Norwegian School of Economics
Bergen, Fall 2023

Sailing into the Storm?

Utilizing machine learning to predict defaults in the Norwegian shipping sector

Anders Viseth and Espen Rysjedal

Supervisor: Jonas Andersson

Master thesis, MSc in Economics and Business Administration,
Finance and Business Analystics

NORWEGIAN SCHOOL OF ECONOMICS

Acknowledgments

This Master Thesis was written in 2023 as a part of our degree in Economics and Business Administration, with a specialization in Financial Economics and Business Analytics. Throughout the process of completing this study, we have gained valuable insights into the Norwegian shipping industry and the work of modeling defaults.

This journey has been both challenging and rewarding, particularly given the extensive data we navigated. Most importantly our gratitude goes to Jonas Andersson for his valuable assistance. With his support and input the quality of our thesis has greatly improved. Furthermore, we would like to express our appreciation to the individuals at the Centre for Applied Research at NHH (SNF), including Endre Berner, Aksel Mjøs, and Marius Olving, for providing access to the database, contributing to the completion of our research.

Norwegian School of Economics

Bergen, December 2023

Anders Viseth

Espen Rysjedal

Abstract

In this thesis, we develop multi-year models to predict defaults in the Norwegian shipping industry. Our primary objective is to create a model suited for Norwegian shipping companies with high predictive accuracy of default. By incorporating shipping specific and macro variables in the model we aim to better capture the dynamics of this highly volatile and globally influenced industry. In the study we utilize two different machine learning techniques and the more traditional logit method and investigate the difference in accuracy between them. To further assess the performance of our model's, we compare them with the SEBRA-model, used by the Norwegian Central Bank to predict defaults of Norwegian companies. We base our analysis on a dataset retrieved from the Norwegian Corporate Accounts which after thorough cleaning contains 889 shipping companies whereof 19 are defaulted.

Our best performing model is the Random Forest, yielding an AUC of 87%, predicting defaults one year in advance, a performance comparable to the original SEBRA-model. For predictions two years prior, our AUC reduce to 76%. While the results from the other two models are slightly inferior, they are both better than our replicated SEBRA-model. Further findings indicate that oil price is the most important macro variable in our Random Forest model, a variable neglected in earlier research. Our prediction model is intended to be used by investors, banks, and other stakeholders involved in the Norwegian shipping industry. Although the models yield a high AUC they are estimated on an imbalanced dataset with few defaults, and this is a limitation which need to be considered when utilizing the models.

Contents

List of Figures

List of Tables

1. Introduction

Maritime shipping serves as the main mode of transportation for global trade, with approximately 90% of all traded goods transported across the world's oceans (OECD, n.d.). The world's total fleet is registered in over 150 nations and without shipping the import and export of goods necessary to sustain the modern world would not be possible (IMO, 2015). Moreover, the dynamic landscape of the shipping industry requires companies to navigate and address challenges arising from technological innovations, environmental and geopolitical challenges to ensure their survival (Katsomitros, n.d.).

The shipping industry is characterized by its volatility, where shipping companies can experience years of profitability followed by long periods with losses (Careratings, 2018). Moreover, ships tie up a lot of capital and the most expensive ships can cost as much as $225 million each. As a result of being a capital-intensive industry, capital can constitute to as much as 80% of the costs of running a shipping company (Stopford, 2009). To finance its operations shipping companies rely on bank loans as the most important source. The combination between volatility and high levels of debt, can increase the likelihood of default during bear market conditions, by failing to make the payments required under the loan agreement (Alexandridis, et al., 2018).

Hanjin Shipping was once one of the world's biggest shipping companies, but did not manage the downfall in the shipping industry. For many years the global economic downturn had affected profits in the shipping industry, led to overcapacity, lower freight rates and increasing debt (Illmer, 2017). With a debt of $5,4 billion Hanjin failed to get more money from their creditors, and ports started to refuse docking for Hanjin ships, in the fear of not getting paid docking fees (The Guardian, 2016). Hanjin went bankrupt in 2016 and immediately ceased operations (GAO, 2020). The collapse was not to be avoided and led to ripple effects for the whole economy (Graham, 2016).

Events such as Hanjin shipping are rare, but shipping defaults also occur in Norway. Norway is a global maritime power and currently the world's fourth largest shipping nation measured by value (Stautland, 2021). The maritime industry alone contributed to value creation equivalent to 8 percent of the Norwegian GDP in 2018 (NHO, n.d.). However, in the period between the peak year of 2014, and low year of 2017, there was a reduction of more than 25%

in value creation in the industry (Norwegian Shipowners' Association, 2021). During such economically challenging times companies experience negative profit, excess capacity, capital shortage, which in worst case can lead to defaults. Preventing defaults is crucial due to the substantial costs and extensive repercussions they impose at the individual, firm and regional levels. Just in our dataset we found that defaults lead to a total loss of equity amounting to 4.96 billion NOK[1], and a total asset loss of 7.67 billion NOK[2] as shown in Figure 1.1.

