

Master thesis submitted in partial fulfillment of the requirements for the degree  
Master of Science  
at Technische Universität München

## Cross-sectional predictability of stock returns in Nordic stock markets using machine learning methods

Reviewer	Prof. Dr. Christoph Kaserer Department of Financial Management and Capital Markets TUM School of Management Technische Universität München
Advisor:	Noorhan Elkhayat
Study program:	TUM-BWL
Composed by:	Jesse Keränen Motorstraße 64 80809 Munich Tel.: +49 (0) 1628410926 Matriculation number: 03748837
Submitted on:	March 6, 2024

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Literature</b>	<b>3</b>
2.1	US stock market anomalies . . . . .	3
2.2	European stock market anomalies . . . . .	4
2.3	Nordic stock market anomalies . . . . .	5
<b>3</b>	<b>Data</b>	<b>7</b>
<b>4</b>	<b>Methodology</b>	<b>12</b>
4.1	Benchmark factors . . . . .	14
4.2	Linear regression . . . . .	16
4.3	Random forest . . . . .	17
4.4	Neural networks . . . . .	18
4.5	Variable importance . . . . .	20
4.6	Sample splitting . . . . .	21
<b>5</b>	<b>Empirical results on Nordic equities</b>	<b>22</b>
5.1	Benchmark factor performance . . . . .	22
5.2	Predictive performance . . . . .	23
5.3	Portfolio performance . . . . .	27
5.4	Variable importance . . . . .	27
<b>6</b>	<b>Conclusion</b>	<b>27</b>
<b>A</b>	<b>Data</b>	<b>35</b>
<b>B</b>	<b>Benchmark factors</b>	<b>39</b>
<b>C</b>	<b>Additional Information</b>	<b>41</b>
	<b>References</b>	<b>43</b>

## List of Figures

1	Illustrative regression tree . . . . .	18
2	Illustrative neural network . . . . .	19
3	Sample splitting scheme . . . . .	21
4	Benchmark factor performance . . . . .	24
5	Cumulative return of equal weighted machine learning portfolios . . . . .	29
6	Cumulative return of equal weighted zero investment portfolios . . . . .	30
7	Cumulative return of value weighted machine learning portfolios . . . . .	31
8	Cumulative return of value weighted zero investment portfolios . . . . .	32
9	Variable importance . . . . .	33
10	Relative variable importance . . . . .	34
A.1	Number of companies . . . . .	37
A.2	... . . . .	38
B.1	Factor autocorrelation . . . . .	40
C.1	Time series of out-of-sample $R^2$ . . . . .	41
C.2	Turnover of highest expected return portfolios . . . . .	42
C.3	Optimized random forest hyperparameters . . . . .	42

## List of Tables

1	Country summary statistics . . . . .	8
2	Variable definitions . . . . .	9
3	Descriptive statistics . . . . .	11
4	Hyperparameters . . . . .	22
5	Benchmark factor summary statistics . . . . .	23
6	Prediction accuracy . . . . .	25
7	Expected return regression summaries . . . . .	27
8	Machine learning portfolio performance . . . . .	28
9	Zero investment portfolio performance metrics . . . . .	29
A.1	Constituent lists and keywords . . . . .	35
A.2	Static screens . . . . .	35
A.3	Common keywords . . . . .	36
B.1	Benchmark factor correlation matrix . . . . .	39

# 1 Introduction

In 1808 world was in many ways different compared to what it is today. In 1808 Napoleon was the Emperor of the French Empire and Maximilian I was the King of Kingdom of Bavaria. In 1808 Finnish war broke between Kingdom of Sweden and the Russian Empire which would ultimately lead to establishment of autonomous Grand Duchy of Finland. It would still take more than 100 years for Finland to gain its independence. Same year began the Dano-Swedish war between Sweden and Denmark-Norway. Something historically far less remarkable, but essential for this study happened in 1808 as well. First stock exchange in a Nordic country was opened in Copenhagen. Slowly after that rest of the Nordic countries would open their own stock exchanges as well. Upon facilitated change of ownership of assets, investors were left with a question how to price these assets.

Major breakthrough in this topic happened when capital asset pricing model was developed in the sixties (Sharpe, Treynor etc...). In the eighties performance of capital asset pricing model was questioned and scholars came up with so called stock market factors (Earnings-to-price = Basu(1977), Earnings-to-price = Basu(1977), Profitability = e.g., Basu, 1983). During these time machine learning gained large interest and artificial neural networks were popularized. Next big step in asset pricing happened when Eugene Fama and Kenneth French combined these factors to three factor asset pricing model (Fama French 1993). Two years after that machine learning method called random forest was introduced (Ho in 1995). Although three factor model was remarkable improvement compared to capital asset pricing model it was not able to explain variation in stock returns completely. In recent years lot of research has applied machine learning methods to capture abnormal returns in stock markets.

Objective of this study is to apply set of machine learning methods to well established asset pricing factors to capture abnormal stock return patterns. This study will focus on four Nordic stock markets namely Denmark, Finland, Norway and Sweden. These four markets are relatively homogenous in many aspects. They are geographically close, politically stable and economically interconnected. Denmark, Finland and Sweden all belong to European union. Additionally, stock exchanges of all these three countries are operated by Nasdaq. Therefore, investors could view them as a single market. Some of the features that are characteristic for Nordic markets make them fertile ground for stock market anomaly studies. As mentioned Nordic countries are geographically closely located in northern Europe and therefore relatively distant from large European and especially American markets. European market integration is emphasised also by Fama and French (2012).

So called periphery effect has been studied by the scholars a lot. It refers to investor behaviour where during times of a crisis investors tend to liquidate their investments first from the markets more distant to them. This increases the volatility of periphery markets and can challenge the efficient market theory. Another common feature that Nordic markets share is a high level of

foreign ownership. Share of foreign investments in Nordic stock markets can reach more than 50%<sup>1</sup>. Given the remote location of Nordic stock markets and their high share of foreign ownership it is likely that Nordic countries could be subject to periphery effect. Which again can result abnormal returns.

This study contributes to the existing literature in several ways. First of all it applies machine learning framework from Gu, Kelly and Xiu (2020) to new market. Machine learning approach has been applied to European markets, but this study focuses on arguably even more coherent nordic submarket (Drobetz & Otto, 2021; Fieberg, Metko, Poddig, & Loy, 2023). Objective is to examine if investor investing only in Nordic markets could benefit from implementing the machine learning framework of Gu et al. (2020). Dataset of study is unique in a sense that lot of machine learning approach studied have been conducted in more wider markets such as European stock markets, where as Nordic stock market anomaly studies mainly focuses on single markets. Pooling the four Nordic markets ensures us sufficient amount of data to train complex machine learning model, but also allows us to focus on homogenous clearly defined submarket.

Additionally, existing stock market anomaly literature focusing solely to Nordic markets is rather limited. Section 2.3 introduces the existing stock market anomaly literature. Characteristic for studies in this section is that they mainly focus on one to two stock market anomalies. As this constructs 23 stock characteristic anomalies, it allows us examine anomalies that have not been applied to Nordic markets previously. Meaning that this study can not only reveal the profitability of machine learning framework in Nordic market setting, but it can also reveal existence of certain stock market anomalies in Nordic markets. As mentioned lot of Nordic stock market anomaly research focuses only up to two anomalies at once. Since this study includes 23 anomalies simultaneously, it allows us also to examine performance of already discovered Nordic anomalies while controlling for many other variables. Applying sophisticated machine learning models allows us to also to control for complex interactions.

Objective of this study is slightly more ambitious than in existing Nordic stock market anomaly literature. Existing literature mainly examines existence of anomalies by uni- or multivariate portfolio sorts. Studies predefine variable of interest and form portfolios based on that variable. Then the historical excess returns are investigated. This study goes one step further and attempts to predict stock level out-of-sample returns based on the predefined set of variables. This allows us to evaluate which portion of the return variability these variables are able to capture in addition to the profitability of the strategy.

Final contribution of this study is to expand the explanatory variable set. This study includes variable called on-balance volume. On-balance volume is a technical trading indicator which has not been studied in great extend as cross-sectional stock return predictor. Due to strong performance of momentum indicators, this study includes several momentum indicators. Extended variable set allows us to examine whether on-balance anomaly exists in Nordic stock markets or whether including on-balance volume affects performance of well established

momentum indicators.

Structure of this paper goes as follows. In second chapter introduction to related literature is provided. In this chapter performance of different methods and persistence of different anomalies in different regions is discussed. Third chapter introduces the methodology and data. Fourth chapter presents the empirical study and final chapter provides as conclusion.

## 2 Literature

Being the largest and most prominent stock market in the world US stock market has been subject to majority of asset pricing studies. Despite the dominance of US markets in capturing the the attraction of the academics lot of anomaly asset pricing literature has been conducted in international setting as well. Characteristic for international asset pricing literature is that instead of focusing on single countries they aggregate stock market data to a certain regional level such as Europe or Asia-Pacific. Following chapter provides an overview for pioneering asset pricing anomaly literature. Focus will mainly be on literature on US and European markets. US stock markets are chosen because of their significance on international stock markets and because most anomalies have been discovered there and therefore majority of the initial studies of these anomalies have been conducted there. European studies provide interesting perspective since in many of them Nordic countries are included.

Chapter introduces the most important anomalies in these markets and how they have been exploited with different methods. This works as starting point to define set of factors that will be used in this study. It can be argued that this kind of process when the set of variables are chosen based on their performance in previous studies is one sort of forward looking information if we try to mimic investors information set. On the other hand Jacobs and Müller (2020) only find reliable post-publication decline in long/short returns in US which emphasizes the practical potential of this study.

### 2.1 US stock market anomalies

Many of the recent cross-sectional stock return studies use framework of Lewellen (2015) as base model. He runs 10-year rolling Fama-MacBeth regressions using lagged firm characteristics to predict out of the sample stock returns. He studies cross-sections of US stock return between 1964 and 2013 using different model settings up to 15 company characteristics. He finds strong positive correlation between expected return derived from rolling Fama-MacBeth regressions and realised returns. Additionally Lewellen shows that spread between realised return of portfolio formed from stock with lowest expected returns and portfolio with highest expected return is up to 2.36%. In his study logarithmic market value of equity, logarithmic book-to-market value,

---

<sup>1</sup>Butt and Hogholm (2020) calculate share of foreign ownership from IMF Coordinated Direct Investment Survey CDIS data. Foreign ownership share of Butt and Hogholm is 52% for Denmark, 42% for Finland, 35% for Norway and 56% for Sweden.

momentum and accruals show the strongest statistical power in explaining monthly returns using lagged variables.

Gu, Jelly and Xiu (2020) contribute to the literature by applying machine learning methods to exploit stock market anomalies. By deploying sophisticated models that do not suffer from over parameterization as heavily as OLS Gu et al. are able to include 94 stock characteristics and their interactions as well as eight aggregated time series variables to their models. Gu et al. use large variety of statistical methods including linear regression, generalized linear models with penalization, dimension reduction via principal components regression and partial least squares, gradient boosted regression trees, random forest and different settings of neural networks. Out of these gradient boosted regression trees and neural networks <sup>2</sup>explain the monthly out of sample stock return the best reaching out of sample  $R^2$  0.33 and 0.44 correspondingly where as three factor OLS model introduced by Lewellen (2015) only reaches out of sample  $R^2$  of -3.46.

Similar to Lewellen (2015) Gu et al. construct portfolios based on predicted return of different models. Monthly spread in realized return between portfolio constructed from decile of companies with lowest expected return and decile of stocks with highest expected return <sup>3</sup>is 0.94, 1.62 and 2.12 for models based on OLS, random forest and three layer neural network correspondingly. Gu et al. also show that all methods they examine show somewhat similar patterns on variable importance on return predictability. Most important factors are price trends such as momentum followed by stock liquidity, stock volatility, and valuation ratios.

## 2.2 European stock market anomalies

As mentioned US market environment is different in many ways compared to Nordic markets. Fortunately lot of stock market studies have been conducted in Europe and since Nordic markets are usually just a subset of European markets it can be beneficial to have a look on European studies. Tobek and Hronec (2021) study machine learning based anomaly strategies in international setting. Their study includes 153 different equity anomalies and they only include anomalies to their data after documented discovery of corresponding anomaly. This way they can mimic the information set investor would have had and avoid forward looking information. Tobek and Hronec examine five different models including weighted least squares, penalized weighted least squares, gradient boosting regression trees, random forest and neural networks. Their data set spans from 1990 to 2018. Similar to Gu et al. (2020) in US, Tobek and Hronec find that strategy using neural networks provides highest returns on quintile long-short portfolios. Mean return for neural network long-short portfolio in Europe was 0.7. Interestingly penalized weighted least square method provided mean return of 0.651 which is higher than return of random forest based portfolio's return of 0.396. Tobek and Hronec find that Industry momentum, lagged momen-

---

<sup>2</sup>Gu et al. (2020) use five different settings of neural networks differing by number of hidden layers. Neural network with three hidden layers reaches the highest  $R^2_{oos}$  and is reported here.

<sup>3</sup>Portfolio returns are average value weighted returns.

tum, liquidity shocks, 52 week high, book-to-market value and return on equity are the most important variables for neural networks model.<sup>4</sup>

Exploiting stock market anomalies using machine learning methods is also studied by Drobetz and Otto (2021). Their data set contains all companies listed in at least one of the 19 Eurozone countries on December 2020 and spans from 1990 to 2020<sup>5</sup>. Drobetz and Otto examine performance of ordinary least squares, penalized least squares, principal components regressions, partial least squares, random forests, gradient boosted regression trees and neural networks on predicting monthly stock level return exploiting a set of 22 predictors, their two-way interactions and second- and third-order polynomials. Findings of Drobetz and Otto are similar to Gu et al. (2020). They show that with large number of explanatory variables simple linear regression is not able to explain out of the sample stock returns.

Findings of Drobetz and Otto (2021) are also similar to Tobek and Hronec (2021) in a sense that least squares methods where dimensionality is restricted can actually perform better than tree based methods. Like in majority of other literature, Drobetz and Otto find out that neural networks provide superior framework for stock return prediction model measured in both explanatory power and profitability. Neural network method reaches out of the sample  $R^2$  value of 1.23 and long-short portfolio formed based on expected returns derived from neural networks model provide average value weighted monthly return of 1.94%. Similar to Gu et al. (2020) Drobetz and Otto find that same variables show the most importance across the different models, most notably earnings-to-price ratio and 12 month momentum.

Fieberg et al (2023) study stock market anomalies in 16 European stock markets using machine learning methods over almost the same period as Drobetz and Otto (2021)<sup>6</sup>. Nevertheless they choose a slightly different approach where instead of including vast set of anomalies, they only consider six prominent equity factors. Their conclusion endorses findings of Drobetz and Otto (2021) and Tobek and Hronec (2021) as they shown that more complex machine learning models beat linear approach in terms of both economic and statistical performance.

## 2.3 Nordic stock market anomalies

This chapter provides an overview of observed stock market anomalies in different Nordic stock markets. Many studies in this chapter have slightly different objective than this study. Studies show the existence of the anomalies by constructing a portfolio heavily weighted on certain factor. Nevertheless, they do not describe the magnitude of the relationship between the factor and the stock returns. This study has slightly more ambitious objective and tries to derive return

---

<sup>4</sup>Tobek and Hronec (2021) discover possibilities training models either only using historical data from US, using historical data from local markets or using international historical data. Only results for models trained using local data are reported here because that is closest to the setting of this study. Additionally, Tobek and Hronec state that difference between model trained on US data and local data are small.

<sup>5</sup>Finland is the only country included in the study of Drobetz and Otto (2021) that is also included in this study, since it is the only country belonging to Eurozone.

<sup>6</sup>Dataset of Fieberg et al (2023) contains Denmark, Finland, Norway and Sweden.



expectations from predefined stock market factors. This literature review serves as starting point for choosing most promising stock market factors that have already been studied.

Magnitude of value and momentum anomalies in Nordic stock markets are examined in the paper by Grobys and Huhta-Halkola (2019). They combine information from companies listed in main lists of Danish, Finnish, Norwegian and Swedish stock exchanges between 1991 and 2017. Grobys and Huhta-Halkola measure value with book-to-market value and momentum with past 12-month total shareholder equity. Grobys and Huhta-Halkola show that momentum effect exists in Nordics markets and profitability of momentum strategy is not related to size factor. Value factor yields also significant excess return, but according to Grobys and Huhta-Halkola it could be partly driven by the size factor, since value premium reduces when accounted for the size. Among all stocks monthly average equally weighted long-short return is 1.73% and 1.25% for momentum and value strategies correspondingly. Both of the excess returns are statistically highly significant. Grobys and Huhta-Halkola also test combination strategies using signals from both momentum and strategy which yield even stronger results.

