Tampere University

Janne Vikelä

ACCRUAL ANOMALY AND PREDICTING STOCK RETURNS WITH MACHINE LEARNING

Master's Thesis
Faculty of Information Technology and Communication Sciences
Examiners: Prof. Juho Kanniainen,
Prof. Moncef Gabbouj
October 2022

ABSTRACT

Janne Vikelä: Accrual Anomaly and Predicting Stock Returns with Machine Learning
Master's Thesis
Tampere University
Master's Programme in Information Technology
October 2022

This master's thesis investigates the accruals anomaly and if machine learning methods can be utilized to predict stock returns based on the anomaly. This thesis first verifies using a panel regression that the accrual anomaly exists in the US stock market sample from 1980-2020. Then, three machine learning methods are tested: A naive model, linear regression and a recurrent neural network.

Accrual anomaly is a a stock market anomaly where stocks with high accruals in their companies' income statements seem to deliver worse returns in the long run. The contrary also apply to company stocks with low accruals. The accrual anomaly was first presented by Sloan (1996). The anomaly is primarily explained with investor attention bias or possible earnings management.

This thesis tests three machine learning models to test if the accrual anomaly can be utilized in investing. In each model, the stocks are pooled into ten portfolios based on the forecast. Then, the realized equally weighted returns of each portfolio are calculated. Finally, the realized returns are regressed against the Carhart factors to see if a positive alpha can be obtained.

The first machine learning model tested is a naive model that pools stocks based on their scaled accruals into ten portfolios. Then, the portfolio returns are examined in the three following years. It is shown that the model can deliver a positive long-short alpha between the low and high accrual deciles.

The second machine learning model is a linear regression that uses the three previous years scaled accruals as explanatory variables. Additionally, controls for the logarithm of the market capitalization and the book-to-market ratio are used as controls. The model is estimated for each year with previous three years of return data. The linear model delivers similar performance compared to the naive model.

The third machine learning model is a recurrent neural network (RNN) using a three-year time series of explanatory variables. The main explanatory variables are the scaled accruals with the logarithm of market capitalization, the book-to-market ratio and the stock's own return acting as controls. The training data includes three previous years' return time series and is estimated for each year. The RNN model also delivers some moderate performance, however the performance is weaker than with the first two.

Keywords: Accrual anomaly, Recurrent neural network, Linear models, Stock market predictions, Stock market anomalies

---

TIIVISTELMÄ

Janne Vikelä: Jaksotusanomalia ja osakkeiden tuottojen ennustaminen koneoppimisen avulla
Diplomityö
Tampereen yliopisto
Tietotekniikan DI-ohjelma
Lokakuu 2022

Tämä diplomityö tutkii koneoppimismallien käyttöä osakkeiden tuottojen ennustamiseen hyödyntäen jaksotusanomaliaa. Tämä diplomityö ensiksi vahvistaa käyttäen paneeliregressiota, että jaksotusamonalia on läsnä Yhdysvaltain osakemarkkinoilla vuosien 1980-2020 otoksessa. Sitten tämä työ testaa kolmea koneoppimismallia: Naiivia mallia, lineaarista regressiota ja takaisinkytkettyä neuroverkkoa.

Jaksotusanomalia on osakemarkkinoiden anomalia, jossa osakkeet, joiden yritysten tuloslaskelmassa on suuri jaksotuskomponentti, tuottavat pitkässä juoksussa huonommin. Vastakkainen tilanne pätee yritysten osakkeille, joilla on matalat jaksotuskirjaukset. Jaksotusanomalian löysi Sloan (1996). Anomaliaa selitetään pääsääntöisesti sijoittajien huomioon liittyvällä harhalla ja tulosmanipulaatiolla.

Tämä diplomityö testaa kolmella koneoppimismallilla, että pystyykö jaksotusanomaliaa hyödyntämään sijoittamisessa. Jokaisen mallin kanssa osakkeet kootaan kymmeneen portfolioon niiden tuottoennusteen perusteella. Sitten, jokaiselle portfoliolle lasketaan keskimääräiset toteutuneet tuotot. Lopuksi keskituotoista estimoidaan lineaarinen regressio Carhart-faktoreita vasten ja tutkitaan, että saavutetaanko positiivinen alfa.

Ensimmäinen testattu koneoppimismalli on naiivi malli, jossa osakkeet jaotellaan skaalattujen jaksotuskirjausten avulla kymmeneen portfolioon. Sitten portfolioiden tuottoja tarkastellaan kolmena seuraavana vuotena. Näyttää, että malli pystyy tuottamaan positiivisen alfan matalien ja korkeiden jaksotuskirjausten desiilien välille.

Toinen koneoppimismalli on lineaarinen regressio, joka käyttää kolmen edeltävän vuoden skaalattuja jaksotuskirjauksia selittävinä muuttujina. Lisäksi kontrolleina käytetään markkina-arvon logaritmia ja kirjanpito- ja markkina-arvojen suhdelukua. Estimointi tehdään joka vuodelle käyttäen kolmen edellisen vuoden tuottodataa. Lineaarinen malli tuottaa samanlaisia tuloksia kuin naiivi malli.

Kolmas koneoppimismalli on on takaisinkytketty neuroverkko (RNN), joka käyttää kolmen edeltävän vuoden selittävien muuttujien aikasarjaa. Päämuuttujat ovat skaalatut jaksotuskirjaukset, joiden lisäksi kontrollit ovat markkina-arvon logaritmi, kirjanpito- ja markkina-arvojen suhdeluku sekä osakkeen oma aiempi tuotto. Opetusdata koostuu kolmen aiemman vuoden tuottojen aikasarjasta ja malli estimoidaan joka vuodelle. RNN-malli myös tuottaa kohtalaisia tuloksia, mutta kuitenkin tulokset ovat heikompia kuin kahdella edellä mainitulla mallilla.

Avainsanat: Jaksotusanomalia, Takaisinkytketty neuroverkko, Lineaariset mallit, Osakkeiden tuotot, Osakemarkkinoiden ennustaminen, Osakemarkkinoiden anomaliat

---

PREFACE

According to Vikelä (2021b), four years is a short time. On the other hand, according to Vikelä (2021a), eight years is a long time. Thus logically, we can only conclude that nine years is just a longer time. Or is it?

Well, honestly speaking nine years ago could not have imagined that I would complete my studies somewhere else than I planned, not to mentioned that would have tripled the number of master's degrees from the planned number. That just happened. Even though after coding my first lines of Scala, I once promised I would never write a single line of code again. That is what I thought. That is what they always say.

Well, the gateway opened after all. I blame R. How dare they start indexing at 1?

Vikelä (2021b) calls completing a second master's thesis a side project. However, if the second one was a side project, then what is a third one? Maybe it is a hobby or just another unhealthy way of passing time during the COVID-19 pandemic. Or could it be something even worse - a sign of addiction?

To be fair, master's degrees are like beer. After completing the second one, you need a third. It seems that all the signs of studyholism are in the air. Indeed yes. It is evident that my blood's master's degree concentration is soon to exceed any legal studying limit. Thus, I will have to face all the consequences.

The justice in this case is swift an merciless. My studying rights will soon be revoked and I will have to enter a post-study rehabilitation by getting a job. They say that substitutes for student life exist. Let us hope so.

I will walk before they make me run.

In Hämeenkyrö, 28th October 2022

Janne Vikelä

P.S. I would like to thank Juho and Moncef for instructing this thesis.

CONTENTS

1. INTRODUCTION

This master's thesis examines accrual anomaly and utilizing machine learning methods to predict stock returns with machine learning. Accrual anomaly was first presented by Richard Sloan (1996) in his article *Do stock prices fully reflect information in accruals and cash flows about future earnings?*. This thesis examines the accrual anomaly in the years 1980-2020 in the US stock market.

This thesis first verifies the existence of the phenomenon in the US market. Second, this thesis applies three machine learning methods to forecast the stock returns based on the accruals. The methods used are a naive model, linear regression and recurrent neural network. Then, the performance of the models is examined by investigating the generated returns and the Carhart alphas of the forecast returns.

This introduction first presents the accrual anomaly. Then, the stock return forecast methods are briefly introduced. Finally, the thesis structure is presented.

1.1 Accrual anomaly

The concept of accrual anomaly is the following. Accounting is accrual based. This means that the revenue and cost component in income statements do not represent actual cash flows but instead they represent accruals. In the case of revenues, the revenues are first written into receivables at the time the actual transaction is made and the actual payment is only received afterwards. Then, in the case of expenses, the expenses are first written into account payables and the payment is only made afterwards. The accrual accounting practice thus does not always give the correct picture on the actual cash flows of the company.

According to Sloan (1996), the investors may be mislead by the accruals. For example, an investor may look into the accruals and ignore the actual cash flows. Such case would be a temporary increase or decrease in profit that is driven by accruals. This kind of accrual driven profit change can become priced in the company stock prices temporarily.

In practice, an accrual driven profit increase may raise share price temporarily or an accrual driven profit decrease may drop the share price temporarily. However, as the accruals tend to revert in the long run, the accrual driven stock price change is expected

to revert also. Sloan (1996) shows that companies with a high accrual component in their income statement seem to deliver weaker stock return in the three following years. Also, the contrary applies as the companies with low accrual components seem to outperform the stock returns of the companies with high accruals in the three following years.

The literature generally acknowledges two factors contributing to the accrual anomaly. These factors are investor attention and earnings management.

The first factor or investor attention is that the investors may simply focus too much on the income statement instead of realized cash flows. Thus, the investors may overprice a temporary accrual driven profit increase or underprice a temporary accrual driven profit increase. This view is supported by Sloan (1996).

The second factor is earnings management. Dechow et al. (1995) show that company management may use the discretionary accruals to manipulate earnings. For example, the managers could use aggressive accounting to recognize revenue faster an the income side. Then, on the expense side, the managers could make expenses look lower by recognizing cost slower or decreasing the depreciation rate of property, plant and equipment.

1.2 Machine learning methods used to predict stock returns

This thesis tests three machine learning methods to forecast stock returns based on the accrual anomaly. The three methods are a naive model, linear regression and recurrent neural network. The goal is to test and compare which kind of methods can be utilized to predict stock returns utilizing accrual anomaly and how well they perform.

The first model is a naive model and is repeated from Sloan (1996). Sloan (1996) calculates the scaled accruals each year for each company and sorts the companies into ten portfolios each year based on the scaled accruals. Then, the equally weighted portfolio returns are examined in the three subsequent years. It is shown that the stocks with the lowest accruals outperform the stocks with high accruals in the long run.

The second model is linear regression. The regression is used to predict the stock returns using three explanatory variable sets. The first explanatory variable set consists of scaled accruals from the three preceding years. The second explanatory variable set uses control variables, which are used to test if other variables in addition to the scaled accruals drive the returns. The controls used include the logarithm of market capitalization and the book-to-market ratio. The third explanatory variable set includes both scaled accruals and control variables. The regression model is estimated for each year using three previous years of return data.

The third model is recurrent neural network (RNN). The RNN is tested as the RNN is able to capture time series properties such as trend. The RNN is tested with three explanatory

variable sets. The variable sets are the scaled accruals, control variables and scaled accruals together with the control variables. Each variable is constructed of a three-year time series of the particular explanatory variable. The RNN is implemented using Keras and Tensorflow libraries. As with the regression model, the RNN is estimated for each year using three previous years of return data to train the neural network.

Each forecast method is tested by sorting the stocks each year based on the forecasts. First, the naive model sorts the stock each year based on the scaled accruals. Second, the linear regression and third, the RNN are used to predict returns for each stock with a sufficient available explanatory variable time series.

When testing the models, each year the stocks are sorted into ten portfolios based on the forecast returns. The realized returns are calculated for each portfolio with equal weighting and the return performance is assessed. First, the annual mean returns are calculated. Then, the portfolio excess returns against the risk-free benchmark returns are regressed against the Carhart factors and the Carhart alpha is used to assess the risk-adjusted performance.

1.3 This thesis structure

This thesis first presents the research questions in Chapter 2. Second, literature related to the accrual anomaly along with other literature related to this thesis is reviewed in Chapter 3. Then, the thesis methodology and required data are presented in Chapter 4 and Chapter 5, respectively. The results are in Chapter 6. Finally, the conclusions are presented in Chapter 7.

2. RESEARCH QUESTIONS

This chapter presents the two research questions of this thesis. The first research question verifies the existence of the accrual anomaly in the sample. The second research question tests the machine learning methods utilized in predicting the stock returns utilizing the accrual anomaly.

2.1 Existence of the accrual anomaly

This thesis first verifies the existence of the accrual anomaly in the US stock market in the years 1980-2020. The original Sloan (1996) research focused on the years 1961-1991. So, the time period of this thesis is only partially overlapping with the original Sloan (1996) research. It is possible that market dynamics has changed between the Sloan (1996) paper and this thesis. It also may be that the changed reporting regulations e.g., the Sarbanes-Oxley act of 2002. This possibility is supported by Chambers and Payne (2008) who find an inverse relation between accrual anomaly and audit quality and attribute the Sarbanes-Oxley act for this finding. Also, the activity of market participants may have changed the way how the stock prices incorporate earnings information. For example, Ali, Chen et al. (2008) and Green et al. (2011) note that some funds are able to capitalize the anomaly in investing. However, it is known that the ability to capitalize stock market anomalies my also lead into arbitraging them away.

