FACULTY
OF SOCIAL SCIENCES
Charles University

# Thesis Evaluation Report

| | |
|---|---|
| **Author:** | Lukáš Petrásek |
| **Advisor:** | Jozef Baruník |
| **Title:** | On the Utilization of Machine Learning in Asset Return Prediction on Limited |
| **Opponent** | Jiří Novák |

## Summary

The author performs a simplified version of the empirical framework of Gu et al. (2018). He analyzes whether modern machine learning techniques (e.g. elastic nets, random forests, neural networks) trained on data samples of limited size outperform the simple linear 5-factor model Fama & French (2015). He concludes that (in principle) they do not.

> *We found that while several methods were able to slightly outperform the benchmark model containing the 5 factors from the well-known model introduced by Fama & French (2015), the improvement over the benchmark was not big enough for us to be able to assign it meaningful statistical significance via Diebold-Mariano tests. We then proceeded to perform several robustness checks of our findings. By performing each of these checks, we attempted to undermine our conclusions from the previous chapter. But eventually, we failed to provide satisfactory evidence about a significant contribution of the examined machine learning methods over the traditionally used linear approaches.*

## Evaluation

I am afraid I am highly skeptical about the value of the author's analysis/contribution. It is very difficult to understand how the thesis extends Gu et al. (2018). It is not clear how is the machine learning techniques can contribute to asset pricing. The author does not motivate the benefit of using small data samples. The literature review does not convince me that the author actually acquire in-depth understanding of the individual machine learning techniques. The main result is essentially limited to two tables in section 6. Some of the statement the author makes are contradictory. Therefore even though the seasons is written in very good English and addresses a modern feature finance I find it less good that most of the theses I have reviewed in the past.

# Contribution

In a recent unpublished working paper Gu et al. (2018) evaluate the relevance of machine learning techniques for asset pricing (see below for the abstract). I view the study as essentially a naive and simplified replication of Gu et al. (2018) using smaller data samples. It is not clear to me how the author contributes to what we already know from Gu et al. (2018). I suggest the examination committee asks the student to motivate his contribution relative to Gu et al. (2018) at the oral defense.

> *We synthesize the field of machine learning with the canonical problem of empirical asset pricing: measuring asset risk premia. In the familiar empirical setting of cross section and time series stock return prediction, we perform a comparative analysis of methods in the machine learning repertoire, including generalized linear models, dimension reduction, boosted regression trees, random forests, and neural networks. At the broadest level, we find that machine learning offers an improved description of expected return behavior relative to traditional forecasting methods. We identify the best performing methods (trees and neural networks) and trace their predictive gains to allowance of nonlinear predictor interactions that are missed by other methods. Lastly, we find that all methods agree on the same small set of dominant predictive signals that includes variations on momentum, liquidity, and volatility. Improved risk premium measurement through machine learning can simplify the investigation into economic mechanisms of asset pricing and justifies its growing role in innovative financial technologies.*
> (Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2018. Empirical Asset Pricing via Machine Learning. National Bureau of Economic Research. https://doi.org/10.3386/w25398.)

In contrast to Gu et al. (2018) The author of the thesis concludes that the modern machine learning techniques do not outperform the simple 5-Factor Model estimating using OLS. The thesis lacks a comparing reconciliation of the two findings. I recommend the examination committee to ask the author at the oral defense why the two studies differ in their conclusions.

> *Examples of these are Gu et al. (2018) and Krauss et al. (2017) who have recently shown that the application of deep learning methods like, for example, neural networks, or tree-based methods like gradient boosting or random forest can be highly beneficial for the predictive power of asset pricing models.*

On the fundamental level, I have doubts about how much the machine learning techniques can contribute to asset pricing. My understanding of asset pricing is that it is a field of study that attempts to explain the cross-sectional variation in stock returns and provide compelling conceptual argument(s) why the documented empirical factors are fundamentally related to systematic risk and expected stock returns. The machine learning techniques remain silent about the strengths of the individual factors and about the potential reasons why they are related to stock returns. Hence, while I believe that machine learning techniques can potentially be useful for traders trying to achieve success returns I also believe that they have limited (if any) potential to help us understand the factors driving the cross-sectional variation in stock returns. I did not find any discussion on this issue in the phases. I recommend the examination committee to ask the author at the oral defense.