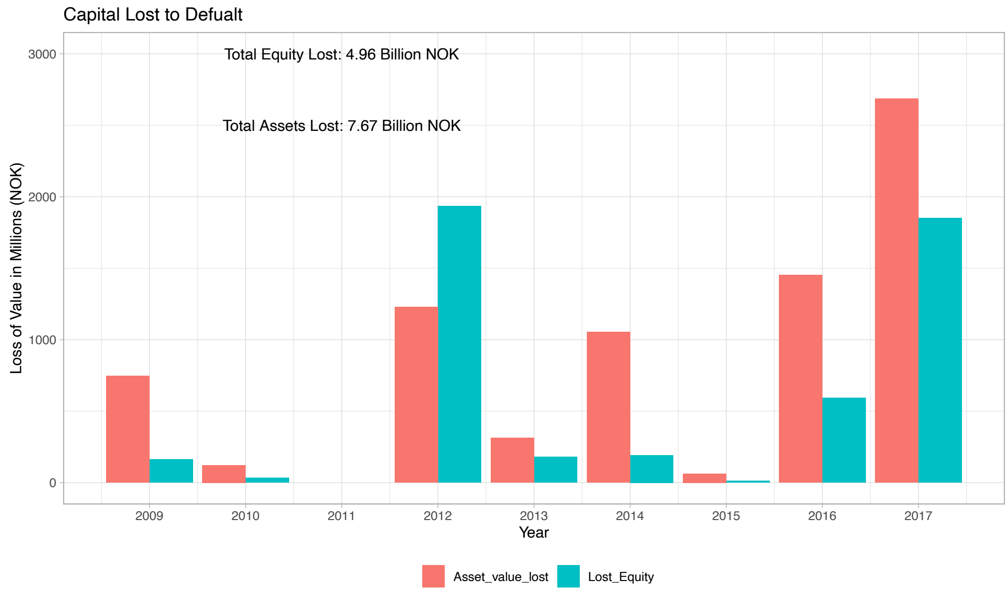

Figure 1.1: Asset and Value loss of companies by the year of default, measured by maximum assets and value in their lifetime.

To mitigate the costs associated with defaults and in an attempt to prevent future defaults in one of Norway's most important industries, we aim to estimate a prediction model suited for Norwegian shipping companies. Previous research regarding shipping defaults has focused on the international shipping companies, but the Norwegian shipping industry is still largely unexplored. Our study therefore makes a specific contribution to the literature on shipping defaults by exploring the Norwegian industry and the importance of macroeconomic and shipping specific variables in default prediction. In this study, unlike earlier studies, we will also utilize machine learning methods to predict shipping defaults, under the assumption that

[1] Calculated by difference of invested equity and equity at observed default, as equity and debt must equal assets, it is often negative when the debt is not repaid and asset depreciate, leading to an inflated measurement of lost equity.

[2] Calculated by maximum total asset value of default company in the dataset.

there might be non-linear relationships in the sector that are better estimated with machine learning methods. In this study we present a model that achieves high accuracy in predicting shipping defaults by utilizing the Random Forest classification method. We further observe that non-financial variables and the oil price are of most importance in our prediction model. Our research holds relevance for banks, investors, and stakeholders, offering insights to proactively manage or take action regarding companies at risk of future defaults.

The paper is structured as follows: In section 2 we provide a literature review which presents prior research regarding default, shipping defaults and the SEBRA-model. Section 3 describes the methods used to predict default and their accuracy measurement. In section 4 we present the data sample and in sector 5 describe the variables included in the SEBRA-model and our own model. Section 6 presents the results followed by a case in section 7. The results are discussed and summarized in section 8 and 9, respectively.

2. Literature Review

The first part of the literature review involves a definition of default. Further, we aim to establish an understanding of the broader predictive methodologies used across various industries by reviewing earlier research on default prediction. Lastly, we look closer at the literature concerning prediction of defaults in the shipping industry and the SEBRA-model, used by the Norwegian Central Bank.

2.1 Definition of Default

In this study we have defined default as instances where a company has entered liquidation, filed for bankruptcy, or faced equivalent situations. This is consistent with the study from Lozinskaia et al. (2017) which we further describe in section 2.3. The term “default” is used as a broader term than “bankruptcy”, a choice made to better capture the characteristic of our response variable. This more comprehensive definition ensures that our analysis covers not only bankruptcies but also a broader range of events such as mergers resulting from financial distress, or closures, all of which is captured in our response variable.

2.2 Early Research

Interest in forecasting financial default by using financial statement has existed for decades. The literature on bankruptcy prediction dates back to the 1930’s, where the initial studies utilized financial statement and ratio analysis to predict bankruptcy. These studies focused on ratios and compared the ratios of failed companies with ratios of successful firms. The studies conducted univariate analysis, which focus on the use of a single variable, and this form of analysis established the foundation for future development of bankruptcy models and the use of multivariate analysis (Bellovary, et.al, 2007)

One of the earliest and pioneering studies in the field of bankruptcy predictions is Beaver’s (1966) research. The study applies a univariate model to classify between bankrupt and non-bankrupt firms, based on the mean of 30 financial ratios for all five years before failure. His sample comprised 79 failed and 79 non-failed firms over 38 industries. Compared to earlier research Beaver took his study a step further and tested the predictive abilities of each individual ratio to distinguish between failed and non-failed companies. Of the ratios Beaver examined he found that cash flow to total debt was best to predict failure, followed by net income to total assets and then debt to assets. Although the study indicates that ratio analysis can be useful in the prediction of failure for at least five years prior to failure, the study has certain limitations. One limitation is that the study may be understating the usefulness of ratios. The reason is that if ratios can detect failure before it occurs, firms can initiate action and prevent failure and the sample would contain bias due to missing values for firms were ratios detected failure. The use of univariate model also represents a limitation as multivariate analysis could yield better predictions than single ratios alone, something Beaver himself suggest for future analysis.

Two years later Altman (1968) introduced a model which extended and improved upon earlier univariate models such as Beaver's and analyzed the differences between bankrupt and non-bankrupt firms by considering multiple ratios simultaneously. Altmann's study included a sample of 66 firms, evenly split into 33 firms classified as bankrupt and 33 firms categorized as financially healthy. Through the application of a multiple discriminant analysis, he developed a five-factor model to predict bankruptcies. Altman's model more commonly referred to the Z-score assigns a value to each firm, where a Z-score exceeding 2.99 indicates a healthy firm, while a firm with a score below 1.81 is considered to have a higher probability of experiencing financial distress or failure. A score between 1.81 and 2.99 are considered as in the "grey area", where the model can't distinguish between healthy and bankrupt firms, but where it exists a possibility of the company going bankrupt. Although Altman's model has some limitations, such as potential challenges for new firms with low earnings, he finds that his model has high predictive power one year before failure and compared to Beaver's univariate model the errors are significant lower.