Value premium has shown consistency in Finnish stock markets. Davydov, Tikkanen and Äijö (2017) examine profitability of different value investing strategies between 1991 and 2013. Davydov et. al. investigate set of value indicators which included earnings to price, book to price, cash flow to price, dividends to price and earnings before income and taxes to enterprise value ratios. Additionally they test performance of investing strategy developed by Greenblatt (cite here) where portfolios are formed based on combined ranking of company's return on invested capital and earnings before income and taxes to enterprise value ratio. They show that returns of all of the value portfolios not only beat the market return, but can also not be explained by the four factor model of Carhart (?).

Similar to Grobys and Huhta-Halkola (2019) Leivo and Pätäri (?) combine value anomaly with momentum anomaly in Finnish stock market for data set between 1993 and 2008. They show that two step portfolio sort that first allocates stocks to three portfolios based on their value indicators and subsequently based on momentum indicator can capture extraordinary stock returns. Leivo and Pätäri show that including momentum further increases returns of already recognised value sorting. Strategy performs even better when authors allow for long position in high value high momentum portfolio and short position on low value low momentum portfolio. Excess returns resulting from the two-fold portfolio construction can not be explained by CAP-model or two factor model including also size factor. It is not a surprise that value and momentum premium show existence in Nordic markets. Value and momentum anomalies are among the most well documented factors showing persistence in multiple cross-sectional studies (e.g, Gu, Jelly and Xiu (2020), Lewellen (2015), Drobetz and Otto (2021), Tobek and Hronec (2021)).

Nordic stock markets have several characteristic features. One is that all Nordic stock markets are considered to be developed, but also small. Especially market capitalization of

companies listed in Nordic stock exchanges are on average much smaller than in US. Therefore, it is reasonable to ask whether liquidity of the stock could be driving factor of the stock returns. Impact of illiquidity risk to stock returns in Nordic market setting has been studied by Butt and Hogholm (2020). Butt and Hogholm test variety of different illiquidity measures and find that dollar zero returns is the most profitable illiquidity anomaly measure across all four Nordic market. Dollar zero return measurement is calculated by dividing number of days stocks return in US dollars is zero by total number of trading days. Butt and Hogholm construct five quintile portfolios based on liquidity of the stocks with data spanning from 1988 to 2013. They show that in all Nordic markets there exists large illiquidity premium as annual difference in equal weighted return of most illiquid portfolio and least illiquid portfolio is more than 18% for Finland, Norway and Sweden. For Denmark premium is slightly smaller 8.8%.

Jokipii and Vähämaa (2006) investigate free cash flow anomaly in Finnish stock markets between 1992 and 2002. They construct portfolios from stocks listed in Finnish stock exchange based on predefined thresholds for free cash flow ratios. These ratios include market value to free cashflow and total debt to free cashflow ratios. High free cashflow portfolio yields higher returns than market on average and the excess returns can not be completely explained by weightings in Fama and French (1993) risk factors.

### 3 Data

Main datasource for this study is Thomson Datastream. Company fundamentals data is collected from Worldscope database. Dataset spans from 1990 January to 2022 December which is shorter than in many previous studies. Reason why time period is limited to 1990 is that the amount of publicly listed companies in Nordic markets was rather low in the 1980's and finding reliable data for period before 1990 is difficult. Dataset contains all stocks listed in primary markets of corresponding countries including also companies that went bankrupt or were de-listed for any other reason. Therefore, dataset is not subject to survivorship bias. Panel A of table A.1 in Appendix shows the constituent lists used in data collection. As highlighted by Ince and Porter (2006) data in Datastream can be noisy and uncleaned data could lead to a false statistical inference. Therefore, several static and dynamic screens are applied to the data. Static screens include filtering non-equity securities, securities that are not listed in respective country and securities that are quoted in currency other than respective country's currency. Abnormal returns are also cleaned from the dataset. All applied static and dynamic screens are comprehensively explained in appendix.

On average number of companies with large market capitalization is more limited in nordic countries than in US or Europe. The smallest companies can be numerous, but still only account for fraction of total market capitalization. Liquidity of these companies is often also quite low. To avoid results to be driven by such a stock, approach of Hanauer and Kalsbach (2023) applied

**Table 1: Country summary statistics**

Table provides summary statistics for pooled Nordic market and separate country specific Nordic markets. Minimum number of companies tells the amount of companies included to the data set in a month that the value was lowest for respective country. Maximum number of companies tells the amount of companies included to the data set in a month that the value was highest for respective country. Mean number of companies is the time series average of monthly number of companies for each country. Total number of companies is the number of unique companies in whole data set. Time series averages for monthly mean, median and total market values are also presented. Total market value is the sum of market values of respective country in each month. All marked values are converted to USD. Only companies in the final dataset are included in calculation of the figures. Micro stocks are excluded from the dataset.

Market	Number of companies				Market value		
	Min	Max	Mean	Total	Mean	Median	Total
Denmark	42	106	70	235	2399.78	810.32	141657.9
Finland	26	83	62	186	1893.43	634.47	124389.2
Norway	44	132	79	408	1520.75	506.17	124689.5
Sweden	45	256	132	593	2115.65	616.59	308952.1
Nordic	200	527	343	1422	1946.46	583.02	699688.8

for emerging markets and exclude companies with smallest market value that account 3% of the aggregated market value. On the other hand we do not want few extremely large companies to drive the results either. Therefore, market value of the companies is winsorized monthly to 99%. If company's market value is among 1% biggest market values in corresponding month, the market value will be replaced by the 1% threshold.

Table 1 presents the number of companies and their sizes in separate and pooled nordic markets after above described filters. The total number of non-micro cap stocks in the dataset is 1422 where as the monthly number of stocks in the dataset is 343 on average. Figure A.1 in Appendix shows the development of the company count that passed the statistical screens over time. Sweden is clearly the biggest of the four markets both in regards off the number of companies and total market value of the companies. Even though Sweden is the biggest market it is not dominating. On average Sweden accounts less than half of the total market value of the pooled Nordic market. In regards of average and median market value, Denmark has the biggest companies. Finland on the other hand is clearly the smallest of the markets included in the study measured both in number of companies and market value of the companies.

Total 23 characteristics are derived from stock level data. All the models use same set of explanatory variables which includes book-to-market ratio (BEME), investment (INV), earnings-to-price ratio (EP), cash-to-total assets ratio (CA), capital turnover (CTO), cashflow-to-price ratio (CFP), leverage (DEBT), sales-to-price ratio (SP), return on assets (ROA), return on equity (ROE), one month momentum (MOM<sub>2</sub>), momentum from  $t-12$  to  $t-7$  month (MOM<sub>7</sub>), momentum from  $t-12$  to  $t-2$  (MOM<sub>12</sub>), momentum from  $t-36$  to  $t-12$  (MOM<sub>36</sub>), industry momentum (MOM.IND), log scaled market capitalization (LOG.MV), standard deviation (SD), ratio of current price and 52 week high price (HIGH52), beta coefficient (BETA), idiosyncratic volatility (IDVOL), turnover (TO) and on balance volume (OBV).

**Table 2: Variable definitions**

Tables provides definitions and initial authors for all anomalies used in the study. Construction of variables follows mainly Green et. al (2017) and Hanauer and Kalsbach (2023) and can deviate from variable definitions of initial authors.

Variable	Author	Definition
Cash-to-Assets	(2012)	Cash-to-Asset ratio is calculated by dividing cash and short-term investments by total assets.
Capital Turnover	i	Capital turnover is calculated by dividing total sales by one year lagged total assets.
Investment	i	Investments are defined as a yearly change in total assets.
Book-to-Market Equity	Rosenberg (1985)	Book-to-Market value is calculated by dividing company's book value of equity by company's market capitalization end of previous year. Book value of equity is calculated by summing common equity and deferred taxes of the company.
Cash Flow-to-Price	(1994)	Cash flow to price ratio is calculated by dividing company's cash flow from operating activities by the asset's market capitalization end of previous year.
Debt-to-Price	i	Debt-to-price value is calculated as difference between total assets and common equity divided by the asset's market capitalization end of previous year.
Sales-to-Price	i	Sales-to-price ratio is calculated by dividing total sales by asset's market capitalization end of previous year.
Earnings-to-Price	i	Earnings-to-price ratio is calculated by dividing net income before extra items and preferred dividends by asset's market capitalization end of previous year.
Return-on-Assets	i	Return-on-assets is calculated as net income before extra items and preferred dividends divided by one year lagged total assets.
Return-on-Equity	i	Return-on-equity is calculated as net income before extra items and preferred dividends divided by one year lagged book value of equity. See book-to-market equity for definition of book value of equity.
Tobin's Q	i	tbd
Momentum <sub>7</sub>	i	Cumulative return in US dollars between $t - 7$ and $t - 12$ months.
Momentum <sub>12</sub>	i	Cumulative return in US dollars between $t - 2$ and $t - 12$ months.
Momentum <sub>36</sub>	i	Cumulative return in US dollars between $t - 12$ and $t - 36$ months.
Momentum <sub>2</sub>	i	Prior month return in US dollars.
Industry Momentum	Moskowitz and Grinblatt (1999)	12 month cumulative equal weighted industry momentum. Industry is defined using INDG attribute from Datastream.
Standard deviation	Ang, Hodrick, Xing and Zhang (2006)	Standard deviation of unadjusted weekly price for last 52 weeks.
52-week high price	George and Hwang (2004)	Calculated from weekly unadjusted prices by dividing current price by past 52-week high price.
Beta	Fama and MacBeth (1973)	Beta coefficients are estimated by regressing unadjusted weekly returns noted in US dollars with equally weighted market returns. Minimum 15 observations is required.

Table continues in the next page ...

... Continued from the previous page

Variable	Author	Definition
Idiosyncratic volatility	Ali, Hwang and Trombley (2003)	Idiosyncratic volatility is estimated by standard deviation of regression residuals from regressing unadjusted weekly US dollar returns by equally weighted market return.
Log. market value	Banz (1981)	Natural logarithm of the market value of the company end of previous month.
Turnover	Datar, Naik and Radcliffe (?)	Turnover is defines as end of previous month weekly trading volume divided by the shares outstanding.
On-balance volume	Tsang and Chong (?)	First on-balance volume is the weekly trading volume multiplied by corresponding week return's sign. Following on-balance volumes are calculated by adding the product of trading volume and sign of return to previous on-balance volume.

In this study Nordic markets are examined as one market. In the introduction it was argued that in the eyes of foreign investors Nordic markets can appear quite homogenous. There is also more practical reason. Table 1 shows that individual Nordic stock markets hold limited amount of large market capitalization stocks. This leads to situation where the performance of the whole market or portfolios formed from the market can be driven by very few large market capitalization stocks even after winsorizing the market values. Later in the study we will allocate stocks to portfolios based on their expected returns and we want to ensure that there exist reasonable amount of companies to diversify each portfolio. Unfortunately, Nordic countries have different currencies. In order to ensure comparability of companies from different countries, we have to convert certain variables to common currency which in this case is US dollar using historical currency rates. Variables that are converted to US dollars include for example return index and market capitalization. Majority of the variables are ratios which can be calculated from local currency values.

Data consists of variables that are available on three different frequencies. Majority of these variables are ratios calculated from accounting data. Usually income statement or balance sheet information are available annually and therefore majority of accounting based variables are updated only once a year. To account for possible reporting delay associated with accounting data, accounting data from year  $t$  is considered to be available end of June  $t + 1$ . Detailed descriptions how each of these variables is calculated is provided in the table 2. Dataset contains also variables calculated from monthly data. These include momentum variables and market value. Even though the frequency of the return prediction will be monthly, some variables are calculated from weekly data. These include standard deviation, ratio between price and 52-week high price, beta coefficient, idiosyncratic volatility, turnover and on-balance volume. Standard deviation is calculated as rolling 52-week standard deviation of the stock returns.

Beta coefficient and idiosyncratic volatility are calculated by regressing returns of the stocks by the excess market return. As described above, in order to pool the dataset certain variables

**Table 3: Descriptive statistics**

Table provides time series average of cross-sectional means and standard deviations of all variables used in this study. Values are reported separately for pooled Nordic market and four Nordic markets Denmark, Finland, Norway and Sweden. EXC.RET is monthly excess return calculated from total return index. Risk free rate used to calculate excess returns is US dollars one-month Treasury bill rate.

Variable	Nordic		Denmark		Finland		Norway		Sweden	
	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
EXC.RET	0.007	0.100	0.007	0.088	0.009	0.089	0.006	0.106	0.009	0.098
CA	0.117	0.148	0.107	0.159	0.108	0.119	0.132	0.159	0.115	0.144
CTO	0.813	0.710	0.739	0.670	0.970	0.647	0.651	0.674	0.868	0.741
INV	0.161	0.422	0.118	0.305	0.097	0.273	0.203	0.481	0.179	0.443
BEME	0.699	0.709	0.688	0.558	0.757	0.661	0.680	0.630	0.753	0.827
CFP	0.080	0.122	0.086	0.140	0.090	0.099	0.093	0.145	0.065	0.102
DEBT	2.574	5.231	3.025	5.249	2.415	4.646	3.330	6.456	2.353	4.443
SP	1.585	2.279	1.356	1.636	2.066	2.105	1.316	1.719	1.940	2.849
EP	0.045	0.143	0.044	0.114	0.051	0.106	0.026	0.163	0.058	0.160
ROA	0.045	0.103	0.043	0.096	0.053	0.072	0.029	0.110	0.052	0.112
ROE	0.088	0.224	0.095	0.203	0.103	0.158	0.058	0.280	0.093	0.212
Q	0.696	0.326	0.597	0.391	0.735	0.286	0.655	0.345	0.759	0.273
MOM <sub>7</sub>	0.080	0.229	0.069	0.198	0.072	0.201	0.080	0.245	0.089	0.230
MOM <sub>12</sub>	0.171	0.382	0.149	0.329	0.153	0.320	0.171	0.409	0.190	0.387
MOM <sub>36</sub>	0.397	0.751	0.400	0.679	0.363	0.625	0.357	0.797	0.432	0.754
MOM <sub>2</sub>	0.015	0.093	0.013	0.081	0.014	0.085	0.015	0.097	0.017	0.093
MOM.IND	1.144	0.284	1.132	0.248	1.142	0.285	1.148	0.293	1.148	0.279
SD	0.047	0.030	0.045	0.028	0.042	0.028	0.051	0.030	0.051	0.029
HIGH52	0.684	0.284	0.722	0.266	0.621	0.294	0.684	0.271	0.695	0.261
BETA	0.863	0.525	0.703	0.391	0.732	0.495	0.909	0.530	0.999	0.509
IDVOL	0.046	0.030	0.044	0.025	0.040	0.028	0.051	0.032	0.048	0.029
LOG.MV	6.331	1.340	6.400	1.328	6.402	1.272	6.124	1.246	6.452	1.414
TO	0.043	0.109	0.053	0.126	0.025	0.065	0.028	0.073	0.056	0.119
OBV	0.170	0.507	0.151	0.544	0.098	0.390	0.174	0.529	0.216	0.524

are converted to US dollars. One of these variables is weekly unadjusted stock price which is used to calculate the weekly stock returns used in the regression. Market return is constructed as equal weighted weekly market return following Green, Hand and Zhang (2017). Because the returns are noted in US dollars one-month Treasury bill rate, which is obtained from Kenneth French's database, is used as a risk free rate proxy. The regression is run for each company separately for each month on rolling basis. For each regression up to three years of weekly historical data is considered, but minimum 15 weeks of data is required. Finally, the beta is simply the sensitivity of stock returns on the market return changes and the idiosyncratic volatility is the standard deviation of the regression residuals.

Turnover is calculated by dividing weekly trading volume by number of shares out standing. Calculation of on balance volume consists of two steps and it is also calculated from weekly data. First current weekly turnover is multiplied by the sign of corresponding weeks return. Then this product is added to the cumulative sum of the historical on previous on-balance volumes.