If I understand correctly the author argues that his methodology differs because he uses smaller data sets for training the algorithms than in prior studies.

> *Most of the successful applications of machine learning in finance have one thing in common. They utilize enormous datasets compared to what a usual financial analyst or junior researcher*

*is able to collect. But what happens when the size of the data is small? Because the modern machine learning techniques are rarely applied to datasets of limited size and dimension, we believe that the employment of these highly complex methods on small data should receive some attention. In this thesis, we attempt to investigate the consequences of applying machine learning techniques which have been heavily promoted in the finance literature lately to a dataset of limited size.*

I do not recall another study that claims a small date sample an advantage or actually the main feature of the author's contribution. As the name itself suggests the machine learning algorithms are performed by machines that can hardly be intimidated buy large datasets. Especially when it comes to the easily observable financial measures such as stock returns it is remarkable that the author only collects data for the last 20 years. I do not understand why there are missing observations for variables like stock returns. It is well known that some empirical pricing factors such as firm size are much weaker in the last 20 years. In addition, long estimation windows (5 years) are needed to obtain testable measures for the individual factors. Therefore it is difficult for me not to see the short data sample a weakness rather than an advantage of the study.

*While our dataset cannot be considered large with respect to other ones used in the literature, we see this as an advantage. The lower relative size of our dataset allows us to expose the advanced methods to unfavorable conditions in which we can test their performance.*

*Also, no missing values in the data are imputed, all of them are skipped instead. We restrain from imputing missing values because the excess return data is well known to contain a lot of noise and hence we consider imputing missing values not to be appropriate.*

*Nevertheless, the author experimented with training the models on very small samples, starting at 1 year of training data, but the results proved to be very poor and training sample sizes of less than 60 months were not included in the analysis.*

# Literature

Even though the literature list is rather long I do not like the way the author reviews prior literature. Most of the literature review is a textbook-like review of asset pricing literature starting from the capital asset pricing model and leading up to the more recent empirical models. Such a review can be easily copied from a finance textbook or a recent paper in asset pricing however its relevance for the Thesis is limited. If I understand correctly what the author aims to achieve with the Thesis the main focus of the literature review should be the machine learning techniques and their applications in finance and in particular in predicting stock returns. In that section the author frequently refers two prior study without providing at least a rudimentary discussion how the individual methods work and why we should expect differences in their predictive abilities in the context of stock markets. This raises doubts about the level of the author's understanding of these techniques.

*We suggest the reader to address Section 3.4 of Chapter 3 or Gu et al. (2018) for more information about the techniques used.*

Related to the point above, I am somewhat amused by the following statement. If the author considers the discussion of machine learning techniques redundant because it has already been done in prior literature, I am surprised that he does not apply the same argument to the asset pricing literature that has been discussed at great length before.

> *In what follows, we provide a short overview of machine learning methods which comprise the building blocks of our analysis. We will be concise by purpose, because most of the details and technicalities have been described many times in various scientific materials.*

# Methodology

Related to my evaluation of the author's contribution above the thesis essentially replicates a recent unpublished working paper Gu et al. (2018).

It is not clear to me from the methodology section whether dividends are included in the computation of monthly stock returns. If not then this would be a significant mistake that can introduce noise to the dependent variable and possibly be the underlying reason why the author documents weak results the machine learning techniques (in contrast to Gu et al. (2018)).

From the methodology section it is not clear to me whether the author make sure that the accounting data is legend by 6 months to make sure that the data is available to the investors at the time of the investment decision. If not this would cause a hindsight bias.

I am not convinced about the benefit of using shorter datasets (see above).