To make this thesis meaningful, it is vital to verify the existence of the anomaly before the machine learning methods can be utilized in predicting the returns. In literature, at least Mashruwala et al. (2006) and Hirshleifer et al. (2006) have documented that the anomaly exists in more recent samples after Sloan (1996).

The first research question is:

1. Does the accrual anomaly exist in the US stock market sample between the years 1980-2020?

This thesis uses a panel regression to verify the existence of the accrual anomaly in the 1980-2020 sample. The dependent variable is the annual stock return of an individual company and the explanatory variables are the accruals of the same year and the three preceding years. Also, controls include the company's logarithm of the market capitaliza-

tion and the book-to-market ratio. Moreover, fixed effects for year and stock are used to control for the omitted variable bias.

2.2 Predicting stock returns with machine learning

This thesis tests if machine learning methods can be used to predict stock returns utilizing the accrual anomaly. The questions of interest are if the dependencies between returns and accruals are linear or nonlinear or if some relation to time series properties such as trend exist. No much literature exists that has tested applying machine learning methods for forecastbing stock returns based on accrual anomaly.

The second research question is:

2. Can machine learning models be used to predict stock returns utilizing the accrual anomaly?

The machine learning methods tested include three methods and they are: 1) A naive model classifying stock by their scaled accruals, 2) A linear regression model predicting returns based on the scaled accruals and controls and 3) A recurrent neural network predicting stock returns based on a time series of scaled accruals and controls.

3. LITERATURE REVIEW

This chapter first presents literature related to the accrual anomaly. Then, literature related to the machine learning models is presented.

3.1 Literature on the accrual anomaly

The accrual anomaly is a stock market anomaly where stocks with high accruals in their income statement seem to deliver worse returns in the coming years and vice versa. The anomaly was first presented by Sloan (1996). The idea is that income statements are based on accrual accounting instead of actual cash flows. So, when revenue is recognized, the revenue is first recorded as receivable and the cash flow is realized later. Also, expenses are first recorded as a payable and only paid later. Similarly, the depreciation of property, plant and equipment is a discretionary accrual and not an actual cash flow.

The Sloan (1996) claim is that investors may focus too much on the accrual component and ignore the actual cash flows, which should be more important to valuation. Thus, the investors may overprice accrual driven profits and also underprice net result that is pushed down by accruals. However, the accruals tend to revert in the long run. This causes the stocks with high acccruals to perform worse in the long run compared to stock with low accruals. Also, Chan et al. (2006) shows a similar result to Sloan (1996), where the quality of earnings in terms of low accruals entails better future returns.

Two main reasons to explain the accrual anomaly exist. The first one is investor attention, i.e. the investors fixate too much on the accrual component of earnings, mispricing the stocks. The second explanation is earnings management, according to which managers use discretionary accruals to manipulate earnings and misleading investors. Also, risk based explanations have been provided, but they have not gained support in literature.

3.1.1 Investor attention based explanations

Sloan (1996) represents the investor attention based explanations. The Sloan (1996) conclusion is that the investors fixate too much on the accrual component of income statement and ignore the actual cash flow impact. Thus, as accruals revert in the long run, the high accruals predict low future returns and vice versa.

Hirshleifer et al. (2006) examine if the accrual anomaly is risk-based or simply an investor attention based misvaluation. They do not find evidence on the risk premium and conclude that the most likely cause is investors' misvaluation of stocks.

Bradshaw et al. (2001) examine how independent auditors and equity analysts examine accruals. Bradshaw et al. (2001) find that the auditors do not seem to signal the problems associated with the accrual component of earnings. Also, they note that the equity analysts do not seem to take into account the accrual effect in their forecasts. This also provides evidence on the investor attention based explanations.

Ali and Gurun (2009) claim that during high investor sentiment times, the small cap stocks on average seem to exhibit stronger accrual anomaly. They attribute the effect to two reasons. First, the investors may have limited attention. Second, the managers may consciously drive the stocks' overvaluation by using discretionary accruals during high-sentiment times.

3.1.2 Earnings management based explanations

Dechow et al. (1995) provide evidence that discretionary accruals are used for earnings management. Also, Dechow et al. (1995) notes that the discretionary accruals used in earnings management revert in the long run.

Pincus et al. (2007) examine the accrual anomaly internationally and find that the accrual anomaly is the highest in the countries allowing the extensive use of accruals in accounting. According to Pincus et al. (2007), the main reasons for the accrual anomaly are earnings management and also the barriers to arbitrage the anomaly away.

3.1.3 Decomposition of accrual components and effect on accrual anomaly

Lewellen and Resutek (2016) break the accrual anomaly into working capital accruals, investment accruals and non-transaction accruals. The non-transaction accruals include depreciation, deferred taxes and extraordinary items. Thus, the non-transaction accruals are the most discretionary component of accruals. Therefore it also may be the strongest driver of the accruals anomaly if the investors fail to price this component of accruals correctly. Also, Lewellen and Resutek (2016) claim that investment does not seem to relate to the accrual anomaly.

Detzel et al. (2018) also find that accrual anomaly can be broken into two components. They are investment accruals and non-investment accruals. Detzel et al. (2018) claim that exposure to investment accruals is a risk factor that cannot be arbitraged away. However, they claim that the non-investment accruals are driving the mispricing phenomenon.

Zhang (2007) investigate the relationship between accrual anomaly and investment. He notices that the accruals co-vary with various measures of corporate investment and growth. He concludes that at the same time the accruals and measures of investment and growth seem to predict worse one year ahead returns. However, in the long run, these investment related measures seem to predict better long run returns. Thus, Zhang (2007) conclusion is that accrual anomaly may in fact be explained by investment and corporate growth.

Ball et al. (2016) compare the accruals' and operating profit's predictive power for stock returns. They first verify a similar pattern to Sloan (1996) where low accruals predict better future returns and vice versa. However, they note that operating profit has even stronger predictive power than the accruals. The higher the operating profitability is, the better the future returns are and vice versa.

3.1.4 Investors and accrual anomaly

Ali, Chen et al. (2008) provide evidence that some mutual funds are able to capitalize the accrual anomaly and seem to focus on stock with low accruals. Ali, Chen et al. (2008) also show that these funds tend to be smaller, less diversified and have higher volatility than their peers.

Mashruwala et al. (2006) find that the the accrual anomaly is difficult to arbitrage away. According to Mashruwala et al. (2006), that the stocks where the accrual anomaly is present are often low-price and low-volume stocks with high idiosyncratic volatility. Idiosyncratic volatility is the component of the volatility that cannot be explained by systemic risk. These factors may inhibit arbitrageurs from utilizing these stock for capitalizing the accrual anomaly.

Green et al. (2011) claim that the hedge fund are utilizing the accrual anomaly. Also, they claim that the accrual anomaly is being arbitraged away by hedge funds and also the changed regulation such as the Sarbanes-Oxley act of 2002 controlling the earnings disclosures may inhibit the accrual anomaly.

3.1.5 Accrual anomaly internationally

According to LaFond (2005), the accrual anomaly is present in 17 countries in the years 1989-2003. Also, he notes that managerial discretion of accruals is a driving factor of the anomaly. Moreover, equity analyst following seems to mitigate the accrual anomaly. Additionally, concentrated ownership also contributes to the accrual anomaly.

The results in LaFond (2005) are very similar to Pincus et al. (2007). Pincus et al. (2007) claims that in the countries that allow more accrual accounting, the anomaly is

also stronger.

3.2 Literature on machine learning methods

This section resents an overview of machine learning methods. First, simple linear regression presented. Then, an overview of neural networks is presented including explanation of feed-forward and recurrent neural networks.

3.2.1 Linear regression

The linear regression method is a method of fitting a linear function into a data set. The equations of a multivariate linear regression generally follow the equations:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{p1} + \epsilon_i, \tag{3.1}$$

where the parameters are:

- y_i The y values
- β_i Coefficient for term i
- x_{ji} The explanatory variable j
- ϵ_i The difference between y and estimate

Also, a matrix form can be used:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}. \tag{3.2}$$

According to Stigler (1981), the linear regression and the least squares method is credited to Adrien-Marie Legendre (1805) and Carl Friedrich Gauss (1809). Legendre and Gauss invented a method of fitting the regression called the least squares method. The idea is to fit and estimated the parameters by minimizing the sum of squared errors ϵ_i. The SSE is calculated in matrix form as:

$$\begin{aligned} SSE &= \boldsymbol{\epsilon}^T \boldsymbol{\epsilon} \\ SSE &= (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^T (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \\ SSE &= \left(\boldsymbol{Y}^T\boldsymbol{Y} - \boldsymbol{Y}^T\boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{Y} + \boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta}\right). \end{aligned}$$

Then, a derivative against $\boldsymbol{\beta}$ is taken to minimize the sum of squared errors:

$$\frac{\partial SSE}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X}^T\boldsymbol{Y} + 2\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta}.$$

The derivative is set to zero at minimum:

$$\begin{aligned} -2\boldsymbol{X}^T\boldsymbol{Y} + 2\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta} &= 0 \\ \boldsymbol{X}^T\boldsymbol{Y} &= \boldsymbol{X}^T\boldsymbol{X}\boldsymbol{\beta}. \end{aligned}$$

Then, the optimum coefficient matrix $\boldsymbol{\beta}$ is solved:

$$\boldsymbol{\beta} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{Y}.$$

The estimated coefficients $\boldsymbol{\beta}$ can be used to predict y values based on the explanatory variables.

3.2.2 Artificial neural networks

This section briefly covers the concept of artificial neural networks. First, the very foundation of neural networks or the perceptron is introduced. Moreover, activation functions for the neural networks are presented. Then, the "vanilla" feed forward neural networks are introduced. Finally, the recurrent neural networks are introduced.

The perceptron

Often, the perceptron introduced by Frank Rosenblatt, is considered a starting point of neural networks and deep learning. Rosenblatt (1958) and Rosenblatt (1961) introduce a perceptron that can be perceived as a simple binary classification with a mathematical formulation:

$$f(x) = \begin{cases} 1 & \text{if } \mathbf{w} \cdot \mathbf{x} + b > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{3.3}$$

where:

- The **x** denotes the input as vector

- The **w** denotes the weights as vectors
- The b is the bias
- **w** $\cdot$ **x** is a dot product between the input and the weights.

The formulation has been adapted from the original Rosenblatt (1961) book for simplification.

Intuitively, the concept of perceptron means that a perceptron classifies an input if a certain condition is met. However, the Rosenblatt formulation of the perceptron has some limitations to its applications as for example the perceptron is not differentiable function. Thus, different alternative for approximation must be used. These are called activation functions.

Activation functions

Activation functions can be perceived as approximations of the perceptron and generally they are functions that pass and modify the input signals and forward them. Here, different kinds of activation functions are presented. Graphs of these function are shown in Figure 3.1.

Heaviside function

The Heaviside step function of a simple for of activation function. Intuitively, it works as a binary classifier that activates if a certain threshold is met. The Heaviside function is defined as:

$$\mathcal{H}(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0. \end{cases} \tag{3.4}$$

As is the case with the perceptron, the Heaviside function is not differentiable. Thus, often approximations or alternatives are preferred over the Heaviside step function.

Sigmoid function

Sigmoid or logistic function is a function that is commonly used to approximate the Heaviside step function. The sigmoid activation is defined as:

$$f(x) = \frac{1}{1 + e^{-x}}. \tag{3.5}$$

The output of the sigmoid function is an S curve. The output is also restricted between values 0 and 1. Thus, the sigmoid is suitable for modeling probabilities or classification. Also, the advantages compared to the Heaviside function include the differentiability of the sigmoid.

