



**MODELLING OF STOCK RETURNS IN MARKET CRASH AND RECOVERY
PERIOD USING STOCK- AND FIRM LEVEL CHARACTERISTICS**

Case of Covid-19 and Financial crisis

Lappeenranta–Lahti University of Technology LUT

Master's Programme in Strategic Finance and Analytics, Master's thesis

2021

Petrus Halkola

Examiners: Professor Eero Pätäri

Junior Researcher Mahinda Mailagaha Kumbure

ABSTRACT

Lappeenranta–Lahti University of Technology LUT

LUT School of Business and Management

Business Administration

Petrus Halkola

Modelling of stock returns in market crash and recovery period using stock- and firm level characteristics – Case of Covid-19 and Financial crisis

Master's thesis

2021

75 pages, 18 figures, 15 tables and 4 appendices

Examiners: Professor Eero Pätäri and Junior Researcher Mahinda Mailagaha Kumbure

Keywords: stock returns, market crash, characteristics, linear regression, machine learning

This thesis studies cross-sectional stock returns in market crash and recovery periods. Modelling is performed by using stock- and firm level characteristics and three modelling approaches: linear regression, random forest and support vector regression. The aim is to examine which characteristics are important determinants of the returns and whether machine learning algorithms can add value to modelling accuracy. Data over Covid-19 and financial crisis periods from Nordic markets is employed. These two periods are modelled individually but financial crisis models are also tested on Covid data.

The results differ between the two periods. Better modelling accuracy is documented for financial crisis data but applying financial crisis -based models on Covid data does not appear useful. In financial crisis, dividend yield, earnings yield and share turnover were the most important drivers of returns, whereas in Covid crash, they were operating leverage, volatility and share turnover. Machine learning methods outperform linear regression, but the advantage is minor. If an investor had followed the models, total returns of up to 75-96% could have been achieved, while market returns were 0%. There are similarities in important characteristics found in this study with the presented previous research, but the results are not a repetition of previous studies. This indicates that important characteristics vary over time and context of the crash.

TIIVISTELMÄ

Lappeenrannan–Lahden teknillinen yliopisto LUT

LUT-kauppakorkeakoulu

Kauppätieteet

Petrus Halkola

Osaketuottojen mallinnus markkinaromahduksissa osake- ja yritystason tunnusluvuilla – Covid-19 ja Finanssikriisi

Kauppätieteiden pro gradu -tutkielma

75 sivua, 18 kuvaa, 15 taulukkoa ja 4 liitettä

Tarkastajat: Professori Eero Pätäri ja Nuorempi tutkija Mahinda Mailagaha Kumbure

Avainsanat: osaketuotot, markkinaromahdus, tunnusluvut, lineaarinen regressio, koneoppiminen

Tämä tutkielma tarkastelee osakkeiden poikkileikkaustuottoja markkinoiden romahdustoipumisjaksoilla. Mallinnuksessa käytetään osake- ja yrityskohtaisia tunnuslukuja sekä kolmea mallinnustapaa: lineaarista regressiota, random forest -menetelmää ja tukivektoriregressiota. Tavoitteena oli selvittää, mitkä tunnusluvut ovat tuottojen mallinnuksessa tärkeitä ja voivatko koneoppimismenetelmät tuoda lisäarvoa mallinnustarkkuuteen. Tutkimuksessa käytettiin dataa Covid-19 romahduksesta sekä finanssikriisistä pohjoismaisilla markkinoilla. Näitä kahta ajanjaksoa mallinnettiin erikseen, mutta finanssikriisipohjaisia malleja testattiin myös Covid-romahduksen datalla.

Tulokset eri ajanjaksoille eroavat toisistaan ja parempi mallinnustarkkuus saavutettiin finanssikriisidatalla, mutta finanssikriisipohjaisten mallien soveltaminen Covid-dataan ei vaikuta hyödylliseltä. Finanssikriisissä osinkotuotto, tulostuotto ja osakkeen kiertonopeus olivat tärkeitä tunnuslukuja mallinnuksen kannalta. Covid-datassa taas operatiivinen velkaantuneisuus, volatilitiiteetti ja osakkeen kiertonopeus olivat tärkeitä. Koneoppimismenetelmät toimivat lineaarista regressiota paremmin, mutta hyöty on vähäistä. Jos tutkimuksen malleja olisi hyödyntänyt näillä periodeilla, olisi voitu saavuttaa jopa 75-96% kokonaistuottoja, markkinatuoton ollessa 0%. Tässä tutkimuksessa löydettyissä tärkeissä tunnusluvuissa on yhtäläisyyksiä aikaisempaan tutkimukseen, mutta tulokset eivät ole toisintoa aikaisemmista tutkimuksista. Tämän perusteella tärkeät tunnusluvut vaikuttavat vaihtelevan ajan ja romahduksen kontekstin mukaan.

Table of contents

Abstract

1. Introduction	5
1.1 Research hypotheses and questions	7
1.2 Structure and limitations	8
2. Literature review	9
2.1 Financial modelling and forecasting.....	9
2.2 Stock return determinants on stock- and firm level	13
2.3 Machine learning in finance	17
3. Data and research methodology	20
3.1 Case introduction	20
3.1.1 Financial crisis of 2007-2009	21
3.1.2 Covid-19 Crash.....	22
3.2 Detailed data analysis and description.....	24
3.3 Models.....	28
3.3.1 Linear regression	28
3.3.2 Random Forest Regression.....	29
3.3.3 Support vector regression.....	32
3.3.4 Model assessment and variable importance	35
4. Empirical results.....	39
4.1 Financial crisis.....	39
4.2 Covid crash.....	51
4.3 Applying financial crisis -based models on Covid period	63
5. Discussion	65
6. Conclusions	68
References	70

Appendices

Appendix 1. Distributions of variables in financial crisis data.

Appendix 2. Distributions of variables in Covid crash data.

Appendix 3. Residual plot and distribution in linear regression on financial crisis data.

Appendix 4. Residual plot and distribution in linear regression on Covid data.

1. Introduction

Stock markets are an instrumental vehicle for distributing capital between those that have an excess of it and those who are in the need of it. Specifically, the secondary markets have been vital for functioning capital distribution by ensuring that investors are able to sell their shares if needed. Thus, they have been an important part of the capitalistic economy and have provided economic development and welfare to many economies globally. However, because the principal nature of these markets is that prices of assets are determined based on mutual agreement on the price between a buyer and a seller, prices tend to fluctuate significantly. Sometimes these fluctuations can be extreme, and asset prices can experience great declines and increases in short time periods.

Just by looking at the history, it can be seen that significant market-wide declines are experienced from time to time and commonly these can be referred as market corrections or in more extreme events as crashes. Definition of market crash varies, and Mishkin & White (2002) suggest that a 20% drop in stock price index within a period lasting up to one year should be defined as a crash. This definition is used in this thesis and examples of such events include the black Monday of 1987, financial crisis of 2007 and the most recent one as 2020 Covid-19 related market panic. If we think of the pinnacle financial theory of efficient markets hypothesis (EMH) by Eugene Fama (1970), its main explanation for these significant market movements is new and unexpected information that is absorbed into asset prices.

The problem is that EMH relies on assumption of constant investor rationality, and this is widely disputed, even though in many cases, such as the examples of financial crisis and 2020 market panic caused by coronavirus pandemic, unexpected new information has been the igniter of market turbulence. Perhaps it's precisely the lasting turbulence and panic-like movements in these events that cast shadows on investor rationality. Problems with this assumption have even sparked a research field of its own, known as behavioural finance.

Vogel (2018) lists key behavioural elements in bubbles and crashes to be widespread uncertainty and speculation, investor irrationality in the form of FOMO (fear of missing out) and FOSI (fear of staying in), and that in these events considerations of price and fundamental values are replaced by considerations of quantity. He also states that in the end, the dominant cause for the turbulence becomes irrelevant in the markets and in this case deep fundamental analysis and knowledge of companies and industries could provide above average returns.

Perhaps one of the most successful investors, Warren Buffett, has stated that his philosophy is that in the financial markets “one should be fearful when others are greedy and greedy when others are fearful” (Buffett, 1987). Still when investors know this, they usually end up doing the opposite (Zweig & Graham, 2003). The question is then, how can one spot the good investment opportunities in times of turbulence? One approach to this is to ask, which stocks or companies will exhibit highest returns during a crash and recovery period?

This thesis aims to study this question by approaching the topic with linear and machine learning (ML) methods during the stock market crash and recovery caused by Covid-19 pandemic in 2020 and during the financial crisis market crash and recovery period in 2007-2015. The research method is to use a group of stock- and firm level characteristics of individual stocks and apply linear regression and machine learning approaches to model the crash and recovery period returns. In addition, the models’ predictive ability is tested on out-of-sample data and on Covid period using models built on the financial crisis data. The stock markets being analyzed are two Nordic stock exchanges of Helsinki and Stockholm.

Advantages of ML methods in stock selection and modelling returns are that a vast variety of factors can be supplied to ML algorithms (MLAs) from which the algorithms themselves determine the ones that matter and their relationship to the dependent variable, thus offering a way to combine weaker information into stronger investment signals. ML methods can be used to find hidden and dynamic relationships between the variables, that are difficult to identify with traditional statistical and linear approaches. In addition, ML methods are usually more effective in cases involving multicollinearity, for example. A distinct

advantage of MLAs in practical uses is that they are not subjected to the same statistical assumptions, or restrictions, as linear regression. (Rasekhschaffe & Jones, 2019)

In addition to the aim of this study, the objective is also to compare the performance of linear regression to more complex machine learning approaches and broaden the landscape of machine learning applications in finance. Using financial indicators that are easy to understand and available to all investors offer a way to use ML so that the results are still easily understandable and depending on the results, the research should provide information, of which investors can take advantage in future market turbulences when prospecting for investment opportunities. Empirical contribution of defining the stock- and firm-level characteristics in the Nordic markets that are associated with larger returns, and thus faster recovery, is also a target and this would be intriguing information to investors.

1.1 Research hypotheses and questions

Research hypotheses of this thesis are: stock- and firm level characteristics are important indicators of how much a stock produces returns in market crash and recovery period, and that machine learning models will perform better at predicting crash and recovery period returns than a traditional statistical model.

Main research questions:

How accurate predictions on returns can be made using machine learning and linear regression?

What are the most important characteristics in modelling the returns?

Sub-questions:

What are the positive and negative drivers of crash and recovery period returns?

Which of the selected machine learning methods performs the best?

1.2 Structure and limitations

This thesis consists of a review of previous research and literature and an empirical part of modelling crash and recovery period returns with three different modelling approaches. Literature review focuses on the topics of financial modelling and forecasting, machine learning in finance and stock return determinants on stock- and firm level. Empirical part is divided into two parts: First the reader is familiarized with the data and methodology and after that, results of the modelling are presented. After that the results and their relevance is discussed and finally, the conclusions are presented.

Limitations to this study are that even though interesting results could be achieved, it is still problematic to justify their extrapolation to future events. This is because research will be carried on data which is unique to both of the research periods. However, if similarities are found in this and previous studies of financial ratios and performance indicators relationship on stock return, reasonable conclusions could be drawn. In addition, the research is limited to include only Helsinki and Stockholm exchanges.

2. Literature review

Previous research and literature giving a base to the study is presented in this section, by briefly exploring the topics of financial modelling, stock return determinants and machine learning. In the empirical section these topics are combined and applied to crash periods to see whether new insights can be produced.

2.1 Financial modelling and forecasting

Financial modelling can be seen as an umbrella term for various differing modelling activities in the broad field of finance, financial management and economics. Avon (2015) describes financial modelling as a construction of a theoretical model that attempts to depict a project, process, transaction or an investment by identifying and presenting key variables and their logical and quantitative relationships.

Avon (2015) writes of financial modelling as a variety of tasks including data- and scenario analysis, financial information processing, project management and software development. His examples include, among others, risk modelling, pricing models and other quantitative models in investment banking and capital budgeting, financial statement analysis and investment project modelling in corporate finance. However, his approach is perhaps inclined on the accounting perspective and thus for the purpose of this study it is useful to also investigate the definition of econometrics as another definition of financial modelling.

Financial econometrics is an extension of the traditional econometrics, which means “measurement in economics”, to applications in the financial markets and can be defined as application of statistical techniques to quantitative problems in finance. Financial econometrics is a tool, or an approach, commonly used for testing different financial theories, determining asset prices and returns, testing and finding out of different relationships between economic and financial variables, quantitative analysis of financial

markets under different conditions and forecasting of future values for financial variables. (Brooks, 2008)

It is clear, that this kind of definition works better for this thesis, as here the perspective is also prediction and modelling of financial asset prices, in this case stocks. One aspect of financial econometrics is also that the used data is often of high frequency, and based on historical events, thus measurement error is usually not a problem. However, high frequency data also creates a problem of noisy data, i.e., that in financial data it is more difficult to distinguish underlying relationships or patterns from random features. Data used for this kind of financial modelling is either cross-sectional, focusing on one or several variables at a single point in time, time-series, with observations collected over time, or panel data, which combines cross-sectional and time-series aspects, for example monthly price data of several stocks over a 2-year period. (Brooks, 2008) Cross-sectional approach is used in the empirical analysis of this thesis.

Forecasting or predicting is strongly linked to the previous definition of financial modelling, but it is not exactly the same, as financial modelling can be conducted without the purpose of prediction, for example to study relationships of variables. It is needed to distinguish that forecasting in finance can be applied to many different targets, for example default prediction, forecasting of companies' financial figures, forecasting volatility or risk of a stock, predicting asset returns or prices, forecasting macroeconomic conditions and so on. Forecasting landscape is wide and the used methods vary depending on the target.

Penman (2010) defines that from a pure statistical sense forecasting means that the target of prediction is drawn from a conditional distribution as the expected value determined by transitional parameters that are applied to the variables used for prediction, and the error of the forecast are determined by some distribution of unpredictable realizations around the expected value. Generally, this is called the generating process and in a statistical sense the parameters of the process are estimated from behaviour in the used data.

Penman (2010) also makes the important note, that the generating process is usually governed by some laws of nature (or man-made), which can be utilized in forecasting. Financial processes, be it accounting or the determination of an asset's price, are also governed by some laws, and thus should be possible to predict with some degree of error if principles of the generating processes are known at least to some extent. For example, in this thesis the question is whether or not stock- and firm-level characteristics form a significant part of the return generating process.

To further specify financial forecasting Brooks (2008) divides it into time series forecasting, where future values of a variable are predicted using previous values and possibly previous values of the error term, and structural forecasting, where a dependent variable is predicted using independent variables. Forecasting returns based on arbitrage pricing models and using long-run relationships of variables arising from market efficiency framework are examples of structural forecasting and where these models typically perform well. Evidently, structural forecasting, or modelling, is where this thesis categorizes better, but it is notable that the division of these forecasting categories is sometimes blurred. In the context of this thesis an example includes using lagged values and previous cumulative returns, but still structural forecasting is a better depiction.

The modelling approach of this study is explaining the cross-sectional equity returns with various indicators, or factors, that are suspected to have an identifiable relationship with individual returns. The theoretical roots of this type of research are found in Markowitz's modern portfolio theory (MPT), capital asset pricing model (CAPM), and specifically the arbitrage pricing theory (APT) which can be seen as an extension of the CAPM. The asset pricing theories build on the MPT's idea of optimal mean-variance efficient portfolios and diversification benefits. In CAPM theory, it is assumed that as idiosyncratic risk can be diversified away, only the systematic risk component "beta", depicted as the co-variation with market returns, is important for modelling expected returns (Perold 2004). However, in many empirical tests, CAPM has been unsatisfactory in explaining the returns (Fama & French 2004).

The issues with CAPM led to the development of APT by Ross (1976), which extends the idea by incorporating more explanatory factors to the linear return function. Originally it is suggested that these factors should have undiversifiable elements, thus being mostly macroeconomical. (Chen, Roll & Ross 1986) The APT theory and its idea of explanatory factors, which can be time- and market varying, has still sparked wide research on potential factors to explain asset returns and these are not limited to macro-level. Famous models incorporating idiosyncratic characteristics are for example Fama-French three-factor model (Fama & French 1993) and its four-factor extension by Carhart (1997). The search for return-explaining factors continues to this day, and in the next chapter research from employing idiosyncratic characteristics to return modelling is presented.

Final note on the topic of financial modelling is of assessing the forecasts. In order to assess the effectiveness and feasibility of using the predictions, it is needed to have suitable forecast benchmarks, prediction accuracy measures and evaluation of effective model implementation i.e. for example testing for assumptions of linear regression and out-of-sample testing. Guerard (2013) However, Brooks (2008) states that it is sometimes argued, that if a model produces accurate predictions, but contains insignificant variables or violates model assumptions, such problems of statistical significance are largely irrelevant.

Important aspect of assessing the modelling is the difference between out-of-sample and in-sample fit. In-sample fit refers to the case where the assessment of model or forecasts is conducted on the same data which was used to build the model i.e., the full sample is used to build the models. In the case of out-of-sample fit, a holdout sample is separated, and the model is built on the rest of the sample. Assessing model performance is then conducted by applying the model on the holdout sample. (Inoue & Kilian 2004) In the field of predicting stock returns using financial variables Rapach & Wohar (2006) note that evidence of predictability is typically based on in-sample tests and comment it to be somewhat contradictory to the commonly accepted principle, where out-of-sample tests are considered as important merits of significant, reliable and repeatable results. Out-of-sample fitting is seen as guarding against model overfitting and the perils of data mining. Both types of sample fit will be studied in this thesis.

2.2 Stock return determinants on stock- and firm level

The relationship of different factors and performance indicators, such as financial ratios, on the performance of stocks and their predictive ability on stock return is a quite vastly studied topic in the field of financial research. One of the reasons for this is that for example financial ratios are simple to understand and easily available making them popular among investors in attempt to evaluate investment opportunities and creating investment strategies. Important example of research in this field is the three-factor model by Fama & French (1993) which was then later expanded to a five-factor model (Fama & French, 2015) with the findings that firm characteristics of size, value, profitability and investment activity are important in explaining returns.

Research made by Pech, Noguera & White (2015) studied the real-world use of financial ratios in investment analysis through equity analysts' recommendation reports. They found that most popular ratio types used were profitability and margins, leverage, price multiples and cash flow ratios. Five most used single ratios were EPS, P/E, Firm value to EBITDA, Sales growth and Dividend yield. Pech et al. also found empirical evidence of predictive power on 1-year returns based on estimates of the ratios most used by analysts.

Important empirical evidence of predictive power of ratios is widely available, for example Musallam (2018) finds strong positive results between asset returns and EPS, earnings yield and dividend yield in Qatari listed stocks, Chairakwattana & Nathaphan (2014) study on Thai markets singled out book-to-market as the most important ratio in predicting returns and Petcharabul & Romprasert (2014) also contributed on the Thailand markets and found that ROE and P/E have a significant relationship to stock returns.

Evidence of the predictive power of financial ratios is also found in the US markets by for example Lewellen (2004). He finds supporting evidence to Fama's & French's (1988) study that dividend yield can predict stock returns, and also that B/M and E/P ratios have predictive power. It is important to note however, that the findings of previous studies are in many

cases conducted on data that does not include market crashes, thus the importance of financial ratios and other indicators do not imply that the same factors would be important during times of extreme market movements and times of distress. In these times it could be possible that for example measures of liquidity, leverage and volatility become more important.

Wang, Meric G., Liu & Meric I. (2009) approached return modelling during crash periods with linear models and event study methodology using company or stock specific characteristics and also industry related characteristics. Their argument was that even though traditional pricing models such as CAPM and Fama & French three factor model have empirical evidence backing them, the tests on these have not been conducted on crash period data. Their findings on previous research also implied, that factors like company-specific idiosyncratic risk and illiquidity characteristics can be important determinants of return during crash periods.

Wang et al. (2009) found that lowest returns in market crashes are found in stocks with high betas, large capitalization, low illiquidity, high volatility prior to crash. Financial ratios connected with lower returns were also high debt ratios, high levels of liquid assets, low cashflow per share and low profitability. They also found a notable reversal effect for cumulative returns earned three months and -years prior to crash. It is notable that their research is mainly focused on the period of market decline, and not the recovery. However, they did find that size is an important factor for short period returns after the crash, and high-cap firms recovered faster than small-cap firms.

Relationship between stock prices and financial ratios during a time of financial distress has been researched for example by Džikevičius & Šaranda (2011). They studied whether stock prices could be forecasted with financial ratios of the given company by measuring correlations and covariances between ratios and stock returns in the Lithuanian markets. The research period is particularly interesting considering the topic of this thesis, as it was 2007-2010 i.e., the financial crisis period. Džikevičius & Šaranda used 20 financial ratios, related to profitability, capital structure, liquidity, solvency and turnover and found that the used

ratios and stock returns were in all cases dependent values with varying strength of dependence.