In 1980, Ohlson's study emerged and challenged earlier studies in the field of bankruptcy predictions. He argues that the multi discriminant analysis relies on several assumptions which may not hold. These assumptions encompass the distributional properties of the predictors, the way failed, and non-failed firms are matched and the requirement that the variance-covariance matrices of the predictors are equal between the two groups. Ohlson (1980) therefore introduced the conditional logit analysis, where no assumptions need to be made regarding prior probabilities of bankruptcy or distribution of predictors. The logit model therefore overcomes the problems related to the multi discriminant analysis. As prior research Ohlson use financial ratios for predicting financial distress, and the result indicated that there were four underlying factors derived from the financial statement, which were statistically significant for the purpose of assessing failure. The logistic regression method proved to be superior to earlier methods and therefore became widely used by later research.

2.3 Research on Shipping Defaults

Over the years, in addition to the earlier studies done by Beaver, Altman and Ohslon, there have been a lot of research considering the topic of bankruptcy prediction. The topic is well documented and after the work of Altman the number and complexity of bankruptcy predictions models have increased dramatically (Gissel et al., 2007). A common characteristic of earlier research is their general focus, rather than on specific sectors, which can explain some of the scarcity in research concerning shipping defaults predictions.

The study by Grammenos et al. (2008) used a binary logit analysis to predict the probability of default, by the time of issuance, of high yield bonds offered by shipping companies. The study was conducted using a dataset consisting of 50 high yield bonds, issued by shipping companies between 1992 and 2004. Of the 50 bonds 13 of them defaulted, while the remaining 37 were active still active or had expired. In the analysis numerous variables, categorized into bond-related, industry-specific, and financial characteristics were tested, in order to best predict the probability of defaults. The result suggests that the primary indicators of bond defaults were financial ratios such as the gearing ratio and the amount raised over total assets ratio, as well as the industry specific variable considering the time-charter rates.

Mitroussi et al. (2016) used a binary logit model to examine the primary determinants of shipping loan credit risk and defaults, during times with financial turbulence in the shipping industry. The sample consisted of a portfolio of 30 shipping loans issued by a Greek bank for the period from 2005 to 2009. Out of these 30 loans, 12 loans had problems with their repayment, whereas the remaining loans were fully paid. Through assessing both financial and non-financial factors they find that both categories are important drivers of credit risk and defaults. Their result also suggest that shipowners experience, employability and market risk indicators are considered as the best criteria for evaluating shipping loans during turbulent market conditions.

Another study from 2016 emphasize the unique characteristics of the shipping industry and concludes that the financial factors are not the important factors that drives default in the shipping sector. Kavussanos and Tsouknidis (2016) argued that the shipping market is affected by the very volatile freight rate and other international markets the companies are exposed to. Hence, data extracted from financial statements is often considered "old" by the time a decision needs to be made. Based on this, their results suggest that the most important factors for predicting defaults are factors measuring the current and expected market conditions. The study estimates a logit model on panel data consisting of 128 shipping loans issued between 1997-2011.

Unlike the other studies, Lozinskaia et al. (2017) does not limit their analysis to defaulted companies, but also include firms considered as financially distressed. A sample of 192 listed shipping companies between 2001-2016 are used, and a logit model is estimated in order to predict the probability of default. Their key findings are in line with prior research and finds that both financial and non-financial factors should be considered when predicting financial default. Out of the factors used in their final model they find that Tobins Q[3], earnings before interest, taxes, depreciation, and amortization (EBITDA), GDP and total assets are statistically significant. An overview of previous studies is included in Table 2.1:

	Grammenos et al., 2008	**Mitroussi et al., 2016**	**Kavussanos,and Tsouknidis, 2016**	**Lozinskaia et al., 2017**	**Our study**
Time period	1992–2004	2005–2009	1997–2011	2001–2016	2005-2020
Sample	50 bonds	30 loans	128 shipping loans	192 shipping companies	889 Norwegian shipping companies
Defaults	13	7	12	41	19
Definition of default	Non-payment of interest to bondholder	Loan not repaid at maturity	Delay in payment of interest on the loan for more than 90-days	Bankruptcy, liquadation or reorganization	Liquadation process starts for the first time
Method	Binary logit model	Binary logit model, linear probability model	Binary logit model	Binary logit model, linear probability model, ordered	Binary logit model, Random Forest and SVM

Table 2.1: Summary of earlier research

[3] A ratio for how much the company's assets can fall in value when going bankrupt

The studies concerning shipping defaults all have in common that they use some form of the binary logit model to predict defaults. Despite variations in the specific factors included in each study, all studies incorporate both financial and non-financial factors as a result of the complexity of the shipping industry. Variations in findings can be attributed to differences in how default is defined, the specific time periods covered by their samples, as well as the fact that some studies predict loan default and others bankruptcy. Our study, like the previously mentioned studies, will utilize a binary logit model and consider both financial and non-financial factors. However, what sets our study apart from prior studies is our focus on the Norwegian industry as well as the incorporation of the machine learning methods Random Forest and Support Vector Machines. By comparing our approach to prior findings, we aim to enhance the accuracy of default predictions and provide insights into the dynamics of the Norwegian shipping industry.

2.4 Local Adaption: The SEBRA-Model

The SEBRA model is an empirical model used by the Norwegian Central Bank to evaluate and predict default risk of Norwegian enterprises. The model, initially developed in 2001 by Eivind Bernardsen, incorporated 12 explanatory variables, encompassing financial ratios and company characteristics (Syvertsen, 2004).

The original SEBRA model was created based on a sample consisting of around 400 000 observations, after removing companies without significant assets. Out of the total sample 8436 companies were bankrupt. In the model factors such as profitability, liquidity, solidity, age, size, and industry characteristics were included to cope with different aspects of the businesses. A generalized logit model was used to predict the probability of a company going bankrupt and the final model achieved an accuracy of 82% and an AUC of 89,73% (Bernhardsen, 2001).