Hyperbolic tangent function

Hyperbolic tangent is a substitute for the sigmoid activation function. The difference is that the hyperbolic tangent maps the input values on -1 to 1 interval. The hyperbolic tangent is defined as:

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \tag{3.6}$$

Linear activation function

Linear activation function is an activation function that simply passes the input forward. The linear activation function is defined as:

$$f(x) = x. \tag{3.7}$$

Rectified Linear Unit (ReLU) activation function

The Rectified Linear Unit (ReLU) is an activation function that passes forward the positive values. The ReLU activation function is defined as:

$$f(x) = max(0, x). \tag{3.8}$$

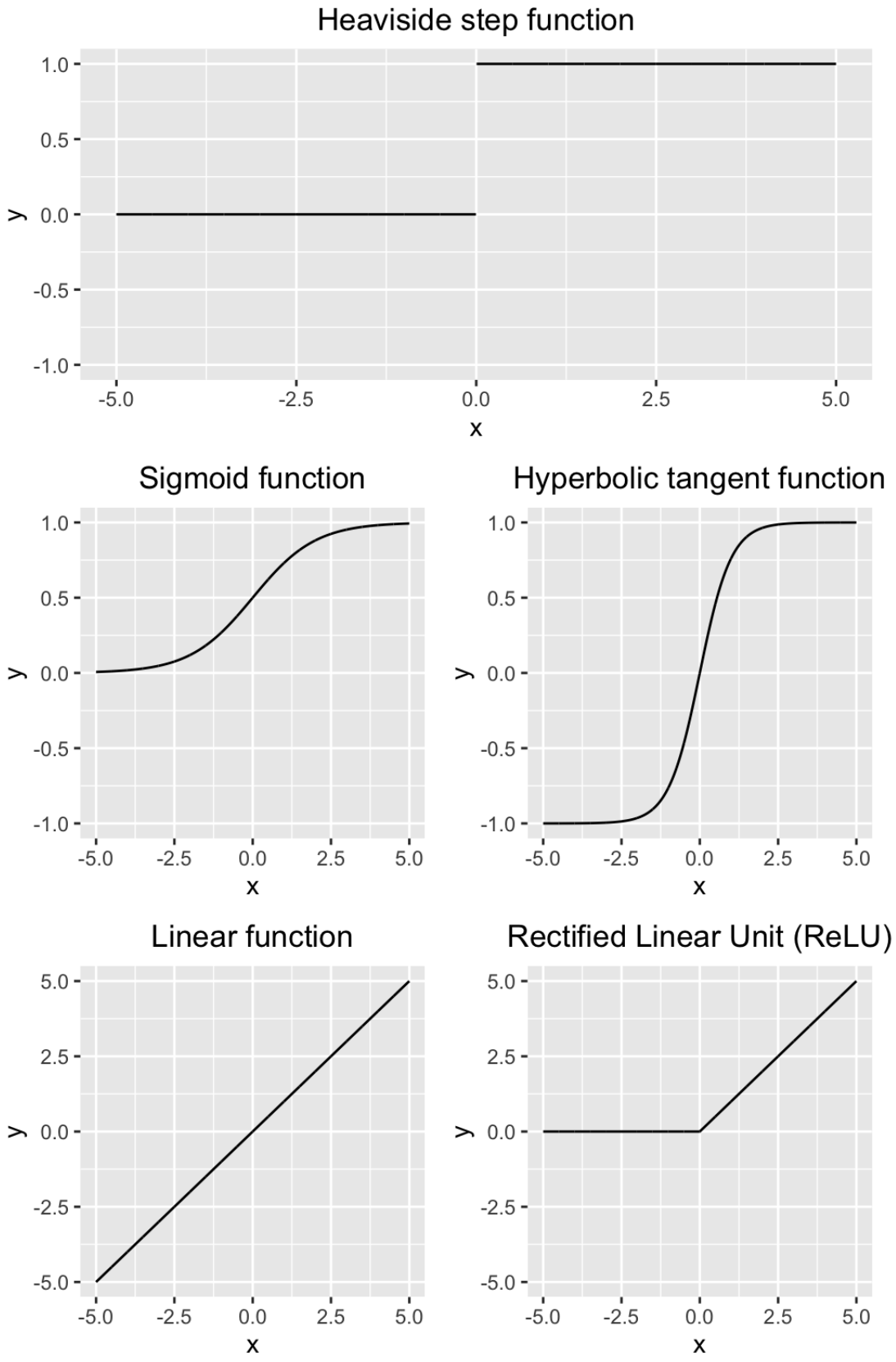

Figure 3.1. *Activation function examples: Heaviside step function, Sigmoid function, Hyperbolic tangent, Linear function and Rectified Linear Unit (ReLU)*

3.2.3 Feed-forward neural networks

For example, Bishop (1995), provides an overview of the feed-forward neural networks. Concept of feed-forward neural network is the following. Inputs are take from the input layer and then they are fed into the hidden layer. Figure 3.2 shows an illustration of a fully connected feed-forward neural network.

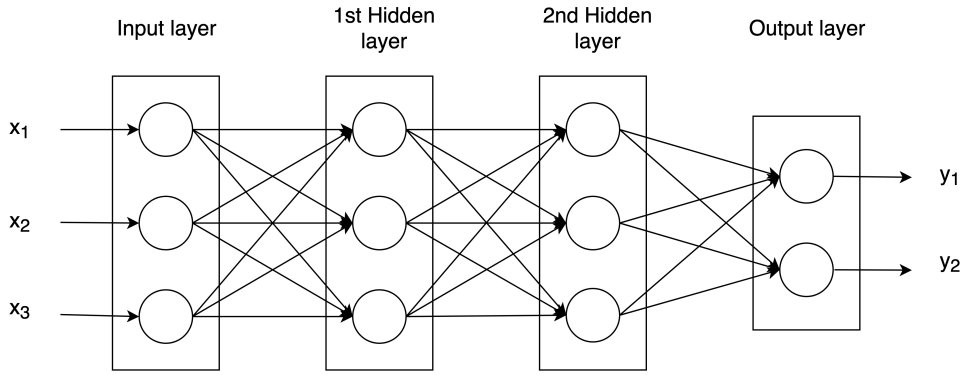

Figure 3.2. *A simple example of a fully connected feed-forward neural network with three input variables, two hidden layers and two output variables.*

In the hidden layer cells, weights w_i are assigned on the inputs x_i according to the formula:

$$a = \sum w_i * x_i + b, \tag{3.9}$$

and a weighted linear sum a is formed. The bias b may or may not be included. Then, the weighted linear sum a is passed to the activation function $g(a)$ which forms the output y_1:

$$y_1 = g(a). \tag{3.10}$$

The output y_1 obtained can be either returned or passed to another hidden layer. The neural network cell is presented in Figure 3.3.

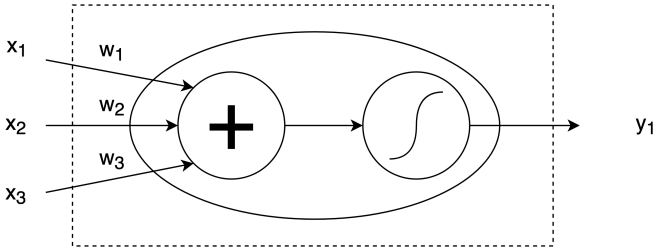

Figure 3.3. *An illustration of a feed-forward neural network cell. The input variables form a weighted sum and an activation function is used to forward the output.*

3.2.4 Recurrent neural networks

A pitfall with feed-forward network is that they are unable to handle time series properties. Thus, recurrent neural networks (RNN) with a state were introduced. For example, Goodfellow et al. (2016) present the RNN. The RNN is presented as figure in Figure 3.4.

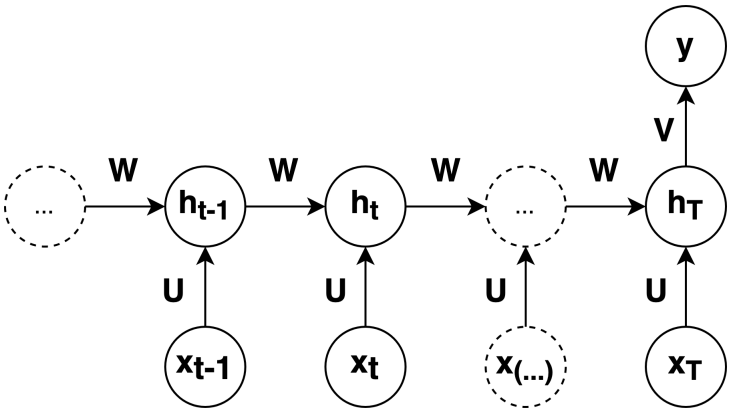

Figure 3.4. *An illustration of an unfolded many-to-one recurrent neural network. Variables* $\boldsymbol{x_t}$ *represent input,* $\boldsymbol{h_t}$ *the state of the neural network and* **U***,* **W** *and* **V** *are weight matrices.*

The RNN consists of equations:

$$\mathbf{a}_t = \mathbf{b} + \mathbf{W}\mathbf{h}_{t-1} + \mathbf{U}\mathbf{x}_t \tag{3.11}$$

$$\mathbf{h}_t = tanh(\mathbf{a}_t) \tag{3.12}$$

$$\mathbf{y}_t = \mathbf{c} + \mathbf{V}\mathbf{h}_t, \tag{3.13}$$

where:

- **U**, **W** and **V** are weight matrices
- $\mathbf{h}_t$ is the state
- Function $tanh$ is the hyperbolic tangent
- **b** and **c** are the biases.

3.2.5 Training methods

Training of neural networks is based on the minimization of the loss function that is the difference between the training values and the modeled output of the neural network. First, the loss functions are presented and then optimization methods used in training neural networks are presented.

Loss functions

Generally, the neural networks are trained using a loss function. Here, the loss function is presented from the viewpoint of estimation theory. The concept of loss function is that basically the loss function is an aggregated difference between estimated model outputs and true values. For example, Aggarwal (2018) gives an overview of loss functions.

In regression type problems, the loss function choice is usually the sum of squared error or mean squared error. Here, the sum of squared errors (SSE) loss function is presented:

$$SSE = \sum (y - \hat{y})^2 , \tag{3.14}$$

where the y is actual value and $\hat{y}$ is the modeled output. The difference to mean squared error (MSE) is that the MSE averages the SSE.

The minimization of these loss function is usually performed using numerical optimization methods as the models tend to be too difficult for analytical solutions. The most popular choice of optimization methods is the gradient descent.

The concept of gradient descent is intuitively simple. The idea is that the gradient shows the direction in which the function decreases the most. In each iteration step, the algorithm should converge towards the optimum. The gradient descent method along with some of its variants is presented below.

Gradient descent methods

The gradient descent methods are first order numerical optimization methods that are used to find a local minimum of a function. The gradient descent method family is very popular in training neural networks by optimizing the loss functions.

The gradient of a function is defined as a vector of the function's partial derivatives and is defined as:

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} . \tag{3.15}$$

The gradient descent method was first proposed by Cauchy (1847). The method is pre-

sented in Algorithm 1.

Algorithm 1: Gradient descent algorithm

Parameter: Learning rate α

Parameter: Initial weights $\mathbf{w}_0$

$i \leftarrow 0$;

while *Stopping criterion not met* **do**

Compute gradient: $\hat{\mathbf{g}} \leftarrow \nabla_w L\left(f(\mathbf{x}, \mathbf{w}), \mathbf{y}\right)$;

Update weights: $\mathbf{w}_i \leftarrow \mathbf{w}_{i-1} - \alpha * \hat{\mathbf{g}}$;

$i = i + 1$;

The idea of the gradient descent is that on every update, the function argument is updated towards the minimizing argument by a set step size towards the negative gradient direction, i.e. the direction of the greatest descent.

Stochastic gradient descent

The stochastic gradient descent (SGD) variant is a popular choice for machine learning applications. The difference between the gradient descent and the stochastic version is that the stochastic version samples a set minibatch from the observations to calculate the gradient. This practice relieves the computational burden compared to a full batch. Of course, this may be at the expense of the convergence speed. For example, Goodfellow et al. (2016) explain the stochastic gradient descent algorithm. The algorithm is described in Algorithm 2.

Algorithm 2: Stochastic gradient descent algorithm (SGD)

Parameter: Learning rate α

Parameter: Initial weights $\mathbf{w}_0$

$i \leftarrow 0$;

while *Stopping criterion not met* **do**

Sample a minibatch of *m* examples from the training set $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}\}$ with corresponding targets $\mathbf{y}^{(i)}$;

Compute gradient estimate: $\hat{\mathbf{g}} \leftarrow \frac{1}{m} \nabla_w \sum_i L\left(f(\mathbf{x}^{(i)}, \mathbf{w}), \mathbf{y}^{(i)}\right)$;

Update weights: $\mathbf{w}_i \leftarrow \mathbf{w}_{i-1} - \alpha * \hat{\mathbf{g}}$;

$i = i + 1$;

Stochastic gradient descent with momentum

The stochastic gradient descent algorithm is often augmented with momentum. The concept of momentum is intuitively a moving averaging of gradients. Often, an exponentially decaying weighting is used over previous gradients. The averaging of gradients is helpful when dealing with functions with noisy or small gradients and also with high curvature

functions.

The concept of momentum was introduced by Polyak (1964). The momentum stochastic gradient method is generally defined, according to Goodfellow et al. (2016), is Algorithm 3.