Machine learning algorithms can be sensitive to outliers in the data. Therefore, all explanatory variables are winsorized between 1st and 99th percentiles. In case value of the certain variable is less than 1st percentile of the corresponding months values it will be replaced by

the 1st percentile threshold value. In case value of the certain variable is above 99th percentile of the corresponding months values it will be replaced by the 99th percentile threshold value. Additionally, any missing value in explanatory variables will be replaced by zero.

Table 3 provides descriptive statistics of company characteristics. For each explanatory variable time series average and standard deviation of the cross-sectional mean is reported. Values are reported for pooled nordic markets as well as individual markets.

## 4 Methodology

This study will evaluate machine learning methods from two perspectives. First perspective that is evaluated is the profitability of the methods. Profitability is estimated via portfolio construction following Lewellen (2015). First expected returns are derived from each model. This process is introduced in more detail in following subsections. After obtaining the expected returns, each month all stocks are distributed to ten portfolios based on the magnitude of the expected return. Allocation is univariate and does not consider other variables than the expected return of the stock for the next month. Even though models are trained only once a year, expected returns are re-calculated every month as the most recent available data is inserted to the model. Therefore, also the portfolio allocation is repeated monthly. Each month all stocks are allocated to one of the ten expected return portfolio, but to avoid result to be mainly driven by small stocks approach from Hanauer and Kalsbach is applied (2023) and the breakpoints for the allocation are calculated only from the large market value stocks. Large market value stocks are the biggest stocks that account for 97% of the aggregated total market value of respective month.

In addition to the ten expected return portfolios, for each method zero investment portfolio is formed. Zero investment or long-short portfolio is simply the spread between return of the highest expected return portfolio and return of the lowest expected return portfolio. Both value and equal weighted returns will be reported for each portfolio. Performance of the machine learning portfolios is backtested and evaluated in multiple ways. For all expected return portfolios historical realized mean returns are reported together with Sharpe ratios. For long-short portfolios also maximum drawdown and maximum 1 month loss will be reported. Maximum 1 month loss is simply the largest negative monthly return of each portfolio. Maximum drawdown is define as

$$MaxDD = \max_{0 \leq t_1 \leq t_2 \leq T} (Y_{t_1} - Y_{t_2}) \quad (1)$$

where  $Y_t$  stands for cumulative return from the beginning of the period until date  $t$ . Turnover will be calculated only for long side of the long-short portfolios. In order to examine risk adjusted returns long-short portfolio returns will be regressed against Fama-French (2015) six

factor model factors <sup>7</sup>. From this regression alphas, which can interpret as the excess return that the models are able to generate that cannot be explained by the loadings in the six risk factors. Also  $t$ -statistics for the alphas and  $R^2$  values are reported. Regression formula for risk adjusted performance

$$\hat{r}_{i,t} = \alpha + \beta_1 RMRF_t + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 CMA_t + \beta_5 RMW_t + \beta_6 MOM_t + \epsilon_t \quad (2)$$

where  $RMRF$  is the excess market return,  $SMB$  is the spread in the return between small market value stocks and large market value stocks,  $HML$  is the spread in the return between high book-to-market value stocks and low book-to-market value stocks,  $CMA$  is the spread in the return between conservatively investing stocks and aggressively investing stocks,  $RMW$  is the spread in the return between stocks with robust profitability and stocks with weak profitability and  $MOM$  is the between returns of stocks that had highest return in period  $t - 1$  and the stocks that had lowest return in period  $t - 1$ . Factors are constructed from the same dataset as machine learning portfolios, except that the micro-cap stocks are not excluded. Construction of these factors is described in more detail in Appendix.

Machine learning models are also evaluated based on their prediction accuracy. Prediction accuracy will be evaluated using out-of-sample  $R^2_{os}$  and Diebold-Mariano tests. Two out-of-sample  $R^2$  figures are presented. Traditional out-of-sample  $R^2$  uses historical mean return as the benchmark estimation. Traditional out-of-sample  $R^2$  is defined as

$$R^2_{os \text{ Trad.}} = 1 - \frac{\sum_{t=1}^T \sum_{i=1}^N (r_{i,t} - \hat{r}_{i,t})^2}{\sum_{t=1}^T \sum_{i=1}^N (r_{i,t} - \bar{r}_{i,t})^2} \quad (3)$$

where  $r_{i,t}$  is the realized return of stock  $i$  in month  $t$ ,  $\hat{r}_{i,t}$  is the predicted return of the same stock for month  $i$  and  $\bar{r}_{i,t}$  is the historical mean return of the same stock excluding month  $i$ . Nevertheless, Gu et al. (2020) argue that the historical mean return is so noisy estimator that it underperforms compared to static estimation of zero and therefore artificially improves the out-of-sample  $R^2$ . Instead they propose alternative out-of-sample  $R^2$  measure where the squared sum of returns in denominator is not demeaned.

$$R^2_{os} = 1 - \frac{\sum_{t=1}^T \sum_{i=1}^N (r_{i,t} - \hat{r}_{i,t})^2}{\sum_{t=1}^T \sum_{i=1}^N r_{i,t}^2} \quad (4)$$

$R^2$  presents the prediction accuracy as a single figure, whereas Diebold-Mariano allows for pairwise comparison of different models. Diebold-Mariano value is calculated as:

---

<sup>7</sup>Fama and French (2015) introduced the five factor model. Factors used to regress machine learning portfolio returns include five factor model factors and momentum factor from Carhart (1997).



$$\begin{aligned}
d_{12,t} &= \frac{1}{N_t} \sum_{i=1}^N ((r_{i,t} - \hat{r}_{i,t,1})^2 - (r_{i,t} - \hat{r}_{i,t,2})^2) \\
\bar{d}_{12} &= \frac{1}{T} \sum_{t=1}^T d_{12,t} \\
DM_{12} &= \frac{\bar{d}_{12}}{\hat{\sigma}_{d_{12}}}
\end{aligned} \tag{5}$$

where  $\hat{r}_{i,t,1}$  is the return prediction of first model for company  $i$  at time  $t$  and  $\hat{r}_{i,t,2}$  is the return prediction of the second model for company  $i$  at time  $t$ .  $N_t$  is number of observations in prediction period  $t$ . Therefore,  $d_{12,t}$  is a time series of differences in average squared prediction errors between model 1 and model 2.  $\bar{d}_{12}$  is the mean of  $d_{12,t}$  and  $\hat{\sigma}_{d_{12}}$  is the Newey and West (source) standard error of  $d_{12,t}$ . Diebold-Mariano test allows us to estimate statistical significance of the prediction accuracy of two models. Under assumption that there does not exist difference in prediction accuracy between models Diebold-Mariano test statistic follows normal distribution with mean of 0 and standard deviation of 1,  $DM \sim \mathcal{N}(0, 1)$ . The significance of the difference is reported both for traditional 5% level as well as for 3-way comparisons with Bonferroni adjustment.

In the spirit of Lewellen (2015) expected returns are also estimated by regressing realized returns with the expected returns. This regression follows

$$r_{i,t} = \alpha + \beta_1 \hat{r}_{i,t} \tag{6}$$

where  $r_{i,t}$  is the realized return of company  $i$  at time  $t$  and  $\hat{r}_{i,t}$  is the expected return of corresponding model for company  $i$  at time  $t$ . For these regressions betas,  $t$ -statistics for betas and  $R^2$  values will be reported. Ideally the beta coefficient or the slope for the predicted return should be 1 and highly significant. Magnitude of the beta coefficients can provide information of possible over or undershooting of the models.

## 4.1 Benchmark factors

Benchmark factor construction follows  $2 \times 3$  portfolio sort approach of Fama and French (1993, 2015) and Carhart (1997). Fama and French (1993) use NYSE breakpoints for size and book-to-market value sort. Since on compared to US markets Nordic markets have less companies with high market value. Using NYSE breakpoints could lead to highly un-diversified portfolios especially among the high market value portfolios. On the other hand breakpoints should not be driven by the small stocks that are numerous, but only account for small part of the total market capitalization. Therefore approach of Fama and French (2012) is applied.

In the end of each June stocks are first distributed to two size portfolios. Companies with

biggest market value that account for 90% total market value are classified as big stocks. All the rest of the stocks are considered to be small stocks. Next stocks are allocated to three value, investment, profitability and momentum portfolios. For all of above variables 30th and 70th percentiles are used to calculate breakpoints. Breakpoints are calculated using only big companies from the size allocation, but the breakpoints are used to allocate all stocks to a portfolio.

$$\begin{aligned}
SMB_{B/M} &= \frac{1}{3}(Small.High + Small.Neutral + Small.Low) \\
&\quad - \frac{1}{3}(Big.High + Big.Neutral + Big.Low) \\
SMB_{OP} &= \frac{1}{3}(Small.Robust + Small.Neutral_{OP} + Small.Weak) \\
&\quad - \frac{1}{3}(Big.Robust + Big.Neutral_{OP} + Big.Weak) \\
SMB_{INV} &= \frac{1}{3}(Small.Conservative + Small.Neutral_{INV} + Small.Aggressive) \\
&\quad - \frac{1}{3}(Big.Conservative + Big.Neutral_{INV} + Big.Aggressive) \\
SMB_{MOM} &= \frac{1}{3}((Small.Winner + Small.Neutral_{MOM} + Small.Loser) \\
&\quad - \frac{1}{3}(Big.Winner + Big.Neutral_{MOM} + Big.Loser) \\
SMB &= \frac{1}{4}(SMB_{B/M} + SMB_{OP} + SMB_{INV} + SMB_{MOM}) \\
HML &= \frac{1}{2}(Small.High + Big.High) - \frac{1}{2}(Small.Low + Big.Low) \\
RMW &= \frac{1}{2}(Small.Robust + Big.Robust) - \frac{1}{2}(Small.Weak + Big.Weak) \\
CMA &= \frac{1}{2}(Small.Conservative + Big.Conservative) \\
&\quad - \frac{1}{2}(Small.Aggressive + Big.Aggressive)
\end{aligned} \tag{7}$$

Book-to-market value is used as indicator of value characteristic of a company. Book-to-market value is calculated as ratio between sum of common equity and deferred taxes and market capitalization on December  $t - 1$ . Profitability is defined as net income before extra items/preferred dividends divided by the book equity of the company. Investment variable is calculated as annual change in total assets. Momentum is defined as cumulated return from  $t - 12$  to  $t - 2$ . Returns are calculated using total return index that is converted to US dollars for comparability between different countries. Market value used in size allocation as well as to weight portfolio returns is also converted to US dollars.

Equation 7 shows formula for each factor. Acronym for each variable is derived from how they are calculated. Value factor is called high minus low (HML), profitability factors is called

robust minus weak (RMW), investment factor is called conservative minus aggressive (CMA). Only momentum factor is exception to this rule and more intuitive naming is used. Portfolio allocation results six portfolios for value, investment, profitability and momentum factors and 24 two-fold size portfolios. After portfolio construction portfolio returns are calculated as difference on value weighted average returns on portfolios formed based on respective variable. E.g. value factor return is difference between average of value weighted returns of two high book-to-market portfolios and average of value weighted returns of two low book-to-market portfolios. Market factor is the average value weighted excess return of the whole market. Risk free rate is obtained from Kenneth French's website.

## 4.2 Linear regression

Benchmark model of this study is Fama-MacBeth (1973) regression. First step of the method is to run rolling cross-sectional regressions with lagged variables. Second step of the method calculates means of the factor loadings obtained from the cross-sectional regressions. Finally expected stock return can be obtained by multiplying the mean factors loading with latest available stock characteristics. Below formulas show one model specification for Fama-French (1993) three factors model.

$$f(x_{i,t}; \theta) = \theta^T x_{i,t} \quad (8)$$

$$\bar{\theta}_j = \frac{1}{T} \sum_{t=1}^T \theta_{j,t} \quad (9)$$

$$E_t[r_{i,t}|x_{i,t-1}] = \bar{\theta}_{t-1}^T x_{i,t-1} \quad (10)$$

One benefit that linear regression models have is that they do not require hyperparameter tuning. Therefore data does not have to be split to three sub-samples for separate validation of hyperparameters and testing. To obtain the expected return mean of 120 historical regression coefficients is calculated. Due to their high computing cost machine learning models are usually trained only once a year and then used for the rest of the year. Each month recent information is just inserted to the model. Computing requirements for linear model is far lesser than for non-linear models. Nevertheless to ensure comparability between different models also the linear model is trained only once per year. That means that no more recent stock returns than  $t-1$  are used to train the model to predict stock return  $t$ , but the gap between predicted return and last return used to train the model can grow up to 12 months. Since we use lagged variables, this means that for prediction of stock return  $t$  we always use stock characteristics from  $t-1$ , but some factors are only updated yearly. To mimic information set investor would have had available in historical periods we have to account for the delay in reporting balance sheet information.

Therefore timeline of Fama and French (1993) is followed and models are trained each year at end of June.

### 4.3 Random forest

Decision trees are one example of nonparametric machine learning algorithms. Idea of the decision trees is to split data into the most homogenous groups. Decision trees can be used for both classification and regression tasks. Starting point of the decision tree is called a root node. At each iteration of decision tree algorithm finds the optimal threshold to split the data to the nodes to minimize the objective function value. Then iteratively these nodes can be further split and the tree grows. This process is repeated until predefined tree size, set by the user, is reached or objective function cannot be improved anymore. Regression tree nodes that are not further split are called leaves. Final prediction of the regression tree leaf is the average of the dependant variable values of training set observations inside it. Gu et al. (2020) formulate prediction of a regression tree with  $K$  leaves as

$$f(x_{i,t}; \theta, K, L) = \sum_{k=1}^K \theta_k 1_{\{x_{i,t} \in C_K(L)\}} \quad (11)$$

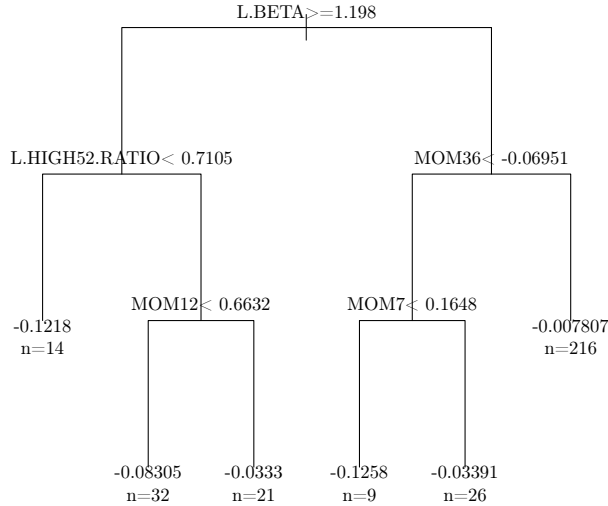
Where  $C_k(L)$  represents one of the  $K$  splits the tree consists of.  $L$  is the indicator of the depth of the leaf.  $\theta_k$  indicates average return within leaf  $k$  and  $1_{\{x_{i,t} \in C_K(L)\}}$  indicates whether observation  $x_{i,t}$  belongs to leaf  $k$ . Since observation can only belong to one leaf, partition  $C_K(L)$  is the product of the above partitions.

Advantage of the regression trees is that they are rather simple and intuitive, but still they are able to model even complex interactions and non-linear relationships among the predictors. One common problem with regression trees is that they easily overfit the data and would require heavy regularization. Random forest models aims to avoid this problem by deriving the predictions from ensemble of regression trees. As the name might suggest random forest consists of multiple decision trees.

Idea of the random forest is to randomly generate set of decision trees and then use the average outcome of the decision trees as the final output. This way the model is less likely to overfit the data. Nevertheless to avoid the overfitting trees inside random forest should not be too correlated and this is ensured including randomness in the construction of the decision trees. Randomness in the generation of the decision trees is applied by restricting the set of observations used in the training of the model. Number of the variables model considers in each split as well as maximum depth of the decision tree and number of trees in the random forest can also be limited. Setting these parameters correctly is a crucial part of the training. These are the hyperparameters which require input from the user, but which also can be optimized for different tasks. Table 4 in Appendix shows which values were considered for each hyperparameter that were optimized for random forest.

**Figure 1: Illustrative regression tree**

Tree is trained from the actual dataset for 30th of July 2004 and then pruned to show only few most important leaves. Figure serves only illustrative purposes and random forest models used in the study do not necessarily contain identical trees.