> *The excess return data span from December 1997 to November 2017, covering exactly 20 years, i.e. 240 months, of time.*

> *While our dataset cannot be considered large with respect to other ones used in the literature, we see this as an advantage. The lower relative size of our dataset allows us to expose the advanced methods to unfavorable conditions in which we can test their performance.*

> *The remaining RoSSS specifications, RoSSS(10, 3, 1) and RoSSS(10, 5, 1), show improved performance over RoSSS(10, 1, 1) or those RoSSSs with small training windows. We consider this as a sign of how important it is to feed the models with as much data as possible, and that the regularization is generally more successful when the validation window size increases.*

I do not understand why the authors uses a combination of two highly regulated (NYSE, NASDAQ) and one unregulated US market (BATS) that is plausibly much less liquid and efficient.

> *The dataset contains data on 3219 stocks traded on the largest exchanges in the United States of Amer-ica, these are National Association of Securities Dealers Automated Quota-tions (NASDAQ), New York Stock Exchange (NYSE) and Better Alternative Trading System (BATS).*

# Form

In general the Thesis follows the standard structure.

I believe the introduction should include a more compelling argument motivating the contribution of the study. I also believe that the introduction should contain a preview of the results.

The results section is very short. The main results can essentially be summarized in two main tables in section 6.

The thesis is written in very good English. However, the author sometimes uses informal language that is inappropriate for a thesis.

> *This reminds us to remind the reader of another important aspect of our estimation procedures, the act of standardizing training and validation data each time we roll the samples forward.*

# Other

Some of the statement the author makes are incorrect. For example, the main purpose of the pricing factors is to explain the cross-sectional variation in stock returns. Given the stability of these factors over time one can say that the models aim at predicting expected stock returns.

> *While the five Fama and French factors included were not originally used to make forecasts of asset returns (Fama & French 2015), we feel that it is appropriate to add them among our regressors because of the proclaimed contribution.*

In a similar vein, the first sentence below suggests that the results in this Thesis may not be quite reliable because of the limited data set. In contrast, the second sentence is based on the premise that the results are reliable and using his results as the conclusion the author challenges the popularity the neural networks. In my opinion these two statements cannot be valid at the same time.

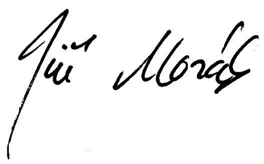
> *This does not, however, mean that NNs are useless for asset price prediction, it can merely suggest that perhaps we did not provide them with enough data, or that the selection of factors we used was insufficient. Nevertheless, the results might be surprising due to the prominence the NNs achieved in the literature and their widely accepted qualities.*

# Conclusion

The author examines a modern phenomenon but I find the way of organizing the analysis rather unfortunate. It is hard for me to see what we can learn from the thesis. I recommend the evaluation committee to ask the author about the takeaways from the study and about its contribution relative to prior literature.

# Awarded Points and Grade

| | |
|---|---|
| Contribution (max 30) | 20 |
| Methods (max 30) | 22 |
| Literature (max 20) | 14 |
| Form (max 20) | 18 |
| **Total** (max 100) | **74** |
| **Grade** (A – B – C – D – E – F) | **C** |



Digitally signed by Jiri Novak
Date: 2019.09.06 12:36:35 +03'00'

**Referee's Signature**

6 September, 2019

**Evaluation Date**

Jiří Novák

**Referee's Name**

# Grading Scale

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

| Strong | Average | Weak |
|---|---|---|
| 20 | 10 | 0 |

**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.

| Strong | Average | Weak |
|---|---|---|
| 30 | 15 | 0 |

**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.

| Strong | Average | Weak |
|---|---|---|
| 30 | 15 | 0 |

**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.

| Strong | Average | Weak |
|---|---|---|
| 20 | 10 | 0 |

**OVERALL GRADING:**

| **Total Points** | **Grade** |
|---|---|
| 91 – 100 | A |
| 81 – 90 | B |
| 71 – 80 | C |
| 61 – 70 | D |
| 51 – 60 | E |
| 0 – 50 | F |