On the individual stock level, highest positive correlations (>0.70) between returns and ratios were found in liquidity measures of cash ratio, quick ratio and current ratio and also operational profitability ratios of gross profit margin and operating margin. Intuitively it seems a plausible idea that in a time of financial distress, investors value these aspects in companies. Highest negative correlations are found in liabilities to assets or equity ratios, which again seems intuitive, high leverage is unappealing to investors during market turbulence and hence correlates negatively with stock returns. However, when looking at the overall results of the study, Dzikevičius & Šaranda have found that average form or stronger correlations are most frequently found in asset turnover and capital structure ratios. Notable is that the sample size of the study is small and perhaps their most important finding is that during a time of financial distress, financial ratios seem to have been related to stock returns and could possibly be used in prediction of returns.

Of particular interest in this study is what indicators can explain the returns in a crash period. Fauzi & Wahyudi (2016) approached this question by studying three crashes with a multivariate regression methodology. They incorporate both stock- and firm-level characteristics in their model. Stock level characteristics used were size, beta, book-to-market ratio, stock illiquidity, lagged returns, and volatility. Firm level characteristics studied were leverage, asset liquidity, cashflow per share and profitability. Most prominent characteristics for predicting crash-period returns were found to be beta, market cap, volatility, leverage, firm asset liquidity and profitability.

Baker and Wurgler (2006) studied the cross-sectional stock returns through market sentiment approach. Their hypothesis, for which evidence was found, is that even though many firm characteristics do not seem to exhibit predictive power at first, actually hide strong patterns that are conditional to the prevalent market, or investor sentiment. They used market cap, firm age, volatility, book-to-market ratio, earnings-to-book equity, dividend-to-book equity, fixed assets-to-total assets, R&D costs-to-assets, external finance-to-assets and sales growth

to explain returns, finding that when the beginning-of-period sentiment is high, i.e. peak of market cycle, the following period (decline or even a crash), returns are low for young, small, unprofitable, non-dividend paying, high volatility, and extreme growth stocks.

Explaining cross-sectional returns with firm- and stock-level factors is a widely studied subject that has both evidence for and against predictability. Perhaps the most well-known and widely accepted research is provided by Fama & French (1993 & 2015) with their three- and five-factor models. However, for example McLean & Pontiff (2016) study 97 different variables that are shown to explain cross-sectional returns in previous research and find that these anomalies tend to erode after being published, thus having little relevance for generating excess return in the future. Similarly, Welch & Goyal (2008) study the most prominent variables explored in previous literature, of which firm- and stock-level ones are dividend yield, earnings yield, dividend payout ratio, book-to-market ratio, volatility and issuing activity. They find that the previously presented models or variables have poor out-of-sample performance and were unable to find any model on an annual or shorter period that would inhibit good in-sample or out-of-sample performance. However, they propose to attempt using more sophisticated models for better predictability.

While the bulk of return determinant research is conducted on the US markets, a brief visit in different markets is provided by Artmann, Finter & Kempf (2012) in German markets, Hahn & Yoon (2016) in Korea and Bannigidadmath & Narayan (2016) in India. In Germany, Artmann et al. find that among beta, size, book-to-market, earnings yield, leverage, return-on-assets, asset growth, momentum and reversal indicators the most important drivers are book-to-market ratio, earnings yield and momentum. In Korea, the examined variables are beta, size, book-to-market, leverage, earnings yield, share turnover, share illiquidity, momentum, growth and foreign ownership factor. Most prominent variables according to Hahn & Yoon are size, book-to-market, earnings yield and share turnover. Lastly, in the Indian market, out of dividend payout ratio, earnings yield, dividend yield, dividend-to-current price, and book-to-market, the most important determinants are book-to-market ratio, dividend-to-current price and earnings yield. (Bannigidadmath & Narayan, 2016)

Research has also been carried out on more novel measures. Widely notified indicators of this category are ILLIQ measure proposed by Amihud (2002) and Operating leverage (Novy-Marx, 2011 and García-Feijóo & Jorgensen, 2010). Amihud presents strong evidence of predictability of returns with stock illiquidity, and his proposed illiquidity measure is interpreted as price impact, or how strongly the price responds to one dollar of trading volume. Operating leverage has various proposed methods for calculation, however, at its core it measures how much in relative terms a company has operational costs and assets, which are difficult to adjust in the short term. Operational leverage is proposed as a partial explanation for the value premium, especially in market downturns.

Supporting evidence for return predictability of many previously presented variables is found by Haugen & Baker (1996). They employ several risk, liquidity, price level, growth, technical and sector factors for predicting cross-sectional returns in five major markets over varying timeframes. Findings are that several variables, different periods of prior returns, book-to-price ratio, cashflow-to-price, earnings yield, sales-to-price, debt-to-equity, volatility and return on equity give robust predictions of returns, challenging the efficient markets hypothesis.

2.3 Machine learning in finance

Machine learning can be divided into three categories: supervised learning, unsupervised learning and reinforcement learning. Supervised learning involves feeding readily labelled data to the used algorithm from which the algorithm attempts to learn relationships and then create a model for prediction of new unlabelled data. Common supervised learning methods include support vector machines, decision trees and neural networks but also linear and logistic regression-based methods fall in this category. Unsupervised learning means feeding unlabelled data to the algorithm and letting the algorithm itself learn from the data. Unsupervised learning is used in the context of clustering models, such as hierarchical and k-means clustering, self-organizing maps but also some neural network models fall in this category. Reinforcement learning is a generalization of supervised learning, an approach to

Markov Decision Process, meaning that instead of considering a single action at each point in time, it considers the optimal sequence of actions. (Bilokon, Halperin & Dixon, 2020)

Machine learning is a topic of interest for many practitioners looking for an edge in stock picking. Rasekhschaffe & Jones (2019) conducted interesting research on stock selection using machine learning. They tested the performance of multiple ML methods against ordinary least squared regression -based forecasting, using for example Fama-French-Carhart factors and other financial ratios such as growth in EPS, 1-month reversal and ROE, in the end totalling to 194 factors or company characteristics. Their findings were that ML based strategies performed significantly better than traditional linear methods. Algorithms used in their study were AdaBoost (decision tree based), gradient boosted regression tree, a neural network and a support vector machine based bagging estimator.

Research on stock performance applying ML approach seems to be commonly done by employing decision tree and support vector machine -based algorithms. For example, Delen, Kuzey & Uyar (2013) employed four different decision tree algorithms using 31 financial ratios such as liquidity ratios, turnover ratios, growth ratios and asset structure ratios. Their dependent variables were ROE and ROA, and findings were that two most important factors for predicting were earnings before taxes-to-equity ratio and net profit margin. Example of SVM (Support Vector Machine) usage is found for example in Sun, Jia & Li (2011) where they model financial distress prediction using AdaBoost with single attribute test, AdaBoost with decision tree and a support vector machine.

One important thing to address is that while ML methods are good for finding subtle patterns and useful with collinear variables, they are also susceptible to overfitting. To account for this, forecast combinations and feature engineering can be used. Forecast combinations can be done with combining different algorithms, training windows, horizons, subsetting factors, and by using boosting and bagging methods. Feature engineering means for example predicting discrete variables, which MLAs are usually effective in, e.g. outperformer or underperformer. Other aspects include standardization of factors, choosing training window etc. (Rasekhschaffe & Jones, 2019)

Important and widely used methods in similar type of research problems as in this thesis are classification and regression tree models (CART). These models are based on the decision tree framework, where the algorithm goes through observations of the sample item and draws conclusions of the target value. Observations are represented as branches and target values as leaves, hence the name decision tree. CART methods are types of supervised learning, which means that the model is trained on pre-labelled data, and then it is employed on unlabelled data to test performance and make predictions. In classification trees, the model predicts discrete classes based on simple decision thresholds on variable values. On the other hand, regression trees predict continuous values, and therefore, they can be used to predict directly the amount of stock return. Problem with decision tree methods is that they are prone to overfitting, and this is why ensemble methods such as Random Forest are usually employed. Random Forest uses an applied method of bootstrap aggregating (bagging) to build several decision trees and draw conclusions on the predicted value based on all of the built trees. Using Random Forests also enables to assess feature importance, which variables have highest predictive power. (Joshi, 2020)

Support vector machines are also a common technique for financial prediction models. SVM models are also based on supervised learning, and similarly to decision trees, can be employed to classification of binary or multiple classes and also to regression in the form of support vector regression proposed by Drucker, Burges, Kaufman, Smola & Vapnik (1997). In SVMs, the algorithm maps points in space based on the points' values, which can be in multiple dimensions. Then based on the learning data, the algorithm builds a model that fits lines to separate the point clouds as well as possible. New unlabelled data is then labelled by the model based on to which cloud, separated by the lines, they fall or in regression the fitted lines are used similarly to linear regression to predict continuous values. One distinctive advantage of SVM over linear regression methods is that it can also be applied for nonlinear data with the kernel function approach proposed by Boser, Guyon & Vapnik (1992), to allow support vector lines to be nonlinear by mapping the data to higher dimension based on the kernel function used.

3. Data and research methodology

Data and research methodology are presented in this section as follows: An introduction to the two “cases” of which the data is collected is presented first, then data collection and pre-processing is explained. Descriptive analysis of the used dataset is provided, along with data splitting methodology. Three modelling approaches are presented in Models -section and their assessment methods are summarized.

The research setting is following; Individual stock returns of crash and recovery period of the overall markets are collected, and pre-crash indicators are used to model and predict the returns. The idea is to find out which of the chosen indicators are related with crash and recovery period returns and whether an investor amidst a crash could use the indicators to draw conclusions on where to invest and where not. Region of research is the Nordic markets.

Research is carried out using statistical software R as it is well equipped with readily available packages offering tools and methods needed for this analysis. Excel is used for fetching data from Refinitiv Eikon database and for pre-processing of data.

3.1 Case introduction

Two separate cases of market crashes are included, the first one being the financial crisis of 2007 and the second being the crash induced by Covid-19 pandemic. The purpose of including two different crashes is to potentially amplify the results in either case, in the event of finding similarities between the two periods. This would provide more robust evidence and encourage to use the findings of this study in future crash periods. In addition, the two separate periods also allow us to build a model on the data from earlier crash and test this on the data from the Covid crash, again bringing us to the question of whether we can “learn” from the previous crashes or not.

3.1.1 Financial crisis of 2007-2009

Financial crisis of 2007-2009 is perhaps the most well-known market crash among common public and refers to the widespread realization of large systemic risks around the year 2008 in the global stock markets and banking system. The crisis began in the US already in 2007, when the US stock markets began to have a downturn. It was rooted in the loose risk management in the market for subprime loans, especially in the housing segment, and is thus often referred to as the Subprime crisis. (Acharya, Philippon, Richardson & Roubini, 2009) Already in 2007 Federal Reserve chairman Alan Greenspan predicted a recession to take place at the end of the year (Twin, 2007).

A pivotal point happened in September 2008, when one of the largest American investment bank, Lehman Brothers, was filed for bankruptcy. This led to global panic among the banking system and stock markets. While US authorities responded relatively quickly to manage the crisis, the response in Europe was slower, and it is speculated that this was one reason for slower recovery in Europe. The effects of the financial crisis are also considered the contributors of the European sovereign debt crisis, which also spilled to the stock markets. (Vogel, 2018) Thus the research period financial crisis is relatively long for the exchanges of Helsinki and Stockholm.

Stock market development in Helsinki and Stockholm is depicted in Figure 1. Indices used are OMXH- (Helsinki) and OMXS GI (Stockholm), which are gross return indexes that include dividends, splits and delistings. It can be observed that the financial crisis had longer term effects in Helsinki, with stronger initial decline and longer recovery period than in Stockholm. The index peak and recovery dates are shown in Figure 1, and these dates represent the sample period over which the return data is collected.

Financial crisis period

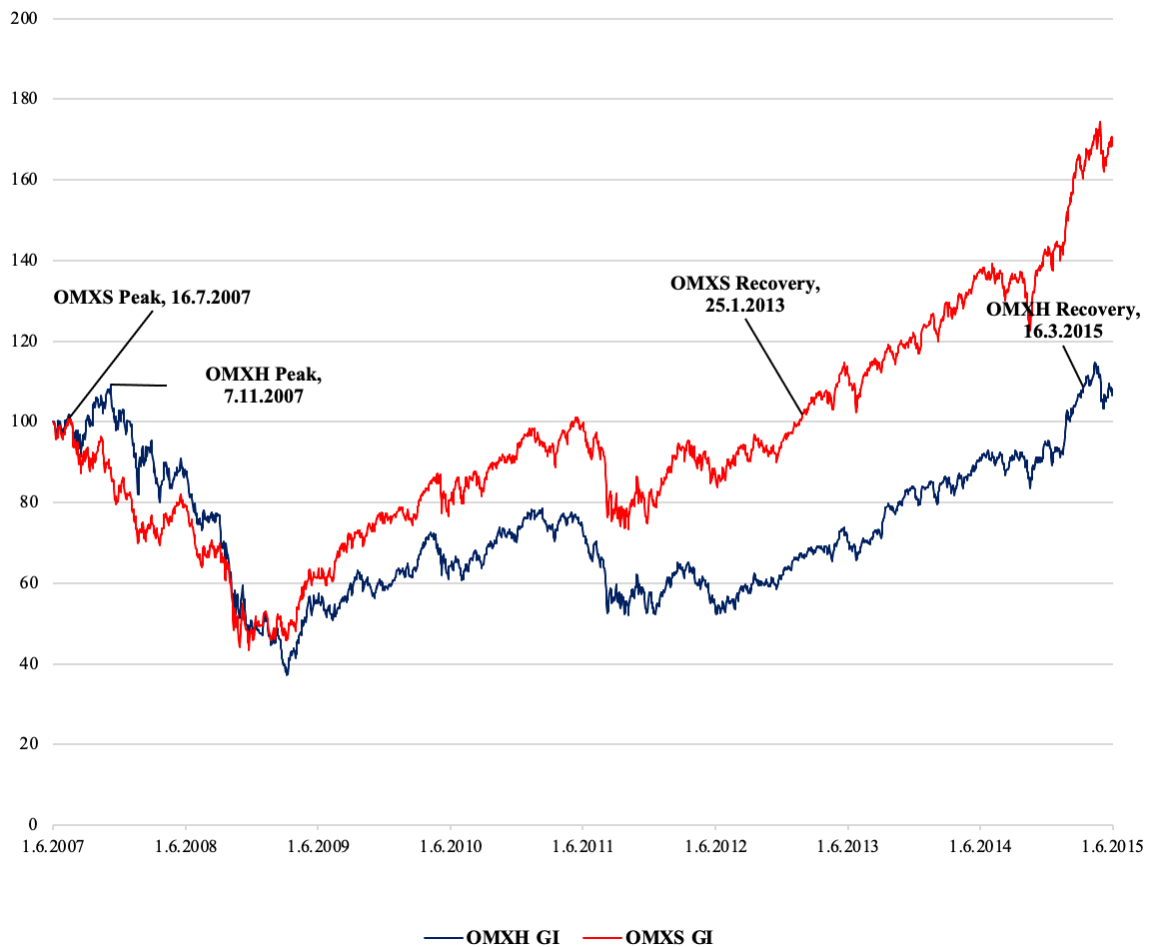


Figure 1. OMXH and OMXS indices during financial crisis period.

3.1.2 Covid-19 Crash

Covid crash is currently the latest market crash experienced. It began in early 2020, induced by the global pandemic from the spreading of Coronavirus-19. Early views were that the prevalent reasons for the Covid crash were external shocks i.e., repercussions of the virus spreading, but also evidence pointing out that the roots of the crash began growing already in the strong market upturn from 2018 onwards has been found. Government policies to limit the spread of the pandemic, mainly lockdowns and recommendations to limit all social contacts caused severe real economy issues seen for example in the deterioration of economic growth globally. (Shu, Song, Zhu, 2021).

Response to the crisis was very strong, in both fiscal and monetary terms. Many countries took unprecedented amounts of debt to support national economies and central banks released vast amounts of liquidity to the markets. This undoubtedly contributed to the strong recovery of stock markets, but the long-term effects are yet to be seen. (Siddik, 2020)

Similarly to the financial crisis, development of the Helsinki and Stockholm markets are depicted in Figure 2 with the same indices in 2.1.2020 – 31.12.2020. Even though the overall decline was not as great as with financial crisis, the market downturn was much sharper, as the period of downturn only lasted slightly over one month, followed by strong recovery period of about 6-7 months. Again, the research period for both markets is displayed as the peak and recovery dates in the graph. Notably for the Covid period, Helsinki and Stockholm seem to move very similarly, and the research timeframes are very close to each other.

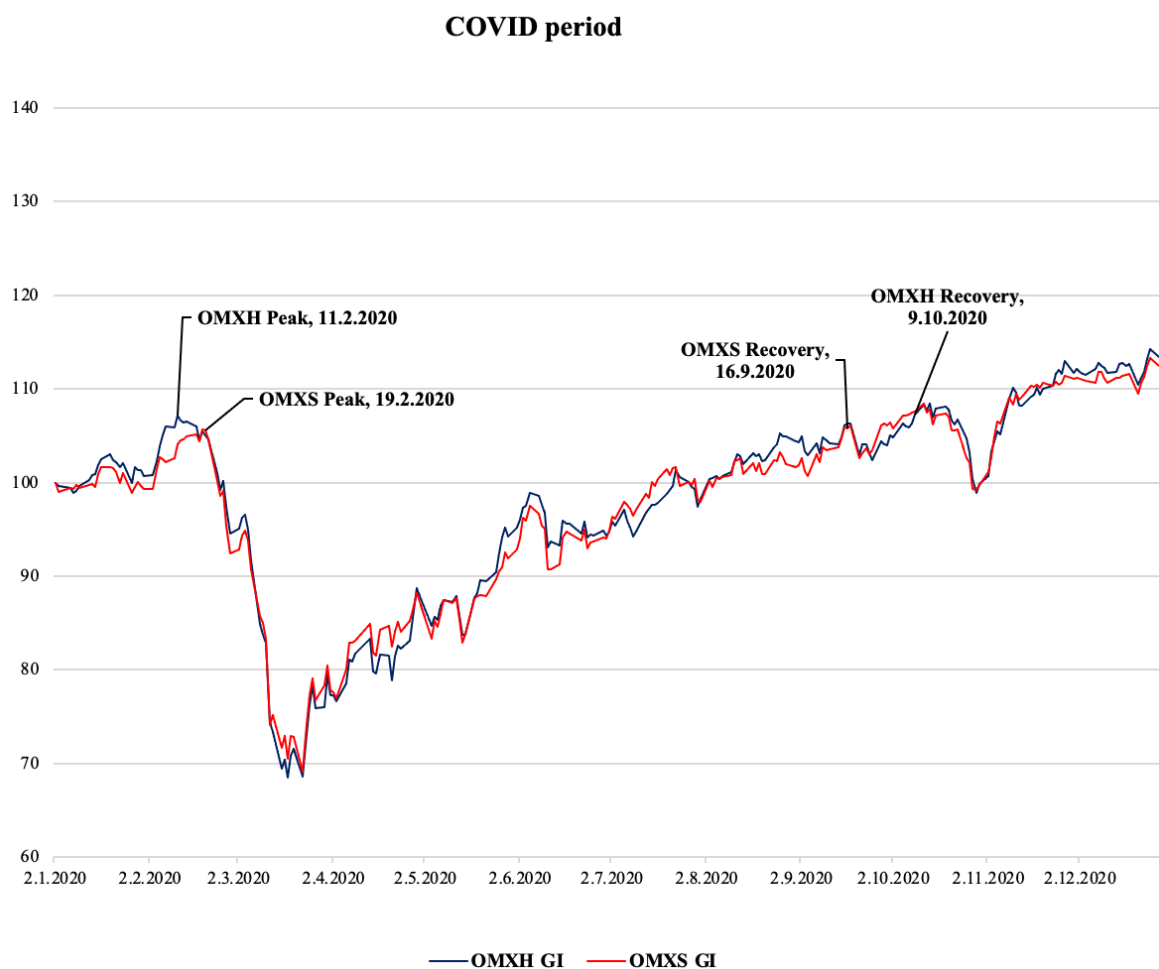


Figure 2. OMXH and OMXS indices during the Covid period.

3.2 Detailed data analysis and description

Data for this thesis is comprised of, as previously explained, two distinct events, the Covid of 2020 and financial crisis of 2007. Data for both crashes includes stocks listed in the exchanges of Stockholm and Helsinki. These were chosen to be a good representation of the Nordic markets as they are the two largest Nordic exchanges, with the exclusion of Oslo, which is not operated by Nasdaq Nordic. One aspect of this study is to focus on the Nordic markets and produce new research from this region, as it is less researched than for example US or major European markets.