Algorithm 3: Stochastic gradient descent algorithm with momentum

Parameter: Learning rate α, momentum parameter β

Parameter: Initial weights $\mathbf{w}_0$, initial velocity $\mathbf{v}$

$i \leftarrow 0$;

while *Stopping criterion not met* **do**

Sample a minibatch of *m* examples from the training set $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}\}$ with corresponding targets $\mathbf{y}^{(i)}$;

Compute gradient estimate: $\hat{\mathbf{g}} \leftarrow \frac{1}{m} \nabla_w \sum_i L\left(f(\mathbf{x}^{(i)}, \mathbf{w}), \mathbf{y}^{(i)}\right)$;

Compute velocity update: $\mathbf{v} \leftarrow \beta * \mathbf{v} - \alpha * \hat{\mathbf{g}}$;

Update weights: $\mathbf{w}_i \leftarrow \mathbf{w}_{i-1} + \mathbf{v}$;

$i = i + 1$;

Adaptive Moment Estimation (Adam)

One of the most popular stochastic gradient descent methods used in deep learning is the adaptive moment estimation or Adam presented by Kingma and Ba (2014). The method uses first and second moment estimates which are corrected for bias. According to Kingma and Ba (2014), the Adam method performs well against other similar methods in terms of convergence and is also suitable for operations on large and multidimensional data sets. The Adam algorithm is presented in Algorithm 4.

Algorithm 4: Adaptive moment estimation - Adam

Parameter: Learning rate α

Parameter: Exponential decay rates for moment estimates

Parameter: Small constant δ used for numerical stabilization (Suggested default: 10^{-8})

Parameter: Exponential decay rates for the moment estimates: $\beta_1, \beta_2 \in [0, 1)$

Parameter: Initial weights $\mathbf{w}_0$

Parameter: Initialize 1st moment vector: $\mathbf{m}_0 \leftarrow 0$

Parameter: Initialize 2nd moment vector: $\mathbf{v}_0 \leftarrow 0$

$i \leftarrow 0$;

while *Stopping criterion not met* **do**

$i = i + 1$;

Sample a minibatch of *m* examples from the training set $\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(m)}\}$ with corresponding targets $\mathbf{y}^{(i)}$;

Compute gradient estimate: $\hat{\mathbf{g}} \leftarrow \frac{1}{m} \nabla_w \sum_i L\left(f(\mathbf{x}^{(i)}, \mathbf{w}), \mathbf{y}^{(i)}\right)$;

Update biased first moment estimate: $\mathbf{m}_i \leftarrow \beta_1 \cdot \mathbf{m}_{i-1} + (1 - \beta_1) \cdot \hat{\mathbf{g}}_i$;

Update biased second moment estimate: $\mathbf{v}_i \leftarrow \beta_2 \cdot \mathbf{v}_{i-1} + (1 - \beta_2) \cdot \hat{\mathbf{g}}_i^2$;

Correct bias in first moment: $\hat{\mathbf{m}}_i \leftarrow \frac{\mathbf{m}_i}{1-\beta_1^i}$;

Correct bias in second moment: $\hat{\mathbf{v}}_i \leftarrow \frac{\mathbf{v}_i}{1-\beta_2^i}$;

Update weights: $\mathbf{w}_i \leftarrow \mathbf{w}_{i-1} - \alpha * \frac{\hat{\mathbf{m}}_i}{\sqrt{\hat{\mathbf{v}}_i} + \delta}$;

3.3 Over-fitting prevention

Neural networks are always optimized with their training data sample. This leads into a risk that the trained neural network performs well in-sample but does not generalize out of sample. This phenomenon is called over-fitting and this phenomenon is for example presented in Bishop (1995). Here, two methods to prevent over-fitting, Dropout layers and early stopping using validation data, are presented.

3.3.1 Dropout layers

The concept of dropout layer is that during training, neural network connection is randomly set to zero at every step at a defined probability. The dropout practice, also known as dilution in some sources, was introduced by Hinton et al. (2012). According to Warde-Farley et al. (2013), use of dropout significantly reduces the risk of over-fitting.

3.3.2 Validation data

The concept of early stopping using validation data is the following, and is for example presented in Bishop (1995). The training sample is split into two sets. The first set is the actual training set and the second one is the validation set. The training set is used for optimizing the loss function. At the same time, the loss value is also calculated for the validation data set that is independent of the training sample.

Usually, the case in optimization is that the training loss improves by every step, but the loss with validation data starts to increase after some point as a sign of over-fitting. Thus, it is possible to stop the training earlier if the validation loss does not improve as lot as wished or if the validation loss does not improve after a number of steps.

3.4 Literature on performance measurement

For common investors, the most familiar way to compare investment performance between assets is the return. However, the return in itself does not reveal all the stock performance related aspects. Usually, finance literature breaks the asset performance into systemic components that are common between the assets market-wide and idiosyncratic risk that is the risk that cannot be explained by the market wide factors. Also, the return of assets is often decomposed into systemic market risk and the so-called alpha that is the part of the return that cannot be explained by the market-wide factors.

The simplest model describing systematic risk is the Capital Assets Pricing Model (CAPM) that is often credited to William Sharpe (1964) among other authors. The CAPM is presented as:

$$R_i - R_f = \alpha + \beta_{Mkt-Rf}(R_{Mkt} - R_f) + \epsilon_i, \tag{3.16}$$

where $Mkt-Rf$ represents the market factor or the market excess return on the risk-free rate Rf. The β represents systemic risk or intuitively the sensitivity of the asset's return against the market return. The α represents the systemic return component that cannot be explained by the market-wide factors. The ϵ represents the error term.

Fama and French (1993) augmented the model with two additional factors. The two factors are small-minus-big (SMB) and high-minus-low (HML). The concept of the SMB is related to the stock market phenomenon where small capitalization stocks seem to outperform the large capitalization stocks. Then, the concept of HML is related to the stock market phenomenon where the stock with a high book-to-market equity ratio seem to outperform the stocks with low book-to-market ratio. The Fama-French three-factor

model is defined as:

$$R_i - R_f = \alpha + \beta_{Mkt-Rf}(R_{Mkt} - R_f) + \beta_{SMB}SMB + \beta_{HML}HML + \epsilon_i, \tag{3.17}$$

where the factors are:

- Mkt - Rf — Market excess return
- SMB — Small-minus-big (size)
- HML — High-minus-low (value).

Finally, Carhart (1997) added a momentum factor to the model. The momentum is a stock market phenomenon where the stocks with past high returns seem to outperform stocks with low past returns also in the future. The Carhart model is defined as:

$$R_i - R_f = \alpha + \beta_{Mkt-Rf}(R_{Mkt} - R_f) + \beta_{SMB}SMB + \beta_{HML}HML + \beta_{UMD}UMD + \epsilon_i, \tag{3.18}$$

where the factors are:

- Mkt - Rf — Market excess return
- SMB — Small-minus-big (size)
- HML — High-minus-low (value)
- UMD — Up-minus-down (momentum).

The most common models in financial literature to measure asset performance are the Fama-French model and the Carhart model. The alphas in the models are used to measure how the assets performs against a benchmark.

4. METHODOLOGY

This chapter explains the methodology of this thesis. First, measuring the accruals is presented. Then, the machine learning models used to predict the stock returns are presented, i.e. the naive model, linear regression and the recurrent neural network.

4.1 Defining accruals

The main components of accruals are related to the current assets, current liability and depreciation. Here, the components are presented following the methodology of Sloan.

4.1.1 Current asset component

The current assets mainly drive the accruals by the increase or decrease in accounts receivable. The current assets component of accruals is defined as:

$$\text{Current asset component of accruals} = \Delta\text{CA} - \Delta\text{Cash}, \tag{4.1}$$

where ΔCA is the annual change in current assets and ΔCash is the annual change in cash. The definition of the current asset component of accruals is indirect as Compustat does not directly provide necessary data items such as accounts receivable to directly calculate the current asset component.

4.1.2 Current liability component

The current liability component is mainly driven by the change in accounts payable. As with the current liability component, the calculation is indirect due to the structure of the Compustat items. The current liability component is defined as:

$$\text{Current liability component of accruals} = \Delta\text{CL} - \Delta\text{STD} - \Delta\text{TP}, \tag{4.2}$$

where ΔCL is the change in current liability, ΔSTD is the change in short-term included in current liabilities and ΔTP is the change in tax payable.

4.1.3 Depreciation component

The final component of accruals is the depreciation. Depreciation includes the depreciation of property, plant and equipment. Depreciation is a discretionary accrual which can be adjusted by managerial discretion and thus used in earnings management.

4.1.4 Scaling accruals

To make the accruals comparable between different companies, the accruals are scaled by the amount of total assets by each company. Then, the total scaled accruals are obtained by deducting the current liability component and the depreciation component from the current asset component:

$$\text{Scaled accruals} = \frac{\Delta\text{CA} - \Delta\text{Cash}}{\text{Total Assets}} - \frac{\Delta\text{CL} - \Delta\text{STD} - \Delta\text{TP}}{\text{Total Assets}} - \frac{\text{Depreciation}}{\text{Total Assets}}, \quad (4.3)$$

where the variables are:

- ΔCA — Change in current assets
- ΔCash — Change in cash
- ΔCL — Change in current liabilities
- ΔSTD — Change in debt in short-term liabilities
- ΔTP — Change in tax payable

The scaled accruals used as explanatory variables in the later phases of this thesis.

4.2 Verifying the existence of the accrual anomaly - Panel regression

This thesis first shows that the accrual anomaly exists in the US stock market sample in the year 1980-2020. The method used is panel regression. The regression is implemented so that the stock returns of an individual stock in the year T are regressed against its scaled accruals from the years T, $T-1$, $T-2$ and $T-3$. Also, the panel regression includes controls for the logarithm of the stock's market cap and the stock's book-to-market ratio in the beginning of the year T. Finally, fixed effects are used for year and company.

The regression equation is:

$$\begin{aligned}\text{Return}_T = {} & \lambda_{\text{AC}_T} * \text{AC}_T + \lambda_{\text{AC}_{T-1}} * \text{AC}_{T-1} \\ & + \lambda_{\text{AC}_{T-2}} * \text{AC}_{T-2} + \lambda_{\text{AC}_{T-3}} * \text{AC}_{T-3} \\ & + \lambda_{\text{Mkt cap}_T} * log(\text{Mkt cap}_T) + \lambda_{\text{Book-to-market}_T} * \text{Book-to-market}_T,\end{aligned} \tag{4.4}$$

where parameters are:

- λ_i — Coefficient for regression term i
- AC_t — Scaled accrual component at time t
- Mkt cap — Market capitalization of the stock
- Book-to-market — Book-to-market ratio of the stock.

Also, fixed effects are added for time and stock to control for the omitted variable bias.

4.3 Predicting stock returns

The goal of this thesis is to predict stock returns based on the accrual data. Three methods to predict stock returns are used in this thesis.

The first one is a naive model which sorts the stock in year T based on accruals and then the return are examined in the three following years. The method is similar to the Sloan (1996). The second method used is linear regression. The regression predicts the returns using scaled accruals and controlling for log of market cap and book-to-market ratio. The third and final method is the recurrent neural network. The recurrent neural network uses a three-year time series of accrual in addition to a time series of log of market cap, book-to-market and stock's own previous returns.

4.3.1 Naive model

The naive model directly follows the methodology of Sloan (1996). In year T, the stocks are sorted based on their scales accruals that year into ten portfolios. The portfolio P1 includes the stocks with the lowest accruals and the portfolio P10 the highest accruals. The portfolios are equally weighted and the stock returns are examined in the three following years $T+1$, $T+2$ and $T+3$. A filter for returns is used and stock returns exceeding 500 % per annum are treated as outliers and removed.

Both mean annual absolute returns are calculated and also, the annual returns of the portfolios are regressed against the Carhart factors to see if a significant Carhart alpha is

obtained. Also, a long-short portfolio between portfolios P1 and P10 is examined to see if the difference between the portfolio returns in different accrual deciles is statistically significant.

4.3.2 Linear regression

A linear regression model is used to predict the stock returns. The explanatory variables of the regression predicting the stock return in the year T are the scaled accruals in the years $T-1$, $T-2$ and $T-3$. Also, controls include the log of market cap and the book-to-market value in the year $T-1$. Three different forecast sets from different variable sets are used:

1. Scaled accruals only
2. Controls only
3. Scaled accruals and controls.

The different forecast sets are made to compare if the accrual's performance can be matched by using some external set of explanatory variables or if the accruals alone can provide the forecast performance.

The full linear regression equation for the model with scaled accruals and controls is:

$$\begin{aligned} \text{Return}_T = {} & \lambda_{AC_{T-1}} * AC_{T-1} + \lambda_{AC_{T-2}} * AC_{T-2} \\ & + \lambda_{AC_{T-3}} * AC_{T-3} + \lambda_{\text{Mkt cap}_{T-1}} * log(\text{Mkt cap}_{T-1}) \\ & + \lambda_{\text{Book-to-market}_{T-1}} * \text{Book-to-market}_{T-1}, \end{aligned} \tag{4.5}$$

where parameters are:

- λ_i — Coefficient for regression term i
- AC_t — Scaled accrual component at time t
- Mkt cap — Market capitalization of the stock
- Book-to-market — Book-to-market ratio of the stock.

The regression forecast is made for every year. The training sample for the linear model consists of the three preceding years of individual company return data. So, the same regression model is used for every company in that year.