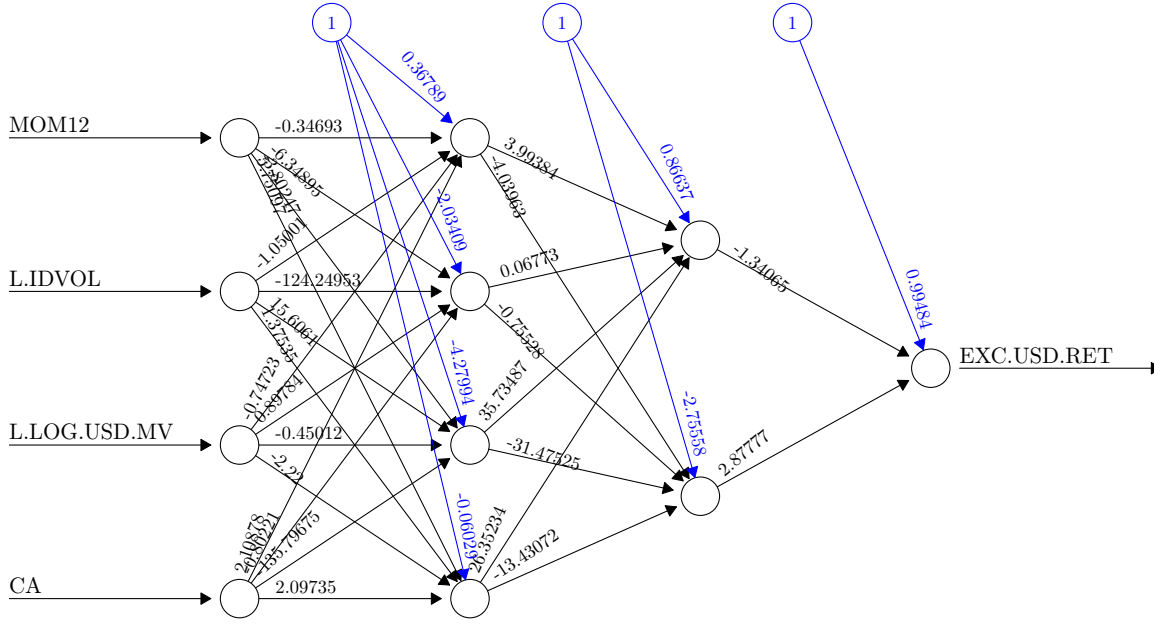


#### 4.4 Neural networks

Artificial neural networks are powerful machine learning method category. Currently neural networks are popular approach to many real world prediction issues. Due to their strong performance in multiple domains, neural networks are often considered as state of the art machine learning method. Despite their popularity, for many users neural networks are sort of black box tools because of their complexity. Compared to linear regression and tree based models, neural networks are far less interpretable. Another weakness of the neural networks is that they are highly parameterized and highly sensitive for parameter initialization. Some of the parameter, such as learning rate, are usually optimized during the training of the model while others such as architecture of the model are usually fixed.

One of the first things user has to decide while training a neural network is the architecture of the model. This study focuses on feedforward neural networks which consists of input layer, hidden layers and an output layer. Input layer consists of the predictive variables whereas output layer produces the final prediction. In between thee exist 1 to N hidden layers. Hidden layers again consists of so called neurons. Similar to number of hidden layers, user has to also decide number of neurons in each of the hidden layers. Number of the hidden layers is often referred as the deepness of the model where as number of the neurons in each hidden layer is referred

**Figure 2: Illustrative neural network**



as the width of the model. While lot of previous literature simultaneously examine multiple different architectural forms, due to computing capacity in this study only one architecture will be examined (Gu et al., 2020; Hanauer & Kalsbach, 2023; Tobek & Hronec, 2021). Neural network of this study has two hidden layers. First hidden layer has 16 neurons and following common geometric pyramid rule second hidden layer has 8 neurons. Rather shallow and narrow architecture is chosen because they usually perform better with smaller datasets (Gu et al., 2020). In order to improve and fasten the converging of the model batch normalization is implemented between all layers.

Idea of the neural network is that each neuron, using weights and biases terms, aggregates information from previous layer and subsequently feeds the information to the activations function. Neural network model used in this study is fully connected, meaning that each neuron is connected to all neurons in previous layer. Output of the activation function will be the input for the next layer. Neural network model is trained by optimizing these weights and biases terms. There exists many options for the activation function, which is again one choice user has to make. Activation function used in this study is rectified linear unit

$$ReLU(x) = \max(0, x) \quad (12)$$

Since model is trained for a regression task final neuron in the output layer has different

activation function than the neurons in the hidden layers. Activation function for the output neuron is linear function.

As mentioned neural networks include numerous hyperparameters that can be optimized during the training of the model. Training neural network is computationally demanding. Due to limited computing capacity hyperparameters are not optimized in this study, but predefined values are used. Hyperparameters and their values are presented in table 4. Additionally, to further limit the computational demand and simultaneously avoid overfitting early stopping algorithm is applied. Early stopping is implemented so that training of the model is terminated after five epochs where the loss function value does not reduce for validation set. Instead of inserting whole dataset to the model at once data is inserted to the model in smaller subsamples so called batches. Epoch on the other hand measures how many times the whole dataset is run through the model.

Neural networks learn by adjusting weights to the direction of gradient. This is done in repetitive iterations. In each iteration size of the change is defined by hyperparameter called learning rate. Since learning rate is a hyperparameter it needs an input from the user. It can also be optimized. Setting correct learning rate is crucial, since too big learning rate might prevent algorithm from converging to optimal solution, but too small learning rate makes converging slow. For above described reasons learning rate is not optimized in this study, but using learning rate scheduler it is adjusted during the training of the model. In order to ensure efficient training learning rate is set 0.01 in the beginning of the training and after ten epochs learning rate will be multiplied by 0.9.

Neural networks are also sensitive to the weight initialization, where the initial weights are set which the model starts to optimize. Depending on the initialization of the weights neural networks can converge to different results. To reduce model variance caused by this, an ensemble method is applied. Ensemble is implemented by training five separate models with different initial weights. Final prediction will be then average of the predictions of the five models.

## 4.5 Variable importance

One challenge in dealing with various statistical methods is that they lack common metrics for explanatory inference. Many of the models have metrics for variable importance, but comparability of these metrics can be questioned. Therefore, approach of Gu et al. (2020) is implemented to define variable importance metrics for model applied in this study. Approach consists of following steps. First one variable at a time is set to zero. Then the reduced model is retrained and new predicted returns are derived using the reduced model. Process of training and predicting returns is identical to the reduced model as for the full model. After obtaining the predicted returns from reduced model, out-of-sample  $R^2$  values are calculated for these returns. Then change compared to out-of-sample  $R^2$  of full model is calculated. Finally, to obtain relative

variable importance metric sum of changes in out-of-sample  $R^2$ s is normalized to one within model. Same process is applied to each variable and all models.

## 4.6 Sample splitting

It is common while training machine learning models to split the data to three sets. Training set will be used as the name suggest to train the model. In case machine learning model includes hyperparameters these can be optimized with validation set. Finally the true out-of-sample predictions can be performed for testing data. Because we want to mimic situation and information set of an investor we have to take into a consideration timeseries nature of the data.

In stock return prediction literature it is common to split the data as described above, but considering the chronological order. For example Fieberg et al. (2023) use rolling 10 year rolling scheme where they first train model using the first seven years of the data and then optimize hyperparameters using last three years of the rolling window. Finally they train the model with optimal hyperparameter initialization with whole ten year window to predict returns for the next year. Gu et al. (2020) use slightly different approach. Instead of using rolling window they increase the training window size after each training period by one year. Common for these two approaches is that they both train the model only once a year.

### Figure 3: Sample splitting

Illustration sample splitting used in training of the machine learning models. Machine learning models are trained once a year and after each training model is used for next 12 months to predict the stock returns. Each year training period is extended by 12 months. Training period is further split into training data and validation data randomly allocating 80% of the data to training set and 20% of the data to the validation set. Validation set is used to optimize hyperparameters. Minimum length of the training period is 50 months.



The sample splitting scheme applied in this study is slightly different from above described ones. Above approaches use disjoint time periods to mimic the out-of-sample setting in the hyperparameter optimization. In this study training and validation set are separated from the testing data based on time. Approach is closer to approach of Gu et al. (2020) in a sense that training data window is increased each year. Nevertheless, the difference is that the data is distributed to training data and validation randomly instead of using the disjoint periods. Reason why this scheme is chosen is that we want to avoid the retraining of the model after the hyperparameter optimization which is necessary if the most recent data should be included to the model. Sample splitting scheme is illustrated in figure ??.

Since linear regression does not require any hyperparameter optimization there is no need for validation set and all data can be used to train the model. For random forest model we actually



**Table 4: Hyperparameters**

..

	FM	RF	NN
Hyperparameter	-	ntree = 300 mtry = $\in (2, 3, 5, 7)$ max.depth = $2 \sim 6$ sample.fraction = 0.5	Learning rate = 0.001 Batch size = 502 Epochs = 100 Patience = 5 Ensemble = 5

optimize the hyperparameters. Therefore, training window is split to training and validation data so that 80% of the data is used in the training and 20% is assigned to the validation set. For neural network model we do not optimize any actual hyperparameters, but we still need a validation set for the early stopping algorithm. Therefore, for neural network 15% of the training window is assigned to the validation set. Approach of this study follows the common approach to only train the models once a year.

## 5 Empirical results on Nordic equities

Performance of the machine learning models will be evaluated from two aspects. First profitability of the models will be evaluated by investigating performance of univariate expected return sort portfolios. Construction of these portfolios is explained in more detail in section 4. Second aspect that is examined is the prediction accuracy of the machine learning models. Prediction accuracy is evaluated by out-of sample  $R^2$ . Additionally, in the spirit of Lewellen (2015) relationship between expected and realized excess returns are examined by regressing the realized excess returns with the individual stock level predicted returns. Finally, the variable importance for different methods are calculated to see effect of each explanatory variable to the prediction accuracy of the model. Process to define variable importance is described in section 5.4.

### 5.1 Benchmark factor performance

This section works as prerequisite for the machine learning portfolios as it shows the performance of Fama and French five factor (2015) model factors extended by the momentum factor. Later these factors are used to evaluate the risk adjusted performance of the machine learning portfolios, but prior to that it is interesting to observe whether simpler factor construction shows persistence in Nordic markets.

Table 5 provides time series average factor returns, standard deviation of the factor returns, corresponding  $t$  and  $p$ -values as well as monthly minimum and maximum returns for all six factors. It is clear that momentum factor shows strongest performance measured both by the magnitude of the return as well as the statistical significance. Monthly momentum factor return

**Table 5: Benchmark factor summary statistics**

Table presents the mean returns and standard deviations of the benchmark factors together with two-sided  $t$ -statistics and corresponding  $p$ -values. For each factor minimum and maximum monthly return is reported. RMRF is the average value weighted excess return of the pooled Nordic market. Portfolio returns are calculated based on  $2 \times 3$  sorts on size and one other factor. HML is the difference in average of value weighted return of two high value portfolios and average of value weighted return of two low value portfolios. RMW, CMA and MOM are calculated in similar manner, but portfolio sort are done based on investment, profitability momentum factors. SMB is the average of the value weighted returns of the 12 portfolios of small stocks minus the average of the value weighted returns of the 12 portfolios of big stocks. Returns are calculated in US dollars.

	Mean	Std.	$t$ -stat.	$p$ -value	Min	Max
HML	0.0014	0.0022	0.6299	0.5291	-0.2662	0.2508
RMW	0.0013	0.0015	0.8711	0.3842	-0.1251	0.1640
CMA	0.0014	0.0015	0.9102	0.3633	-0.1077	0.1704
MOM	0.0090	0.0021	4.2990	0.0000	-0.1501	0.1828
SMB	-0.0001	0.0014	-0.1091	0.9132	-0.1204	0.1042
RMRF	0.0074	0.0032	2.3051	0.0217	-0.2576	0.2072

is 0.9% with  $t$ -statistic of 4.3. Table B.1 in appendix shows the correlations between factor returns. In Nordic markets correlation of momentum factor with other factors is only minor. Interestingly in Nordic markets momentum factor seems to negatively correlate with market factor.

Strong performance of the momentum factor is inline with previous literature. Many previous studies have documented excess momentum returns either in pooled or individual Nordic markets (e.g. Grobys and Huhta-Halkola (2019) and Leivo and Pätäri (?)). Slightly more surprising is the poor performance of the value factor. Some of the previous studies document value premium in Nordic markets. Grobys and Huhta-Halkola (2019) find statistically significant value premium in Nordic markets, but Grobys and Huhta-Halkola construct equal weighted portfolios where as benchmark factors reported here are formed from value weighted portfolios.

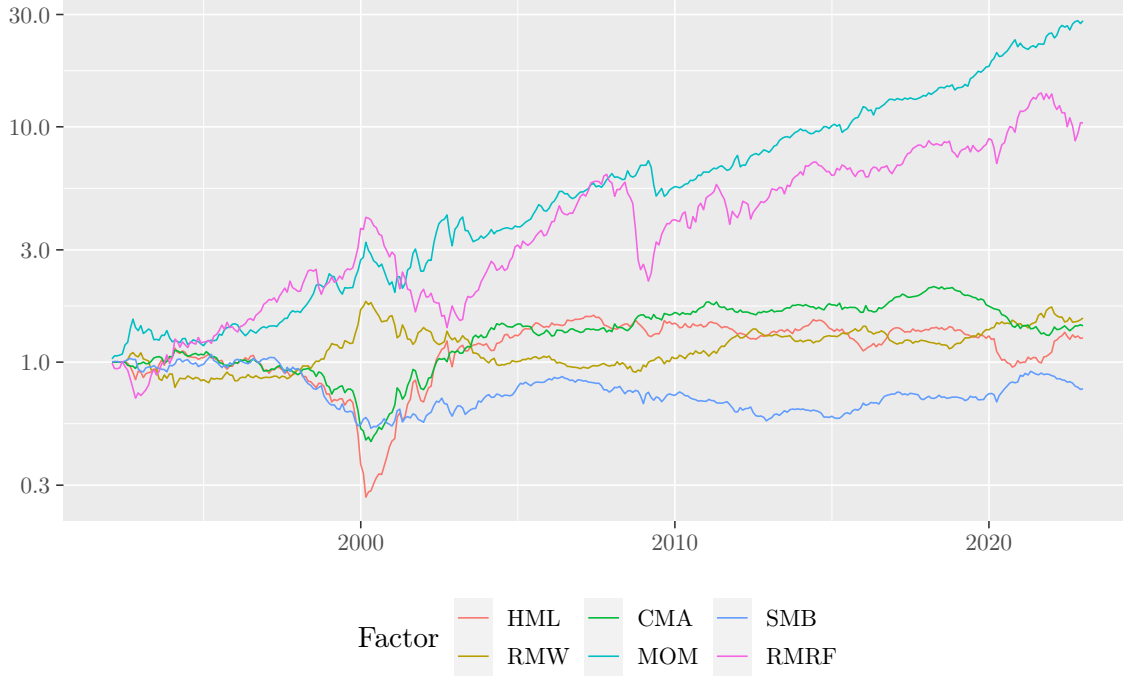
Figure 4 shows the historical performance of the benchmark factors. It shows that until 2008 even momentum factor could not exceed the market return, but still seemed to be less volatile. In post financial crisis period momentum factor has been clearly the strongest performing factor. Interestingly late nineties, which was strong period for market return, was extremely difficult period for value factor. Contrary in the early 2000's value factor performed strongly when market factor was decreasing steeply. Until 2003 performance of value factor is almost opposite to the market factor. Figure 4 also shows that the performance of the size, investment and profitability factors has been poor through out the examined period.

## 5.2 Predictive performance

Panel A of table 6 presents the out-of-sample predictive accuracies for all models. In panel B it presents the pairwise Diebold-Mariano statistics. It can clearly be seen from table 6 that random forest model produces the most accurate out-of-sample predictions. Out of the three models it

**Figure 4: Benchmark factor performance**

Plot presents the cumulative return of the benchmark factors. RMRF is average value weighted excess return of pooled Nordic market. Portfolio returns are calculated based on  $2 \times 3$  sorts on size and one other factor. HML is the difference in average of value weighted return of two high value portfolios and average of value weighted return of two low value portfolios. RMW, CMA and MOM are calculated in similar manner, but portfolio sort are done based on investment, profitability momentum factors. SMB is the average of the value weighted returns of the 12 portfolios of small stocks minus the average of the value weighted returns of the 12 portfolios of big stocks. Returns are calculated in US dollars.



is the only one that produces positive out-of-sample  $R^2$  value even with the more strict version where the benchmark model is prediction of zero. While linear and neural network models both produce negative out-of-sample  $R^2$  of -0.0191 and -0.0039, it can be seen that neural network model performs better than the linear model.