Bernardsen and Larsen (2007) revised and extended the original SEBRA model aiming to enhance the accuracy of classifying bankrupt firms even if it meant accepting slightly lower overall accuracy. They introduced two new models called SEBRA-basic and SEBRA-extended[4]. SEBRA-basic closely resembled the original model but included a set of sector variables which varied more over time than in the initial model. SEBRA-extended, based on the basic model, included three additional variables: the aggregated value of assets, accounts payable over total assets and payable fees over total assets. The revision was developed based on a new sample consisting of one million companies, with 20 000 of them identified as bankrupt. The result from Bernardsen and Larsen showed that the models performed almost equally well with an AUC of 88% for the basic and 89% for the extended.

While the two new models achieve nearly identical AUC scores, the basic model is slightly better when used to approximate potential bank loan losses.[5] On the other hand, the extended model outperforms the basic model in predicting defaults on a company level. (Bernhardsen & Kai, 2007)

[4] We will in the rest of the study focus on the extended SEBRA-model

[5] Analysis is conducted by weighting bankruptcy probability with the debt in each company and then sum over all companies, which give expected bank loan losses.

3. Methodology

For this thesis we have chosen three models for default prediction: Binary logit, Random Forest, and Support Vector Machines. The reason for using several approaches is that there might exist non-linear relationships between predictors and default risk that could possibly be better estimated through machine learning algorithms. Random Forest and SVM algorithms both separate the predictor spaces in different ways, and as we will come to see, the Random Forest model is more simplistic, whereas the SVM offers a range of options creating highly complex models. In our dataset there is a high degree of class-imbalance, with 7495 yearly observations, of which only 19 results in default. As 99.8% of observations belong to the class of healthy observations, traditional accuracy measures are of little use. For this reason, we will first introduce the confusion matrix, true and negative positive rates, the ROC curve, and the need for resampling techniques for model and accuracy estimation, before introducing the methods.

3.1 Performance Measures

3.1.1 Confusion Matrix

A confusion matrix can be used to assess the general performance of a classification model by presenting the amount of correct and incorrect predictions, to their true classes (James et al., 2021). An important part of the confusion matrix is the degree of true positive (TP) and true negative (TN) values, which are the degree of correctly predicted outcomes, to actual outcomes, as illustrated in Table 3.1:

	Actual	
Predicted	**Default**	**Healthy**
Default	**TP**	**FP**
Healthy	**FN**	**TN**

Table 3.1: Confusion Matrix, default and healthy predicted observations are shown in a table according to their true class.

The confusion matrix is especially useful in our case, as traditional performance measures are inappropriate due to the class imbalance of our problem. Predicting all observations as healthy would lead to an "accuracy" of 99.8%, preventing no defaults. Defaults can be costly events, and falsely predicting a default company as healthy, is much more costly than an opposite misprediction of a healthy company. A model in this case must be measured by its ability to predict true defaults, without a high degree of noise from false positive predictions, leading to the true and false positive rates.

3.1.2 True Positive Rates and False Positive Rates

$$True\ Positive\ Rate\ = \frac{TP}{TP + FN} \qquad False\ Positive\ Rate\ = \frac{FP}{FP + TN}$$

The True Positive Rate (Sensitivity) indicates the share of accurate default predictions, while the False Positive Rate indicates the amount of falsely predicted defaults at the same threshold. The models we will introduce can all generate probabilities of default in the range between 0 and 1, and a class prediction is made by the value predicted relative to a threshold (James et al., 2021). For instance, given a threshold of 0.20, an observation with a predicted probability of default of 0.22 will be classified as default. If the threshold was higher at 0.25, the same probability would result in a healthy prediction. A high TPR and low FPR is preferred as it shows the model predicts a high share of the true defaults, without predicting many false defaults, and can therefore distinguish well between the two classes. By lowering or increasing the threshold, the TPR and FPR will both increase or decrease respectively, and an optimal threshold can therefore be found depending on a user's preference of the rates. If the model cannot distinguish between the classes, it will only increase its ratio of correctly predicted defaults, by simultaneously increasing the ratio of false default predictions (James et al., 2021), creating equal rates for all thresholds. A classification model can therefore be evaluated by its ratio of true positives to false positives for each threshold. This dynamic can be illustrated graphically through the Receiver Operating Characteristic curve (ROC).

3.1.3 ROC Curve

The ROC curve provides a graphical representation of the trade-off between true predicted defaults along the y-axis, and falsely predicted defaults along the x-axis, as the threshold is changed (James et al., 2021). For a high threshold there will be fewer observations classified as default, as the predicted probabilities must be high, resulting in low true and false positive rates. When the threshold is lowered, the number of default predictions increase, and therefore also the rates. If the model holds no predictive power, it is essentially random, and the rates will increase equally along the diagonal line as the threshold decreases. If the model holds good predictive power, it will separate the classes well, and the True Positive rate will increase more than the False Positive rate as the threshold changes. An illustration of the ROC curve is shown in Figure 3.1:

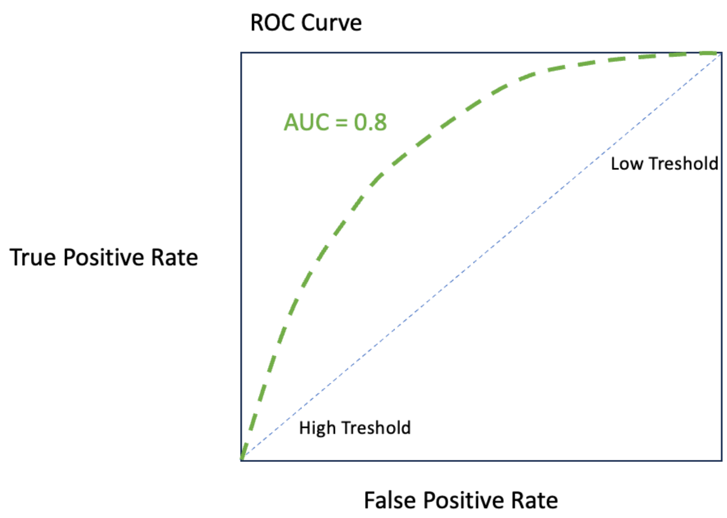

Figure 3.1: ROC Curve, with an Area Under Curve of 0.8 (illustrative). The FPR and TPR are on 0 to 1 interval, as it cannot predict more correct or false observations than there is in the dataset.

