolio Selection via Machine Learning
Type of Defense:	**DEFENSE**
Date of Pre-Defense	October 30, 2024
Opponent**:**	Prof. Jan Hanousek Ph.D. (Mendelova univerzita v Brně)

Address the following questions in your report, please:

a) Can you recognize an original contribution of the author?
b) Is the thesis based on relevant references?
c) Is the thesis defendable at your home institution or another respected institution where you gave lectures?
d) Do the results of the thesis allow their publication in a respected economic journal?
e) Are there any additional major comments on what should be improved?
f) What is your overall assessment of the thesis? (a) I recommend the thesis for defense without substantial changes, (b) the thesis can be defended after revision indicated in my comments, (c) not-defendable in this form.

(Note: The report should be at least 2 pages long.)

The submitted thesis consists of three essays on using machine learning techniques for optimal portfolio allocation. It emphasizes using anomalies and mispricing signals as a specific factor or employing empirical valuation methods of particular assets or portfolios.

Paper 1 (Chapter 2): Does it pay to follow anomalies research?

The first paper aims to show significant profitability gains in long-short portfolios formed using predictions derived from machine learning techniques. Compared to other existing studies, the author examines the usefulness of global data (outside of the US markets) in expanding possible gains in managed portfolios.

I found the paper interesting, yet I have several questions and comments, which are (I understand) irrelevant, as the chapter has been already published:

1) Several other papers have presented the idea of analyzing returns on anomalies outside the US. For example, the authors refer to Jacobs and Müller (2018, 2020) in the chapter. Authors referred to the main difference using a broader set of assets, including small-cap stocks, and a more global view. It would be worth more discussion, maybe extending the model by using weights stemming from links or (dynamic) correlation of the regional markets and the US to shed more light on differences in the results. The discussion and arguments about using global data sources for a higher number of observations (and out-of-the-sample confirmation) should be better structured. Some of the assumptions, such as the normality of the results, seem relatively strong to me. Nevertheless, the law of large numbers should work independently of the normality assumption.
2) In the previous presentation (pre-defense), I had the following question/suggestion to compare the existing results: Authors should try to disentangle the effect of small cap vs. large (liquid) assets and regional vs. global anomalies. In other words, what would be the results if authors kept only liquid assets and used global anomalies, and what would be the effect if they used small-cap and regional anomalies? Maybe I misunderstood the described differences since the authors use the universe of the most liquid stocks later in the chapter.
3) It would be beneficial to use a diagram/scheme of the methods used, their order (and their variations); the same would apply to the training/tuning setup. Let me note that comparing the out-of-the-sample predictions is difficult when the training sample grows in each step. Also, an exciting approach could be to accept the dynamic and

changing features of financial markets (and their anomalies) and to use moving windows for the training, say the last five years.

4) I did not understand the importance of the variables in the presentation (Table 1.x). In most cases, why do the local markets show lower importance than the global market? I expected the international market to use all markets together, right?
5) Technical notes on fundamental data availability. It was unclear in the paper if the authors controlled the availability of the fundamental financials used as additional explanatory variables. I mean that the variables for the firm's financials are available with a delay. The minimum delay is determined by the date of the quarterly announcements (referring to the previous quarter's results)
6) Overall, and this applies to all chapters, machine learning training/tuning, and the issue of overfitting should be discussed a bit more.
7) As concluding remarks, the authors said: "Most of the predictability of the expected stock returns in all the regions can be captured solely with the U.S. training sample as the benefits of a larger estimation sample are offset by the region-specific differences." This statement can be grounds for future research, analyzing which factors offset the benefits and what selection of certain regional anomalies can be helpful from the information and portfolio returns point of view.

Minor suggestions:

a) The tables presented should have a better legend. For example, in Panel D, I can only guess that the second line below each row is the t-statistics. Similarly, it could bring a better outcome by presenting formal tests of differences between methods and approaches, both means and coefficients.

Paper 2 (Chapter 3): Predicting the distributions of stock returns around the globe

This chapter proposed an alternative method for predicting/modeling the distribution of returns. Contrary to the classical approach, the authors used a forecasting method combining a two-stage quantile neural network with cubic B-splines interpolation. Their approach has the advantage of predicting the other distributional characteristics, such as skewness and kurtosis, or even estimating the entire distribution of the returns.

Some of my comments and suggestions are common to the entire thesis, so let me mention only those specifically appearing in this chapter (and not already mentioned above). Again I

understand that it is a waste of time, since the paper is likely published, and the author will not change a word in the chapter. Here are my comments:

1) In the pre-defense report, I raised my concern about the training setup, and the possible overfitting discussion is still valid here. Moreover, the long period for training and estimation requires an assumption of distribution stability over a longer period. Which is not generally accepted. There is, for example, observed and accepted increased investor demand for lottery-like returns, e.g., Kumar, A. (2009). Who gambles in the stock market?. The Journal of Finance, 64(4), 1889-1933. Blau, B. M., Bowles, T. B., & Whitby, R. J. (2016). Gambling preferences, options markets, and volatility. Journal of Financial and Quantitative Analysis, 51(2), 515-540, to name a few articles about changes in investor preferences, suggesting instability of the studied distribution over a long period. (The suggested
2) When authors present their estimates of distributional characteristics, one can ask what their error is. For example, are the numbers for the entire sample and subsamples statistically different?
3) As before, I would love to see the comparisons, maybe using confidence intervals across subsamples and parameters, or a table with a composite summary of rejected equality hypotheses. It also carries our interesting interpretations.

Paper 3 (Chapter 4): Portfolio selection under quantile preferences

The fourth chapter builds on the results and techniques introduced in the previous part(s). In particular, this section focuses on the out-of-sample performance of portfolios chosen by investors, maximizing the τ-quantile of the future returns.

Additional comments to this chapter:

1) The authors find that "A portfolio often yields higher out-of-sample τ-quantile returns compared to approaches to direct quantile maximization." Using a formal test (based on the multinomial distribution associated with the Monte Carlo simulation), authors should be specific if "often" means a statistically significant finding.
2) The technical setup and interesting results would deserve a deeper statistical analysis of the Monte Carlo study's results, like various tests of (linear) trends, increasing/decreasing patterns, differences, etc., since they provide attractive, non-technical, and practical results.

Overall statement:

The thesis is an exciting collection of essays that makes several contributions to the financial literature, and I recommend it for the PhD defense. I highly value a detailed list of market anomalies in Chapter 2 and a broad coverage of markets and assets. My comments should not alter the high quality of the thesis, as they were meant to provide additional feedback for future publication and possible research extension.

Date: April 11, 2024	
Opponent's Signature:	
Opponent's Affiliation:	Prof. Jan Hanousek Ph.D. (Mendelova univerzita v Brně)