Timeline for both events is from the date of market peak before the crash to the date of recovery above that same peak value. These dates are observed from OMXH- and OMXS GI indices, which represent their respective overall markets and account for dividends, splits and de-listings. Dates are visible in figures 1 & 2. It is important to note that the timelines from which data is gathered differs for the two exchanges in accordance with what was presented in the case introduction. This means that for financial crisis, the timeline of Helsinki stock exchange is around 2 years longer and for Covid crash, around a month shorter than the same events in Stockholm. In OMXS GI, the index value has a one-day peak in 2011 where the value is equal to period start value, but this is not regarded as the recovery date of the market as it is not a long-term recovery.

The most important consequence of this is that the returns of all stocks are not directly comparable as they are artefacts of different time periods depending on the exchange. However, because different markets have their own characteristics and recover differently, it is justifiable to take the time periods for both markets as they are, because the interest is in the return generated by a stock in the period of its home market's crash and recovery.

Variables and their abbreviations used in this study are summarized in Table 1, with descriptive statistics presented later. Variables were chosen on the basis of previous research, along with author discretion so that the collected set of variables provide a good overview

of stock- and firm-level characteristics. All data was sourced from Refinitiv Eikon database and most of the used variables are readily available in Eikon. All stock-level, i.e. price-, return- or trading volume based variables are from the date of market peak and all firm-level variables are from the latest published financial statements, i.e. 2006 or 2019.

Table 1. Research variables.

Return	Return in crash and recovery period, %	Stock level, Dependent variable
MVA	Market Value – EUR million	Stock level
BTM	Book-To-Market	Stock level
DY	Dividend Yield, %	Stock level
EY	Earnings Yield, %	Stock level
DTA	Total debt to Total assets, %	Firm level
ROA	Return on total assets, %	Firm level
CR	Current ratio	Firm level
TAT	Total asset turnover	Firm level
GRO	1Y Net revenue growth, %	Firm level
OPM	Operating profit margin, %	Firm level
DOL	Degree of operating leverage, %	Firm level
RET	1Y Return before research period, %	Stock level
STO	Share turnover volume, 1Y daily average	Stock level
BET	Historical Beta	Stock level
VOL	Annualized daily volatility, prior 1Y, %	Stock level

Variables that were not readily available in Eikon were Return, RET, BTM, DOL, STO and VOL. Return and RET were calculated as simple percentage return for the whole research period and 1 year prior to market peak respectively, using RI data type available in Eikon, which adjusts stock price to corporate actions such as dividends and splits. BTM was calculated using book value per share of previous financial statement and stock price of market peak date. DOL was calculated by dividing Net fixed assets by Total assets, as proposed by García-Feijóo & Jorgensen (2010). STO is calculated as the average of daily trading volume of one year prior to market peak and is used as a measure of stock liquidity. VOL is calculated from daily returns of one year prior and is annualized percentage figure. Market values of Stockholm listed companies are converted from Swedish krona to euro by using the exchange rate on the same date as the market value.

Table 2 presents the summary of data collected from the financial crisis period, so that all stocks from Helsinki and Stockholm which were active in 2007 were included, to avoid survivorship bias and consequently, the minimum return found was -100%. Data pre-processing included removing those stocks which were active only after the index peak, i.e. were listed during the crash. Also, stocks that had missing values for any of the chosen variables were removed and a few stocks which had significant outliers for certain variables (such as OPM of 300 000%) were removed, as these are likely false and would distort the dataset significantly. Some stocks are listed in both exchanges and all secondary listings were removed.

Resulting dataset contains 379 stocks of which 141, or 37,2%, recovered to their pre-crash price level during the research period. Mean return is correspondingly negative. One interesting observation is that average 1Y sales growth and return are significantly positive. As MVA and STO have large ranges and standard deviation, and they are neither relative measures, log-form of these variables will be used. Most variables' distributions are also positively skewed (right tailed) and have high kurtosis indicating leptokurtic distributions. Distributions are visualized in Appendix 1.

Table 2. Descriptive analysis of financial crisis period data.

	<i>Return</i>	<i>MVA</i>	<i>BTM</i>	<i>DY</i>	<i>EY</i>	<i>DTA</i>	<i>ROA</i>	<i>CR</i>	<i>TAT</i>	<i>GRO</i>	<i>OPM</i>	<i>DOL</i>	<i>RET</i>	<i>STO</i>	<i>BET</i>	<i>VOL</i>
Mean	-7.18	2218.3	0.46	1.73	1.71	17.93	4.23	1.97	1.24	44.90	-7.30	18.52	38.44	928.45	0.85	39.44
25th %tile	-70.82	32.67	0.18	0.00	0.49	3.30	1.36	1.16	0.75	5.95	0.52	3.51	-0.77	13.74	0.43	26.65
75th %tile	25.34	790.47	0.57	2.82	7.71	29.34	12.82	2.15	1.62	33.59	11.17	28.41	62.77	386.02	1.17	43.78
Median	-21.09	125.96	0.32	1.57	4.94	14.80	7.91	1.50	1.17	15.49	5.91	11.42	23.89	65.24	0.69	32.92
Std. Dev.	86.73	10441	0.54	1.75	20.80	16.01	19.86	1.70	0.73	153.56	73.64	19.06	64.75	3400.3	0.91	20.47
Kurtosis	3.74	99.36	40.84	0.64	43.10	0.00	10.46	20.87	5.11	89.03	68.46	1.21	5.77	54.11	81.51	5.85
Skewness	1.73	9.27	5.11	0.93	-4.52	0.78	-2.12	4.04	1.49	8.47	-7.44	1.33	1.87	6.89	6.13	2.22
Minimum	-100.00	0.94	0.00	0.00	-207.69	0.00	-134.23	0.29	0.06	-83.64	-828.10	0.11	-82.85	0.00	-2.21	16.07
Maximum	367.07	132058	6.22	8.72	109.62	77.58	75.98	14.26	5.29	2025.3	71.45	89.93	387.09	35764	12.85	134.22
Count	379	379	379	379	379	379	379	379	379	379	379	379	379	379	379	379

Summary of data from Covid period is presented in Table 3. Conducting the same pre-processing steps resulted to 599 stocks of which 296 or 49,4% recovered during the period.

General narrative on the dataset is similar to 2007 data. Mean values are in the same vicinity, with exceptions on period return, EY, ROA and OPM. Distributions are also similar, mostly leptokurtic and right-skewed. Interesting difference in period returns is that for Covid-crash, mean return is positive, maximum return is higher than in financial crisis data and not a single stock lost all value. When considering that the time period is also significantly shorter, this indicates that all stocks recovered rather strongly. The probable cause for this is the strong fiscal and monetary stimulus that could potentially fade the observable effect of indicators to some degree. Histograms of variables are presented in Appendix 2.

Table 3. Descriptive analysis of Covid period data.

	Return	MVA	BTM	DY	EY	DTA	ROA	CR	TAT	GRO	OPM	DOL	RET	STO	BET	VOL
Mean	12.68	1739.0	0.50	1.47	-6.51	21.13	-6.77	2.23	0.95	40.66	-235.29	17.07	31.19	463.28	0.85	50.16
25th %tile	-19.72	20.10	0.16	0.00	-7.79	6.10	-12.82	1.03	0.43	-0.28	-16.26	3.22	-12.74	16.17	0.48	28.92
Median	-0.30	90.41	0.33	0.00	1.79	19.65	3.26	1.48	0.89	8.68	3.46	10.54	15.22	54.04	0.85	38.72
75th %tile	24.67	551.57	0.65	2.67	5.30	31.98	8.25	2.30	1.33	26.50	8.99	23.88	50.44	233.03	1.19	60.54
Std. Dev.	60.96	9940	0.64	2.07	44.96	17.00	29.67	2.81	0.68	241.90	1067.7	18.89	86.46	1567.1	0.7	33.89
Kurtosis	22.22	235.90	39.79	3.53	81.99	0.10	4.34	83.80	2.67	326.53	51.53	2.50	22.03	58.92	4.72	9.65
Skewness	3.65	14.13	5.04	1.69	-4.64	0.71	-1.61	7.23	1.15	16.72	-6.74	1.66	3.83	6.95	0.36	2.58
Minimum	-70.95	0.09	-1.35	0.00	-580.98	0.00	-141.49	0.00	0.00	-95.16	-10635	0.00	-88.39	0.06	-2.94	12.72
Maximum	564.92	186910	7.25	12.28	438.08	87.01	101.66	43.29	4.79	5066.8	66.80	94.76	740.82	19084	5	288.51
Count	599	599	599	599	599	599	599	599	599	599	599	599	599	599	599	599

Final note on the data section is that for all modelling approaches employed, the data will be used according to the “holdout” method. This means that the datasets will be split into model training and testing data, in order to assess the model out-of-sample performance, which is commonly criticized in studies of return determinants. In this thesis, 80% of data will be used as training data and 20% as testing data.

The split is performed with stratified random sampling approach with respect to the recovery of stock, which means that the split data preserves the ratio of recovered and non-recovered stocks. This is important when the data is unbalanced, especially in the case of financial crisis data where 37% of stocks are labelled as recovered. Exception to the holdout method

is using financial crisis data for predicting in Covid-crash, in this case no split is performed, as the data is naturally in the training and testing samples.

3.3 Models

Three modelling approaches are utilized for this thesis, traditional multivariate linear regression, Random Forest and support vector regression. The idea is that using different modelling approaches, more robust conclusions can be drawn of the relationship of the used variables and crash and recovery -period returns. Also, by utilizing two ML methods, nonlinear relationships can be accounted for and an estimate of modelling improvement using more complex methods can be assessed.

3.3.1 Linear regression

Traditional statistical method to be employed is cross-sectional linear regression analysis using ordinary least squares method for estimation. Linear regression, or its multivariate variant in this study, provides an analysis method that is relatively simple and computationally inexpensive. Estimation in multivariate regression is done by fitting a linear hyperplane in the space of dependent and independent variable combinations in a way that it minimizes the sum of squared errors between the plane and actual values. General form of multivariate regression can be described as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \dots + \beta_i x_i + \varepsilon \quad (1)$$

where y = Dependent variable

β_0 = Constant term

β_i = Variable coefficients

x_i = Independent variables

ε = Error term, variation unexplained by the model

One purpose of this method is to provide a base to which compare the ML models' accuracy as interest is placed in whether adopting more complex and computationally expensive methods could provide added predictive power in the framework of predicting crash and recovery returns. In addition, ML models do not offer the same kind of statistical soundness and interpretability as linear regression and due to nonlinear nature, for example variable coefficients cannot be extracted similarly.

One issue with linear regression is that it, by definition, limits the relationship of variables to be linear, which might not be true at all in real-world applications. Other issue is that of relatively strict assumptions that need to be met in order to adequately use the estimated model for statistical inference. Nevertheless, linear regression is very commonly used modelling approach in similar studies, for explaining determinants of equity returns.

Statistical assumptions behind the OLS will be tested and presented, but more emphasis will be based on the out-of-sample performance of the estimated model as this is how different approaches can be compared. Assumptions tested are zero mean of residuals, exogeneity, autocorrelation, heteroscedasticity, multicollinearity, and normally distributed residuals. In addition to these, joint hypothesis testing of model significance is conducted. Joint hypothesis refers to testing whether all the model variable coefficients are jointly zero. The employed test designs are presented in results.

3.3.2 Random Forest Regression

Two models from widely utilized and well-established ML model families in the field of finance were picked for this thesis, the first one being Random Forest from the CART-family. Whilst decision trees are simple, efficient and easy to understand, they suffer from high variance in many cases and are relatively unstable, meaning that a slight change in the dataset may yield significant changes in the optimal tree and thus introducing bias error (Zimmermann, 2008). These troubles can be especially true with regression-type problems as if a full tree is not built, which would lead to overfitting, all instances do not get unique

fitted values. Empirical evidence suggests that ensemble methods are a very robust way to increase performance of tree models (Opitz & Maclin, 1999). Thus, Random Forest for regression proposed by Breiman (2001) is employed, to address for these issues in the simple regression tree.

Random Forest is an ensemble method, which essentially means that a mass of tree models is trained, and their outputs are aggregated to produce more robust and reliable results. Random Forests extend the idea of bagging, or bootstrap aggregation, where bagging is done repeatedly for a random subsample of the data, so that in addition to randomly sampling the data, also the features are subsampled randomly. This is also known as “feature bagging” which has the advantage of decreasing the correlation between each decision tree and thus increase predictive accuracy, on average. To further explain the terminology, bootstrapping refers to random sampling with replacement, meaning that after the random sample’s characteristics are learned, it is returned to the original training set before drawing the next sample. Aggregation then refers to aggregation of the trees from bootstrapped samples to produce robust predictions. In the case of Random Forest regression, this means averaging all the outputs of the “subtrees”. (Breiman 2001)

Simplified structure on the formulation of a Random Forest model is visualized in figure 3. It can be both viewed through the model building perspective or the perspective of prediction from new data. The input, or training data, is first fed to the algorithm and is then bootstrapped to subsamples with random features and instances. Trees are built from each subsample and used for creating a prediction. The subtrees work as independent regression trees, attempting to minimize their prediction error by adjusting the partitioning rules. Lastly the results are aggregated to create the Random Forest model’s prediction. When predicting with new data, it is churned through the same procedure, but the model is not adapted anymore for the new data.

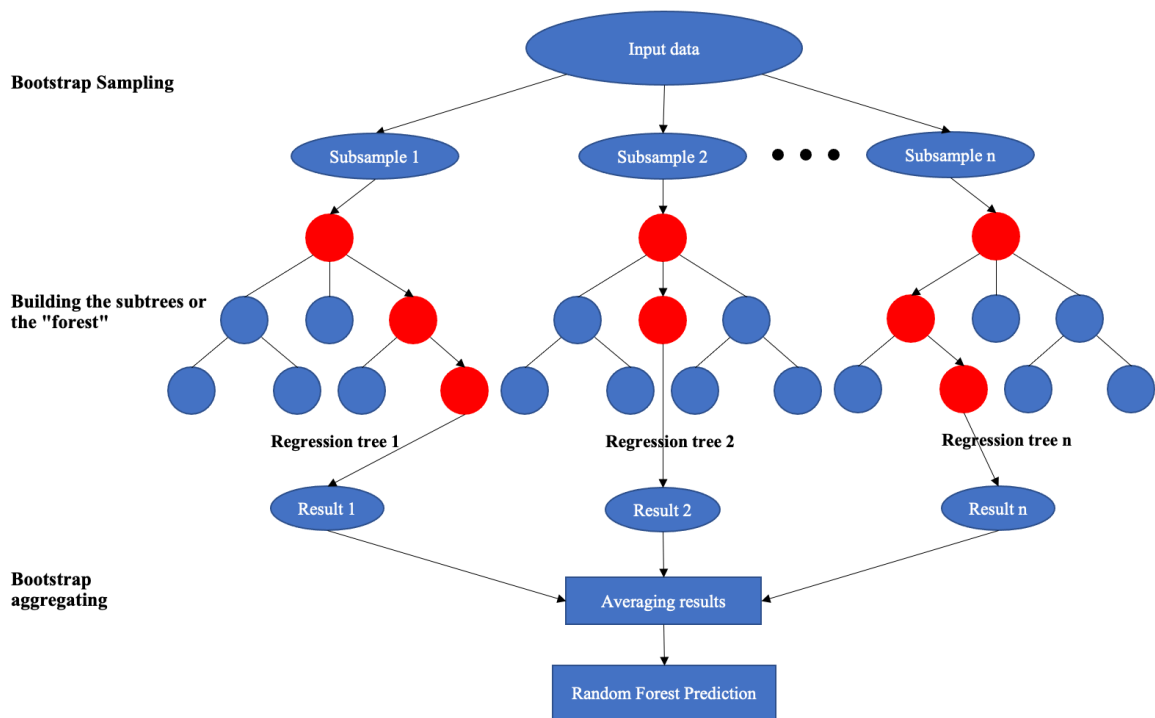


Figure 3. Simplified illustration of the Random Forest modelling approach.

As with other ML methodologies, also Random Forest has adjustable function parameters that can affect the results significantly. These are often called hyperparameters and will be subjected to tuning also here. Instead of using pure instinct, a common method to hyperparameter tuning is called a grid search. In grid search, the tuning parameters are designated to be of a certain range or certain values, and then through a looping procedure, a new model is trained for each combination of parameters, whilst using a k -fold cross-validation approach to minimize the generalization error of the created models. After this, the best parameters, or model, can simply be picked based on the lowest error rate which in the context of this thesis is quantified by mean squared error (MSE). (Hsu, Chang & Lin, 2003)

Cross validation is useful in creating generalizable models. In this study it will be used incorporated with the grid search for parameter tuning. In k -fold cross-validation the used training data will be randomly split into k -number of subsets and tested on the out-of-subsample data k -times. Model accuracy is calculated as the average of the prediction error. It will thus give a better estimate of out-of-sample performance of the built model, and the

best model from grid search is picked on the basis of the cross-validated performance. (Zhang & Yang, 2015)

Hyperparameters of Random Forest are split criteria, size of the bootstrapped dataset, max number of leaf nodes, max depth of individual trees, minimum samples to split at a tree node and the number of random features to include at each split. (Scornet, 2017) In this study, the last two hyperparameters, which are possibly the most important, are tuned with a grid search. The number of trees is determined beforehand to be 250. General principle is that increasing the number of trees lowers variance, creating more robust forests but increases the computational costs. The effect of this parameter can be evaluated afterwards by observing the out-of-bag error. OOB error is the mean prediction error of each subtree using data that was not used in training the subtree. 10-fold Cross-validation is used in the grid search. (Biau & Scornet, 2016)

The split criteria employed is the MSE. Max depth of the individual trees and max number of leaf nodes are not critical parameters for a Random Forest as overfitting is mainly controlled by using a large enough number of trees and minimizing the OOB error. Size of the bootstrapped dataset is kept default at the size of the training dataset, as sampling with replacement selects instances multiple times to the bootstrapped set and thus keeps the set different from the original training set.

3.3.3 Support vector regression

The second ML method to be employed in this study is support vector machine modified for regression problems, i.e., support vector regression. SVMs were first used to solve classification type problems as proposed by Cortes & Vapnik (1995). While vastly popular in this region, SVM can be also generalized to regression problems by incorporating the ϵ -tube, which represents the margin from the function where the misestimates are not penalized, as proposed by Drucker et al. (1997)

In its basic form, support vector regression is somewhat similar to linear regression, with the difference that as linear regression attempts to fit a line so that it minimizes the sum of errors, SVR finds the best fit of a line, or a hyperplane, within boundaries set by the support vectors that contains the largest amount of points. In other words, inside the ϵ -tube, the error term is ignored. Instead, the fit of the hyperplane is adjusted by those points, which are outside of the ϵ -tube. Prediction of values is thus done in a similar manner as in linear regression, i.e. by means of the fitted line or hyperplane. (Drucker et al. 1997)

One important aspect of SVRs are the kernel functions that can be incorporated in this method. As the basic regression-extended version of SVM works in linear space, by using kernel functions the data can be treated as it would be transformed to higher dimensional space, to address for nonlinearity in the underlying relationships of the data. (Kutateladze, 2021) Instead of actually transforming the data, the kernel function calculates the high-dimensional relationships, or dot products, of datapoints and these are utilized. Hence the kernel method is often called the “kernel trick”. (Kwak, 2013) Figure 4 is a visualization of the kernel trick and makes it easier to understand how the use of a kernel can help to better account for underlying relationships in the data. (Jain, 2020). After the kernel transformation, a linear separation can be done.

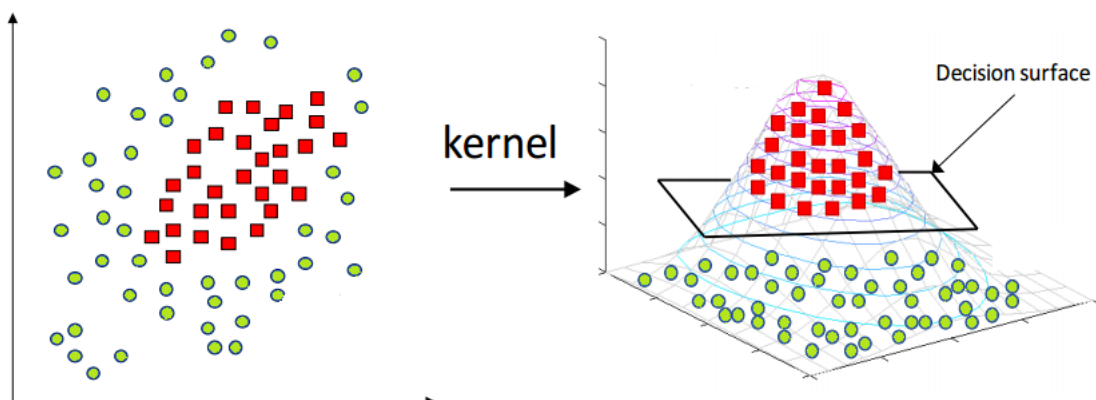


Figure 4. Visualization of the kernel trick. (Jain, 2020)

Commonly used kernels are linear- , polynomial- , gaussian radial basis function- and the sigmoid kernel. These kernels are also easily available in the basic application package of SVM in R. For support vector regression in this thesis, the gaussian radial basis function is used. It is commonly preferred for SVM when non-linearity is addressed and there is no extensive prior knowledge of the data. (Bhavsar & Panchal, 2012)

Whilst the choice of kernels and many other parameters could be done very extensively, such “mining” for the best ML approach to model and predict this data is not of essence in this study. Rather, the interest is in whether these relatively easily employable methods imply that ML methods could provide an edge over the regression analysis, by taking into account the computational efforts in more complex modelling.