3.1.4 Area Under Curve (AUC)

Models with better predictive power will move further away from the diagonal line, as their TPR increases at a rate higher than FPR. The Area Under Curve capture this effect and can therefore be a measurement of a model's performance relative to other models (James et al., 2021). The ROC curves in our thesis are generated by plotting the mean of observed true and false positive rates for 100 thresholds in the range of 0 and 1. Thereby measuring to which degree, the rates increase relative to each other. Due to a low number of observations in total, there will be a low number of observations in the test data and as a result, the ROC curve is likely to be subject to a high degree of variance. To get a more accurate estimate of the true model performances, cross fold validation and bootstrap aggregation is used. By these resampling techniques several estimates of the TPR and FPR will be generated. The AUC and ROC curve will be our main method performance valuation.

3.2 Resampling

In machine learning, repetitive splits of model training and testing data is necessary to evaluate the model's performance accurately (James et al., 2021). Due to highly imbalanced data, with few observations of defaults, we have found it necessary to estimate the model accuracy with the bootstrap method. The Random Forest method is also based on this framework.

3.2.1 Cross Fold Validation

Fitting a model on the entire dataset, and then performing predictions on the same dataset gives an inaccurate estimation of the model's accuracy (James et al., 2021). The reason for this is that fitting the model on the entire dataset might lead the model estimation to include effects that encapsulates noise in the data, and that is not representative for the true distribution of which the sample is drawn (James et al., 2021). When the model is then introduced to new data, it will perform worse as the same noise will not be present, due to the effect of overfitting the model to one data sample. Splitting the data into a training and test set through cross validation, mitigates overfitting of the model (James et al., 2021). Estimating the model on one split of the data, and testing it on the other, will therefore give a more accurate estimation of the true model accuracy (James et al., 2021). In this case however, the model accuracy is dependent on the specific split of observations that is made, and the true accuracy might be different from a single estimation made. Through k-fold cross validation the dataset is randomly split into several different folds, k, each fold consisting of a different train and test split. Increasing the number of different splits, will therefore decrease the variance of the accuracy estimated, giving a more robust estimate of the model's performance (James et al., 2021). The estimated accuracy-rate will then be the average error rate over all splits. The k-fold Cross Validation is illustrated in the following Figure 3.2, for a 4 - fold cross validation:

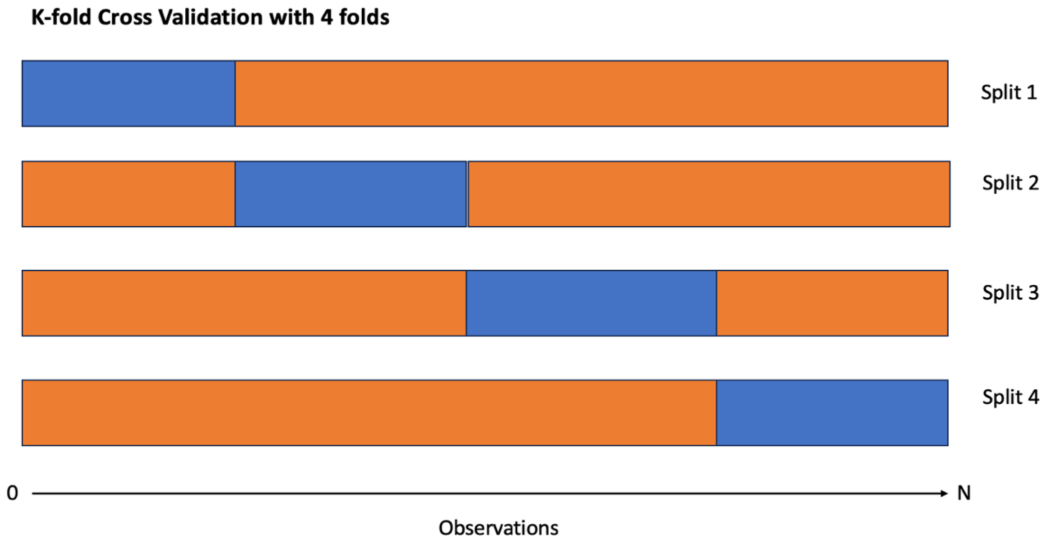

Figure 3.2: Cross Fold Validation split with 4 split.

There is a bias – variance trade-off to be made when selecting the number of folds for cross validation. As mentioned, overfitting the model will lead it to capture noise and have less accuracy on out of sample data. When the number of folds grow large, the test sets become small as the test sets do not overlap, and most of the observations will be a part of the training set. This will then result in an overfitting of the model, leading to a higher variance in the estimates as the model include too much noise. On the other hand, a low number of splits will lead to underfitting. A reduction in the number of training observations might lead the model to be too simple, not capturing important patterns in the data. The model will be biased towards the small number of observations it is fitted to, and as a result it will make systematic errors. According to James et al. (2021), the number of selected folds will involve both bias-variance and computational aspects, but folds between 5 – 10 are usual. In our dataset, we have a highly imbalanced dataset with only 19 observations of one class, which means that a low number of folds might introduce both bias and high variance in our models. Furthermore, a random sample might result in no observations of the default class in the test or training split, to solve this issue stratified k-cold cross validation sets are created (Berrar, 2019). In this way the class balance is kept constant at the ratio of the original dataset, and therefore representative of distribution present in the dataset. A higher number of folds would lead to less bias, but 1 to 2 observations in the test set would result in high variance. To further mitigate problems of a small sample, we will be using the bootstrap method to reduce variance in the accuracy estimations.

3.2.2 The Bootstrap

The bootstrap method introduced by Bradley Efron in 1979 (Singh & Xie, 2010) is a powerful resampling method that can be applied to a range of statistical learning methods (James et al., 2021). The technique is useful in cases of few observations, where several «Bootstrapped» samples are made from the original data by sampling with replacement of observations. The average of a statistic across the samples will provide a better estimate of the statistic than the original sample alone might provide (James et al., 2021).