Another interesting insight table 6 provides is the relationship between traditional out-of-sample  $R^2$  and the modified out-of-sample  $R^2$ . It confirms the hypothesis of Gu et al. (2020) about traditional out-of-sample  $R^2$  metric being too loose and showing unrealistically strong results. The traditional out-of-sample  $R^2$  metric is positive for all three models, which means that compared to the more strict out-of-sample  $R^2$  metric the sign of the metric changes for linear and neural network models. Nevertheless, the order between models does not change based on the definition of the out-of-sample  $R^2$  metric. In this regards results are in line with findings of Fieberg et al. (2023).

Figure C.1 in Appendix shows the out-of-sample  $R^2$  values as a timeseries of prediction

**Table 6: Prediction accuracy**

Table presents the prediction accuracy metrics for different machine learning models. Panel A presents two out-of-sample  $R^2$  values. First one uses zero prediction as a benchmark model. This means that the denominator in the calculation of the metric is squared excess return. Second out-of-sample  $R^2$  figure follows the traditional definition and the realized excess return is demeaned by the historical mean return. Panel B of table presents the pairwise Diebold-Mariano statistics for all the methods. Bolded figure indicated significance at 5% level, whereas asterisk indicates significance at 5% level after three-way Bonferroni adjustment.

<i>Panel A: Out-of-sample <math>R^2</math></i>			
	FM	RF	NN
$R^2_{oos}$	-0.0191	0.0051	-0.0039
$R^2_{oos \text{ Trad.}}$	0.0299	0.0517	0.0431
<i>Panel B: Diebold-Mariano statistics</i>			
	FM	RF	NN
FM		<b>8.4462*</b> (0.0000)	<b>2.1495</b> (0.0316)
RF			<b>-10.0946*</b> (0.0000)

periods. Figure shows that the same trends can be seen from all of the methods. It also reveals that in prediction period starting from July 2011 neural network model produced extremely bad out-of-sample predictions. This period is difficult to other methods as well, but not in the same scale as for neural network model. Modified out-of-sample  $R^2$  of neural network model reaches -0.14% during this period.

Inspecting the timeseries of the out-of-sample  $R^2$  produced by the two definitions further the argumentation that the traditional out-of-sample  $R^2$  is too optimistic metric to evaluate goodness of the stock return prediction model. Figure C.1 shows how the traditional out-of-sample  $R^2$  are not only sifted upwards, but also the variation is smaller. This supports the argument of Gu et al. (2020) that the historical mean as an estimator of future stock return contains so much noise that it actually artificially improves the out-of-sample  $R^2$  values.

Results of the predictive performace of different model measured by out-of-sample  $R^2$  is partially inline with previous literature. Results are inline with findings of Drobetz and Otto's (2021) study in European stock markets in a sense that the linear model offers worst out of sample preditcive power when measured by the out-of-sample  $R^2$ . They also find negative out-of-sample  $R^2$  for linear model. Results are also inline in a sense that random forest model shows strong predictive performance. Results of Fieberg et al. (2023) are slightly more contradictory, since they show positive out-of-sample  $R^2$  also for linear model <sup>8</sup>. Both studies of Drobetz and Otto and Fieberg et al. are conducted in European stock markets, which partially overlap with markets of this study, but the difference is that Drobetz and Otto use twenty-two characteristics as well as their second- and third order polynomials and two-way interactions whereas Fieberg et al. only use six characteristics. Variable selection of this study is in between of these two since we include more variables than Fieberg et al., but we do not include second- and third

order polynomials or two-way interactions like Drobetz and Otto.

Where the results clearly deviation from previous literature is the predictive performance of the neural network model. In studies of Drobetz and Otto (2021) and Fieberg et al. (2023) neural network model produces highest, clearly positive, out-of-sample  $R^2$  values. Naturally studies of Drobetz and Otto and Fieberg et al. are not directly comparable to this study since variable set differs and the datasets of Drobetz and Otto and Fieberg et al. are much wider since they include more countries. The size of the dataset could also partially explain the relative poor performance of the neural network model, since usually neural network models require lot of data.

Panel B of table 6 presents the pairwise Diebold-Mariano statistics for all the models. Calculation of the statistics is described in section 4. Table 6 reports the Diebold-Mariano statistics together with corresponding p-values. Bolding of the Diebold-Mariano figure imply significance in normal 5% level where as asterisk implies more conservative 5% level which is Bonferroni adjusted for three-way comparisons. The three-way Bonferroni adjusted critical one-sided Diebold-Mariano value is 2.13.

Discuss the results.

Next the prediction accuracy is examined by following approach of Lewellen (2015) by regressing the realized excess returns by the return predictions from different models. Table 7 presents the summary statistics for these regressions. Left side of the table presents the univariate properties of expected returns and the right side of the table presents the regression statistics. Comparing univariate properties of the expected returns from table 7 to descriptive statistics in table 3 shows that the mean expected return is really close to actual realized mean excess return for neural network and random forest model. Both mean expected and realized return are calculated as timeseries average of cross-sectional means. Linear model on the other hand seems to predict larger returns on average than what is actually realized. Another remark from the table 7 is that the standard deviation for the return predictions produced by the linear model and neural network model is higher than the standard deviation of the realized excess returns. This indicates that the variation in expected returns from these models is larger than the variation in realized excess returns.

Results from right side of the table 7 support the remarks from out-of-sample predictive performance and univariate properties. For all of the models there exists statistically highly significant positive relationship between expected returns and realized returns. If the expected returns would reflect realized excess returns perfectly regression slope shown in table 7 should be one. Random forest model has the highest predictive slope of 0.58, which means that 1% change in expected return respond to 0.58% change in realized return. Neural network and linear models have slightly smaller slopes of 0.29 and 0.16 correspondingly.  $R^2$  values from the

---

<sup>8</sup>Fieberg et al. (2023) report result for multiple subsets where companies are filtered based on their market capitalization. Linear model produces negative out-of-sample  $R^2$  values when only biggest 20% of the stocks are included, but this is not the setting of this study.

**Table 7: Expected return regression summaries**

Table provides univariate properties of the return predictions for all models and summary statistics for regression where realized excess returns are regressed with expected returns. Mean and standard deviation are reported for expected returns. Mean value reported is the timeseries average of the cross-sectional means and standard deviation is the timeseries average of cross-sectional standard deviations. Right side of the table reports the regression coefficients, standard errors of the coefficients, corresponding  $t$ -statistics and the  $R^2$  values.

	Univariate properties		Slope.	Predictive ability		$R^2$
	Mean	Std.		SE	$t$ -stat	
FM	0.0124	0.0123	0.1638	0.0217	7.324	0.0005
RF	0.0074	0.0093	0.5767	0.0306	18.831	0.0029
NN	0.0070	0.0131	0.2949	0.0210	14.05	0.0016

regressions are also presented as third alternative out-of-sample prediction accuracy metric in addition to two previously introduced out-of-sample  $R^2$  metrics. The third  $R^2$  metric further confirms the message of first two as the ordering of the methods is the same for the their metric as for the two previous.

Given that the mean predicted return matches quite well mean realized excess return, but the standard deviation is higher and the slope is slower, it seems that models seem to overshoot in their predictions. It seems that, especially neural network and random forest models, are able to predict the returns correctly on average, but exaggerate the extreme returns. This could at least partially explain rather low out-of-sample  $R^2$  values discovered in table 6. Further insight for this will be provided in next section where performance of expected return sorted portfolios is examined.

### 5.3 Portfolio performance

This section focuses on backtesting the machine learning portfolios, which are formed based on the expected returns produced by different models. Idea is to mimic of information set of a historical investor. Section describes the historical realized return for an investment strategy build on the machine learning models. Section mainly focuses on evaluating performance of the expected return portfolios. Formation of expected return portfolios is described in section 4. Results are discussed separately for value and equal weighted portfolios.

Table 8 presents the performance statistics for all ten expected return portfolios for all three models. Average predicted return, average realized excess return, standard deviation of the realized excess return, corresponding  $t$ -statistic and Sharpe ratio is reported for each portfolio. Left side of the table presents the values for equal weighted portfolios, whereas right of the table presents the values for the value weighted portfolios. Panel A of the table shows results for linear regression model, panel B shows the results for random forest model and panel C shows the results for neural network model.

**Table 8: Machine learning portfolio performance**

Table reports performance metrics for portfolios formed based on univariate expected return sort. Each month all stocks are allocated to ten portfolios based on their expected returns. Breakpoints for the allocation are calculated only from big stocks, which are the biggest stocks that in current month account for 90 percent of cumulative market value of all stocks in the dataset. H-L is zero investment portfolio which consist of short position in portfolio formed from stocks with lowest expected return and long position in portfolio formed from stocks with highest expected return. Time series average of predicted return and realized return of each portfolio is reported for each portfolio together with standard error of realized return. Additionally, Sharpe ratios are reported. Panel A reports result for equally weighted portfolios and panel B reports results for portfolios where each stock in portfolio is weighted by its lagged market value.

<i>Panel A: Linear regression</i>										
	Equal weighted					Value weighted				
	Pred.	Avg.	Std.	<i>t</i> -stat	SR	Pred.	Avg.	Std.	<i>t</i> -stat	SR
Low	-0.0028	0.0049	0.0773	1.1621	0.0630	-0.0016	0.0081	0.0796	1.8704	0.1014
2	0.0048	0.0035	0.0668	0.9552	0.0518	0.0048	0.0063	0.0666	1.7495	0.0949
3	0.0078	0.0051	0.0645	1.4703	0.0797	0.0078	0.0052	0.0638	1.4971	0.0812
4	0.0098	0.0083	0.0603	2.5384	0.1377	0.0098	0.0089	0.0614	2.6690	0.1447
5	0.0115	0.0097	0.0640	2.7957	0.1516	0.0115	0.0080	0.0662	2.2228	0.1205
6	0.0115	0.0088	0.0609	2.6740	0.1450	0.0131	0.0080	0.0640	2.2946	0.1244
7	0.0148	0.0086	0.0619	2.5691	0.1393	0.0148	0.0080	0.0635	2.3258	0.1261
8	0.0170	0.0094	0.0623	2.7933	0.1515	0.0170	0.0077	0.0646	2.1977	0.1192
9	0.0202	0.0130	0.0636	3.7658	0.2042	0.0203	0.0110	0.0632	3.2209	0.1747
High	0.0328	0.0173	0.0682	4.6742	0.2535	0.0346	0.0127	0.0687	3.4027	0.1845
H-L	0.0356	0.0124	0.0495	4.6193	0.2505	0.0362	0.0046	0.0584	1.4510	0.0787
<i>Panel B: Random forest</i>										
	Equal weighted					Value weighted				
	Pred.	Avg.	Std.	<i>t</i> -stat	SR	Pred.	Avg.	Std.	<i>t</i> -stat	SR
Low	-0.0064	-0.0002	0.0754	-0.0465	-0.0025	-0.0051	0.0036	0.0778	0.8523	0.0462
2	0.0010	0.0039	0.0657	1.0879	0.0590	0.0011	0.0027	0.0674	0.7275	0.0395
3	0.0037	0.0078	0.0621	2.3153	0.1256	0.0037	0.0086	0.0636	2.4876	0.1349
4	0.0059	0.0088	0.0616	2.6241	0.1423	0.0059	0.0094	0.0659	2.6240	0.1423
5	0.0078	0.0102	0.0618	3.0502	0.1654	0.0078	0.0098	0.0628	2.8919	0.1568
6	0.0097	0.0093	0.0602	2.8504	0.1546	0.0097	0.0066	0.0654	1.8736	0.1016
7	0.0114	0.0092	0.0632	2.6902	0.1459	0.0115	0.0071	0.0667	1.9635	0.1065
8	0.0132	0.0118	0.0632	3.4294	0.1860	0.0132	0.0096	0.0672	2.6281	0.1425
9	0.0153	0.0144	0.0619	4.2851	0.2324	0.0153	0.0115	0.0630	3.3713	0.1828
High	0.0220	0.0170	0.0674	4.6541	0.2524	0.0208	0.0136	0.0714	3.5147	0.1906
H-L	0.0284	0.0172	0.0400	7.9379	0.4305	0.0259	0.0100	0.0529	3.4916	0.1894
<i>Panel C: Neural network</i>										
	Equal weighted					Value weighted				
	Pred.	Avg.	Std.	<i>t</i> -stat	SR	Pred.	Avg.	Std.	<i>t</i> -stat	SR
Low	-0.0158	0.0006	0.0791	0.1338	0.0073	-0.0139	0.0048	0.0853	1.0429	0.0566
2	-0.0027	0.0065	0.0623	1.9194	0.1041	-0.0026	0.0097	0.0677	2.6479	0.1436
3	0.0014	0.0078	0.0618	2.3290	0.1263	0.0015	0.0066	0.0646	1.8846	0.1022
4	0.0044	0.0086	0.0633	2.5071	0.1360	0.0044	0.0077	0.0658	2.1447	0.1163
5	0.0068	0.0088	0.0621	2.6022	0.1411	0.0068	0.0081	0.0627	2.3724	0.1287
6	0.0090	0.0091	0.0595	2.8136	0.1526	0.0090	0.0089	0.0624	2.6240	0.1423
7	0.0112	0.0106	0.0579	3.3862	0.1836	0.0112	0.0082	0.0605	2.4928	0.1352
8	0.0138	0.0122	0.0631	3.5661	0.1934	0.0138	0.0094	0.0634	2.7275	0.1479
9	0.0171	0.0105	0.0602	3.2046	0.1738	0.0171	0.0080	0.0632	2.3286	0.1263
High	0.0262	0.0150	0.0704	3.9283	0.2130	0.0251	0.0129	0.0727	3.2685	0.1773
H-L	0.0420	0.0144	0.0439	6.0612	0.3287	0.0391	0.0081	0.0564	2.6347	0.1429

**Table 9: Zero investment portfolio performance metrics**

...

	Equal weighted			Value weighted		
	FM	RF	NN	FM	RF	NN
Max DD(%)	-0.3804	-0.2708	-0.2627	-0.6062	-0.5462	-0.4795
Max 1month Loss(%)	-0.1744	-0.1883	-0.1926	-0.3374	-0.3542	-0.3003
FF Alpha	0.0034	0.0136	0.0110	-0.0048	0.0052	0.0030
t-stats	1.7530	6.8282	4.8320	-1.9326	1.8517	1.0487
$R^2$	0.5482	0.2638	0.2063	0.4547	0.1772	0.2523
Sharpe ratio	0.2505	0.4305	0.3287	0.0787	0.1894	0.1429
Turnover (%)	0.1089	0.1140	0.1258	0.0995	0.1492	0.1417

## 5.4 Variable importance

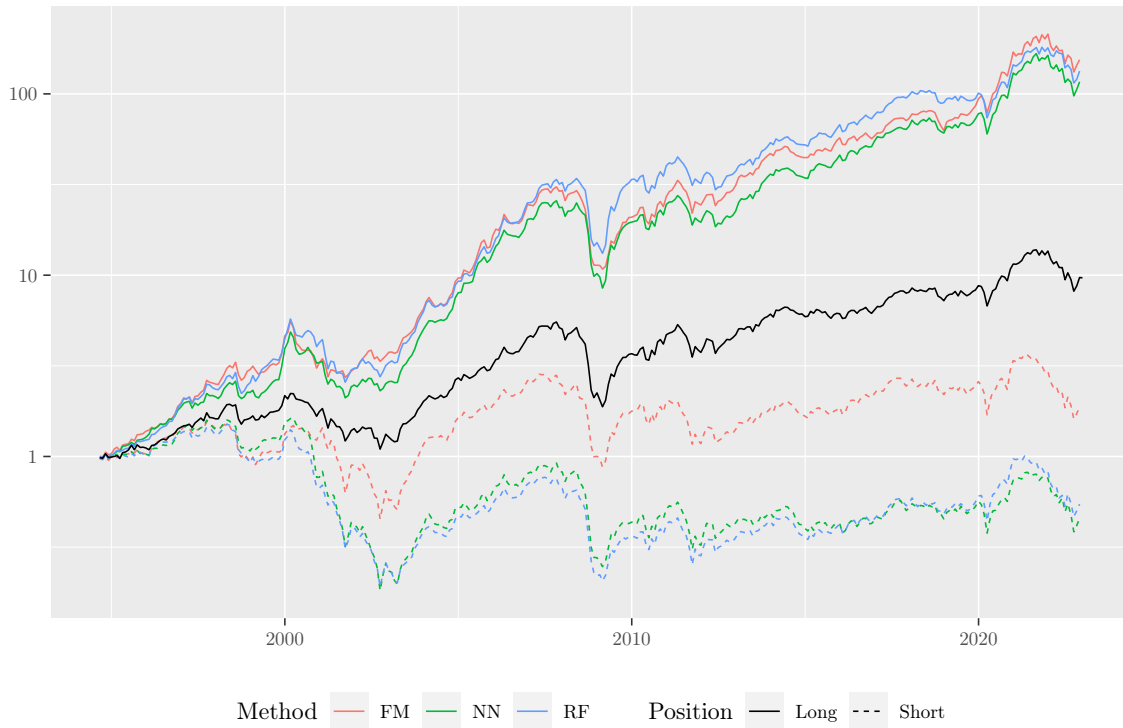
## 6 Conclusion

More trading variables.