Hyperparameters tuned for SVR with grid search are ϵ of the insensitive-zone, cost parameter C and gamma. Similarly to the Random Forest method, 10-fold cross-validation will also be used in the grid search for SVR parameters. ϵ controls the width of the insensitive zone, C is the penalty for the error between actual and predicted value and is proportional to the distance from the decision boundary. Gamma is related to the RBF kernel and is used to control the influence zone of a single training datapoint. The larger the gamma, the closer the points have to be in order to be considered similar. Thus, high gamma exposes to overfitting. (Tay & Cao, 2001)

One way to avoid overfitting, along with cross-validation, is to ensure that the amount of support vector is kept in reasonable limits. This is why Mattera & Haykin (1999) propose to choose ϵ so that the amount of SVs would not significantly exceed 50% of total number of samples, especially when working with larger and noisier datasets. This will be kept in mind, but the cross-validation implemented in grid search has the biggest responsibility of avoiding overfitting.

3.3.4 Model assessment and variable importance

Comparative assessment of the used models accuracy, or ability to explain, will be done using root mean squared error RMSE, mean absolute error MAE and % correct sign predictions. All these performance metrics will be calculated for the training and testing set to grasp an idea of the models' out-of-sample performance. Both RMSE and MAE both measure the average prediction error of the model prediction to the actual values. The formulas for RMSE and MAE are as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}} \quad (2)$$

$$MAE = \frac{\sum_{i=1}^N |Predicted_i - Actual_i|}{N} \quad (3)$$

where N = number of predictions

One point of interest in this study is that what is the practical relevance of the built models. Even though the statistical loss functions RMSE and MAE do tell of the model accuracy, it is speculative whether e.g a decrease in RMSE is practically useful. Evidence of this has been shown by for example Gerlow, Irwin & Liu (1993), who find that statistical criteria of accuracy might not give a clue to what the profitability of using the forecasts in trading is. Thus, also the % correct sign predictions proposed by Pesaran & Timmermann (1992) is calculated for model assessment. It also enables for assessing the model performance on the ground of whether or not a stock is predicted to recover correctly. Calculation is as follows:

$$\% \text{ correct sign predictions} = \frac{\sum_{i=1}^N z_i}{N} \quad (4)$$

where $z_i = 1$ if $(\text{Predicted}_i \times \text{Actual}_i) > 0$

$z_i = 0$ otherwise

$N =$ number of predictions

In addition to the accuracy measures, R^2 and its multi-variable adjusted version are calculated and presented for the trained models. R^2 is a rough measure of how well the model fits the data, or how well the model replicates the already known outcomes. Technically said, it measures the amount of variability in the dependent variable explained by the model and is calculated as the sum of squared residuals divided by the total sum of squares:

$$R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (5)$$

where $y_i =$ known outcomes

$f_i =$ predicted outcomes

$\bar{y} =$ mean of the known outcomes

One problematic aspect of R^2 is that it tends to always increase when adding variables to a model, even if they are irrelevant. Especially in the context of this study, when it is not known beforehand which explanatory variables are important, it is beneficial to also calculate the adjusted R-squared. The adjusted R^2 takes account of the number of explanatory variables when determining the overall explanatory power for the employed model.

$$Adjusted R^2 = 1 - \frac{(1 - R^2)(n - 1)}{n - v} \quad (6)$$

where n = Sample size

v = number of explanatory variables

To determine which of the used variables are most important in predicting the crash period returns, variable importance metrics are calculated. For Random Forest and support vector regression models, a metric called permutation feature importance is calculated. This measure was first introduced for Random Forests by Breiman (2001) and later adopted for a model-agnostic version by Fisher, Rudin & Dominici (2019). Permutation feature importance scores are calculated so that model prediction error is measured and then the feature values are shuffled, destroying the relationship between the outcome and the feature. Then prediction error is measured again and the larger the increase in error, the more important the feature is. Shuffling is done multiple times to obtain more stable results and feature importance is measured as the factor of error increase for the given variable. Loss function used for measuring the error is RMSE.

For linear regression, variable importances are calculated as relative importances, with a method proposed by Lindeman, Merenda & Gold (1980). Relative importance is calculated by determining the relative contribution of each variable to the R^2 . In essence, this method adds variables sequentially to the model and calculates the increase in R^2 . To avoid the problem of correlated predictors, in which case the method would depend on the adding sequence, different orderings of variable sequences are created and the added R^2 is averaged over all the orderings for given a variable. It should be noted that as calculation methods vary, the variable importance scores are not comparable across models. Rather, the order of variables in all models can be observed.

Finally, to investigate the negative-positive relationship between variables and model predictions, i.e. returns, in ML models, a correlation matrix is calculated for the predicted

value and variables. Albeit not a very robust method, this way a rough estimate can be grasped on how the model values rise in e.g., financial leverage in terms of predicted returns. For linear regression models, the variable coefficients and their statistical significance can be directly observed.

4. Empirical results

Results are presented in three parts. The first part contains modelling results of financial crisis data, second part results from the Covid-19 data, and the third part presents the results from modelling the Covid-19 with financial crisis data as model training sample. Model accuracies are presented jointly in the end of each section for easier comparison.

4.1 Financial crisis

Preliminary examination of relationship between the crash and recovery period returns and individual stock characteristics is conducted by calculating a correlation matrix. Correlation coefficients are calculated as the most commonly used measure, Pearson correlation coefficient, which measures the linear correlation of variables with values between -1 and 1. The correlation matrix for the full data of financial crisis period is presented in Figure 5. From the correlation matrix two important things can be observed.

Firstly, no severe multicollinearity is found in the data, the strongest correlation between explanatory variables is with OPM and ROA, 0.63. This relationship seems intuitive, higher profitability in terms of margins also indicates higher profitability in terms of leverage, or vice versa. Other stronger correlations are found in VOL with MVA, DY, ROA and OPM. The negative correlation in these seems to indicate that profitable stocks, in fundamental perspective and relative to price, have lower stock volatility and small cap stocks have higher volatility. Correspondingly, MVA and STO also have an intuitive, relatively strong positive relationship, higher market capitalization stocks have higher turnover and thus, higher liquidity.

Secondly, none of the chosen variables have very strong correlation with the period returns, at least in linear terms. This could give an edge for the ML models, as they work better with weak signals and possible nonlinearity, than linear regression. The strongest positive

correlation is found with DY, STO, EY, MVA, OPM and ROA. Strongest negative correlations are in VOL, GRO, BET and DTA. These imply that period returns are higher with more profitable, larger and liquid companies or stocks, and comparably lower in stocks with higher volatility, market risk, leverage, and pre-crash growth. This is consistent with well-known flight-to-quality phenomenon (see e.g., Kaul & Kayacetin 2017, Opitz & Szimayer 2018), i.e., during a crash period, investors move capital to more stable and profitable “quality” stocks. Interestingly, RET has zero correlation with period returns, which implies that the “momentum effect” would have not been present. CR and TAT also have zero correlation with returns, indicating that short-term liquidity is not related to returns during the financial crisis period.

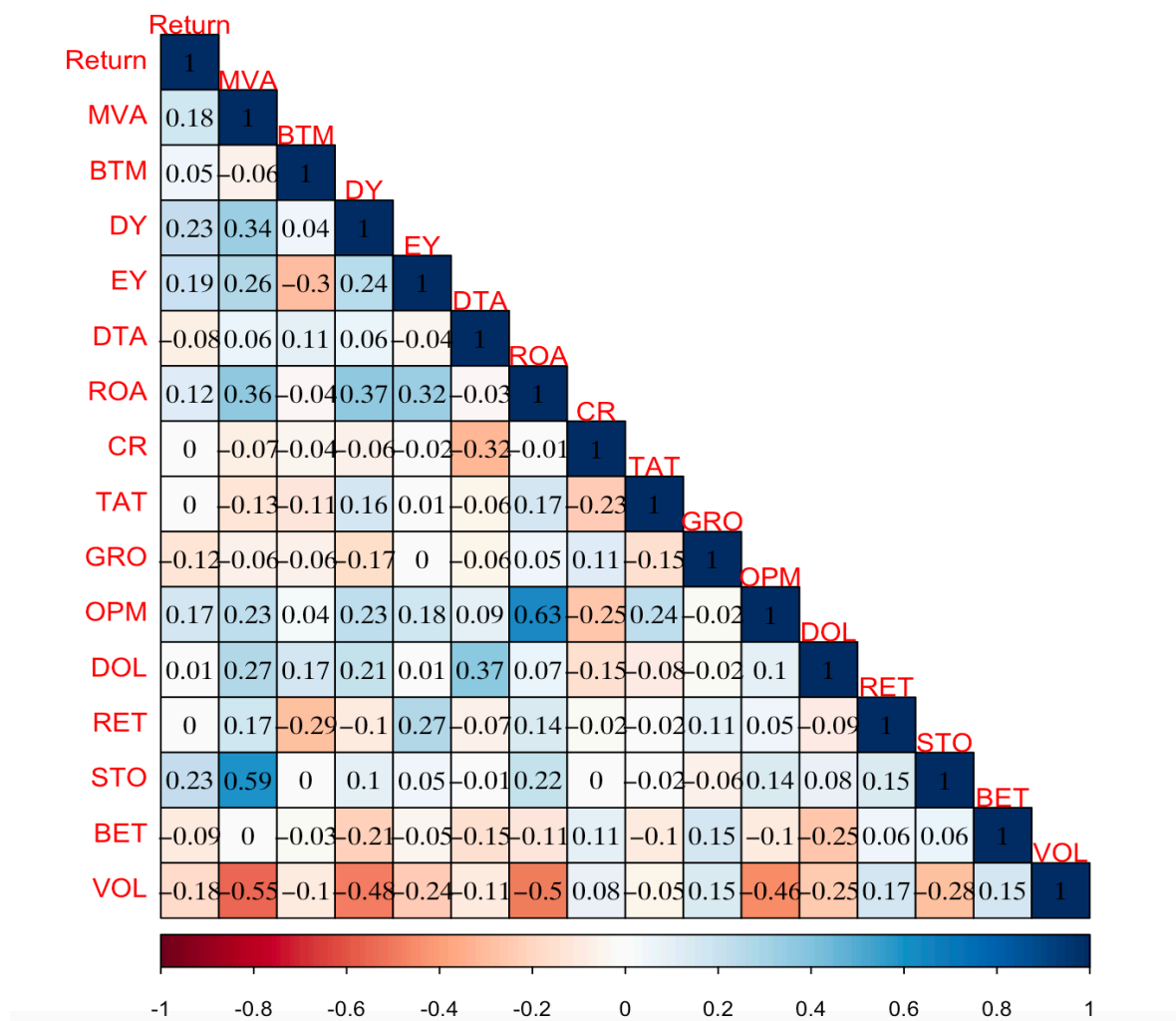


Figure 5. Correlation matrix of financial crisis data.

Division of the financial crisis data to 80/20 ratio means that 303 stocks were included in the training set for model building and 76 stocks are retained for out-of-sample testing. The first model built is the linear regression. Predictors of the linear model are summarized in Table 4 with estimated coefficients and their standard errors, t-statistics and p-values. P-value tells the level of confidence that the given coefficient is statistically different from zero. Statistically significant coefficients are indicated with ***, ** and * for confidence levels of 1%, 5% and 10% respectively. In the estimated model, DY, EY, ROA, OPM and STO are statistically significant, and this indicates that pre-crash profitability metrics, both stock- and firm-level characteristics are relevant in predicting crash and recovery period returns, along with stock liquidity. Interestingly, ROA has a negative coefficient, so high pre-crash return seems to affect negatively to period returns. Table 4 also contains VIF (variance inflation factor) values for multicollinearity testing.

Table 4. Linear regression coefficients in financial crisis data.

	Coefficient	Std. Error	T-stat	P-value	VIF value
Intercept	-30.88	31.17	-0.99	0.32	
MVA	-3.52	3.54	-1.00	0.32	2.45
BTM	12.37	11.43	1.08	0.28	1.18
DY	*** 13.56	3.41	3.98	0.00	1.51
EY	*** 0.99	0.32	3.12	0.00	1.35
DTA	-0.53	0.33	-1.58	0.12	1.34
ROA	** -0.71	0.35	-2.02	0.04	2.37
CR	1.28	3.40	0.38	0.71	1.37
TAT	-10.37	7.76	-1.34	0.18	1.33
GRO	-0.04	0.03	-1.42	0.16	1.12
OPM	*** 0.26	0.09	2.90	0.00	2.09
DOL	-0.19	0.30	-0.65	0.52	1.43
RET	-0.03	0.08	-0.34	0.73	1.41
STO	*** 8.36	2.35	3.55	0.00	1.70
BET	-5.02	5.57	-0.90	0.37	1.22
VOL	0.35	0.35	1.01	0.32	2.15

Whilst the ability of traditional statistical population inference is not the most important aspect of this study, as the ML methods cannot be measured in these terms similarly, it is

still interesting to conduct assessment of the OLS assumptions that need to be valid for statistical inference. If the assumptions are valid, the estimated coefficients can be deemed BLUE, best linear unbiased estimators. However, it still needs to be remembered that most of the variable coefficients do not exhibit statistical significance.

Statistical testing of the linear model is briefly presented in table 5. The Fisher test p-value represents that the independent variables in the model are jointly different from zero with statistical difference. No severe multicollinearity is present, as the highest variable correlation was 0.63, note that correlation matrix for the training data is not separately presented as it is nearly identical to the matrix of the whole set. The same conclusion of no multicollinearity can be seen in VIF values, for which the floor is at 1 and common thresholds of multicollinearity used vary from 5 to 40 (O'Brien 2007). VIF values here are nowhere near those thresholds.

Mean of residuals is zero, meaning that only random error is left in the error term. The Breusch-Pagan test is used for testing residual heteroscedasticity, and no heteroscedasticity is present. Residual autocorrelation is tested with the Durbin-Watson test, and p-value higher than 0.05 implies no autocorrelation, thus the model might contain autocorrelation. Exogeneity was tested with calculating correlation of variables and residuals, and the model does not inhibit endogeneity. Finally, the Jarque-Bera test is used for testing residual normality and p-value of 0.00 indicates that the residuals are not normally distributed.

Table 5. Statistical assumptions of linear regression in financial crisis data.

	Statistic Value	Assumption valid
F-test P-value	0.00	Yes
Multicollinearity	Highest corr. 0.63 & VIF 2.45	Yes
Mean of residuals	0.00	Yes
Breusch-Pagan test P-value	0.25	Yes
Durbin-Watson test P-value	0.05	No
Exogeneity	No correlation	Yes
Jarque-Bera test P-value	0.00	No

The possible violation of autocorrelation and normally distributed residuals reduce the precision of the estimates and non-normally distributed errors mean that confidence and prediction intervals are not reliable. Residual plots are presented in Appendix 3, and visually autocorrelation cannot be detected, but it can be seen that the residual distribution is right tailed, possibly due to outliers in the data. Because the model is not time-series based, the source of possible autocorrelation is not clear, but one reason could be an incorrectly specified model, for example due to uncaptured non-linearity.

One main interest is the variable importances, and for linear regression they were calculated as the relative importances, measured as relative increase in the R-squared when the variable is added. Variable importances are presented in Figure 6. Not too surprisingly, the most important variables in explaining returns in the linear model are the statistically significant predictors. They are in the same order as the magnitude of change in predicted return with 1 unit of change in the explanatory variable. However, it is noteworthy that as STO is logarithmically transformed, increase in return cannot be directly calculated with change in observed STO. An important takeaway is that DY and EY seem usable in the prediction of crash and recovery period returns, especially DY.

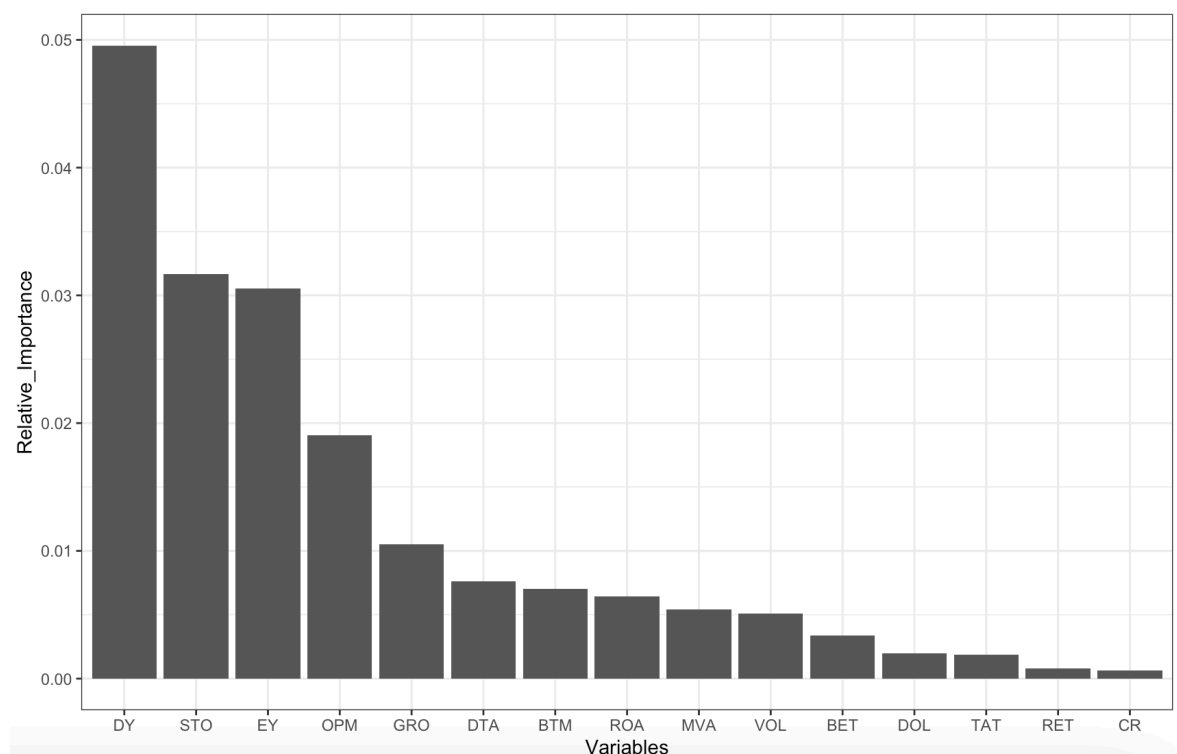


Figure 6. Variable importances in linear regression mode with financial crisis data.

Random Forest model was built so that the grid search for hyperparameter tuning using the training data was first conducted. Two parameters were tuned using the grid search, the number of random features in each tree and minimum node size. Default values proposed for these parameters are 1/3 of used variables i.e. 5, for number of random features and 5 for minimum node size. Explored ranges were [3, 4, 5, 6, 7] for number of features and [5, 10, 15, 20] for minimum node size. Boundaries of ranges were decided on the basis of amount of variables and sample size. Number of trees grown was set to 250, instead of default of 500, in order to save computation time.

Performance evolution of cross-validated accuracy during the grid search is presented in Figure 7, which shows that optimal parameters from the given ranges are 3 random features and minimum node size of 20.

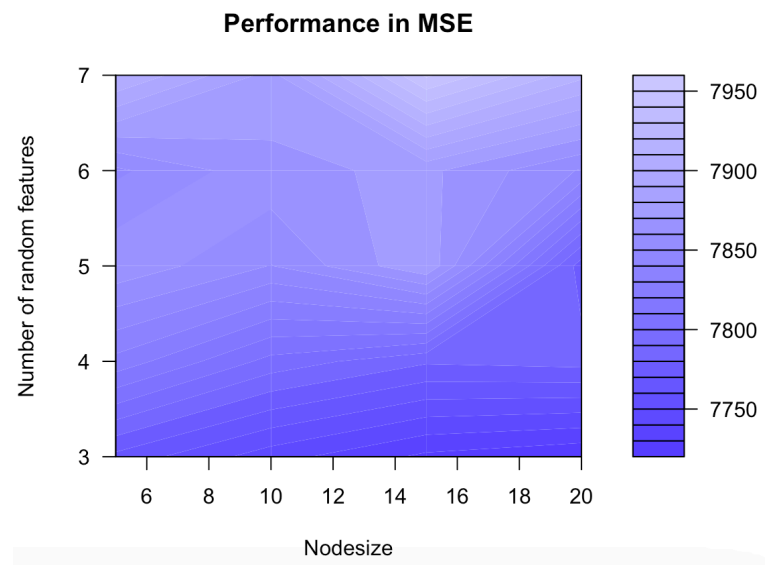


Figure 7. Cross-validated performance in grid search for Random Forest in financial crisis data.