For example, if the forecast is made for the year 2020, the regression used in forecast is estimated from the individual stock returns in the year 2017-2019. And for example, for the annual return data points in the year 2017, the explanatory variables are the scaled

accruals of the corresponding company from the years 2014-2016 and the controls are the log of market cap and the book-to-market ratio of that company in the turn of year 2016-2017.

4.3.3 Recurrent neural network

RNN overview

The recurrent neural network or RNN has been chosen as the model as it is able to model time series properties better than traditional feed-forward neural networks. The purpose of the RNN is to find out if for example the trend component of the accruals plays any role in the accrual anomaly.

The RNN architecture

The neural network in this thesis is implemented using the Keras library and consist of the following components:

1. Input layer
2. Simple RNN layer with 10 units with hyperbolic tangent activation
3. 50 % dropout layer
4. Simple RNN layer with 5 units with hyperbolic tangent activation
5. Dense 1x1 output layer with linear activation

The model is described in Figure 4.1.

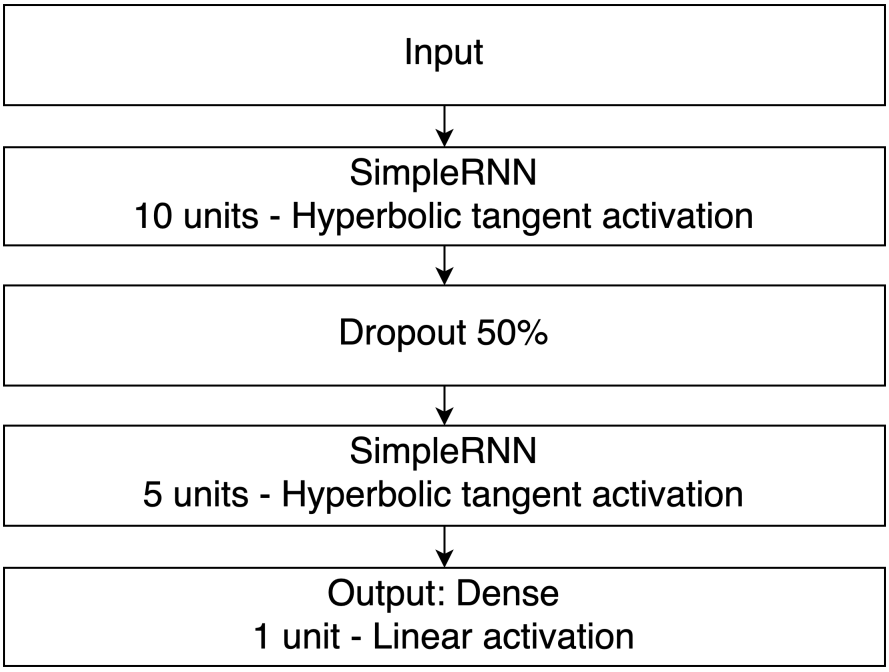

Figure 4.1. *The description of the RNN layers.*

Explanatory variables

The neural network has three variations with different input parameter combinations as explanatory parameters. Each of the variations contain a three-year time series of the variables. The model variations are:

1. Accruals
2. Controls
3. Accruals & controls

The controls include the three-year time series of the log of market capitalization, book-to-market ratio and the stock's historic own return. The input layer parameters are shown in Figure 4.2.

The variables are scaled to 0-1 scale using min-max scaling. For the returns, an outlier rule is used and returns exceeding 500% per annum are removed as outliers.

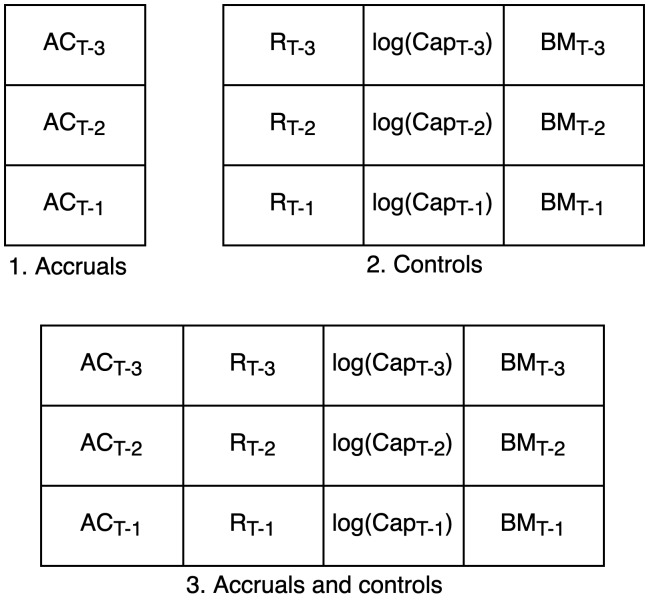

Figure 4.2. *The three types of input layers. Variant 1 contains the accruals as explanatory variables. Variant 2 contains the controls variables as explanatory variables which include the stock's own return, log of the market capitalization and the book-to-market ratio. Variant 3 contains both accruals and the controls. For every variable, a three-year time series is used. The variables are scaled to 0-1 scale using min-max scaling.*

RNN training optimizer parameters

The additional RNN optimizer parameters in used for training are:

1. Loss function: Mean squared error (MSE)
2. Optimizer: Adam
3. Learning rate: 0.0001
4. Number of epochs: 20

Normalizing parameters by min-max scaling

The input parameters for the neural networks are normalized using min-max scaling that normalizes the input parameter to a scale between 0-1, zero representing the smallest and 1 the largest value according to the following equation:

$$y_i = \frac{x_i - \min x}{\max x - \min x}, \tag{4.6}$$

where the y_i represents the normalized value of the original value x_i.

4.4 Pooling stocks by forecast

The linear regression forecast and the RNN forecast both provide a forecast return for an individual company per year. The actual performance of the models is estimated indirectly. The performance estimation is done by pooling the stocks each year into ten portfolios based on the forecast return. Portfolio P10 is for the highest predicted return and portfolio P1 for the lowest. Then, the realized portfolio returns are calculated. The portfolios are equal weight portfolios, so the mean return of all companies in the corresponding portfolio are calculated.

The portfolios are compared for their mean annual return and also their annual returns are regressed against the Carhart factors to see if a statistically significant Carhart alpha can be obtained using this strategy. Also, a long-short portfolio between the portfolios P10 and P1 is examined to see if the difference between the portfolios is statistically significant.

4.5 Measuring stock performance

Two methods are used to estimate the pooled portfolio performance. The first method is the mean return. The second method is using Carhart alphas. In the sample, returns exceeding 500% are excluded from the sample as outliers.

4.5.1 Mean returns

The mean returns are calculated for each realized portfolio returns. The mean return offers a simple way to examine the performance.

4.5.2 Carhart alphas

The Carhart model is defined below with the α in the following equation measuring abnormal returns:

$$\begin{aligned} R_i - R_f = \alpha + \beta_{Mkt-Rf}(R_{Mkt} - R_f) + \beta_{SMB}SMB \\ + \beta_{HML}HML + \beta_{UMD}UMD + \epsilon_i, \end{aligned} \tag{4.7}$$

where the factors are:

- Mkt - Rf — Market excess return
- SMB — Small-minus-big (size)
- HML — High-minus-low (value)
- UMD — Up-minus-down (momentum)

5. DATA

This section explains the data sources used in this thesis. The main data items are the stock price data, company financial data and Carhart factor data.

5.1 Stock price data

Stock price data is obtained from the CRSP database via WRDS. The dataset consists of the US stock market companies listed in NYSE, NASDAQ, NYSE American (formerly AMEX) and NYSE Arca stock exchanges in the years 1980-2020. The returns are calculated annually from the available stock prices.

The market capitalizations used in control variables are obtained by multiplying the share price by number of shares outstanding.

5.2 Financial data

The company financial data is obtained from Compustat via WRDS. The Compustat financial items used include (The Compustat item code is in parentheses):

- Current assets *(actq)*
- Cash and cash equivalents *(cheq)*
- Current liabilities *(lctq)*
- Short term debt included in current liabilities *(dlcq)*
- Tax payable *(txpq)*
- Depreciation of property, plant & equipment *(dpq)*
- Total assets *(atq)*
- Book value of equity *(ceqq)*

The items excluding the book value of equity are used to calculate the scaled accruals. The book value of equity is used to calculate the book-to-market ratio.

5.3 Factor data

The Carhart factor data is provided by the Kenneth French website. French (2021) provides the US market return, small-minus-big size factor, high-minus-low value factor and momentum up-minus-down factor data.

5.4 Descriptive statistics

Table 5.1 contains the descriptive statistics of the data sample. Panel A displays the components of the accrual components along with their mean values, medians and their quantiles. The accrual components are current assets, cash, current liabilities, debt included in current liabilities, depreciation. The total assets are used to scale the accruals so that the accruals between companies become comparable. The scaled accruals are calculated according to Equation (4.3).

Panel B contains summary statistic on the stock returns. Also, the related market capitalization, book value of equity and book-to-market values are shown in the descriptive statistics.

Panel A: Annual accruals and Components

	Mean	St.Dev	25 %	50%	75%	N obs	N companies
Current assets (CA)	933,66	5025,90	13,70	64,88	323,24	98772	8134
Cash	756,97	12118,81	2,27	15,71	90,31	117406	9485
Current liabilities (CL)	709,75	4219,56	5,89	28,74	172,28	99858	8271
Debt in CL	632,35	10365,67	0,10	2,69	27,40	113652	8934
Tax payables (TP)	28,13	226,39	0,00	0,00	2,12	104680	5658
Depreciation (Dep)	42,16	281,85	0,21	1,28	9,56	101530	8747
Total assets	6126,75	62762,50	33,85	196,67	1261,87	117818	9523
Accruals	-51,67	574,63	-11,29	-0,53	2,63	73678	7275
Scaled accruals	-0,013	7,706	-0,042	-0,008	0,027	73653	7275

Panel B: Annual Returns and Market Capitalizations

	Mean	St.Dev	25 %	50 %	75 %	N obs	N companies
Return	0,160	1,710	-0,240	0,010	0,270	264295	28041
Market capitalization	1938,74	12710,08	28,67	126	610,25	287591	29844
Book value of equity	1192,44	7250,12	11,40	66,63	384,22	117398	9516
Book-to -market	2,19	52,22	0,31	0,57	0,97	95605	9318

Table 5.1. *Panel A contains the descriptive statistics for the accrual data. Panel B contains the descriptive statistics for the stock returns, equity book values and market capitalizations. Figures are in USDM. These accrual statistics are from year-end reports. Data interval is 1980-2020. Accruals are defined as:* $Accruals = (\Delta Current\ assets - \Delta Cash) - (\Delta Current\ liabilities - \Delta Debt\ included\ in\ short\ term\ liabilities - \Delta Tax\ Payable) - \Delta Depreciation$. *Scaled accruals are accruals scaled by total assets at the beginning of year.*

6. RESULTS

This chapter introduces the results of this thesis. First, the panel regression results show that the accrual anomaly exists in the US stock market 1980-2020 sample. Then, the results for the machine learning models are presented. Each machine learning model, i.e. the naive model, linear regression and the recurrent neural network exhibit predictive power and ability to generate a positive Carhart alpha.

6.1 Existence of the accrual anomaly - Panel regression

Table 6.1 and Table 6.2 contain the results for a panel regression explaining the relationship between the accruals and stock returns using controls. The panel regression results verify that the accrual anomaly exists in the US stock market 1980-2020 sample. Also, the Sloan (1996) results seem to hold. The accruals in the year T seem to imply high returns in the same year. However, as the coefficients for the accruals in the years T-3 to T-1 are negative, it seems that the high accruals predict worse returns in the long run and vice versa.

In the panel regression, controls and fixed effect were used. The controls included the log of market capitalization and the book to market ratio. Also, fixed effects were used for year and stock to handle the omitted variable bias. The panel regression is explained in detail in Section 4.2.