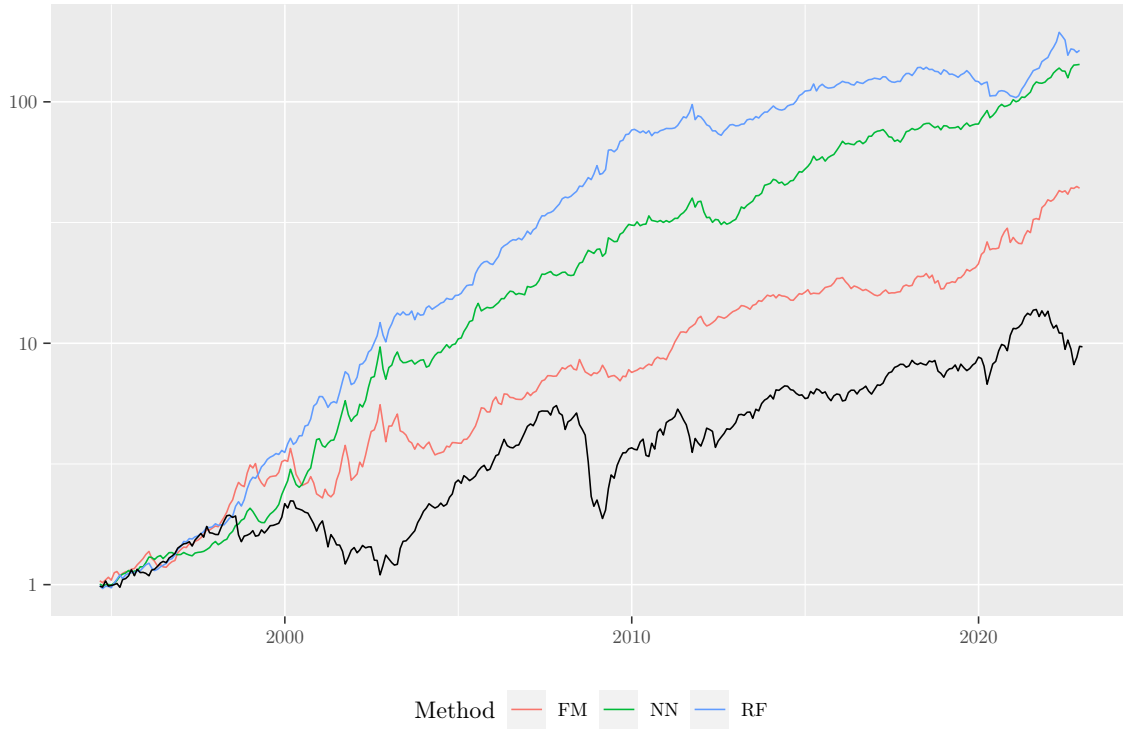
**Figure 5: Cumulative return of equal weighted machine learning portfolios**

Figure plots the realized historical cumulative excess return of the out-of-sample predictions. Figure shows performance of portfolios that are formed allocating all except micro-cap stocks to ten portfolios based on their expected returns. Re-allocation is done monthly. Section 4 describes how expected returns are derived for different models. FM stands for linear model, RF stands for random forest model and NN stands for neural network model. Solid line plots the cumulative performance of the highest expected return portfolio whereas dashed line plots the cumulative excess return for the lowest expected return portfolio. All portfolios are equal weighted. Solid black line shows the value weighted marked return. All returns are converted to US dollars.



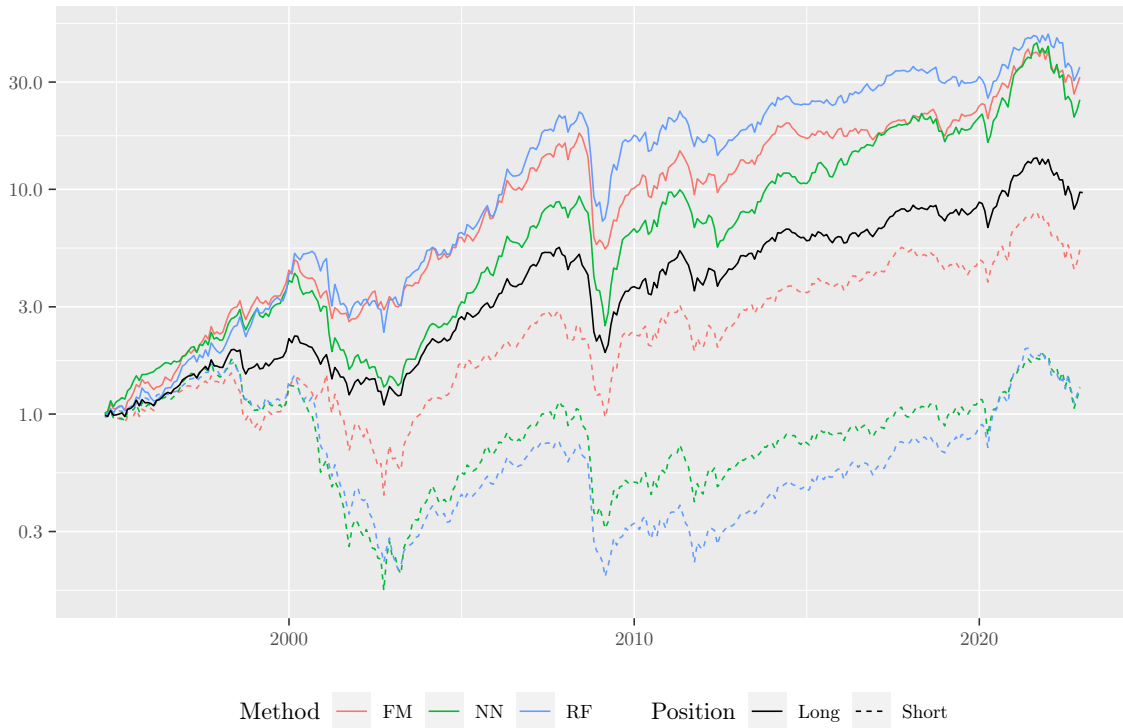
**Figure 6: Cumulative return of equal weighted zero investment portfolios**

Figure presents the realized cumulative spread return between highest expected return portfolio and lowest expected return portfolio. Re-allocation of stocks to portfolios is done monthly. Section 4 describes how expected returns are derived for different models. Both high and low expected return portfolios are equal weighted. FM stands for linear model, RF stands for random forest model and NN stands for neural network model. Solid black line shows the value weighted marked return. All returns are converted to US dollars.



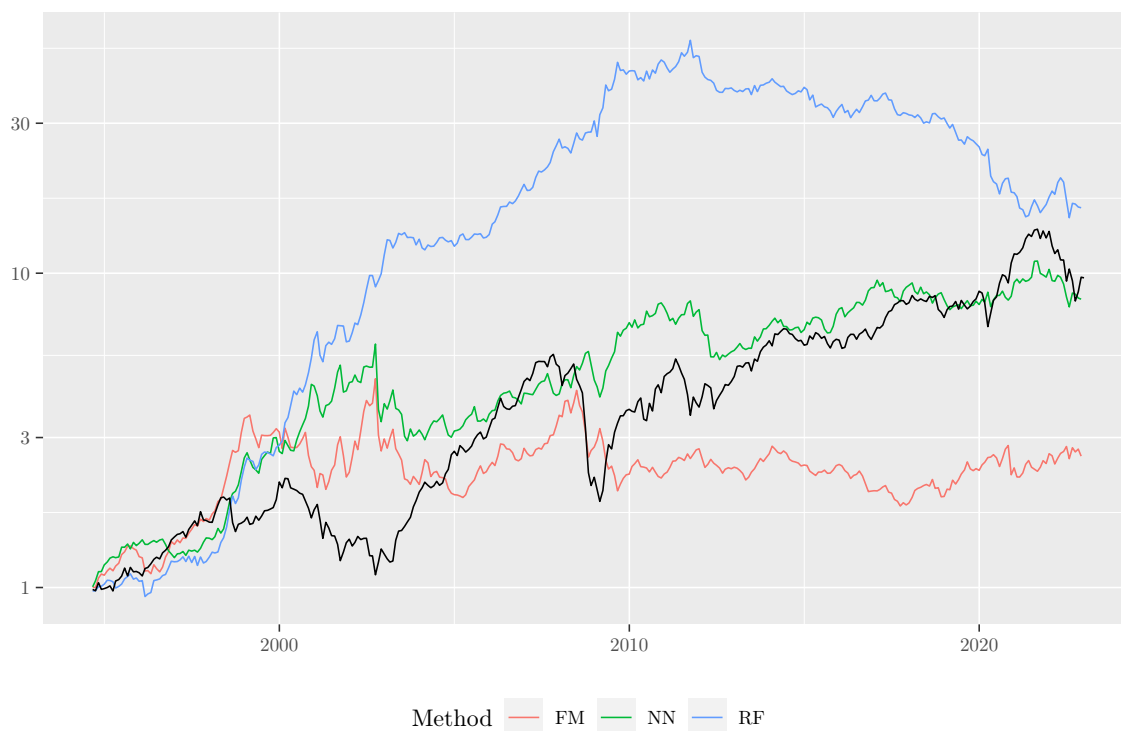
**Figure 7: Cumulative return of value weighted machine learning portfolios**

Figure plots the realized historical cumulative excess return of the out-of-sample predictions. Figure shows performance of portfolios that are formed allocating all except micro-cap stocks to ten portfolios based on their expected returns. Re-allocation is done monthly. Section 4 describes how expected returns are derived for different models. FM stands for linear model, RF stands for random forest model and NN stands for neural network model. Solid line plots the cumulative performance of the highest expected return portfolio whereas dashed line plots the cumulative excess return for the lowest expected return portfolio. All portfolios are value weighted. Solid black line shows the value weighted marked return. All returns are converted to US dollars.



**Figure 8: Cumulative return of value weighted zero investment portfolios**

Figure presents the realized cumulative spread return between highest expected return portfolio and lowest expected return portfolio. Re-allocation of stocks to portfolios is done monthly. Section 4 describes how expected returns are derived for different models. Both high and low expected return portfolios are value weighted. FM stands for linear model, RF stands for random forest model and NN stands for neural network model. Solid black line shows the value weighted marked return. All returns are converted to US dollars.



**Figure 9: ..**

...

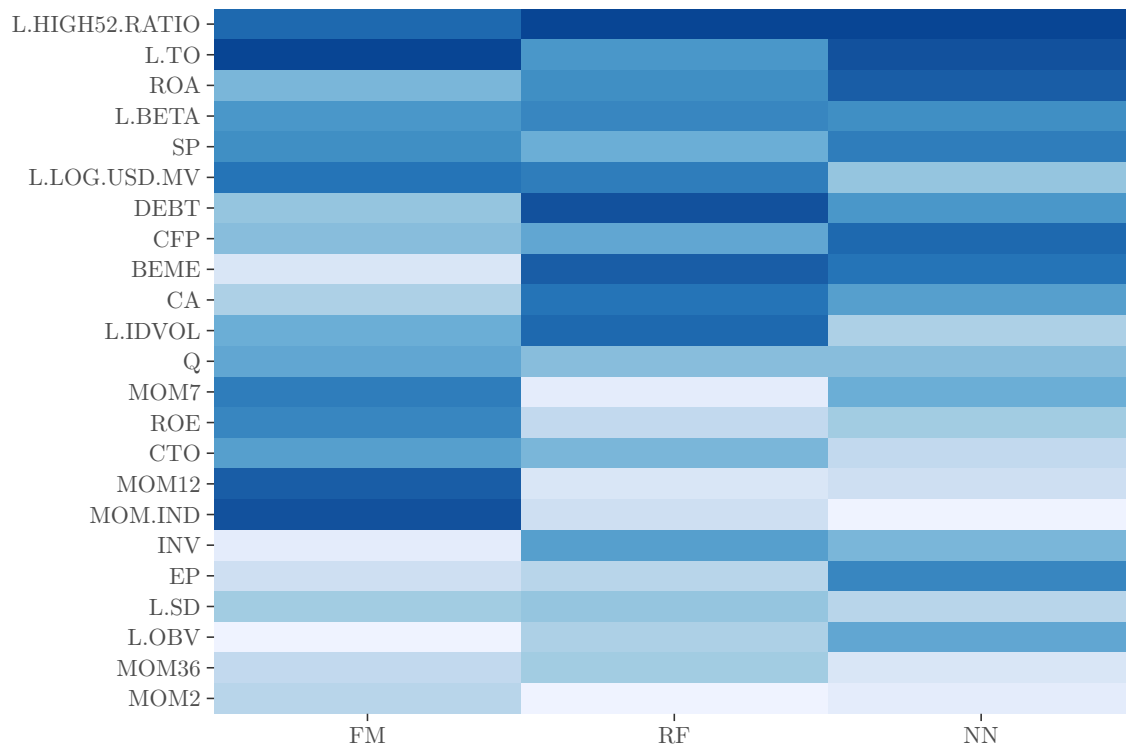
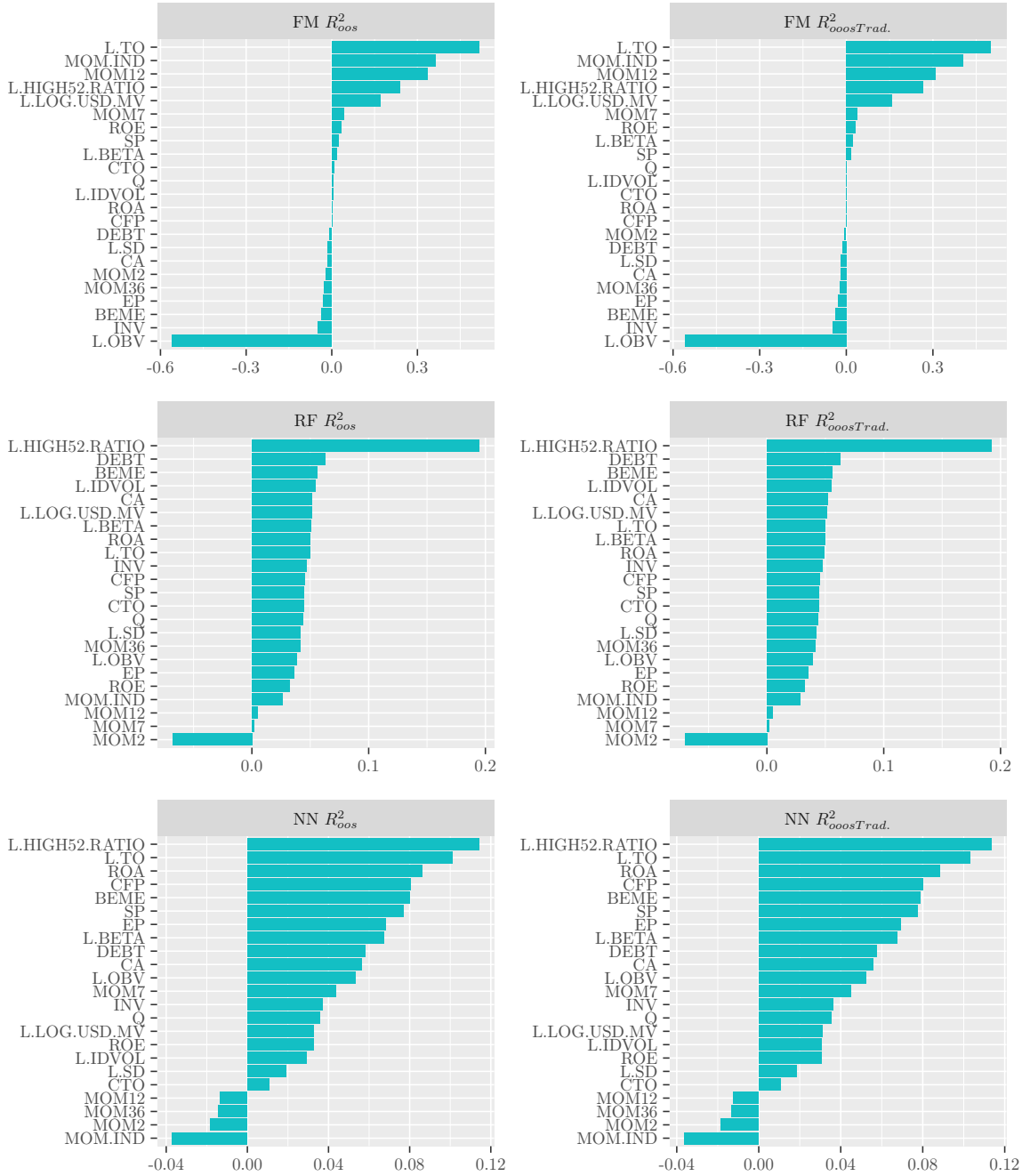


Figure 10: ..