A common use for the bootstrap method, is to get a better estimate of the mean or variance in a small distribution. Considering a set of *n* observations, $Z_1,\ Z_2, \dots, Z_n$, the variance (σ^2) of each observation is defined by its distance from the mean: $\sigma^2 = E[(Z - u)]^2$. As the sample mean $\bar{Z}$ is equal to the sum of all observations over *n,* the variance of the mean is defined by: $\frac{\sigma^2}{n}$. Averaging over a set of observations therefore reduces the variance (James et al., 2021).

In a classification context, estimating the coefficient of a variable with the bootstrap method, means to estimate models on several different bootstrap samples, to obtain an estimation of the coefficient distribution (James et al., 2021). The method is particularly powerful when dealing with a small number of observations, as variance is inversely dependent on the number of observations in the sample, and a small number of observations will result in higher variance. The increase in sample size through the bootstrap method will in turn lead to a reduction in variance, and the bootstrap estimate will provide a more accurate estimate of the true value (James et al., 2021).

In our analysis we produce true positive and false positive rates for a range of thresholds[6], combining them into an ROC curve to evaluate the model's performance. To get a more

[6] The threshold is defined by a sequence of 0 to 1, by 0.01. Given a probability of 0.2, a threshold of 0.19 will provide a false prediction, and 0.21 a true prediction, thereby changing the TPR and FPR without changing the models, or predictions.

accurate estimation of the model's performance, and removing the effect of overfitting, we perform a 5- fold cross validation split. As mentioned, 5 – 10 folds are normal, and a lower split might introduce bias in the models. Selecting a larger number of splits would however lead to few observations of default in the test set, and increased variance in the true positive rate, as well as the following ROC-curve.

The problem of variance in the ROC–curve still exists, as the number of true positives in a test set remains low. In this instance the bootstrap method can be used to get a better estimate of the statistic. In this case the TPR and FPR, the bootstrap can provide a better estimate of the mean and reduce variance of the rates. We produce more estimates of the FPR and TPR, based on different samples by repeating the cross-fold validation method 10 times with different seeds. In this way we are effectively creating new samples with replacement from the data. Across 10 repetitions of 5 cross-fold splits, 50 observations were observed for TPR and FPR for each threshold. An increase in the repetitions was computationally demanding and therefore disregarded. From the 50 observations, a mean TPR and FPR, as well as a confidence interval given by their standard deviation from the mean was made.

The mean ROC-curve, and its confidence interval is presented in the results[7]. Furthermore, the AUC measure was produced for each fold with the AUC function in R, the AUC included in the results for each model, is the mean AUC observed across each fold and repetition. It is therefore important to note that the AUC in the results, does not correspond directly to the mean ROC-curves, as they are computed based on different measurements. The motivation behind the plot is to give a graphical representation of the differences in observed TPR and FPR for each model, and therefore an indication of their reliability in performance, as well as the possibility of selecting and optimal threshold, based on the ratio of TRP and FPR.

3.2.3 Oversampling

In the data sample, we have 7495 observations, of which 19 observations belong to the default category. If a model predicts all observations to the healthy class, the model accuracy will by

[7] The confidence intervals are limited between 0 and 1, as one cannot predict more observations than there is present in the dataset.

approximately 99.8%. This causes a problem as each of the following methods are estimated based on limiting the share of incorrect predictions, calculated by their respective error measures. The models will naturally predict healthy companies for all companies, achieving a high degree of accuracy, assuming that the default values cannot be separated perfectly through the predictors. In this thesis we are interested in accurately predicting default companies, as the cost of misclassifying a default is much greater than misclassifying a healthy company. To mitigate this effect across all models, we choose to oversample the default companies in the training data. Predicting all observations to the healthy companies is now punishing the model accuracy measurement harder, as a larger portion of the data is misclassified. We have chosen to make 50 samples with replacement from the default observations, matched with 200 healthy samples without replacement from the training data to create a training set on which the model will be estimated. It will now have 20% and 80% share of defaulted and healthy observations respectively and will therefore result in a more accurate prediction of defaults. Berg (2007) argued that manipulated data is not representative for the real distribution, and it would therefore limit the true accuracy of the model and rather the threshold for classification should be lowered. It is therefore important to note that the test set folds will remain non overlapping, and each fold randomly sampled from the data for each repetition. In this way the data that the model is evaluated on is not manipulated, and representative of the real distribution, giving non-manipulated model performance measures.

Observations in the train or test set do not include companies that are in the year of default. A defaulting company in 2015, will have a lead variable of default equal of 1 in 2014, but 0 in the year of 2015 as it does not default again in 2016. Observations where the year is equal to an effective default year is not included in train or test set, to not disturb estimation of model or accuracy.

3.3 Logistic Regression

In our study we use a binary variable for default as our response variable which takes value 1 if the company has defaulted and value 0 if the company has not defaulted. Logistic regression is a model used to estimate the probability that the dependent variable belongs to a specific category, based on the independent variables. Mathematically the probability is expressed as $p(X) = \Pr(Y = 1|X)$, where Y represents the response variable, and X represents the independent variables. Using a linear regression, the outcome could yield predictions outside the range of 0 to 1, logistic regression ensures that the predicted probabilities fall within the valid range. In the logistic regression the logistic function is used:

$$p(X) = \frac{e^{\beta_0+\beta_1 X+\cdots+\beta_p X_p}}{1 + e^{\beta_0+\beta_1 X+\cdots+\beta_p X_p}}$$

For any given values of X and β the function will always give a probability between 0 and 1 and this is visible through the S shaped curve the function produce, as shown in Figure 3.3.