Using these parameters, the final forest model was trained, again using the training data. Figure 8 shows that the 250 trees grown is sufficient, as the decrease in the out-of-bag error seems to settle at around 100 trees, and after that point, only minor performance improvement is observed.

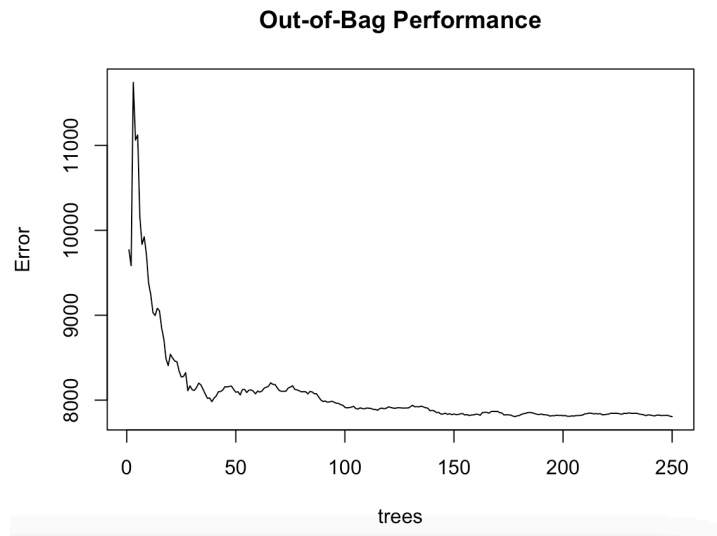


Figure 8. Random Forest Out-of-bag performance in financial crisis data.

Variable importance for the Random Forest model were measured with permutation feature importance method, and the results are presented in Figure 9. Most important variables in Random Forest were found to be DY, EY, STO and VOL. Even though the importance of variables seems to decrease steadily with the order seen on the axis, what is distinctive is that a clear division is seen with stock- and firm-level variables. With the exception of BET having lowest importance, it seems that firm-level variables are unimportant for modelling the crash and recovery period return. Predictive accuracy of Random Forests on training and testing data will be presented later together with the other models.

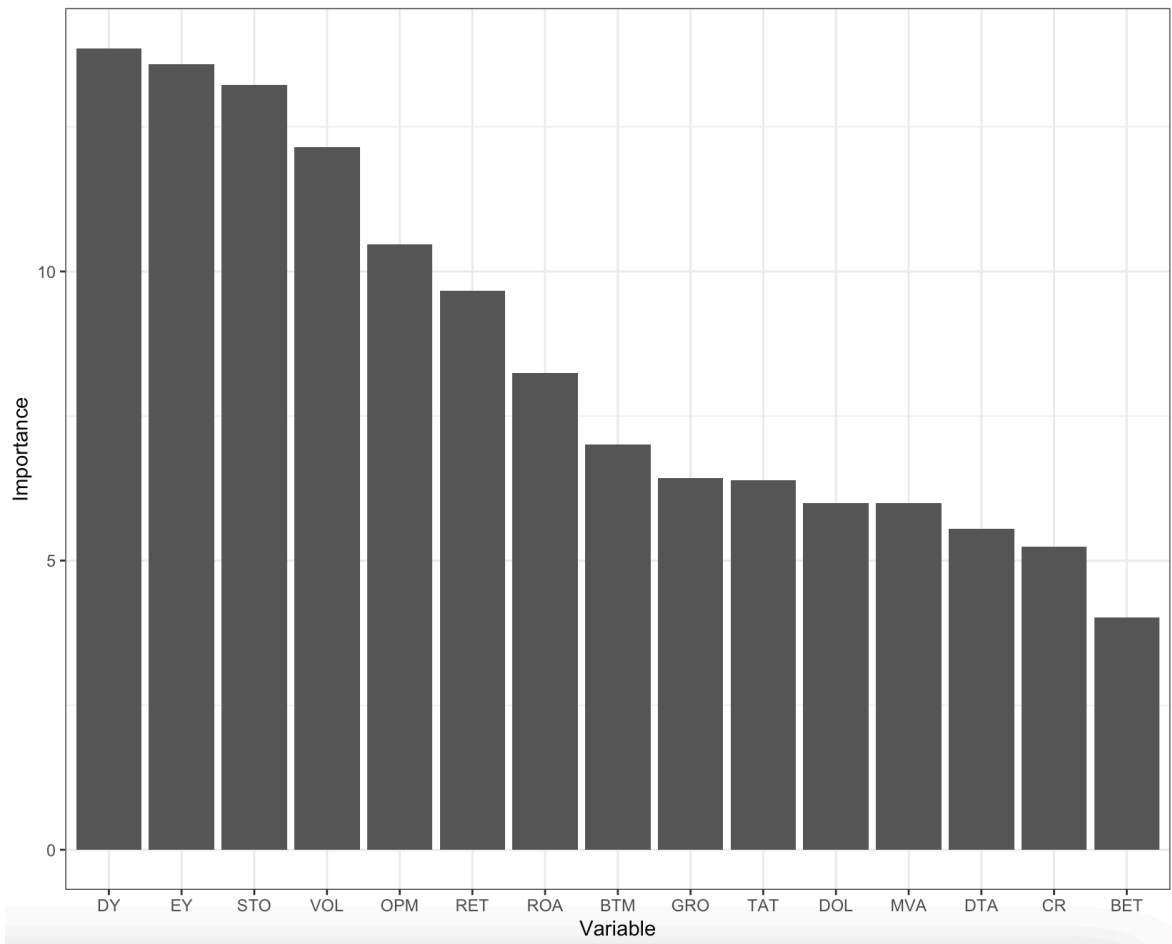


Figure 9. Variable importances in Random Forest model with financial crisis data.

In order to grasp even some understanding of how the variables affect the predicted return, the correlations of model predicted returns and predictors are calculated and presented in Table 6. Based on these statistics, the most important variables DY, EY and STO have positive affiliation with predicted returns and VOL has a negative affiliation. These are very intuitive ideas, and it makes sense that more volatile stocks yield less returns during crash and recovery period, and the affiliation of STO seems to be giving evidence of the flight-to-quality phenomenon.

Table 6. Correlations of predicted returns and variables in Random Forest with financial crisis data.

	Positive/Negative Affiliation	Correlation with predicted return
MVA	Positive	0.21
BTM	Positive	0.14
DY	Positive	0.38
EY	Positive	0.29
DTA	Negative	-0.09
ROA	Positive	0.20
CR	Positive	0.01
TAT	Positive	0.06
GRO	Negative	-0.15
OPM	Positive	0.21
DOL	Positive	0.02
RET	-	0.00
STO	Positive	0.25
BET	Negative	-0.12
VOL	Negative	-0.20

Last approach used is the Support vector regression. Similarly to Random Forest, a grid search is performed to tune the hyperparameters. With SVR, three hyperparameters are explored: epsilon, cost and gamma. Proposed default values for these in R's SVM framework are epsilon of 0.1, cost of 1 and gamma as 1/data dimension, which is 0.0667 with 15 variables. Thus, ranges of variables in grid search are [0.05 – 0.95] with steps of 0.1 for epsilon, [0.1 – 10.1] with steps of 0.5 for cost and [0.02 – 0.2] with steps of 0.02 for gamma. Results of the grid search are presented in Figure 10. Cross-validation performance of different parameter combinations are presented in two contour plots, with Cost on the y-axis on both.

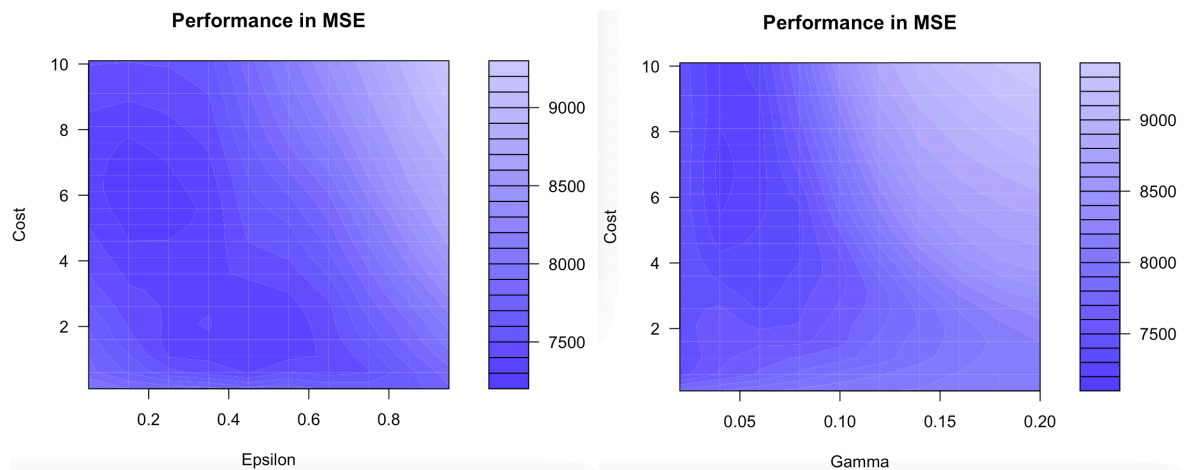


Figure 10. Cross-validated performance in grid search for Support vector regression with financial crisis data.

The optimal parameter combination is determined to be epsilon of 0.65, cost of 5.1 and gamma of 0.02 and the final SVR model is trained accordingly. Using the radial basis function kernel, or RBF, the built model features 107 support vectors, which is roughly 35.3% of the sample size. After fitting the model to training data, variable importances are calculated. Calculation is done using the same methodology as with Random Forest. Permutation feature importance based results are presented in Figure 11. Ranking of the less important variables is different when compared to linear regression or Random Forest, but the most important variables remain the same, as DY, STO and EY are the most important ones based on the SVR model. As with Random Forest, BET is the least important variable in the SVR model as well.

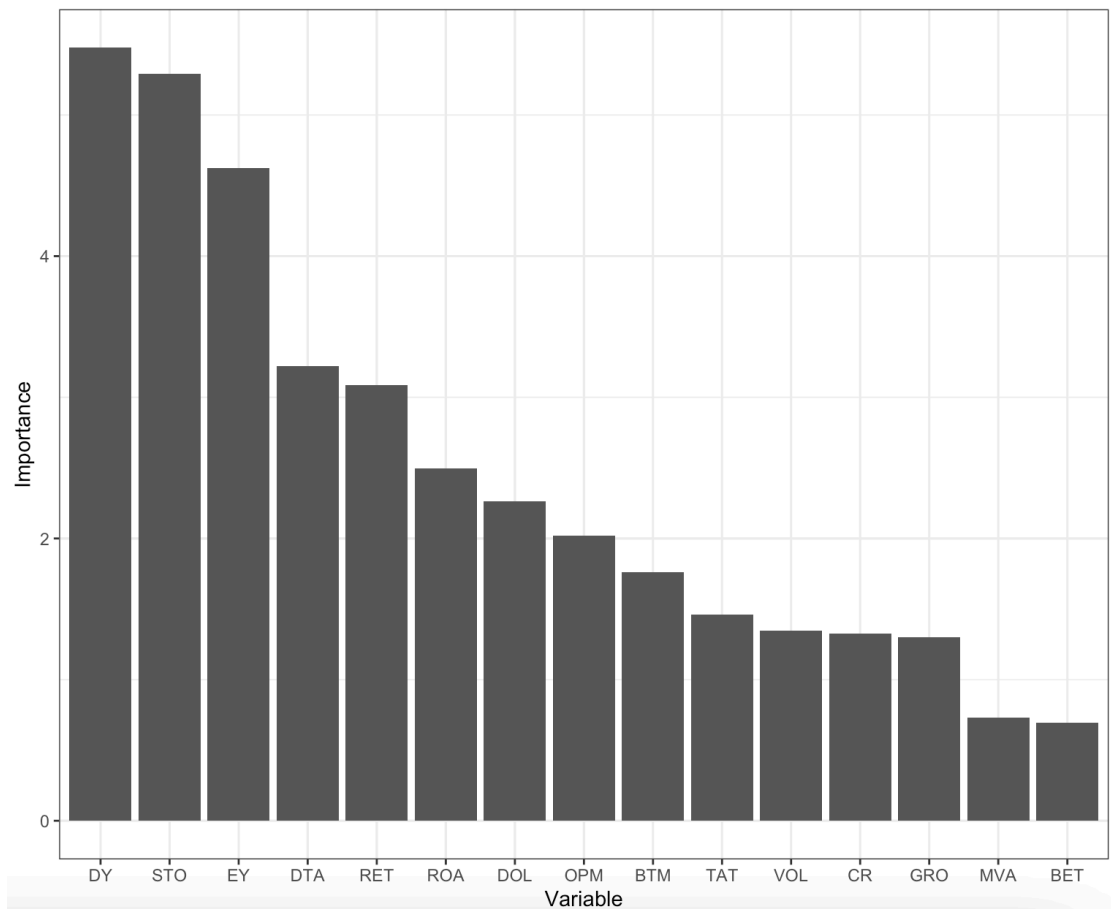


Figure 11. Variable importances in Support vector regression with financial crisis data.

Correlation analysis on the SVR predicted returns and variables yield similar results than with Random Forest, as presented in table 7. Obviously, the magnitude of correlation coefficients varies, but not much interest is placed on those. However, the positive or negative affiliation between variables and predicted returns is exactly the same as with Random Forest. When comparing to the sign of estimated coefficients on the linear regression, less variables seem to have negative affiliation with returns on Random Forest and SVR models. Interesting notion is that in linear regression pre-crash volatility has a positive coefficient which differs from correlation implied affiliation of Random Forest and SVR.

Table 7. Correlations of predicted returns and variables in Support vector regression with financial crisis data

	Positive/Negative Affiliation	Correlation with predicted return
MVA	Positive	0.31
BTM	Positive	0.12
DY	Positive	0.56
EY	Positive	0.36
DTA	Negative	-0.17
ROA	Positive	0.19
CR	Positive	0.02
TAT	Positive	0.07
GRO	Negative	-0.22
OPM	Positive	0.24
DOL	Positive	0.02
RET	-	0.00
STO	Positive	0.45
BET	Negative	-0.12
VOL	Negative	-0.28

Finally, the modelling accuracy of all built models is addressed. As previously presented, modelling accuracy is measured in RMSE, MAE and % correct sign predictions both in training and testing data. In addition, also R^2 figures, or the coefficient of determination, are presented for the models. They are calculated in the model building phase, i.e. when fitting the training data, and measures the proportion of variance of the dependent variable explained by the model, or roughly said, how well the model explains the variability in the returns.

Modelling accuracy is presented in Table 8. Based on training and testing performance, Random Forest appears to be the best performing modelling approach for this data. It has significantly lower RMSE and MAE in training compared to linear regression and SVR. Random Forest also significantly outperforms in % correct sign prediction and R-Squared. However, some overfitting on Random Forest is possible, as both RMSE and MAE nearly double in the testing set. Performance on these is still near to support vector regression, which has slightly lower test data RMSE and MAE. The overall conclusion on model accuracies is that the machine learning approaches perform better than the linear regression.

At first hand, the error measures seem high, but on the other hand, also the standard deviation of realized return is 86.73, as presented in Table 2. Finally, observing the model R^2 , it can be seen that Random Forest significantly outperforms the other approaches, and linear regression is the worst. However, when considering that Random Forest accuracy decreases significantly on out-of-sample data, it is rather likely overfitting.

Table 8. Modelling accuracies in financial crisis data.

	Linear Regression	Random Forest	Support Vector Regression
RMSE, training	80.51	38.05	74.14
RMSE, testing	77.44	76.48	75.73
MAE, training	57.02	26.55	51.98
MAE, testing	60.40	59.63	57.08
% Correct sign predictions, training	65.35	93.73	70.63
% Correct sign predictions, testing	59.21	68.42	56.58
R-Squared, %	18.15	81.73	30.59
Adjusted R-Squared, %	13.88	80.77	26.97

4.2 Covid crash

Similarly to financial crisis data, correlation analysis is conducted for the Covid-crash period data from 2020. Again, correlations are measured with Pearson correlation approach for linear correlation. The whole dataset is used and variables MVA and STO are in the logarithmic form. Correlation matrix is presented in figure 12 with a heat map approach. Strongest correlations between independent variables are found with VOL and MVA, VOL and ROA, STO and MVA, and between ROA and MVA. Highest correlation in absolute term is -0.58 between VOL and MVA, thus no severe multicollinearity seems to be present in the data. Also, the relationship seems intuitive, stocks with larger market capitalization appear more stable with volatility decreasing as size grows.

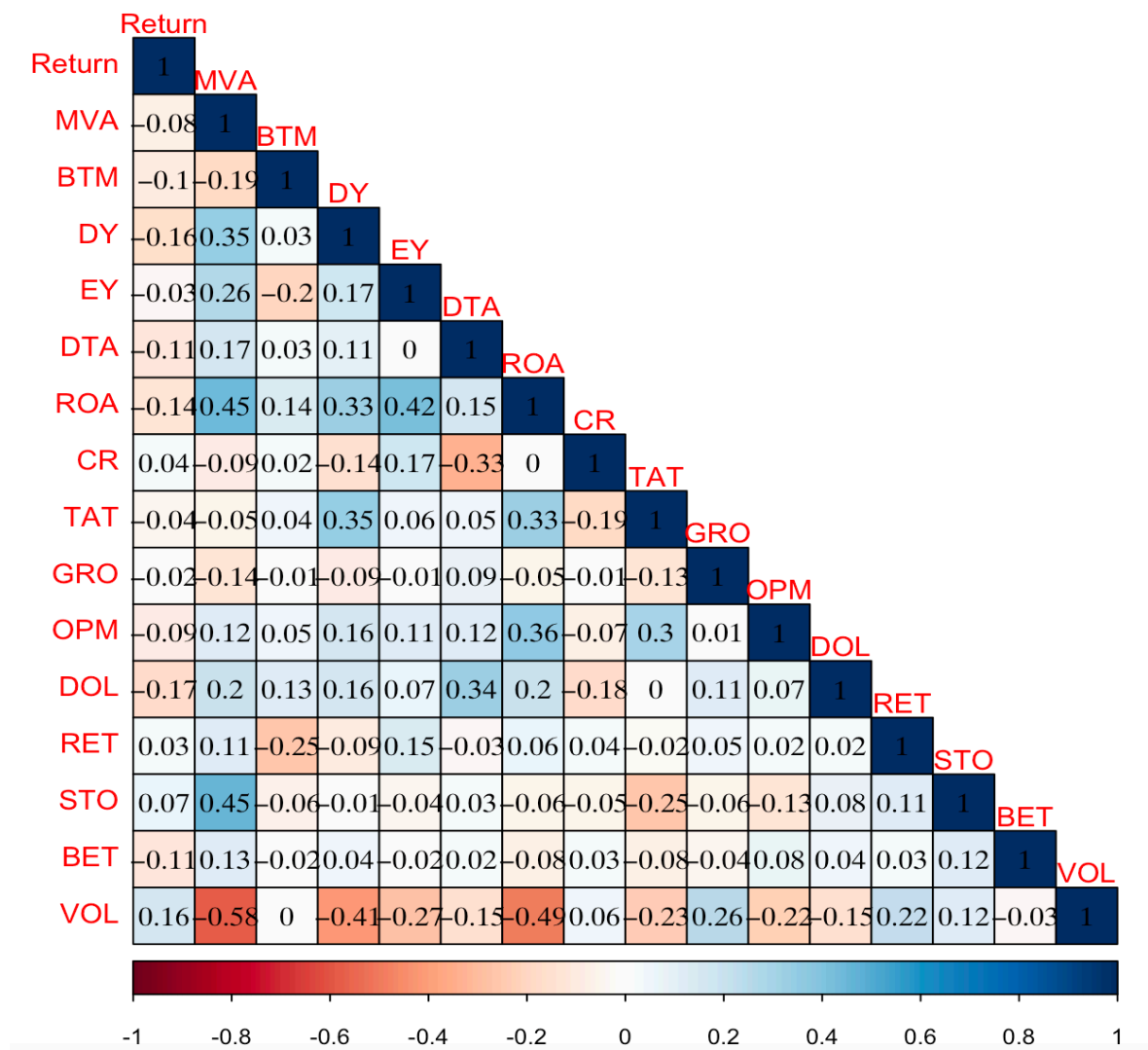


Figure 12. Correlation matrix of Covid crash data.