...



## A Data

**Table A.1: Constituent lists and keywords**

Table provides the constituent lists used in data collection.

Denmark	Finland	Norway	Sweden
FDEN	FFIN	FNOR	FSWD
WSCOPEDK	WSCOPEFN	WSCOPENW	WSCOPESD
DEADDK	DEADFN	DEADNW	DEADSD
			FAKTSWD

**Table A.2: Static screens**

It also provides the country specific keywords that are used to delete entries from the dataset. Panel B provides keywords that were used to delete entries from each market separately. Keyword deletion follows Ince and Porter (2006) and Hanauer and Windmüller (2023). Same logic is applied to remove both country specific and generic keywords. Keyword is searched from Datastream attributes NAME, ENAME and ECNAME. In case if at least one of these attributes contains the keyword security is deleted from the dataset. To avoid deleting proper entries, security is only deleted if keyword occurs at the beginning of the name, at the end of the name or as separate word in the name.

<i>Panel A: Static screens.</i>				
	Denmark	Finland	Norway	Sweden
MAJOR	Y	Y	Y	Y
TYPE	EQ	EQ	EQ	EQ
ISINID	P	P	P	P
GEOGN	DENMARK	FINLAND	NORWAY	SWEDEN
GEOLN	DENMARK	FINLAND	NORWAY	SWEDEN
PCUR	DK	FI, MK	NK	SK
GGSIN	DK	FI	NO	SE
<i>Panel B: Country specific keywords.</i>				
	Denmark	Finland	Norway	Sweden
NAME				CONVERTED INTO, USE,
ENAME	\\)CSE \\	USE		CONVERTED-,
ECNAME				CONVERTED - SEE

**Table A.3: Common keywords**

Table shows the general keywords that were used to delete entries from all markets. Keyword deletion follows Ince and Porter (2006) and Hanauer and Windmüller (2023). Same logic is applied to remove both country specific and generic keywords. Keyword is searched from Datastream attributes NAME, ENAME and ECNAME. In case if at least one of these attributes contains the keyword security is deleted from the dataset. To avoid deleting proper entries, security is only deleted if keyword occurs at the beginning of the name, at the end of the name or as separate word in the name.

Security class	Keywords
Duplicates	1000DUPL, DULP, DUP, DUPE, DUPL, DUPLI, DUPLICATE, XSQ, XETa
Depository receipts	ADR, GDR
Preferred stock	PF, 'PF', PFD, PREF, PREFERRED, PRF
Warrants	WARR, WARRANT, WARRANTS, WARRT, WTS, WTS2
Debt	%, DB, DCB, DEB, DEBENTURE, DEBENTURES, DEBT
Unit trusts	.IT, .ITb, TST, INVESTMENTTRUST, RLSTIT, TRUST, TRUSTUNIT, TRUSTUNITS, TST, TSTUNIT, TST UNITS, UNIT, UNITTRUST, UNITS, UNT, UNTTST, UT
ETFs	AMUNDI, ETF, INAV, ISHARES, JUNG, LYXOR, X-TR
Expired securities	EXPD, EXPIRED, EXPIRY, EXPY
Miscellaneous	ADS, BOND, CAP.SHS, CONV, DEFER, DEP, DEPY, ELKS, FD, FUND, GW.FD, HI.YIELD, HIGHINCOME, IDX, INC. &GROWTH, INC.&GW, INDEX, LP, MITS, MITT, MPS, NIKKEI, OPCVM, ORTF, PERQS, PFC, PFCL, PINES, PRTF, PTNS, PTSH, QUIBS, QUIDS, RATE, RCPTS, REALEST, RECEIPTS, REIT, RESPT, RETUR, RIGHTS, RST, RTN.INC, RTS, SBVTG, SCORE, SPDR, STRYPES, TOPRS, UTS, VCT, VTG.SAS, XXXXX, YIELD,YLD, PF.SHS.



**Figure A.1: Number of companies**

Figure shows the development of total number of securities considered in the dataset from 1990 to 2022 for each Nordic country. Figures counts all securities that passed the static screens. Final amount of companies in each time point can be lower since security will be excluded from the dataset if it is missing certain information.

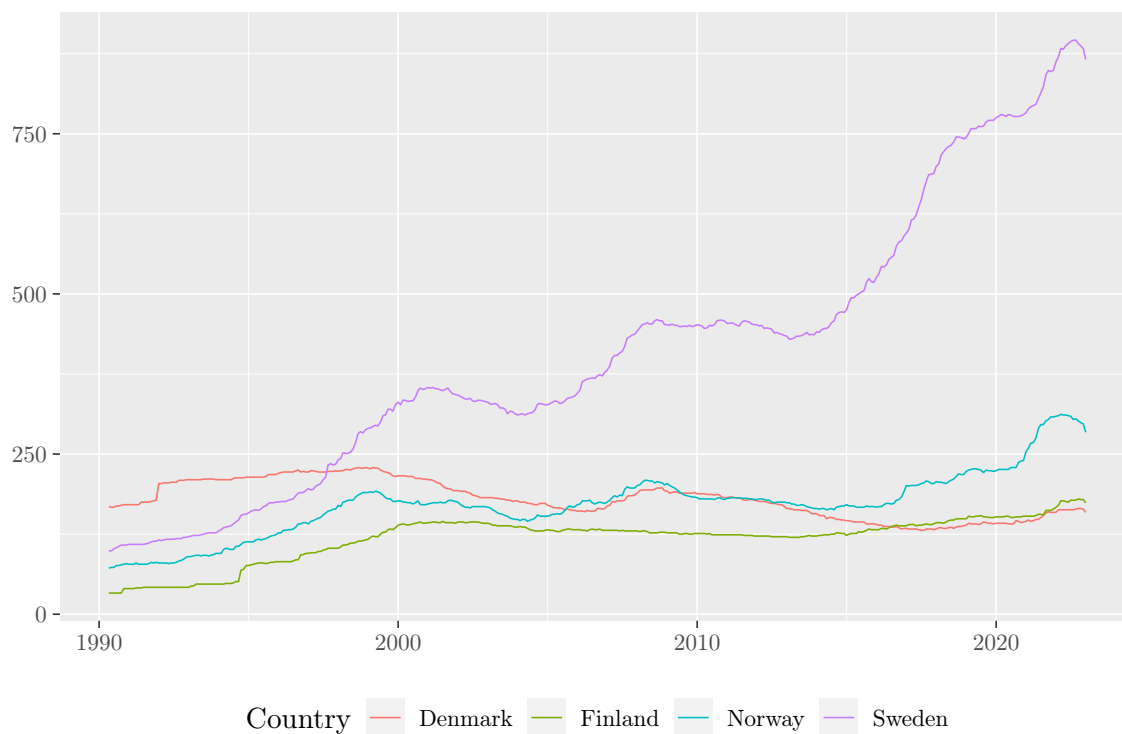
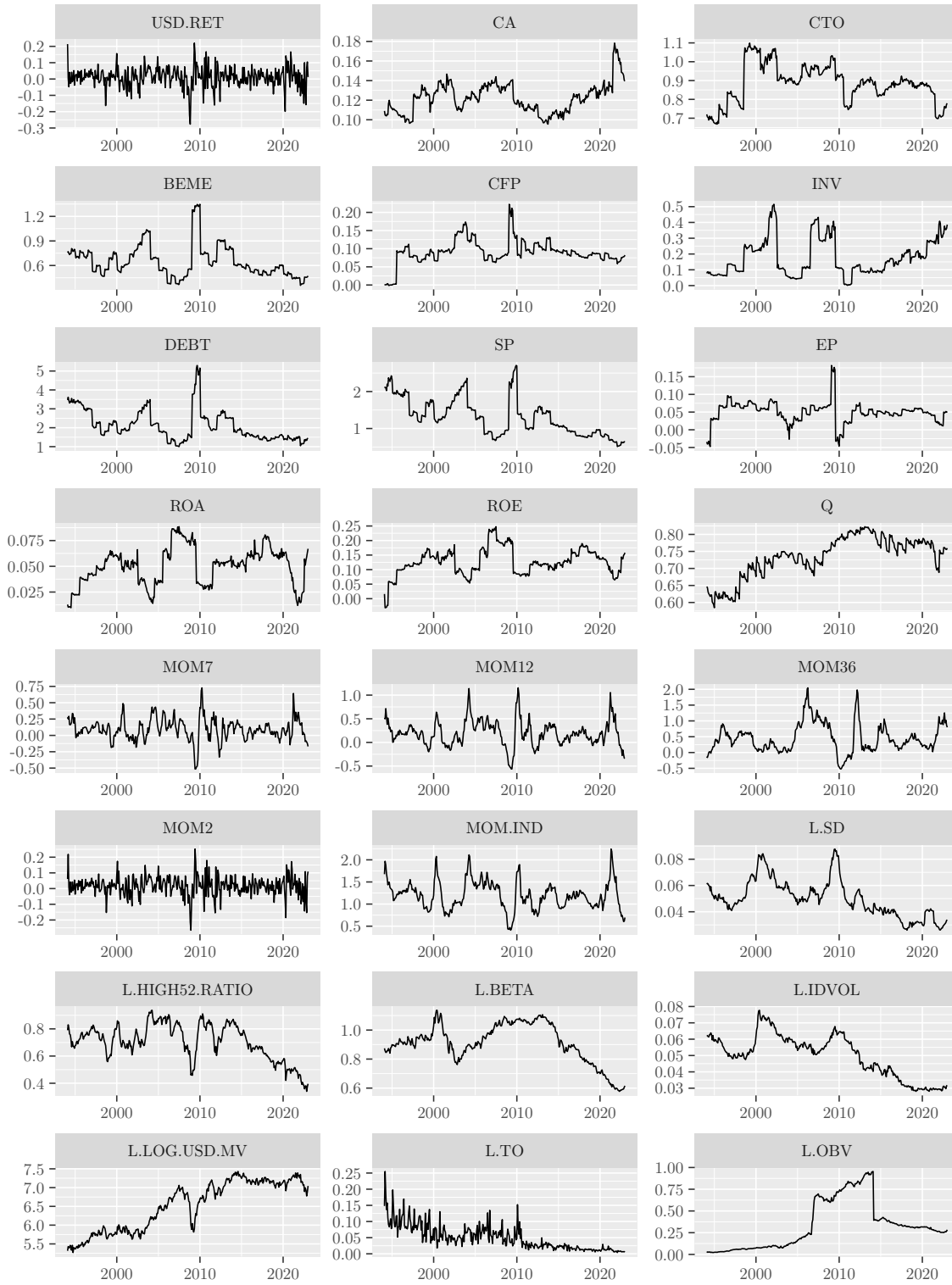


Figure A.2: ....

....



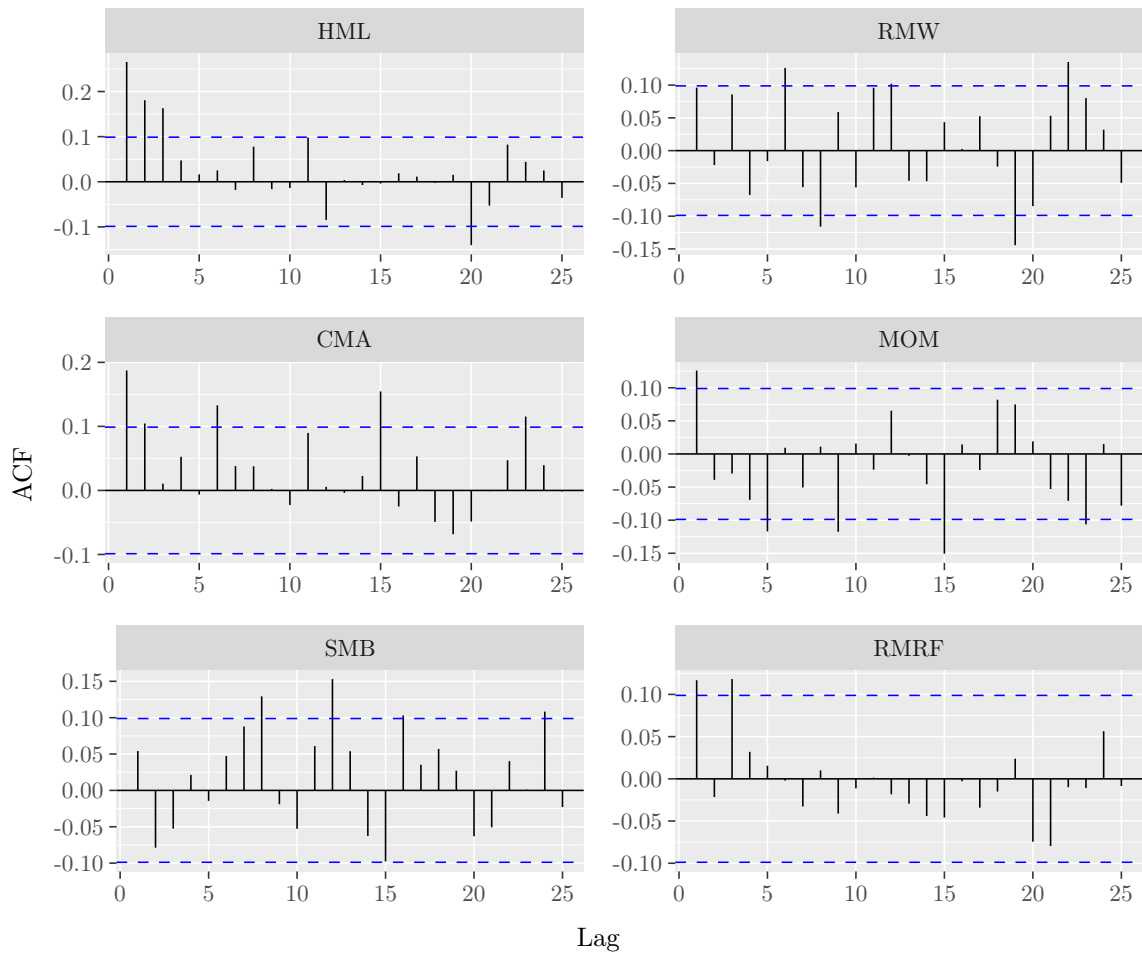
## B Benchmark factors

**Table B.1: Benchmark factor correlation matrix**

Table shows the correlations among the benchmark factors. RMRF is the average value return of the pooled Nordic market. Portfolio returns are calculated based on  $2 \times 3$  sorts on size and one other factor. HML is the difference in average of value weighted return of two high value portfolios and average of value weighted return of two low value portfolios. RMW, CMA and MOM are calculated in similar manner, but portfolio sort are done based on investment, profitability momentum factors. SMB is the average of the value weighted returns of the 12 portfolios of small stocks minus the average of the value weighted returns of the 12 portfolios of big stocks. Returns are calculated in US dollars.