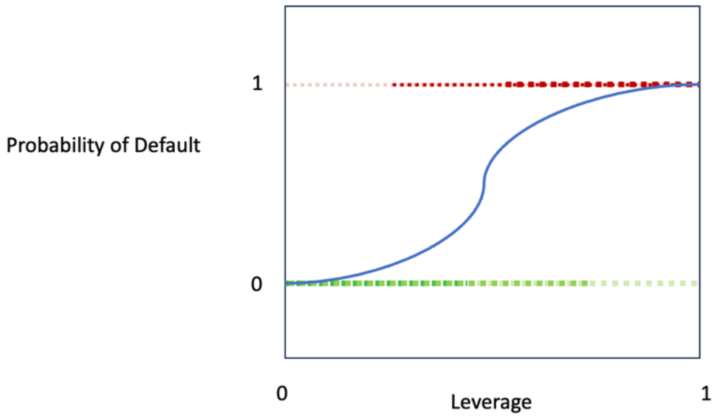

Figure 3.3: Logistic regression illustration for increasing probability of default with increased leverage.

Given the logistic function we need to estimate the unknown β coefficients. We achieve this by using the maximum likelihood method. The intuition behind this approach is that we seek estimates for the coefficients, such that when plugging the β estimates into the equation over, the probability of the observed sample is maximized. In the context of our study, this means getting values close to 1 for defaulted companies, and values close to 0 for healthy observations. One way to formalize this intuition is by using a mathematically equation called the likelihood function (James et al., 2021), as shown in equation 3.1:

$$\ell(\beta_0, \beta_1) = \prod_{i:y_i=1} p(x_i) \prod_{i':y_{i'}=0} \left(1 - p(x_{i'})\right)$$

β_1 and β_2 are unobservable and are chosen such that it maximizes the likelihood function. (James et al., 2021)

3.4 Random Forest

Random Forest is one of the most-used algorithms due to its flexibility, simplicity, and diversity (Donges. N, 2019). The algorithm was developed by Leo Breiman (Breiman, 2001), and further presented in *"An Introduction to Statistical Learning: with Applications in R"* by (James et al., 2021). The method is suitable for addressing both regression and classification tasks. In this study we will focus on Random Forest application in classification problems. To gain a better understanding of Random Forest, it's essential to explore its foundational component, decision trees, which we will discuss before describing the Random Forest method.

3.4.1 Decision Trees

Decisions trees are based on making predictions by segmenting the predictor space into a number of simple regions. A tree-based model consists of two steps, first dividing the possible values of the independent variables into distinct and separate regions, then making the same prediction for all observations in the region (James et al., 2021). In order to make a prediction for a given observation in a classification model, it is assigned the most commonly occurring class value for the training observations in the region it belongs. The splitting rules used to segment the predictor space are sequential, in a top – down approach known as recursive binary splitting (James et al., 2021). All observations belong to one region and are subsequently split into two new branches by the predictor (node) that creates the best split, creating a tree like shape as the process starts at the top of the tree. The approach is illustrated in Figure 3.4, where the segmenting of predictor space is forming the basis of decision tree models:

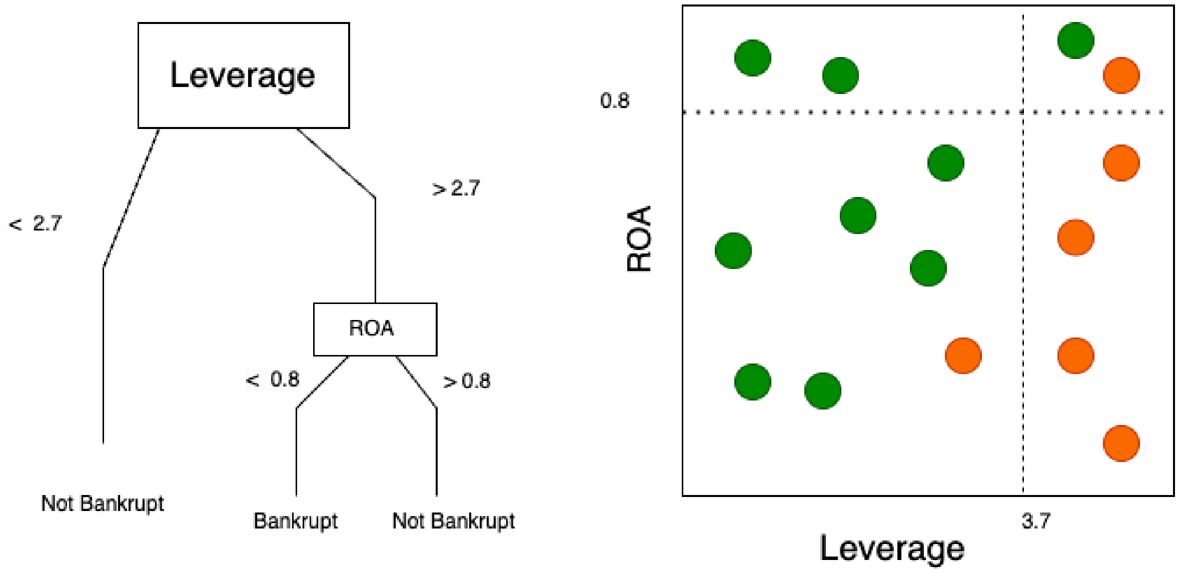

Figure 3.4: The left plot illustrates the tree like sequence, the thresholds separate the predictor space into several spaces in the right plot. Orange = default.

The main advantages of decision trees are that they are easily interpretable, as they mirror human like decision making and as a result can be better at handling qualitative predictors than the other regression models (James et al., 2021). Decision trees are however susceptible to a high degree of variance, where a small change in data can result in a large change in the estimated tree (James et al., 2021). Through the Random Forest method, the variance of the estimated tree is lowered by combining many estimations into a single tree. The objective of the decision tree is minimizing the classification error rate between estimated and actual classifications. The classification error rate is measured by the number of observations in the region that do not belong to the predicted class, through the Gini index (James et al., 2021). When building a classification tree in through Random Forest, the Gini index is used to assess the quality of the split, choosing the predictor resulting in the lowest Gini index. The Gini index is defined by:

$$G = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk})$$

Where K is the total classes, and m the regions which are produced from the predictor thresholds, $\hat{p}_{mk}$ therefore represents the share of observations of k class in the region m. A model separating perfectly will assign only values of 0 and 1 for $\hat{p}_{mk}$, as the share of classes in the space is either 100% or 0%. It is obvious to see that for this instance, the Gini index becomes 0, and have higher values when the classes cannot be separated. The Gini index is a measure of variance over all the classes and measure the node purity of the model, meaning to which degree the classes are correctly separated (James et al., 2021). The lowest Gini index is used to estimate the most accurate model. Another interesting feature of the Gini index is its ability to evaluate the importance of the predictors by measuring the mean decrease in Gini coefficient when each predictor is removed in a Random Forest estimation.