Same intuitiveness goes for positive correlation of 0.45 between STO and MVA, large cap stocks also appear more liquid in terms of trading volume. Analysing correlations between crash and recovery period returns and the used predictors interestingly show that most of the variables have negative correlation to returns. This could imply a weak relationship between the used variables and returns during the period. Overall, the correlations are also lower than in the financial crisis data. Some of the relationships seems more intuitive than others, for example pre-crash DTA, or leverage, is negatively correlated with returns. Similar relationship is also with DOL, high operating leverage affects returns negatively. Conversely, STO is positively correlated with returns, which gives indication of capital movement towards more liquid stocks during the crash. VOL has at first glance an

unintuitive relationship with returns, higher pre-crash volatility is associated with higher returns during the crash and recovery period.

Stratified random splitting of the dataset produces a training set of 479 observations and a testing set of 120 observations. Linear regression model for the Covid data is fitted first. Predictors of the linear model are summarized in Table 9 with estimated coefficients and their standard errors, t-statistics, p-values and VIF values. Statistically significant coefficients are indicated with ***, ** and * for confidence levels of 1%, 5% and 10% respectively. In the estimated model one less significant variable is found when compared to financial crisis. DY and STO are statistically significant as with the financial crisis data, but the other three variables that are significant, DOL, BET and DTA were not found significant earlier indicating different dynamics in the crash returns. However, DTA, STO and BET are only weakly significant at the 10% level, so some caution should be taken when interpreting these.

Table 9. Linear regression coefficients with Covid data.

	Coefficient	Std. Error	T-stat	P-value	VIF value
Intercept	7.70	16.54	0.47	0.64	
MVA	0.07	2.69	0.03	0.98	3.50
BTM	-6.59	4.06	-1.62	0.11	1.38
DY	***-3.16	1.28	-2.46	0.01	1.46
EY	0.11	0.07	1.49	0.14	1.48
DTA	*-0.28	0.16	-1.73	0.09	1.24
ROA	-0.19	0.17	-1.08	0.28	2.33
CR	-0.26	0.74	-0.35	0.72	1.29
TAT	5.65	5.77	0.97	0.33	1.59
GRO	-0.01	0.01	-0.74	0.46	1.17
OPM	-0.00	0.01	-0.15	0.88	1.27
DOL	***-0.40	0.15	-2.72	0.01	1.26
RET	0.03	0.04	0.57	0.57	1.29
STO	*4.58	2.80	1.65	0.10	1.83
BET	*-8.75	4.61	-1.90	0.06	1.09
VOL	0.18	0.19	0.94	0.35	2.72

Statistical testing of the linear model is briefly presented in Table 10. The Fisher test p-value represents that the independent variables in the model are jointly different from zero with statistical difference implying that the model performs better than a bare constant term. No severe multicollinearity is present, as the highest variable correlation in absolute terms is 0.58, again a correlation matrix for the training data is not separately presented as it is nearly identical to the matrix of the whole set. Calculated VIF values for the model also agree on this, no value is near even a threshold of 5.

Mean of residuals is zero, indicating that only random error is left in the error term. The Breusch-Pagan test for testing residual heteroscedasticity indicates possible problem of heteroscedasticity. Thus, the final reported standard errors, t-stats and p-values are estimated with White's heteroscedasticity robust errors -method and F-stat is also calculated using robust errors. Residual autocorrelation is tested with the Durbin-Watson test, and no autocorrelation is present. Exogeneity was tested with calculating correlation of variables and residuals, and the model does not inhibit endogeneity. Finally, the Jarque-Bera test is used for testing residual normality and p-value of 0.00 indicates that the residuals are not normally distributed.

Table 10. Statistical assumptions of linear regression in Covid data.

	Statistic Value	Assumption valid
F-test P-value	0.00	Yes
Multicollinearity	Highest corr. -0.58 & VIF 3.50	Yes
Mean of residuals	0.00	Yes
Breusch-Pagan test P-value	0.00	No
Durbin-Watson test P-value	0.21	Yes
Exogeneity	No correlation	Yes
Jarque-Bera test P-value	0.00	No

Violations in the linear regression assumptions indicated by Breusch-Pagan test and Jarque-Bera test imply that the estimators are not BLUE and residuals are non-normally distributed. As heteroscedasticity can cause the variance and standard errors in the estimators be biased and non-normality leads to unreliable confidence intervals, robust statistical inference of

larger population is difficult with the given model. However, the estimators itself are not biased and estimating with White's robust standard errors helps with the reliability of the errors.

Relative importances of the linear regression variables are presented in Figure 13 which shows that DOL, VOL, STO and DY are the most important. The importance of VOL is surprising as it is not a statistically significant variable. TAT and CR have the same ranking as with financial crisis data, implying that firm-level short-term liquidity has very little to do with crash and recovery period returns, at least in linear models.

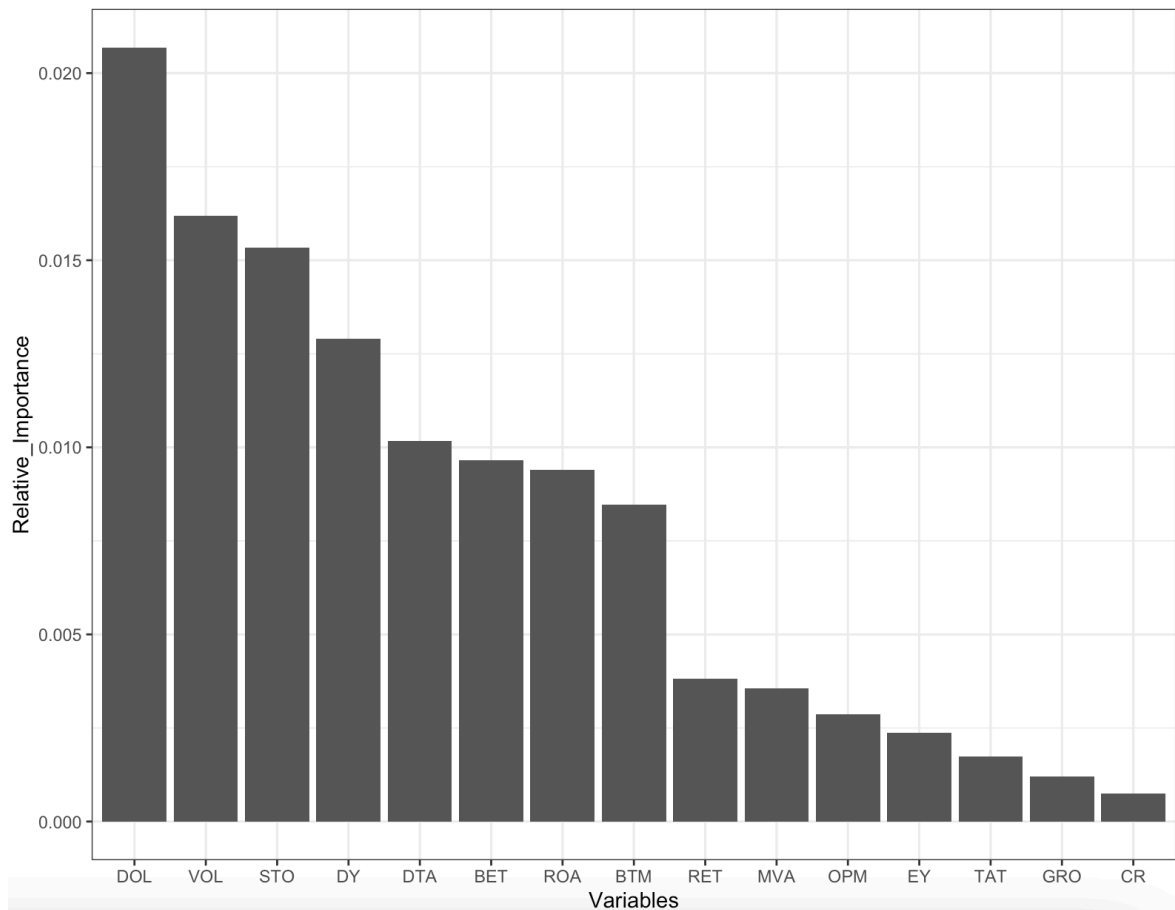


Figure 13. Variable importances in linear regression model with Covid data.

Random Forest model parameters were optimized using grid search and the same hyperparameter ranges as with financial crisis data. Optimal hyperparameters were as follows: the number of random features: 3, minimum node size: 20, and number of built

trees: 250. Number of random features lower boundary was set to 3 and it is not reasonable to explore smaller alternatives as risk of overfitting would grow. Similarly minimum node size was allowed to be 5, but grid search suggests using upper boundary of 20. Being on the boundary is not an issue, as it is only a minimum value accepted and risk of overfitting would increase to the other direction, i.e. lowering the minimum node size. Figure 14 visualizes the cross-validated performance of different parameter combinations in grid search.

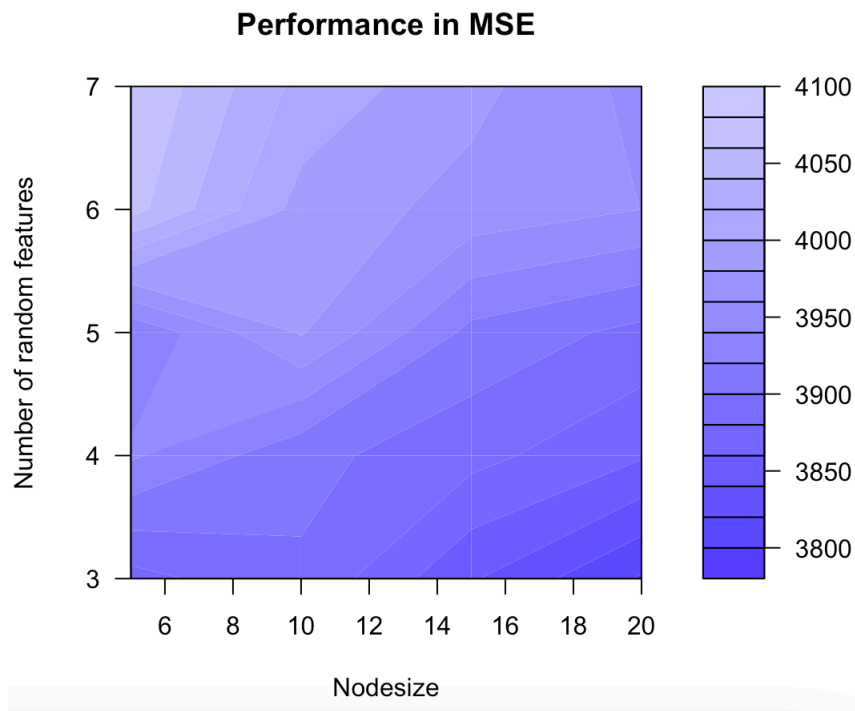


Figure 14. Cross-validated performance in grid search for Random Forest with Covid data.

Final Random Forest model was trained using the optimized parameters and evolution of this model's out-of-bag performance during the training with respect to the number of built trees is presented in Figure 15. 250 trees seem well sufficient for the Covid data as well, as only minor improvements are found after 100 trees and even 50 trees seem to have relatively low OOB performance when observing the OOB performance of the fully trained model.

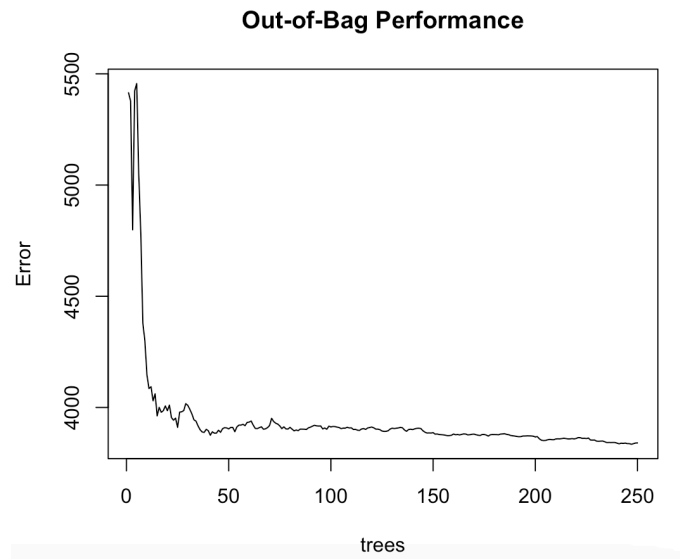


Figure 15. Random Forest Out-of-bag performance with Covid data.

Random Forest variable importances determined with the permutation feature importance method are presented in Figure 16. For Random Forest model, VOL is the most important variable with a considerable margin. Next are DOL and STO, which means that the most important variables are very similar to linear regression results. However, interestingly DY is the least important, which is in sharp contrast to linear regression and also to all models with financial crisis data.

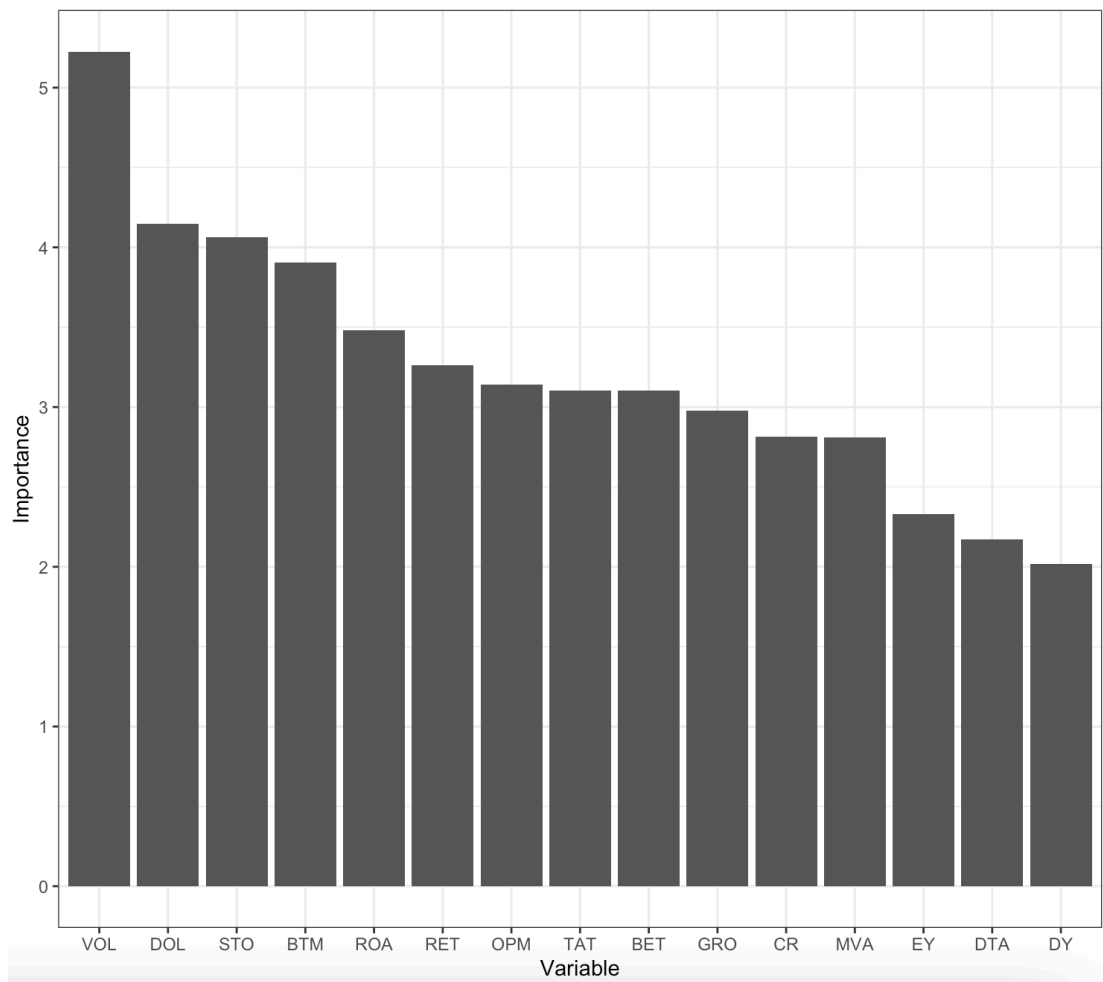


Figure 16. Variable importances in Random Forest model with Covid data.

Correlation of model predicted return and variables in Random Forest is presented in Table 11. 10 out of 15 variables seem to have a negative affiliation with model predicted returns. In addition, some of the affiliations are somewhat counterintuitive. For example, volatility is positively correlated with crash and recovery period returns and ROA is negatively correlated with returns. On the other hand, results are similar to linear regression. Also, it should be kept in mind that the correlation measure is linear, while the model is not.

Table 11. Correlations of predicted returns and variables in Random Forest with Covid data.

	Positive/Negative Affiliation	Correlation with predicted return
MVA	Negative	-0.18
BTM	Negative	-0.20
DY	Negative	-0.33
EY	Negative	-0.06
DTA	Negative	-0.23
ROA	Negative	-0.31
CR	Positive	0.13
TAT	Negative	-0.13
GRO	Positive	0.02
OPM	Negative	-0.27
DOL	Negative	-0.34
RET	Positive	0.23
STO	Positive	0.19
BET	Negative	-0.16
VOL	Positive	0.45

Support vector regression results are presented next. Grid searching for SVR with Covid data was performed using the same parameters as with financial crisis data. Ranges are [0.05 – 0.95] with steps of 0.1 for epsilon, [0.1 – 10.1] with steps of 0.5 for cost and [0.02 – 0.2] with steps of 0.02 for gamma. The iteration ranges are determined to be around the proposed default values in R, in order to see if an increase or decrease in parameters boosts model performance. Results of grid search are presented in Figure 17, with both epsilon and gamma plotted against cost.

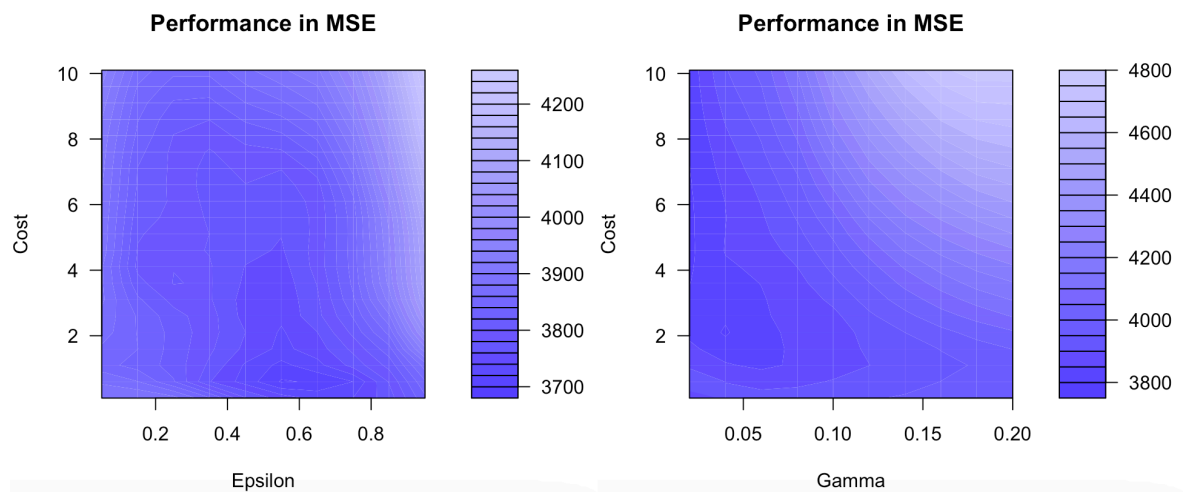


Figure 17. Cross-validated performance in grid search for Support vector regression in Covid data.

The optimal parameter combination is found at epsilon of 0.65, cost of 0.6 and gamma of 0.02 and the final SVR model is trained using these. SVR model is built with RBF kernel and features 141 support vectors, which is roughly 29.4% of the sample size. After fitting the model to training data, variable importances are calculated. Calculation is done using the same methodology as with Random Forest.

Permutation feature importance -based results are presented in Figure 18. By a considerable margin, the most important variable is DOL, which is also an important variable in linear regression and Random Forest. The following variables in importance differ from the other modelling approaches and EY is given a negative importance, indicating that it does not bring any added value to the model. Interesting notion is also that the importance scores are significantly lower than with Random Forest or SVR with financial crisis data. This gives indication that the SVR model with Covid data is weaker with these variables, as shuffling in the permutation feature importance does not lead to significant rise in the error.