	HML	RMW	CMA	MOM	SMB	RMRF
HML	1	-0.5707	0.5542	0.1122	0.2998	-0.2740
RMW	-0.5707	1	-0.5899	0.0857	-0.2568	0.0639
CMA	0.5542	-0.5899	1	-0.0703	0.1777	-0.2078
MOM	0.1122	0.0857	-0.0703	1	0.1544	-0.2040
SMB	0.2998	-0.2568	0.1777	0.1544	1	-0.2695
RMRF	-0.2740	0.0639	-0.2078	-0.2040	-0.2695	1

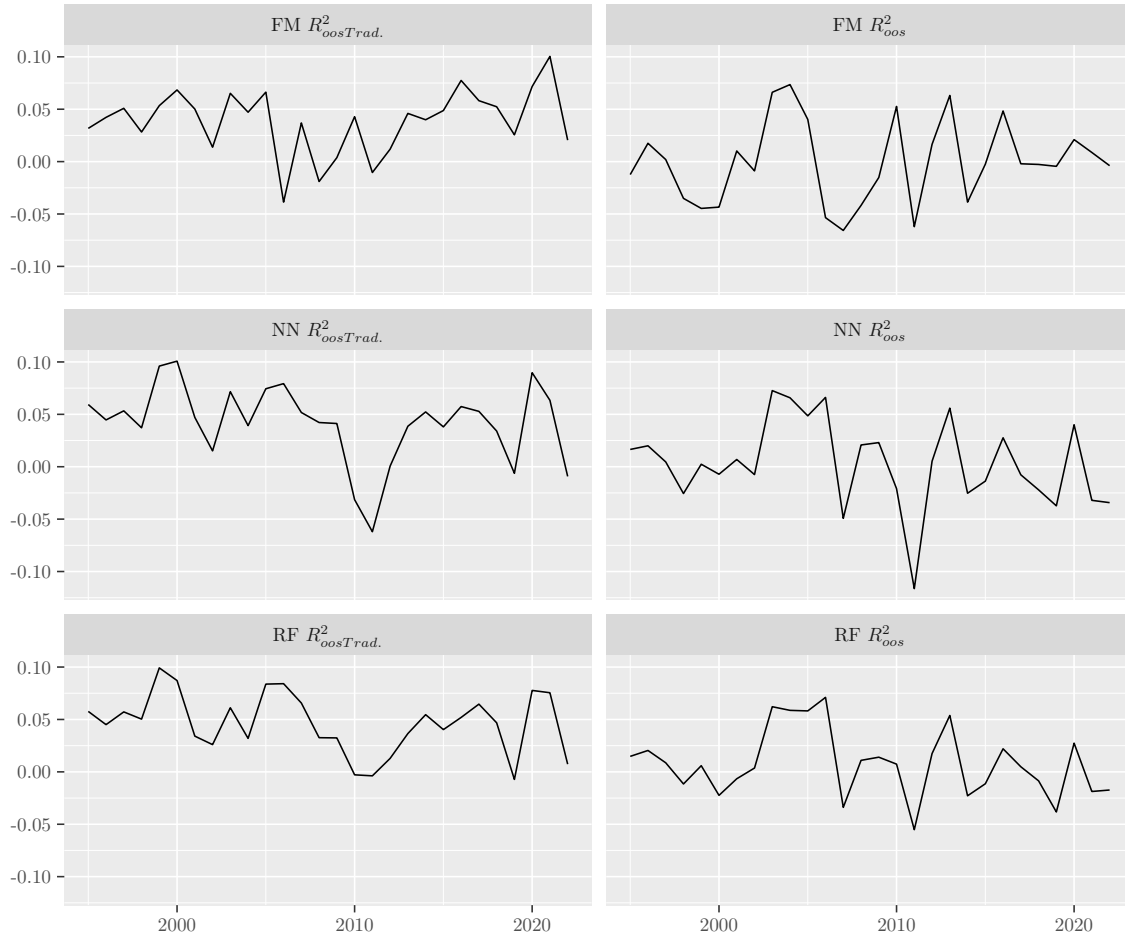
**Figure B.1: Factor autocorrelation**



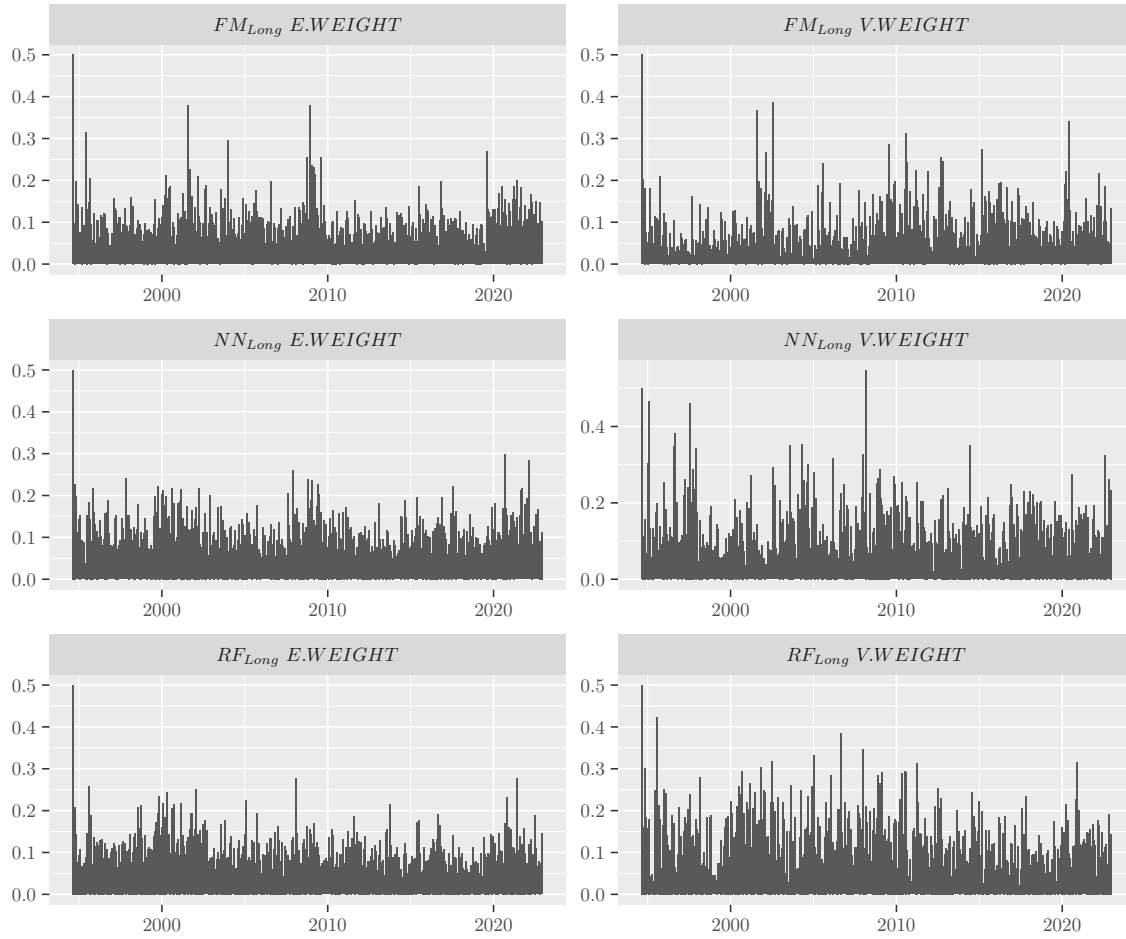
## C Additional Information

**Figure C.1: Time series of out-of-sample  $R^2$ s**

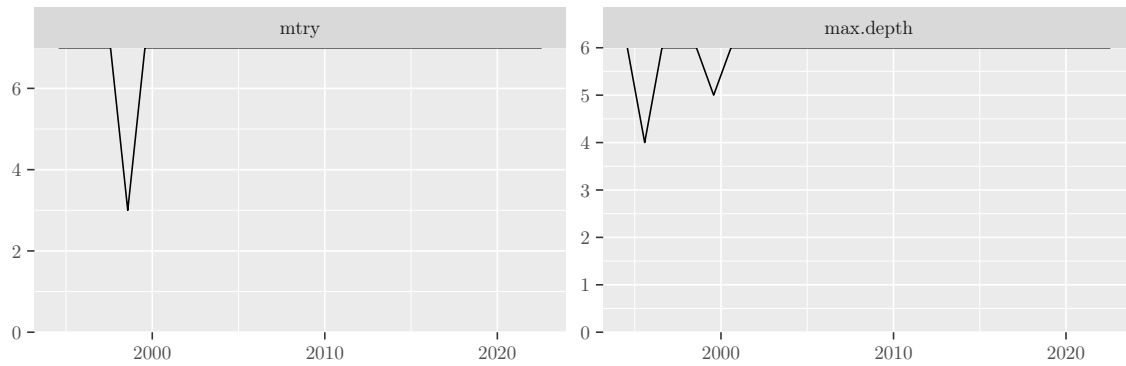
Figures present the out-of-sample predictive performance of different machine learning models. Left side graphs show the out-of-sample  $R^2$  values with benchmark prediction of zero. This method is described in section 4. Additionally, traditional out-of-sample  $R^2$ s are displayed. In traditional out-of-sample  $R^2$  benchmark prediction is the historical mean of corresponding stocks return.  $R^2$ s are calculated for each re-training period.



**Figure C.2: Turnover of highest expected return portfolios**



**Figure C.3: Random forest optimized hyper parameters**



## References

- Ali, A., Hwang, L.-S., & Trombley, M. A. (2003). Arbitrage risk and the book-to-market anomaly. *Journal of Financial Economics*, 69(2), 355-373. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0304405X03001168> doi: [https://doi.org/10.1016/S0304-405X\(03\)00116-8](https://doi.org/10.1016/S0304-405X(03)00116-8)
- ANG, A., HODRICK, R. J., XING, Y., & ZHANG, X. (2006). The cross-section of volatility and expected returns. *The Journal of Finance*, 61(1), 259-299. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2006.00836.x> doi: <https://doi.org/10.1111/j.1540-6261.2006.00836.x>
- Anwar, B. H., & Hogholm, K. (2020). The impact of illiquidity risk for the nordic markets. *Spanish Journal of Finance and Accounting / Revista Española de Financiación y Contabilidad*, 49(1), 28-47. Retrieved from <https://doi.org/10.1080/02102412.2018.1555348> doi: <https://doi.org/10.1080/02102412.2018.1555348>
- Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics*, 9(1), 3-18. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0304405X81900180> doi: [https://doi.org/10.1016/0304-405X\(81\)90018-0](https://doi.org/10.1016/0304-405X(81)90018-0)
- Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1), 57-82. Retrieved 2023-11-24, from <http://www.jstor.org/stable/2329556>
- Davydov, D., Tikkanen, J., & Äijö, J. (2017). Magic formula vs. traditional value investment strategies in the finnish stock market.. Retrieved from <https://api.semanticscholar.org/CorpusID:220593553>
- Drobetz, W., & Otto, T. (2021). Empirical asset pricing via machine learning: evidence from the european stock market. *Journal of Asset Management*, 22(7), 507-538. Retrieved from <https://doi.org/10.1057/s41260-021-00237-x> doi: [10.1057/s41260-021-00237-x](https://doi.org/10.1057/s41260-021-00237-x)
- Fama, E. F., & French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1), 3-56. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0304405X93900235> doi: [https://doi.org/10.1016/0304-405X\(93\)90023-5](https://doi.org/10.1016/0304-405X(93)90023-5)
- Fama, E. F., & French, K. R. (2012). Size, value, and momentum in international stock returns. *Journal of Financial Economics*, 105(3), 457-472. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0304405X12000931> doi: <https://doi.org/10.1016/j.jfineco.2012.05.011>
- Fama, E. F., & French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1), 1-22. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0304405X14002323> doi: <https://doi.org/10.1016/j.jfineco.2014.10.010>
- Fama, E. F., & MacBeth, J. D. (1973). Risk, return, and equilibrium: Empirical tests. *Journal of*

- Political Economy*, 81(3), 607–636. Retrieved 2023-10-29, from <http://www.jstor.org/stable/1831028>
- Fieberg, C., Metko, D., Poddig, T., & Loy, T. (2023). Machine learning techniques for cross-sectional equity returns’prediction. *OR Spectrum*, 45(1), 289–323. Retrieved from <https://doi.org/10.1007/s00291-022-00693-w> doi: 10.1007/s00291-022-00693-w
- George, T. J., & Hwang, C.-Y. (2004). The 52-week high and momentum investing. *The Journal of Finance*, 59(5), 2145–2176. Retrieved 2024-03-03, from <http://www.jstor.org/stable/3694820>
- Green, J., Hand, J. R. M., & Zhang, X. F. (2017, 03). The Characteristics that Provide Independent Information about Average U.S. Monthly Stock Returns. *The Review of Financial Studies*, 30(12), 4389–4436. Retrieved from <https://doi.org/10.1093/rfs/hhx019> doi: 10.1093/rfs/hhx019
- Grobys, K., & Huhta-Halkola, T. (2019). Combining value and momentum: evidence from the nordic equity market. *Applied Economics*, 51(26), 2872–2884. Retrieved from <https://doi.org/10.1080/00036846.2018.1558364> doi: 10.1080/00036846.2018.1558364
- Gu, S., Kelly, B., & Xiu, D. (2020, 02). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*, 33(5), 2223–2273. Retrieved from <https://doi.org/10.1093/rfs/hhaa009> doi: 10.1093/rfs/hhaa009
- Hanauer, M. X., & Kalsbach, T. (2023). Machine learning and the cross-section of emerging market stock returns. *Emerging Markets Review*, 55, 101022. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1566014123000274> doi: <https://doi.org/10.1016/j.ememar.2023.101022>
- Hanauer, M. X., & Windmüller, S. (2023). Enhanced momentum strategies. *Journal of Banking & Finance*, 148, 106712. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0378426622002928> doi: <https://doi.org/10.1016/j.jbankfin.2022.106712>
- Ince, O. S., & Porter, R. B. (2006). Individual equity return data from thomson datastream: Handle with care! *Journal of Financial Research*, 29(4), 463–479. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-6803.2006.00189.x> doi: <https://doi.org/10.1111/j.1475-6803.2006.00189.x>
- Jacobs, H., & Müller, S. (2020). Anomalies across the globe: Once public, no longer existent? *Journal of Financial Economics*, 135(1), 213–230. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0304405X19301618> doi: <https://doi.org/10.1016/j.jfineco.2019.06.004>
- Jokipii, A., & Vähämaa, S. (2006). The free cash flow anomaly revisited: Finnish evidence. *Journal of Business Finance & Accounting*, 33(7-8), 961–978.
- LAKONISHOK, J., SHLEIFER, A., & VISHNY, R. W. (1994). Contrarian investment, extrapolation, and risk. *The Journal of Finance*, 49(5), 1541–1578. Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.1994.tb04772.x> doi:



<https://doi.org/10.1111/j.1540-6261.1994.tb04772.x>

- Leivo, T. H., & Pätäri, E. J. (2011). Enhancement of value portfolio performance using momentum and the long-short strategy: The Finnish evidence. *Journal of Asset Management*, 11(6), 401–416. Retrieved from <https://doi.org/10.1057/jam.2009.38> doi: 10.1057/jam.2009.38
- Lewellen, J. (2015, June). The Cross-section of Expected Stock Returns. *Critical Finance Review*, 4(1), 1-44. Retrieved from <https://ideas.repec.org/a/now/jnlcfr/104.00000024.html> doi: 10.1561/104.00000024
- Moskowitz, T. J., & Grinblatt, M. (1999). Do industries explain momentum? *The Journal of Finance*, 54(4), 1249–1290. Retrieved 2024-03-03, from <http://www.jstor.org/stable/798005>
- Palazzo, B. (2012). Cash holdings, risk, and expected returns. *Journal of Financial Economics*, 104(1), 162-185. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0304405X11002856> doi: <https://doi.org/10.1016/j.jfineco.2011.12.009>
- Rosenberg, B., Reid, K., & Lanstein, R. (1985, Apr). Persuasive evidence of market inefficiency. *Journal of Portfolio Management*, 11, 9-16.
- Tobek, O., & Hronec, M. (2021). Does it pay to follow anomalies research? machine learning approach with international evidence. *Journal of Financial Markets*, 56, 100588. Retrieved from <https://www.sciencedirect.com/science/article/pii/S1386418120300574> doi: <https://doi.org/10.1016/j.finmar.2020.100588>