		Model 1	Model 2	Model 3	Model 4	Model 5
AC_t	Coef	0,403				0,380
	T Stat	12,490				8,748
	P-value	0,000				0,000
	Signif.	***				***
AC_{t-1}	Coef		-0,469			-0,505
	T Stat		-13,534			-10,991
	P-value		0,000			0,000
	Signif.		***			***
AC_{t-2}	Coef			-0,255		-0,332
	T Stat			-6,976		-7,369
	P-value			0,000		0,000
	Signif.			***		***
AC_{t-3}	Coef				-0,174	-0,146
	T Stat				-4,513	-3,472
	P-value				0,000	0,001
	Signif.				***	***
	Stock FE	No	No	No	No	No
	Time FE	No	No	No	No	No
	R^2	0,00265	0,00320	0,000917	0,000422	0,00650
	N obs	58655	56975	53021	48246	40533

Table 6.1. *Panel regression results - Part 1. This is a panel regression where the dependent variable is return of an individual stock in year T. The explanatory variables are accruals scaled by total assets in year T to year T-3. Controls include log of market capitalization and book-to-market value. Additionally, fixed effects are used for Time and Stock. Data is from 1980-2020.*

		Model 6	Model 7	Model 8	Model 9
AC_t	Coef	0,429	0,311	0,475	0,472
	S.e.	0,043	0,048	0,041	0,046
	T Stat	9,907	6,490	11,460	10,307
	P-value	0,000	0,000	0,000	0,000
	Signif.	***	***	***	***
AC_{t-1}	Coef	-0,464	-0,452	-0,421	-0,265
	S.e.	0,046	0,051	0,044	0,049
	T Stat	-10,136	-8,916	-9,607	-5,439
	P-value	0,000	0,000	0,000	0,000
	Signif.	***	***	***	***
AC_{t-2}	Coef	-0,327	-0,303	-0,226	-0,092
	S.e.	0,045	0,049	0,043	0,047
	T Stat	-7,287	-6,161	-5,262	-1,954
	P-value	0,000	0,000	0,000	0,051
	Signif.	***	***	***	.
AC_{t-3}	Coef	-0,166	-0,142	-0,137	-0,032
	S.e.	0,042	0,046	0,040	0,044
	T Stat	-3,964	-3,082	-3,397	-0,718
	P-value	0,000	0,002	0,001	0,473
	Signif.	***	**	***	
Log(Mkt cap)	Coef	-0,036	-0,234	-0,042	-0,296
	S.e.	0,002	0,005	0,002	0,006
	T Stat	-21,344	-49,051	-23,650	-52,302
	P-value	0,000	0	0,000	0
	Signif.	***	***	***	***
Book-to-market	Coef	0,000	-0,001	0,000	-0,001
	S.e.	0,000	0,000	0,000	0,000
	T Stat	-0,898	-3,752	-1,678	-5,115
	P-value	0,369	0,000	0,093	0,000
	Signif.		***	.	***
	Stock FE	No	Yes	No	Yes
	Time FE	No	No	Yes	Yes
	R^2	0,018	0,186	0,111	0,272
	N obs	40478	40478	40478	40478

Table 6.2. *Panel regression results - Part 2. This is a panel regression where the dependent variable is return of an individual stock in year T. The explanatory variables are accruals scaled by total assets in year T to year T-3. Controls include log of market capitalization and book-to-market value. Additionally, fixed effects are used for Time and Stock. Data is from 1980-2020.*

6.2 Naive model results

The naive model is the simplest model used to predict the stock returns based on the accrual anomaly. The naive model is presented in Section 4.3.1.

The stocks are pooled every year based on their scaled accruals into ten portfolios. The stocks with the lowest scaled accruals are pooled into the decile 1 and the highest into the decile 10. Then, the equally weighted realized returns are examined in the year of pooling T0 and the three subsequent years T+1 to T+3.

The returns obtained from the naive model are in Table 6.3 and the Carhart alphas are in Table 6.4. It is shown that the model seems to predict the stock returns quite well. In the year T+1 decile 1 earns a mean 19.9 % return compared to the 6.1 % return in the decile 10. The long-short portfolio between deciles 1 - 10 in the year T+1 return seems to be on average 13.9 % as is the statistically significant Carhart alpha. In the years T+2 and T+3, the performance of the naive model seems to deteriorate, but is still moderate.

Accrual decile	Measure	T0	T+1	T+2	T+3
1	Mean return	-0,011	0,199	0,155	0,145
	St.Dev	0,296	0,294	0,279	0,27
2	Mean return	0,088	0,135	0,13	0,129
	St.Dev	0,229	0,268	0,235	0,236
3	Mean return	0,08	0,117	0,124	0,124
	St.Dev	0,211	0,23	0,239	0,23
4	Mean return	0,098	0,113	0,109	0,125
	St.Dev	0,198	0,193	0,194	0,219
5	Mean return	0,09	0,103	0,1	0,091
	St.Dev	0,199	0,194	0,186	0,191
6	Mean return	0,104	0,086	0,102	0,098
	St.Dev	0,187	0,176	0,202	0,203
7	Mean return	0,104	0,086	0,092	0,105
	St.Dev	0,196	0,185	0,201	0,2
8	Mean return	0,119	0,091	0,097	0,101
	St.Dev	0,205	0,221	0,205	0,209
9	Mean return	0,13	0,075	0,114	0,117
	St.Dev	0,234	0,24	0,231	0,236
10	Mean return	0,173	0,061	0,098	0,102
	St.Dev	0,286	0,26	0,279	0,263
1 - 10	Mean return	-0,183	0,139	0,057	0,043
	St.Dev	0,203	0,151	0,126	0,175

Table 6.3. *Naive model returns. The stock are sorted in year T based on their scaled accruals. The returns are examined in the three following years T+1 to T+3. Decile 1 contains the stock with the lowest and decile 10 the highest accrual stocks. Also, a long-short portfolio between deciles 1 and 10 is formed.*

Accrual decile	Measure	T0	T+1	T+2	T+3
1	Alpha	-0,111	0,092	0,025	0,029
	T Stat	-4,106	3,652	0,779	1,108
	P-value	0,000	0,001	0,441	0,276
	Signif.	***	***		
2	Alpha	-0,018	0,003	0,013	0,012
	T Stat	-1,011	0,144	0,596	0,569
	P-value	0,319	0,887	0,555	0,573
	Signif.				
3	Alpha	-0,020	-0,008	0,002	0,021
	T Stat	-0,996	-0,445	0,081	0,918
	P-value	0,326	0,659	0,936	0,365
	Signif.				
4	Alpha	-0,013	-0,006	0,001	-0,002
	T Stat	-0,786	-0,431	0,036	-0,170
	P-value	0,437	0,669	0,971	0,866
	Signif.				
5	Alpha	-0,011	-0,005	-0,016	-0,006
	T Stat	-0,670	-0,342	-1,036	-0,381
	P-value	0,507	0,734	0,308	0,706
	Signif.				
6	Alpha	-0,008	-0,025	-0,009	-0,011
	T Stat	-0,523	-1,805	-0,569	-0,588
	P-value	0,604	0,080	0,573	0,561
	Signif.		.		
7	Alpha	-0,006	-0,024	-0,005	-0,004
	T Stat	-0,337	-1,571	-0,354	-0,298
	P-value	0,738	0,125	0,726	0,768
	Signif.				
8	Alpha	-0,007	-0,027	-0,018	0,002
	T Stat	-0,398	-1,511	-1,062	0,174
	P-value	0,693	0,140	0,296	0,863
	Signif.				
9	Alpha	0,016	-0,038	0,006	0,002
	T Stat	0,654	-1,938	0,429	0,141
	P-value	0,517	0,061	0,671	0,889
	Signif.		.		
10	Alpha	0,004	-0,046	-0,026	-0,013
	T Stat	0,174	-1,910	-1,457	-0,453
	P-value	0,863	0,064	0,154	0,654
	Signif.		.		
1 - 10	Alpha	-0,115	0,139	0,050	0,042
	T Stat	-3,172	4,641	1,888	1,075
	P-value	0,003	0,000	0,068	0,290
	Signif.	**	***	.	

Table 6.4. *Naive model Carhart alphas. The stock are sorted in year T based on their scaled accruals. The returns are examined in the three following years T+1 to T+3. Decile 1 contains the stock with the lowest and decile 10 the highest accrual stocks. Also, a long-short portfolio between deciles 1 and 10 is formed.*

6.3 Linear regression results

The second machine learning model to predict stock returns based on the accrual anomaly is the linear regression model. Three different explanatory variable sets are used. The first set contains the scaled accruals from the three preceding years. The second set contains the controls. Then, the third set contains the scaled accruals and controls. The regression is estimated for every year using the three previous years returns and the corresponding explanatory variables.

The stocks are pooled every year into ten portfolios based on the forecast return. Portfolio P1 contains the stocks with lowest predicted returns and P10 the highest. Then, the realized returns of each portfolio are examined on an equally weighted basis. Table 6.5 contains the returns from the linear regression models.

		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10 - P1
Accruals	Mean	0,101	0,101	0,103	0,101	0,089	0,103	0,116	0,118	0,141	0,175	0,074
	St.Dev	0,287	0,231	0,213	0,192	0,191	0,189	0,198	0,214	0,243	0,279	0,156
Controls	Mean	0,063	0,067	0,070	0,068	0,085	0,080	0,102	0,114	0,130	0,224	0,161
	St.Dev	0,223	0,219	0,200	0,198	0,215	0,232	0,224	0,221	0,275	0,280	0,196
Accruals & controls	Mean	0,055	0,070	0,078	0,082	0,098	0,123	0,133	0,129	0,153	0,230	0,175
	St.Dev	0,223	0,219	0,200	0,198	0,215	0,232	0,224	0,221	0,275	0,280	0,196

Table 6.5. *The returns from the linear regression prediction model. The returns are predicted based on a linear regression estimated for every year from the three previous years of returns and the corresponding explanatory variables. The variables sets used in prediction include scaled accruals, scaled accruals & controls and only control variables. The control variables are the logarithm of the stock's own market capitalization and the book-to-market ratio. Then, the stocks are each year pooled into ten portfolios based on the forecast returns. The portfolio P1 contains the stocks with the lowest predicted returns and the portfolio P10 the highest. The realized equally weighted mean returns are shown in the table.*

The Carhart alphas from the three regression forecast models are found in Table 6.6, Table 6.7 and Table 6.8. It seems that the model with the accrual forecast in Table 6.6 earns a 7 % long-short alpha with a weak 10 % statistical significance. Then, the control forecast in Table 6.7 seems to yield a 9.4 % alpha with 1 % statistical significance. Finally, the combined accrual and control model in in Table 6.8 earns a 13.7 % alpha with 1 % statistical significance.

Examining the both the returns and the Carhart alphas from the models shows that the regression forecast offers a fairly good prediction power and performance. Also, the return from the combined accrual and control model is on a same level with the naive model.

Accrual regression forecast

		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10 - P1
Alpha	Coef.	-0,019	-0,004	0,002	0,013	0,004	0,009	0,015	0,011	0,014	0,051	0,070
	T Stat	-0,833	-0,246	0,101	0,923	0,329	0,517	1,007	0,710	0,784	2,001	1,966
	P-value	0,411	0,807	0,920	0,363	0,745	0,609	0,322	0,483	0,439	0,054	0,059
	Signif.										.	.
Mkt - Rf	Coef.	1,088	0,903	0,866	0,730	0,741	0,765	0,809	0,896	1,092	1,076	-0,012
	T Stat	9,409	10,224	10,848	10,073	12,319	9,130	10,721	11,800	12,383	8,433	-0,066
	P-value	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,948
	Signif.	***	***	***	***	***	***	***	***	***	***	
SMB	Coef.	1,310	1,054	0,853	0,798	0,733	0,778	0,861	0,896	0,914	1,235	-0,075
	T Stat	7,333	7,727	6,914	7,120	7,879	6,007	7,384	7,632	6,711	6,265	-0,271
	P-value	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,789
	Signif.	***	***	***	***	***	***	***	***	***	***	
HML	Coef.	0,038	0,043	0,151	0,133	0,127	0,159	0,060	0,086	0,193	-0,014	-0,052
	T Stat	0,330	0,493	1,908	1,860	2,135	1,924	0,806	1,151	2,216	-0,112	-0,293
	P-value	0,744	0,626	0,066	0,073	0,041	0,064	0,427	0,259	0,034	0,912	0,771
	Signif.			.	.	*	.			*		
UMD	Coef.	-0,331	-0,261	-0,279	-0,285	-0,341	-0,223	-0,186	-0,210	-0,201	-0,228	0,102
	T Stat	-3,312	-3,415	-4,051	-4,548	-6,557	-3,075	-2,847	-3,206	-2,641	-2,072	0,662
	P-value	0,002	0,002	0,000	0,000	0,000	0,004	0,008	0,003	0,013	0,047	0,513
	Signif.	**	**	***	***	***	**	**	**	*	*	
R^2		0,887	0,899	0,902	0,900	0,932	0,866	0,897	0,911	0,906	0,851	0,029
N		35	35	35	35	35	35	35	35	35	35	35

Table 6.6. *The Carhart alphas from the linear prediction model using the three preceding years' scaled accruals as predictors. The models are estimated each year from the three preceding years of returns and the corresponding predictors. Then, the stocks are each year pooled into ten portfolios based on the forecast returns. The portfolio P1 contains the stocks with the lowest predicted returns and the portfolio P10 the highest. The Carhart factor coefficients are then estimated.*