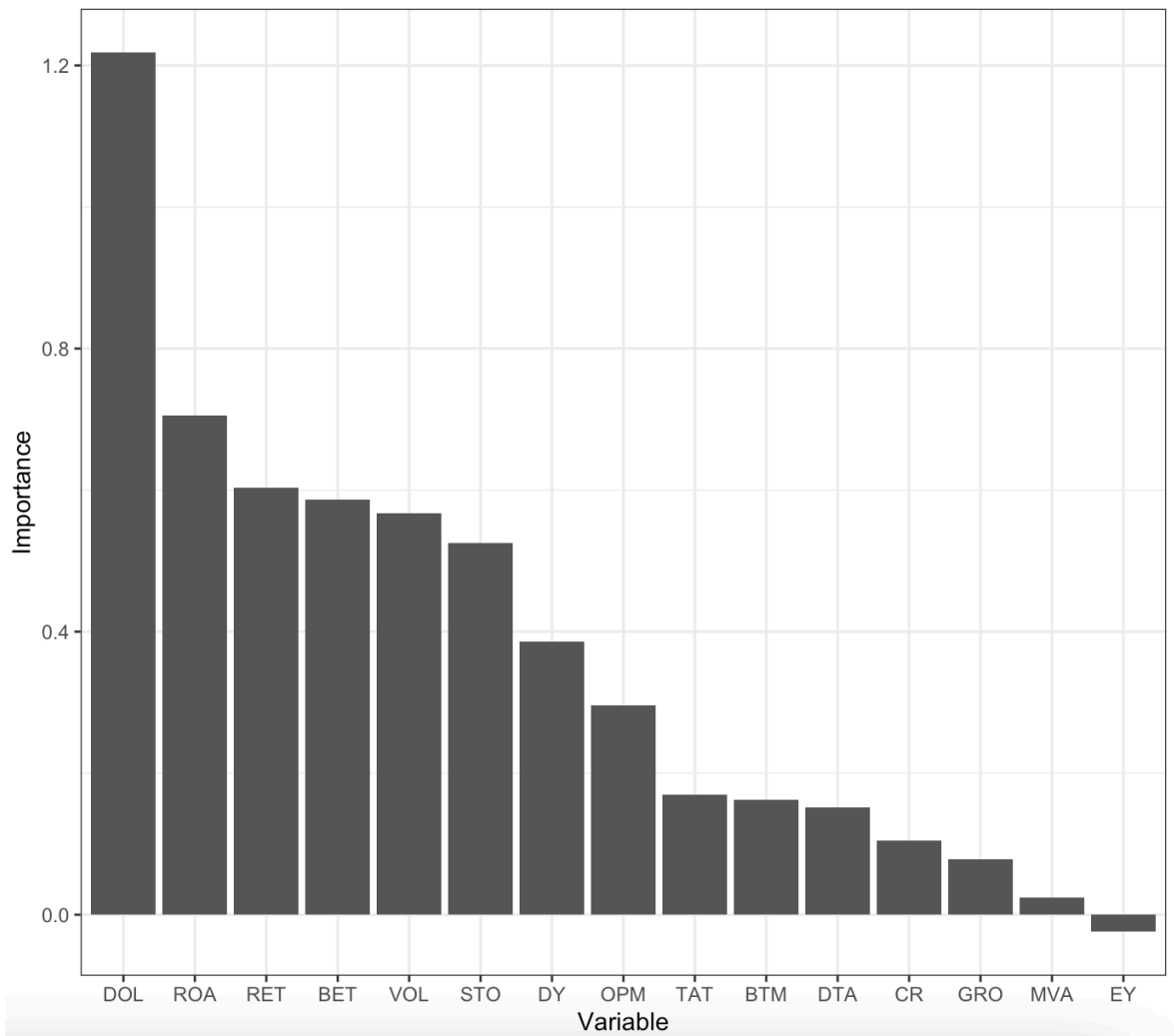


Figure 18. Variable importances in Support vector regression in Covid data.

Correlation analysis of SVR model predicted returns and variables is presented in Table 12. 9 out of 15 variables have negative affiliation with predicted returns based on this measure, which are the same as in Random Forest, with the exception of EY whose correlation is 0. Affiliations are also similar to linear regression coefficients.

Table 12. Correlations of predicted returns and variables in Support vector regression with Covid data

	Positive/Negative Affiliation	Correlation with predicted return
MVA	Negative	-0.02
BTM	Negative	-0.39
DY	Negative	-0.36
EY	-	0.00
DTA	Negative	-0.29
ROA	Negative	-0.28
CR	Positive	0.09
TAT	Negative	-0.17
GRO	Positive	0.01
OPM	Negative	-0.04
DOL	Negative	-0.72
RET	Positive	0.44
STO	Positive	0.19
BET	Negative	-0.11
VOL	Positive	0.31

Table 13 combines the modelling accuracies of all modelling approaches for Covid data. Random Forest appears as the best model in-sample, although its accuracy decreases more than with linear regression or SVR in the testing data. In testing data SVR is the best in RMSE and MAE, but Random Forest is slightly better with correct signs. Noteworthy, all models predict the correct sign of return with less than 50% accuracy in the testing data. This is troubling information, as if an investor is concerned whether the period return is negative or positive, the models produce worse results than a random guess. However, it is not known whether the sign prediction accuracy is the same for positive and negative results. In comparisons of the R-squareds Random Forest performs again significantly better than others, but it is likely explained by overfitting bias. For the Covid data, linear regression explains the variation in return better than SVR, in terms of R^2 .

Table 13. Modelling accuracies in Covid data

	Linear Regression	Random Forest	Support Vector Regression
RMSE, training	58.26	39.77	58.58
RMSE, testing	60.75	60.06	57.47
MAE, training	35.52	22.42	34.33
MAE, testing	40.84	39.48	37.11
% Correct sign predictions, training	61.59	78.91	59.08
% Correct sign predictions, testing	48.33	49.17	48.33
R-Squared, %	11.91	58.95	10.94
Adjusted R-Squared, %	9.05	57.62	8.05

4.3 Applying financial crisis -based models on Covid period

Prediction results of Covid period returns with financial crisis -based models are briefly addressed in this chapter. As can already be seen in the previous sections, the models and thus the dynamics between returns and variables differ between the two periods. Nevertheless, as it is interesting to know whether an investor could have used financial crisis data for predictions to the Covid period, all the models built with financial crisis data are applied to the Covid data. The models were not trained with the full set, as it is not likely to provide any improvement to the models that were built with 80% of data. Thus, the hyperparameters on the ML approaches are also unchanged.

Modelling accuracies are presented in Table 14. As could be presumed, the modelling accuracies decrease in the testing set. The worst performer by far is the linear regression. Interestingly, the ML models do not decrease on accuracy nearly as much and remain in the same regions as on financial crisis out-of-sample testing. Support vector regression's MAE on testing is lower than on training data, which is a rather counterintuitive result. However, as with Covid-models on out-of-sample data, correct sign predictions are again under 50%, indicating that the models perform worse than a random guess, when considering only whether period return is positive or negative.

Table 14. Modelling accuracies with financial crisis models on Covid data.

	Linear Regression	Random Forest	Support Vector Regression
RMSE, training	80.51	38.05	74.14
RMSE, testing	290.67	72.31	75.34
MAE, training	57.02	26.55	51.98
MAE, testing	102.26	44.75	48.02
% Correct sign predictions, training	65.35	93.73	70.63
% Correct sign predictions, testing	47.41	48.75	48.25

5. Discussion

In financial crisis models, the most important variables are quite consistent. Dividend yield, earnings yield and share turnover are found to be the most important, in line with previous research findings, regardless of the research period. It seems that so-called value companies with higher dividend- and earnings yields, and liquid stocks are bought by investors during the research period, assuming that return in those stocks is generated by increasing demand and rising prices. Interestingly, stock-level characteristics seem to dominate in predictive ability, implying that the firm-level characteristics are more rationally priced into stock prices regardless of market turbulence and give less significant predictive edge. However, ROA and OPM are still statistically significant in linear regression. Most important firm-level characteristics are related to profitability, leverage, and prior returns. One interesting notion is that in all models, stock beta is poorly related to returns.

Model accuracies in ML approaches are better than with linear regression. This implies that they are able to catch nonlinear patterns that the linear regression is blind to. One arbitrary example could be leverage, which is generally seen to increase shareholder profits up to a certain point, whereas too high leverage starts to decrease returns. The best model appears to be Random Forest with a significantly high adjusted R^2 of 81%, lowest RMSE and MAE, and the best accuracy in correct signs. However, Random Forest seems to suffer from some overfitting, as accuracy measures decay closer to other models in out-of-sample data. In the end, the differences in out-of-sample accuracies are small, leaving it questionable whether it pays to use more complex methods than linear regression.

Most important variables in Covid data models are not as consistent as with financial crisis. Similarities are still found across models, and degree of operating leverage, prior volatility and share turnover seem to be among the most important on aggregate, with the most important being DOL. Relationship of DOL and STO is intuitive, liquid stocks perform better and companies with less “agile” asset base (i.e. lot of fixed assets) have lower returns. However, in Covid period, previous volatility is positively related to returns which is

interesting to say the least. It seems that capital has flowed to riskier stocks during Covid, at least on this metric.

In Covid period, dominance of stock level characteristics is not as strong as in financial crisis data, but still present. Results differ from financial crisis, as dividend yield and especially earnings yield are not important. DY is still statistically significant in the linear regression, but its effect on the model is reverse to what was documented in financial crisis. The most important firm-level characteristics excluding DOL, are profitability variables.

Modelling accuracy results are slightly more mixed than with financial crisis data, but ML approaches again seem to perform better. Random Forest displays signs of overfitting and the improvement over linear regression provided by ML methods on out-of-sample basis is minor. Interestingly with all models, correct sign predictions are less than 50%, which implies poor predictability with the given models. When financial crisis models are used for prediction on Covid period, the superiority of ML approaches is very clear and their decrease in accuracy is moderate, when comparing to Covid models. Even though the standard deviations of realized cross-sectional returns are 86.73 and 60.96 for financial crisis and Covid period respectively, the level of RMSE and MAE in all modelling approaches make it difficult to justify using these models for predictions or to say that they are good representations of return determinants during the periods. Arguably some aspects are still captured.

Aside from accuracy and statistical measures, practical relevance can be explored by looking into potential returns. The models seem to produce relatively inaccurate predictions, but for example in financial crisis the correct sign predictions encourage to explore the potential returns acquired with model predictions. One thing is relatively clear, the dynamics of the two crashes are different and it would not have been reasonable to employ the characteristics learned from financial crisis to the Covid crash. A brief visit to the realized returns is presented in Table 15. In this simple example, stocks are divided into quintiles based on model predicted returns and average actual returns are calculated for each quintile, as if they

were an equal-weighted portfolio. Only Random Forest is included from ML methods, as it appeared generally better than SVR.

Table 15. Realized returns by quintiles of model predicted returns

	Financial crisis		Covid crash	
	Linear regression	Random Forest	Linear regression	Random Forest
1st quintile	36.38	96.56	36.75	75.07
2nd quintile	15.08	22.08	23.03	18.98
3rd quintile	-11.09	-25.70	7.58	4.73
4th quintile	-16.73	-53.83	5.07	-8.40
5th quintile	-60.22	-75.91	-9.23	-27.34

In this example it can be seen that Random Forest modelling produces far better results than linear regression. The returns are not risk adjusted, but if an investor had been able to follow these models using only a long strategy investing in the top quintile, returns of around 75-96% could have been harvested for given period, whilst the market return of the period is obviously zero. The problem is of course that an investor would have not had these models available. In addition, the modelling results of the two periods were rather different, perhaps due to different root reasons of the crashes. Hence, using previous crash models on Covid period would have not been beneficial either.

6. Conclusions

This thesis studied stock returns in two market crash and recovery periods by modelling the returns with stock- and firm-level characteristics such as volatility, dividend yield and leverage. Research was conducted with data from Nordic stock markets, by using linear regression and two machine learning algorithms: Random Forest and Support Vector Regression. Two research hypotheses were imposed: stock and firm characteristics are important in modelling the period returns and machine learning approaches are better at this modelling.

It was found that stock and firm characteristics can be used to model the returns but it seems that even in a market crash event, the factors are relatively well priced into stocks, or at least they do not exhibit significant predictive abilities at aggregate level, judging on the adjusted R^2 values. Notably, Random Forest model shows relatively high adjusted R^2 s, but also signs of overfitting, which makes its R^2 unreliable. Some of the used variables seem to have explanatory power on future returns, but the models should be specified further to be useful in gaining a good view of return dynamics, e.g. by including industry variables.

Machine learning approaches appear better in the modelling in both periods. They outperform linear regression in both in-sample and out-of-sample predictions, and also in the realized returns examination. However, their superiority is not as great as expected, being in some cases very negligible. Thus, a conclusion can be drawn that they can be recommended for practical applications aiming to find even the smallest edge in predicting returns, but for general use the traditional statistical approach of linear regression is more recommendable, because of its simplicity, ability to draw statistical inferences and depict understandable relationships between dependent and independent variables.

The accuracy measures RMSE and MAE varied from 38 to 77 and 22 to 60, respectively. These imply that the models are relatively inaccurate, but it needs to be kept in mind that also the standard deviations were 81% and 61% for the realized cross-sectional period

returns. For correct sign predictions, over 68% accuracy in test data was found for financial crisis period, but then for Covid period the test sample accuracy was under 50%, meaning worse than a random guess.

In financial crisis, three most important characteristics were dividend yield, earnings yield and share turnover (stock liquidity) when looking at all models' results at aggregate level. These are all positive drivers of returns, while negative drivers are less important. Volatility is one example of negative drivers. In Covid period, the most important characteristics on aggregate inspection are degree of operating leverage (negative driver), volatility (positive driver), and share turnover (positive driver).

In relation to previous literature, the results are not perfectly in line with any of previous studies. However, some of the important variables found across previous studies are found to be important also in this study. The determinants of returns vary in time and across markets, raising questions whether sustaining long term relationships can be found with this type of research. It can also be said that results for the two periods are rather different, implying that it would have not been beneficial to attempt trading strategies in Covid crash based on characteristics learned in financial crisis data. However, one similarity is the role of stock liquidity, measured in turnover, or trading volume.

Future research could be done for example by specifying the models further with e.g. adding industries or macroeconomic variables to the models. One definitely intriguing research topic on these models would be a optimization of the machine learning methods. It would be interesting to see how much the model accuracies could be improved by computational aspects, parameter tuning and feature engineering. Also, more generalizable characteristic-based models could be attempted by taking more crash and recovery periods from different time-spans and generate a mixed dataset from all of these and train models with such data.

References

- Acharya, V., Philippon, T., Richardson, M. & Roubini, N. 2009. "The Financial Crisis of 2007-2009: Causes and Remedies" *Financial Markets, Institutions & Instruments*, Vol. 18, Iss. 2, pp. 89-137.
- Artmann, S., Finter, P., Kempf, A., 2012. "Determinants of Expected Stock Returns: Large Sample Evidence from the German Market" *Journal of Business Finance & Accounting*, Vol. 39, No. 5, pp. 758-784.
- Avon, J. 2015, *The Basics of Financial Modeling*, Apress, Berkeley, CA.
- Baker, M., Wurgler, J., 2006. "Investor Sentiment and the Cross-Section of Stock Returns", *The Journal of Finance*, Vol. 61, Iss. 4, pp. 1645-1680.
- Bannigidadmth, D., Narayan, P., 2016. "Stock return predictability and determinants of predictability and profits" *Emerging Markets Review*, Vol. 26, pp. 153-173.
- Bhavsar, H. & Panchal, M. 2012, "A Review on Support Vector Machine for Data Classification", *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 1, iss. 10, pp. 185-189.
- Biau, G. & Scornet, E. 2016. "A random forest guided tour", *TEST*. Vol. 25, pp. 197-227.
- Bilokon, P., Halperin, I. & Dixon, M.F. 2020, *Machine Learning in Finance: From Theory to Practice*, Springer.
- Boser, B., Guyon, I. & Vapnik, V. 1992, "A training algorithm for optimal margin classifiers", *Proceedings of the fifth annual workshop on Computational learning theory*, ACM, pp. 144-152.
- Breiman, L. (2001) *Random Forests*. *Machine Learning*. Vol. 45, No. 1, pp. 5-32.
- Brooks, C. 2008, *Introductory econometrics for finance*, 2nd edn, Cambridge University Press, Cambridge.
- Buffett, W. 1987, *Berkshire Hathaway Shareholder letter of 1986*. [Website] Available: <<https://www.berkshirehathaway.com/letters/letters.html>>

- Carhart, M.M. 1997, "On Persistence in Mutual Fund Performance", *The Journal of Finance*, Vol. 52, Iss. 1, pp. 57-82.
- Chairakwattana, K. & Nathaphan, S. 2014, "Stock Return Predictability by Bayesian Model Averaging: Evidence from Stock Exchange of Thailand", *International Journal of Economic Sciences*, vol. 3, no. 1, pp. 47-63.
- Chen, N., Roll, R., Ross, S. 1986. "Economic Forces and the Stock Market", *The Journal of Business*, Vol. 59, Iss. 3, pp. 383-403.
- Cortes, C. & Vapnik, V. 1995. "Support-Vector Networks", *Machine Learning*, vol. 20, no. 3, pp. 273-297.
- Delen, D., Kuzey, C. & Uyar, A. 2013, "Measuring firm performance using financial ratios: A decision tree approach", *Expert Systems with Applications*, vol. 40, no. 10, pp. 3970-3983.
- Drucker, H., Burges, C.C., Kaufman, L., Smola, A.J., Vapnik, V.N. 1997, "Support Vector Regression Machines", *Advances in Neural Information Processing Systems 9*, MIT Press, pp. 155-161.
- Dzikevičius, A. & Šaranda, S. 2011, "Can financial ratios help to forecast stock prices?", *Journal of Security and Sustainability Issues*, vol. 1, no. 2, pp. 147-157.
- Fama, E.F. 1970, "Efficient Capital Markets: A Review of Theory and Empirical Work", *The Journal of Finance (New York)*, vol. 25, no. 2, pp. 383-417.
- Fama, E.F. & French, K.R. 1988, "Dividend yields and expected stock returns", *Journal of Financial Economics*, vol. 22, no. 1, pp. 3-25.
- Fama, E.F. & French, K.R. 1993, "Common risk factors in the returns on stocks and bonds", *Journal of Financial Economics*, vol. 33, no. 1, pp. 3-56.
- Fama, E.F. & French, K.R. 2004, "The Capital Asset Pricing Model: Theory and Evidence", *The Journal of Economic Perspectives*, Vol. 18, Iss. 3, pp. 25-46.
- Fama, E.F. & French, K.R. 2015, "A five-factor asset pricing model", *Journal of Financial Economics*, vol. 116, no. 1, pp. 1-22.

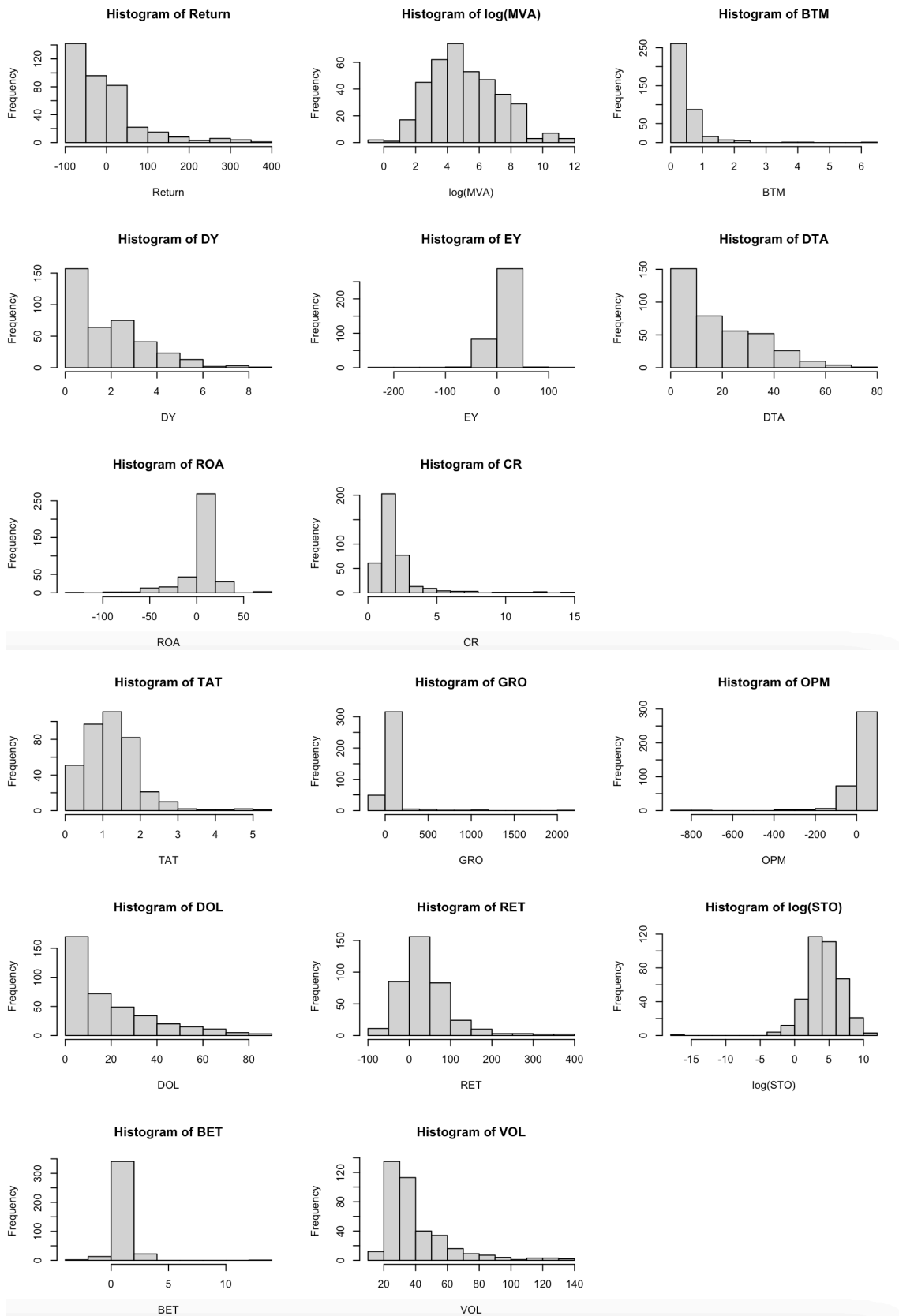
- Fauzi, R., Wahyudi, I., 2016. "The effect of firm and stock characteristics on stock returns: Stock market crash analysis", *The Journal of Finance and Data Science*, Vol. 2, Iss. 2, pp. 112-124.
- Fisher, A., Rudin, C., Dominici, F., 2019. "All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously", *Journal of Machine Learning Research*, Vol. 20, Iss. 177, pp. 1-81.
- García-Feijóo, L. & Jorgensen, R. D. 2010. "Can Operating Leverage Be the Cause of the Value Premium?" *Financial Management*, Vol. 39, No. 3, pp.1127-1154.
- Gerlow, M., Irwin, S., Liu, T. 1993. "Economic Evaluation of Commodity Price Forecasting Models", *International Journal of Forecasting*, Vol. 9, Iss. 3, pp. 387-397.
- Graham, B. & Zweig, J., 2003. *The Intelligent Investor*, Revised edition. 4th Edition, United States. HarperCollins Publishers Inc.
- Guerard, J. (2013) *Introduction to Financial Forecasting in Investment Analysis*. [Online]. New York, NY: Springer New York.
- Hahn, J. & Yoon, H., 2016. "Determinants of the cross-sectional stock returns in Korea: evaluating recent empirical evidence" *Pacific-Basin Finance Journal*, Vol. 38, pp. 88-106.
- Haugen, R., Baker, N. 1996. "Commonality in the determinants of expected stock returns" *Journal of Financial Economics*, Vol. 41, No. 3, pp. 401-439.
- Hsu, C., Chang, C., Lin, C. 2003. "A Practical Guide to Support Vector Classification".
- Inoue, A., Kilian, L., 2004. "In-Sample or Out-of-Sample Tests of Predictability: Which One Should We Use?", *Econometric Reviews*, Vol. 23, No. 4, pp. 371-402.
- Joshi, A.V. 2020, *Machine learning and artificial intelligence*, 1st edn, Springer, Cham, Switzerland.
- Jain, A. "Support Vector Machines (S.V.M) – Classifiers and Kernels", [Website] Available: <<https://medium.com/@apurvjain37/support-vector-machine-s-v-m-classifiers-and-kernels-9e13176c9396>>
- Kaul, A., Kayacetin, N. 2017. "Flight-to-quality, economic fundamentals, and stock returns", *Journal of Banking & Finance*, Vol. 80, pp. 162-175.

- Kutateladze, V. 2021, "The kernel trick for nonlinear factor modelling", *International Journal of Forecasting*, June 2021.
- Kwak, N. 2013, "Nonlinear Projection Trick in Kernel Methods: An Alternative to the Kernel Trick". *IEEE Transaction on Neural Networks and Learning Systems*. Vol. 24, no. 12, pp. 2113-2119.
- Lewellen, J. 2004, "Predicting returns with financial ratios", *Journal of Financial Economics*, vol. 74, no. 2, pp. 209-235.
- Lindeman, R., Merenda, P., Gold, R., 1980. "Introduction to Bivariate and Multivariate Analysis. Glenview, IL: Scott, Foresman and Company.
- Mattera, D. & Haykin, S. 1999, "Support Vector Machines for Dynamic Reconstruction of a Chaotics System", *Advances in Kernel Methods: Support Vector Learning*, pp. 211-241. MIT Press, Cambridge, MA.
- McLean, D., Pontiff, J. 2016. "Does Academic Research Destroy Stock Return Predictability?" *The Journal of Finance*, Vol. 71, Iss. 1, pp. 5-32.
- Mishkin, F., White, E. 2002. "US stock market crashes and their aftermath: implications for monetary policy". NBER Working Paper No. 8992.
- Musallam, S. R. 2018, "Exploring the Relationship between Financial Ratios and Market Stock Returns", *Eurasian Journal of Business and Economics*, vol. 11, no. 21, pp. 101-116.
- Novy-Marx, R. 2011. "Operating Leverage" *Review of Finance*, Vol. 15, No. 1, pp. 103-134.
- O'Brien, R. 2007. "A Caution Regarding Rules of Thumb for Variance Inflation Factors." *Quality & Quantity*. Vol. 41, Iss. 5, pp. 673-690.
- Opitz, D. & Maclin, R. 1999. "Popular Ensemble Methods: An Empirical Study", *The Journal of Artificial Intelligence Research*. Vol. 11, pp. 169-198.
- Opitz, S. & Szimayer, A. 2018. "What drives flight to quality?", *Accounting & Finance*, Vol. 58, pp. 529-571.
- Pech, C.O.T., Noguera, M., White, S. 2015, "Financial ratios used by equity analysts in Mexico and stock returns", *Contaduría y Administración*, vol. 60, pp. 578-592.

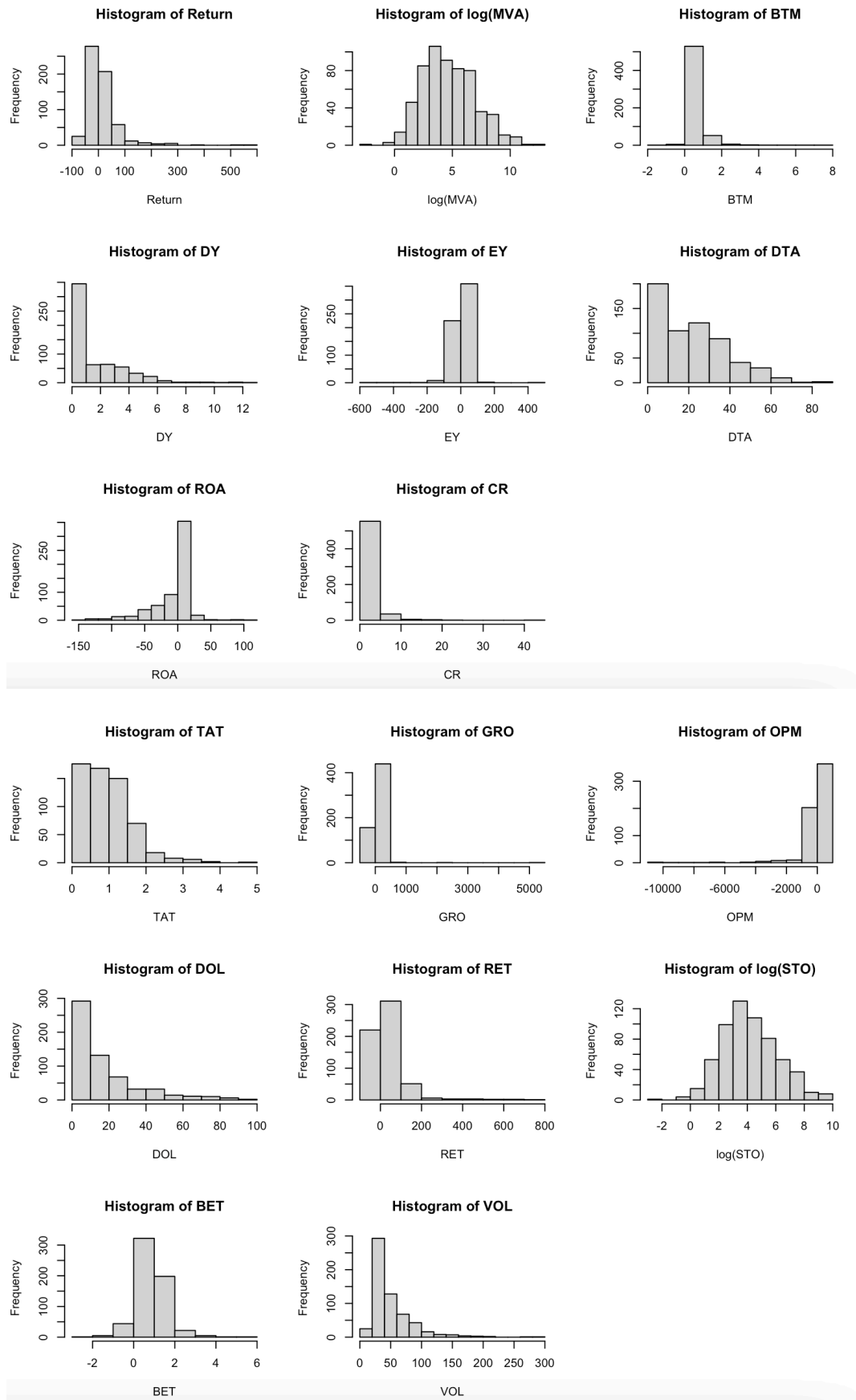
- Penman, S. H. (2010) Financial Forecasting, Risk and Valuation: Accounting for the Future. *ABACUS A Journal of Accounting, Finance and Business Studies*, Vol. 46, iss./no 2, pp. 211–228.
- Perold, A. 2004. “The Capital Asset Pricing Model”, *Journal of Economic Perspectives*, Vol. 18, No. 3, pp. 3-24.
- Pesaran, M. & Timmermann, A. 1992. “A Simple Nonparametric Test of Predictive Performance”, *Journal of Business & Economic Statistics*. Vol. 10, Iss. 4, pp. 461-465.
- Petcharabul, P. & Romprasert, S. 2014, “Technology Industry on Financial Ratios and Stock Returns”, *Journal of Business and Economics*, vol. 5, no. 5, pp. 739-746.
- Rapach, D.E., Wohar, M.E., 2006. “In-sample vs. out-of-sample tests of stock return predictability in the context of data mining”, *Journal of Empirical Finance*, Vol. 13, pp. 231-247.
- Rasekhschaffe, K.C. & Jones, R.C. 2019, "Machine Learning for Stock Selection", *Financial Analysts Journal*, vol. 75, no. 3, pp. 70-88.
- Ross, S. 1976. “The Arbitrage Theory of Capital Asset Pricing”, *Journal of Economic Theory*, Vol. 13, Iss. 3, pp. 341-360.
- Scornet, E. 2017. “Tuning parameters in random forests”, *ESAIM: Proceedings and surveys*. Vol. 60, pp. 144-162.
- Shu, M., Song, R., Zhu, W. 2021. “The ‘COVID’ crash of the 2020 U.S. Stock market”, *North American Journal of Economics and Finance*, Vol. 58.
- Siddik, N., 2020. “Economic stimulus for COVID-19 pandemic and its determinants: evidence from cross-country analysis”, *Heliyon*, Vol. 6, Iss. 12.
- Sun, J., Jia, M. & Li, H. 2011, "AdaBoost ensemble for financial distress prediction: An empirical comparison with data from Chinese listed companies", *Expert Systems with Applications*, vol. 38, no. 8, pp. 9305-9312.
- Tay, F. E. H. & Cao, L. 2001 “Application of support vector machines in financial time series forecasting”. *Omega*. vol. 29, no. 4, pp. 309–317.

- Twin, A. 2007. "Brutal day on Wall Street", CNN. [Website] Available:
<https://money.cnn.com/2007/02/27/markets/markets_0630/index.htm?cnn=yes>
- Vogel, H.L. 2018, Financial Market Bubbles and Crashes, Second Edition Features, Causes, and Effects, 2nd edn, Springer International Publishing, Cham.
- Wang, J., Meric, G., Liu, Z. & Meric, I. 2009, "Stock market crashes, firm characteristics, and stock returns", Journal of Banking & Finance, vol. 33, no. 9, pp. 1563-1574.
- Welch, I., Goyal, A., 2008. "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction" Review of Financial Studies, Vol. 21, No. 4, pp. 1455-1508.
- Zhang, Y. & Yang, Y. 2015. "Cross-validation for selecting a model selection procedure" Journal of Econometrics. Vol. 187, no. 1, pp. 95-112.
- Zimmermann, A. 2008. "Ensemble-Trees: Leveraging Ensemble Power Inside Decision Trees" Discovery Science, DS 2008. Lecture Notes in Computer Science, vol. 5255. Springer, Berlin.

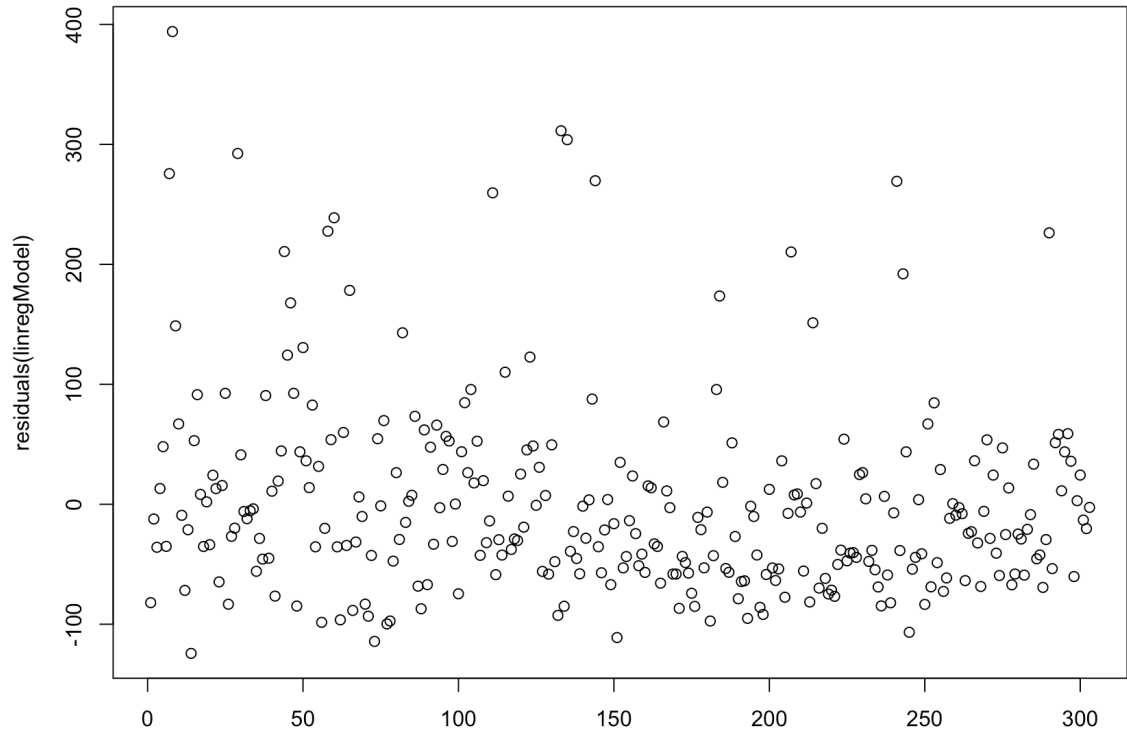
Appendix 1. Distributions of variables in financial crisis data.



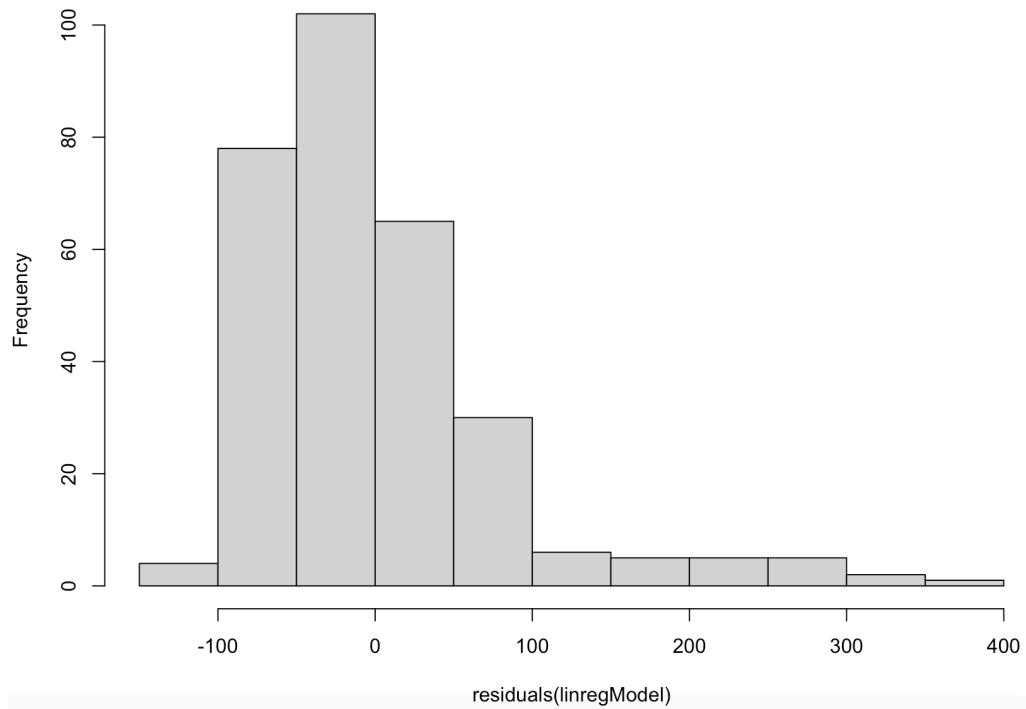
Appendix 2. Distributions of variables in Covid crash data.



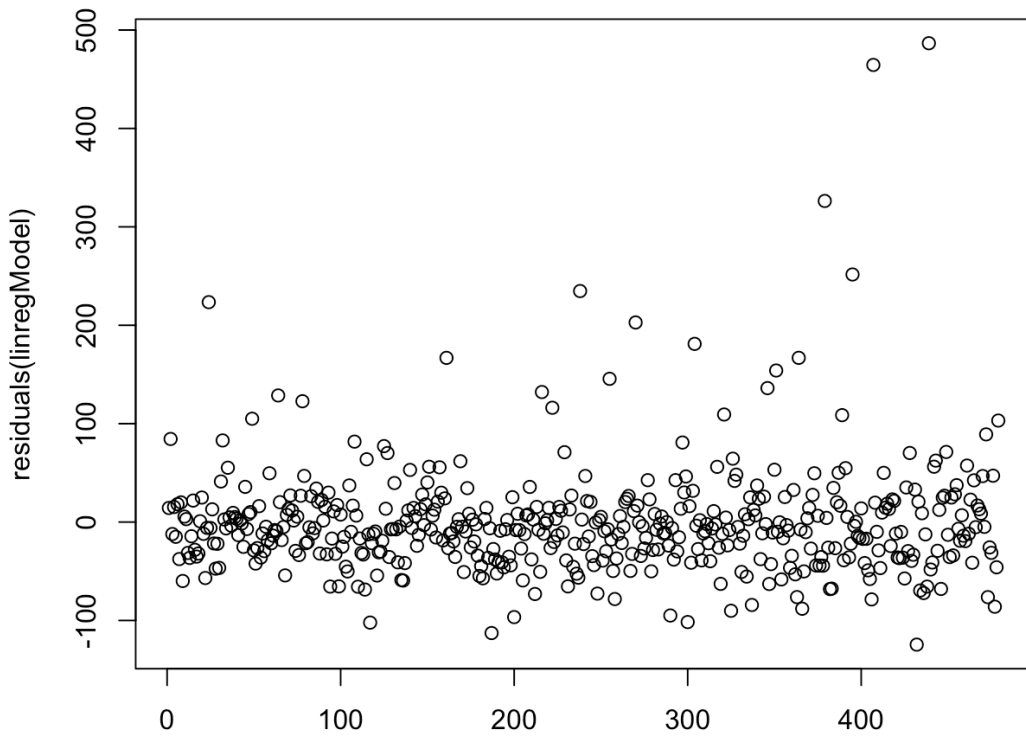
Appendix 3. Residual plot and distribution in linear regression on financial crisis data.



Histogram of residuals(linregModel)



Appendix 4. Residual plot and distribution in linear regression on Covid data.



Histogram of residuals(linregModel)