Control regression forecast

		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10 - P1
Alpha	Coef.	-0,009	-0,030	-0,034	-0,034	-0,034	-0,036	-0,011	-0,009	-0,001	0,085	0,094
	T Stat	-0,572	-2,283	-2,498	-3,029	-2,478	-2,919	-0,865	-0,533	-0,039	3,298	3,022
	P-value	0,572	0,030	0,018	0,005	0,019	0,007	0,394	0,598	0,969	0,003	0,005
	Signif.		*	*	**	*	**				**	**
Mkt - Rf	Coef.	0,783	0,825	0,851	0,875	0,949	0,917	0,879	0,953	1,010	1,081	0,298
	T Stat	9,614	12,589	12,353	15,652	13,631	14,690	14,183	10,937	10,856	8,289	1,890
	P-value	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,068
	Signif.	***	***	***	***	***	***	***	***	***	***	.
SMB	Coef.	0,041	0,535	0,679	0,656	0,982	1,170	1,217	1,262	1,223	1,467	1,426
	T Stat	0,327	5,289	6,382	7,594	9,136	12,137	12,705	9,374	8,502	7,285	5,855
	P-value	0,746	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
	Signif.		***	***	***	***	***	***	***	***	***	***
HML	Coef.	-0,067	0,053	0,104	0,114	0,169	0,227	0,170	0,348	0,250	0,215	0,282
	T Stat	-0,837	0,812	1,532	2,062	2,452	3,685	2,782	4,041	2,722	1,669	1,812
	P-value	0,409	0,423	0,136	0,048	0,020	0,001	0,009	0,000	0,011	0,106	0,080
	Signif.				*	*	***	**	***	*		.
UMD	Coef.	-0,534	-0,250	-0,188	-0,259	-0,126	-0,143	-0,145	-0,104	-0,064	-0,050	0,485
	T Stat	-7,598	-4,413	-3,154	-5,371	-2,089	-2,657	-2,703	-1,388	-0,791	-0,441	3,558
	P-value	0,000	0,000	0,004	0,000	0,045	0,013	0,011	0,175	0,435	0,663	0,001
	Signif.	***	***	**	***	*	*	*				**
R^2		0,886	0,912	0,908	0,943	0,926	0,945	0,945	0,905	0,893	0,841	0,305
N		35	35	35	35	35	35	35	35	35	35	35

Table 6.7. *The Carhart alphas from the linear prediction model using the controls as predictors. The controls include the log of the stock's own market capitalization and the book-to-market ratio. The models are estimated each year from the three preceding years of returns and the corresponding predictors. Then, the stocks are each year pooled into ten portfolios based on the forecast returns. The portfolio P1 contains the stocks with the lowest predicted returns and the portfolio P10 the highest. The Carhart factor coefficients are then estimated.*

Accrual and control regression forecast

		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10 - P1
Alpha	Coef.	-0,040	-0,017	-0,005	-0,012	-0,008	0,009	0,029	0,011	0,030	0,097	0,137
	T Stat	-1,683	-1,005	-0,387	-0,795	-0,510	0,473	1,603	0,643	1,301	3,370	3,476
	P-value	0,103	0,323	0,701	0,433	0,614	0,640	0,119	0,525	0,203	0,002	0,002
	Signif.										**	**
Mkt - Rf	Coef.	0,876	0,851	0,762	0,803	0,884	0,957	0,820	0,905	1,070	1,066	0,190
	T Stat	7,311	10,013	11,613	10,809	11,789	10,041	9,062	10,864	9,353	7,371	0,957
	P-value	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,346
	Signif.	***	***	***	***	***	***	***	***	***	***	
SMB	Coef.	0,554	0,584	0,718	0,751	0,943	0,972	1,185	1,138	1,235	1,421	0,867
	T Stat	2,992	4,445	7,083	6,540	8,134	6,600	8,477	8,841	6,985	6,359	2,829
	P-value	0,005	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,008
	Signif.	**	***	***	***	***	***	***	***	***	***	**
HML	Coef.	-0,122	-0,057	0,002	0,048	0,128	0,202	0,207	0,329	0,222	0,011	0,134
	T Stat	-1,033	-0,678	0,037	0,648	1,733	2,145	2,310	3,996	1,967	0,079	0,682
	P-value	0,310	0,503	0,970	0,522	0,093	0,040	0,028	0,000	0,059	0,937	0,500
	Signif.					.	*	*	***	.		
UMD	Coef.	-0,332	-0,435	-0,378	-0,273	-0,234	-0,227	-0,183	-0,111	-0,273	-0,079	0,253
	T Stat	-3,202	-5,928	-6,678	-4,257	-3,614	-2,755	-2,344	-1,546	-2,764	-0,631	1,475
	P-value	0,003	0,000	0,000	0,000	0,001	0,010	0,026	0,132	0,010	0,533	0,151
	Signif.	**	***	***	***	**	**	*		**		
R^2		0,799	0,896	0,927	0,902	0,915	0,881	0,885	0,901	0,876	0,813	0,239
N		35	35	35	35	35	35	35	35	35	35	35

Table 6.8. *The Carhart alphas from the linear prediction model using the three preceding years' scaled accruals and the controls as predictors. The controls include the log of the stock's own market capitalization and the book-to-market ratio. The models are estimated each year from the three preceding years of returns and the corresponding predictors. Then, the stocks are each year pooled into ten portfolios based on the forecast returns. The portfolio P1 contains the stocks with the lowest predicted returns and the portfolio P10 the highest. The Carhart factor coefficients are then estimated.*

6.4 Recurrent neural network results

The third machine learning model used to predict stock returns is the recurrent neural network. The RNN architecture is described in Section 4.3.3. Three different explanatory variable sets are used. The first variable set is the three-year time series of scaled accruals. The second variable set is the control variable set including three-year time series of stock's own annual return, logarithm of market capitalization and book-to-market ratio. The third variable set is the combination of the scaled accruals and the control variables. The RNN is trained every year using the three previous years of returns and the corresponding explanatory variables.

The procedure is similar as in Section 6.3. The stocks are pooled every year into ten portfolios based on the forecast return. Portfolio P1 contains the stocks with lowest predicted returns and P10 the highest. Then, the realized returns of each portfolio are examined on an equally weighted basis. Table 6.9 contains the returns from the linear regression models.

Predictors		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10 - P1
Accruals	Mean	0,110	0,117	0,107	0,095	0,114	0,106	0,118	0,113	0,119	0,169	0,058
	St.Dev	0,241	0,221	0,211	0,209	0,186	0,180	0,204	0,212	0,211	0,251	0,125
Controls	Mean	0,092	0,079	0,100	0,093	0,115	0,100	0,122	0,116	0,138	0,216	0,124
	St.Dev	0,258	0,219	0,206	0,176	0,194	0,185	0,191	0,216	0,252	0,275	0,181
Accruals & controls	Mean	0,078	0,087	0,104	0,095	0,105	0,115	0,098	0,131	0,160	0,196	0,118
	St.Dev	0,258	0,219	0,206	0,176	0,194	0,185	0,191	0,216	0,252	0,275	0,181

Table 6.9. *The returns from the RNN prediction model. The returns are predicted based on a recurrent neural network estimated for every year from the three previous years of returns and the corresponding explanatory variables. The variables sets used in prediction include scaled accruals, scaled accruals & controls and only control variables. The control variables are the logarithm of the stock's own market capitalization and the book-to-market ratio. A three-year timeseries for each variable is used. Then, the stocks are each year pooled into ten portfolios based on the forecast returns. The portfolio P1 contains the stocks with the lowest predicted returns and the portfolio P10 the highest. The realized equally weighted mean returns are shown in the table.*

The RNN models do not offer as good performance as expected. Compared to the naive and linear model performances, the long-short scaled accrual forecast portfolio in Table 6.10 only yields an alpha of 3.8 % without statistical significance. The control forecast in Table 6.11 offers a 10.1 alpha % performance with 10 % statistical significance. Then, the combined accrual and control forecast in in Table 6.12 offers a 9.7 % alpha with 1 %

statistical significance.

Accrual RNN forecast

		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10 - P1
Alpha	Coef.	0,019	0,016	0,001	-0,006	0,017	0,019	0,016	0,008	0,014	0,057	0,038
	T Stat	0,979	0,898	0,078	-0,413	1,116	1,296	0,866	0,389	0,988	2,721	1,394
	P-value	0,336	0,376	0,938	0,683	0,273	0,205	0,393	0,700	0,331	0,011	0,174
	Signif.										*	
Mkt - Rf	Coef.	0,884	0,874	0,858	0,835	0,780	0,732	0,824	0,839	0,876	0,961	0,076
	T Stat	9,068	9,817	9,710	12,087	10,289	9,790	8,992	8,141	12,579	9,157	0,558
	P-value	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,581
	Signif.	***	***	***	***	***	***	***	***	***	***	
SMB	Coef.	0,882	0,899	0,928	0,950	0,767	0,603	0,881	0,909	0,922	1,139	0,257
	T Stat	5,850	6,534	6,795	8,903	6,549	5,223	6,223	5,704	8,570	7,027	1,222
	P-value	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,231
	Signif.	***	***	***	***	***	***	***	***	***	***	
HML	Coef.	0,009	0,188	0,179	0,132	0,130	0,092	0,115	0,164	0,155	0,161	0,152
	T Stat	0,098	2,133	2,050	1,938	1,737	1,238	1,275	1,612	2,253	1,557	1,128
	P-value	0,922	0,041	0,049	0,062	0,093	0,225	0,212	0,118	0,032	0,130	0,268
	Signif.		*	*	.	.				*		
UMD	Coef.	-0,449	-0,306	-0,203	-0,235	-0,200	-0,275	-0,189	-0,178	-0,226	-0,268	0,181
	T Stat	-5,325	-3,983	-2,665	-3,937	-3,052	-4,258	-2,395	-1,997	-3,766	-2,956	1,536
	P-value	0,000	0,000	0,012	0,000	0,005	0,000	0,023	0,055	0,001	0,006	0,135
	Signif.	***	***	*	***	**	***	*	.	***	**	
R^2		0,889	0,888	0,878	0,923	0,886	0,880	0,859	0,832	0,924	0,876	0,112
N		35	35	35	35	35	35	35	35	35	35	35

Table 6.10. *The Carhart alphas from the RNN prediction model using the three preceding years' scaled accruals as predictors. The models are estimated each year from the three preceding years of returns and the corresponding predictors. Then, the stocks are each year pooled into ten portfolios based on the forecast returns. The portfolio P1 contains the stocks with the lowest predicted returns and the portfolio P10 the highest. The Carhart factor coefficients are then estimated.*

Control RNN forecast

		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10 - P1
Alpha	Coef.	-0,016	-0,021	0,022	-0,002	0,019	0,003	0,019	0,016	0,037	0,085	0,101
	T Stat	-0,604	-1,303	1,253	-0,143	1,234	0,184	1,360	1,152	1,906	2,704	2,445
	P-value	0,550	0,202	0,220	0,887	0,227	0,855	0,184	0,258	0,066	0,011	0,021
	Signif.									.	*	*
Mkt - Rf	Coef.	1,107	0,896	0,701	0,784	0,758	0,763	0,775	0,801	0,810	1,070	-0,037
	T Stat	8,185	10,932	7,881	9,168	9,642	10,263	11,212	11,395	8,170	6,744	-0,176
	P-value	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,861
	Signif.	***	***	***	***	***	***	***	***	***	***	
SMB	Coef.	0,826	0,692	0,616	0,609	0,710	0,791	1,022	0,834	1,097	1,653	0,828
	T Stat	3,951	5,460	4,480	4,608	5,843	6,888	9,565	7,681	7,163	6,743	2,564
	P-value	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,016
	Signif.	***	***	***	***	***	***	***	***	***	***	*
HML	Coef.	-0,203	0,049	0,060	0,153	0,107	0,345	0,247	0,248	0,268	0,043	0,247
	T Stat	-1,521	0,609	0,683	1,811	1,380	4,697	3,610	3,563	2,740	0,276	1,194
	P-value	0,139	0,547	0,500	0,080	0,178	0,000	0,001	0,001	0,010	0,785	0,242
	Signif.				.		***	**	**	*		
UMD	Coef.	-0,474	-0,318	-0,376	-0,232	-0,178	-0,215	-0,134	-0,215	-0,245	-0,140	0,333
	T Stat	-4,054	-4,493	-4,889	-3,143	-2,615	-3,352	-2,240	-3,541	-2,865	-1,025	1,846
	P-value	0,000	0,000	0,000	0,004	0,014	0,002	0,033	0,001	0,008	0,314	0,075
	Signif.	***	***	***	**	*	**	*	**	**		.
R^2		0,849	0,898	0,853	0,851	0,867	0,893	0,913	0,909	0,864	0,811	0,284
N		35	35	35	35	35	35	35	35	35	35	35

Table 6.11. *The Carhart alphas from the RNN prediction model using the three preceding years' controls as predictors. The controls include a time series of the stock's own return, the log of market capitalization and the book-to-market ratio. The models are estimated each year from the three preceding years of returns and the corresponding predictors. Then, the stocks are each year pooled into ten portfolios based on the forecast returns. The portfolio P1 contains the stocks with the lowest predicted returns and the portfolio P10 the highest. The Carhart factor coefficients are then estimated.*

Accrual and control RNN forecast

		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P10 - P1
Alpha	Coef.	-0,031	-0,001	0,015	0,011	0,014	0,013	-0,014	0,023	0,064	0,066	0,097
	T Stat	-1,507	-0,068	0,956	0,643	0,926	0,931	-0,930	1,303	2,654	2,574	2,899
	P-value	0,142	0,946	0,347	0,525	0,362	0,359	0,360	0,203	0,013	0,015	0,007
	Signif.									*	*	**
Mkt - Rf	Coef.	1,042	0,834	0,773	0,684	0,790	0,804	0,800	0,872	0,816	1,056	0,014
	T Stat	10,037	8,552	9,915	8,156	10,486	11,269	10,913	9,647	6,735	8,138	0,086
	P-value	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,932
	Signif.	***	***	***	***	***	***	***	***	***	***	
SMB	Coef.	0,723	0,698	0,851	0,677	0,613	0,725	1,038	0,926	1,277	1,347	0,624
	T Stat	4,508	4,634	7,060	5,219	5,269	6,570	9,170	6,632	6,822	6,717	2,387
	P-value	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,023
	Signif.	***	***	***	***	***	***	***	***	***	***	*
HML	Coef.	-0,289	-0,018	0,046	0,126	0,263	0,263	0,244	0,336	0,096	0,255	0,543
	T Stat	-2,814	-0,184	0,602	1,522	3,529	3,721	3,364	3,762	0,803	1,985	3,249
	P-value	0,009	0,856	0,552	0,139	0,001	0,001	0,002	0,001	0,428	0,056	0,003
	Signif.	**				**	***	**	***		.	**
UMD	Coef.	-0,334	-0,394	-0,314	-0,265	-0,331	-0,182	-0,036	-0,224	-0,302	-0,152	0,181
	T Stat	-3,722	-4,675	-4,657	-3,650	-5,086	-2,955	-0,575	-2,864	-2,882	-1,358	1,242
	P-value	0,001	0,000	0,000	0,001	0,000	0,006	0,569	0,008	0,007	0,185	0,224
	Signif.	***	***	***	***	***	**		**	**		
R^2		0,886	0,864	0,900	0,848	0,896	0,896	0,898	0,878	0,837	0,839	0,353
N		35	35	35	35	35	35	35	35	35	35	35

Table 6.12. *The Carhart alphas from the RNN prediction model using the three preceding years' scaled accruals and the controls as predictors. The controls include a time series of the stock's own return, the log of market capitalization and the book-to-market ratio. The models are estimated each year from the three preceding years of returns and the corresponding predictors. Then, the stocks are each year pooled into ten portfolios based on the forecast returns. The portfolio P1 contains the stocks with the lowest predicted returns and the portfolio P10 the highest. The Carhart factor coefficients are then estimated.*

7. CONCLUSIONS AND DISCUSSION

This chapter concludes the thesis. First, the results are discussed. Then, limitations related to this thesis are discussed. Finally, suggestions for further research are made.

7.1 Conclusions

This thesis examined the existence of the accrual anomaly in the US market in the years 1980-2020. First, Section 6.1 showed that the accrual anomaly exists in the sample. The panel regression that regressed stock returns on the accruals in the same and three preceding years verified that the accrual driven profits seem to drive stock returns upwards in the same year. However, as the coefficients for the three preceding years' accruals are negative, it is evident that the accruals seem to predict lower profits in the long run.

In Section 6.2, Section 6.3 and Section 6.4, the three machine learning models were tested to predict stock returns based on the accrual anomaly. First, the simple naive model where stocks were pooled into ten portfolios based on their scaled accruals and where the stock returns were examined in the three subsequent years provided very good ability to exploit the accrual anomaly. The portfolio with the lowest scaled accruals outperformed the portfolio with the highest accruals by almost 14 % as measured by the Carhart alpha.

Second, the linear regression model which predicted the returns using three years' scaled accruals and controls from the preceding years also displayed predictive power for the returns and ability to generate abnormal returns. The regression model was fitted each year and used three preceding years returns as training data. The stocks then were sorted into ten portfolios based on their forecast return and the realized returns were examined. It was shown that the portfolio with the highest predicted returns based on mere scaled accruals outperformed the portfolio with the lowest predicted returns by a Carhart alpha of 7 %. Adding log of market capitalization and the book-to-market value increased the similar Carhart alpha up to almost 14 %.

Third, three recurrent neural networks were tested if they could be utilized in predicting the accrual anomaly. The first included three years of time series of scaled accruals for each stock as explanatory variables. Then, a control sample using the stock's own three

preceding years' returns along with log of market capitalization and book-to-market was used. Finally, a combination of the scaled accruals and control time series was used in the predictions. Each model was trained for each year using three previous years' individual stock returns. Then, the predictions were made for each year and stocks were pooled each year based on the forecast return.

The mere accrual driven forecast generated a 4 % alpha between the lowest and highest predicted return portfolio. The controls forecast yielded a realized 10 % difference in alpha between the corresponding portfolios. Finally, using both scaled accruals and controls also yielded a 10 % alpha between the highest and lowest realized return portfolio.

To summarize, the three machine learning model exhibit predictive power on the accrual anomaly. Also, all the models seem to provide ability to generate a positive Carhart alpha.

7.2 Limitations

The biggest limitation in this thesis is the recurrent neural performance. The RNN did not offer a superb performance compared to the other models. So, the RNN architecture itself is a limitation. It may be that the RNN is overly complex to model the market misinterpretation of the accrual information. Thus, it may be reasonable to expect that the accrual anomaly driver is a simple over reaction to increased and decreased profits driven by accrual component.

The time span of the data sample is very long, 40 years. This means that the market dynamics has changed over the years and it may be that the nature of the accrual anomaly has changed. It is generally known that the information acquisition of the market has become more powerful over the years. Also, the reporting practices of the market have changed. For example, the Sarbanes-Oxley act of 2002 has made the reporting requirements stricter. Also, the financial crisis of 2007-2009 has changed the market behaviour.

Classic limitations for this trading strategy include trading cost and liquidity. It may be that the stocks causing the anomaly are not liquid enough to capitalize the anomaly. Also, it may be that arbitraging the anomaly away on the short side is not possible because of the stock availability for the short legs. However, the data availability is not a limitation as all information used in this thesis is publicly available and available for trading.

7.3 Suggestions for further research

A clear topic of further research for this thesis is to improve the recurrent neural network performance. As shown in Chapter 6, the current RNN did not yield a superb performance compared to the naive model or the linear regression. Additional improvements to the neural network models could be adding additional explanatory variables and perhaps

breaking down the accrual component in the spirit of Detzel et al. (2018). The additional explanatory variables could for example include volatilities. Also, it could be tested if changing the explanatory variable time series length or using quarterly accruals would change the results.

The architecture could also be enhanced by testing different layer architecture and different training setting. Also, alternative neural network variants such as the long short-term memory LSTM could be implemented.

REFERENCES

Aggarwal, C. (2018). *Neural Networks and Deep Learning: A Textbook*. Springer International Publishing AG. ISBN: 978-3-319-94462-3.

Ali, A., Chen, X., Yao, T. and Yu, T. (2008). Do mutual funds profit from the accruals anomaly?: *Journal of accounting research* 46.1, pp. 1–26. URL: https://doi.org/10.1111/j.1475-679X.2007.00263.x.

Ali, A. and Gurun, U. (2009). Investor Sentiment, Accruals Anomaly, and Accruals Management. *Journal of Accounting, Auditing & Finance* 24.3, pp. 415–431. URL: https://doi.org/10.1177/0148558X0902400305.

Ball, R., Gerakos, J., Linnainmaa, J. and Nikolaev, V. (2016). Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics* 121.1, pp. 28–45. URL: https://doi.org/10.1016/j.jfineco.2016.03.002.

Bishop, C. (1995). *Neural Networks for Pattern Recognition*. Oxford University Press. ISBN: 978-0198538646.

Bradshaw, M., Richardson, S. and Sloan, G. (2001). Do Analysts and Auditors Use Information in Accruals?: *Journal of accounting research* 39.1, pp. 45–74. URL: https://doi.org/10.1111/1475-679X.00003.

Carhart, M. (1997). On Persistence in Mutual Fund Performance. *The Journal of Finance* 52.1, pp. 57–82. URL: https://doi.org/10.1111/j.1540-6261.1997.tb03808.x.

Cauchy, A. (1847). Méthodes générales pour la résolution des systèmes d'équations simultanées. *Comptes Rendus Hebd. Séances Acad. Sci.* 25, pp. 536–538.

Chambers, D. and Payne, J. (2008). Audit Quality and the Accrual Anomaly. *SSRN*. URL: http://dx.doi.org/10.2139/ssrn.1136787.

Chan, K., Chan, L., Jegadeesh, N. and Lakonishok, J. (2006). Earnings quality and stock returns. *Journal of Business* 79.3, pp. 1041–1082. URL: https://doi.org/10.1086/500669.

Dechow, P., Sloan, R. and Sweeney, A. (1995). Detecting earnings management. *The Accounting Review* 70, pp. 3–42. URL: https://doi.org/10.1111/j.1475-679X.2012.00449.x.

Detzel, A., Schaberl, P. and Strauss, J. (2018). There are two very different accruals anomalies. *European financial management: the journal of the European Financial Management Association* 24.4, pp. 581–609. URL: https://doi.org/10.1111/eufm.12162.

Fama, E. and French, K. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33.1, pp. 3–56. URL: https://doi.org/10.1016/0304-405X(93)90023-5.

French, K. (2021). *Current Research Returns, Monthly*. [Online; accessed January 3rd 2022]. URL: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

Goodfellow, I., Bengio, Y. and Courville, A. (2016). *Deep Learning*. MIT Press. ISBN: 978-0262035613. URL: http://www.deeplearningbook.org.

Green, J., Hand, J. and Soliman, M. (2011). Going, Going, Gone? The Apparent Demise of the Accruals Anomaly. *Management science* 57.5, pp. 797–816. URL: https://doi.org/10.1287/mnsc.1110.1320.

Hinton, G., Srivastava, N., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*. URL: https://doi.org/10.48550/arxiv.1207.0580.

Hirshleifer, D., Kewei, H. and Teoh, S. (2006). The Accrual Anomaly: Risk or Mispricing?: *Management science* 42.1–2, pp. 3–33. URL: https://doi.org/10.1287/mnsc.1100.1289.

Kingma, D. and Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*. URL: https://doi.org/10.48550/arXiv.1412.6980.

LaFond, R. (2005). Is the Accrual Anomaly a Global Anomaly?: *SSRN*. URL: http://dx.doi.org/10.2139/ssrn.782726.

Lewellen, J. and Resutek, R. (2016). The predictive power of investment and accruals. *Review of accounting studies* 21.4, pp. 1046–1080. URL: https://doi.org/10.1007/s11142-016-9369-8.

Mashruwala, C., Rajgopal, S. and Shevlin, T. (2006). Why is the accrual anomaly not arbitraged away? The role of idiosyncratic risk and transaction costs. *Journal of Accounting and Economics* 42.1–2, pp. 3–33. URL: https://doi.org/10.1016/j.jacceco.2006.04.004.

Pincus, M., Rajgopal, S. and Venkatachalam, M. (2007). The Accrual Anomaly: International Evidence. *The Accounting Review* 82.1, pp. 169–203. URL: https://doi.org/10.2308/accr.2007.82.1.169.

Polyak, B. (1964). Some methods of speeding up the convergence of iteration methods. *USSR Computational Mathematics and Mathematical Physics* 4.5, pp. 1–17. URL: https://doi.org/10.1016/0041-5553(64)90137-5.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review* 65.6, pp. 386–408. URL: https://doi.org/10.1037/h0042519.

– (1961). *Principles of Neurodynamics - Perceptrons and the Theory of Brain Mechanics*. Cornell Aeronautical Laboratory, Inc. URL: https://apps.dtic.mil/sti/pdfs/AD0256582.pdf.

Sharpe, W. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance* 19.3, pp. 425–442. URL: https://doi.org/10.1111/j.1540-6261.1964.tb02865.x.

Sloan, R. (1996). Do stock prices fully reflect information in accruals and cash flows about future earnings?: *The Accounting Review* 71.3, pp. 163–187.

Stigler, S. (1981). Gauss and the Invention of Least Squares. *The Annals of Statistics* 9.3, pp. 465–474. URL: https://doi.org/10.1214/aos/1176345451.

Vikelä, J. (2021a). Active management, investment style and mutual fund performance - Evidence from the Nordic market. Master's thesis. Aalto University School of Science. URL: http://urn.fi/URN:NBN:fi:aalto-2021121910850.

– (2021b). Earnings announcement premium and institutional holding - Evidence from the US market. Master's thesis. Aalto University School of Business. URL: http://urn.fi/URN:NBN:fi:aalto-202110319850.

Warde-Farley, D., Goodfellow, I., Courville, A. and Bengio, Y. (2013). An empirical analysis of dropout in piecewise linear networks. *arXiv preprint arXiv:1312.6197*. URL: https://doi.org/10.48550/arXiv.1312.6197.

Zhang, X. F. (2007). Accruals, Investment, and the Accrual Anomaly. *The Accounting Review* 82.5, pp. 1333–1363. URL: https://doi.org/10.2308/accr.2007.82.5.1333.