3.4.2 Random Forest

The Random Forest method improves the accuracy of decision trees, by combining estimations of many decision trees. The Random Forest method improves the accuracy by the use of the bagging method and assessing multiple independent decision tree estimates on out of bag observations (James et al., 2021).

Bootstrap Aggregation (Bagging)

Combining the two sampling techniques Cross Fold Validation and bootstrapping, can mitigate problems related to both variance and bias, through a method known as bootstrap aggregation or: Bagging (Breimann, 1966). To reduce variance and increase the accuracy of the test set errors, several models are estimated on multiple samples with replacement of the same training set, and an average model obtained. From the original training set, more samples are generated with the bootstrap technique, resulting in bootstrapped models. From this the average bootstrapped model is estimated through the following formula (James et al., 2021):

$$\hat{f}_{bag}(x) = \frac{1}{B}\sum_{b=1}^{B}\hat{f}^{*b}(x)$$

This method is particularly useful for decision trees, as they suffer from a high degree of variance but are also useful for other classification models (James et al., 2021). Combining bagging with decision trees forms the basis of Random Forest models.

Out of Bag Observations

It can be shown that for Random Forest estimation, only around two-thirds of observations are used in the bootstrapped sample, and the model test error can be estimated on the remaining "out of bag" observations (James et al., 2021). As the method is repeated many times, a class prediction is assigned to each observation based on their most commonly occurring class when included in out of bag predictions. The resulting test error for all observations is a valid test error estimation, as it is based on observations that were not included in the training of the model, similar to Cross Fold Validation (James et al., 2021). The use of the Bagging method has been shown to increase the accuracy of tree-based models (James et al., 2021). Random Forest, however, provides an improvement over bagged trees, as the bagged method is susceptible to correlated estimates. The presence of powerful predictors will naturally position them at the top of the tree, as they offer the model the highest decrease in Gini Index, and consequentially all model estimates will become similar (James et al., 2021). As a result of the similarity, the bagged models will not reduce variance as it averages over many similar observations. By choosing a random subset of predicators to be used for estimation of the model, the Random Forest method will also remove powerful predictors from the estimation. When the number of predictors included in the tree estimation is reduced, strong predictors will no longer always be present at the top of the tree, and the trees will be less correlated. By reducing predictor selection, Random Forest isolates effects of less powerful predictors, and is essentially decorrelating the trees (James et al., 2021). A standard approach for choosing a predictor subset is to estimate trees with the number of predictors, equal to the square root of total predictors (James et al., 2021). A random forest model with $m = (p)^{\frac{1}{2}}$ predictors typically lead to a reduction in both test and out of bag error. As we have a total of 8 predictors, we choose 3 as the initial size of the predictor subset.

3.4.3 Variable Importance Measure

Decision trees are straightforward to interpret when dealing with a single tree. However, when using the bagging method, which involves bagging numerous trees, the outcome becomes complex and cannot be represented in a single tree. It is also no longer clear which variables are most important to the procedure. Thus, the bagging method improves the accuracy of the prediction, at the expense of interpretability. Although the bagging method leads to an outcome which is hard to interpret it is possible to obtain an overall summary of the importance of each predictor using the decrease in Gini index (James et al., 2021). In the case of bagging classification trees, we can add total amount that Gini index is decreased by splits over a given predictor, averaged over all trees. For a high value, it indicates an important predictor, as the node purity is lowered considerably.

3.5 Support Vector Machine

The Support Vector Machine (SVM) is often considered as one of the best "out of the box" classifiers, which means you use it as a pre-configured classifier. It was developed in 1990 and has grown in popularity since then, due to its performance in various settings (James et al., 2021). In this section, we will begin by explaining the concept of a hyperplane, followed by an introduction to the Support Vector Classifier (SVC) and the SVM.

A hyperplane in machine learning is often used as the decision boundary which separates data into classes or groups (James et al., 2021). In two dimensions a hyperplane is simply a straight line, while in higher-dimensional spaces, it becomes a flat linear subspace. To show how the hyperplane can be used as a separator we can write the mathematically equation for a p-dimensional setting:

$$\beta 0 + \beta 1 X1 + \beta 2 X2 + \cdots + \beta p Xp = 0$$

If the left side of the equation equals zero, the point X lays exactly on the hyperplane. However, if the equation is not met, X is positioned on either side of the hyperplane based on whether the equation is greater than or less than zero (James et al., 2021).

A dataset that can be perfectly separated by a hyperplane will have an infinite number of potential hyperplanes (James et al., 2021). The optimal solution is the maximal margin hyperplane which is the one with the farthest distance to the training observations and therefore the widest margin. It relies solely on a subset of observations known as the support vectors, which lies on the margin. If the support vectors are moved slightly, the hyperplane will move as well. A maximal margin hyperplane is effective when a perfect separating hyperplane exists, but this is often not the case (James et al., 2021).

In most cases, data is not perfectly separable. The SVC addresses this challenge by introducing a soft margin that permits violation of the margin, allowing for some misclassifications (James et al., 2021). This misclassification may lead to more robustness to individual observations and better classification of the remaining observations. The degree of tolerance to misclassifications can be fine-tuned using a parameter C which is crucial ensuring optimal bias-variance trade-off (James et al., 2021). Parameter C can be seen as a budget for the number of observations n which can violate the margin. If C is 0 there is no observations violating the margin which is only possible if the two classes are separable. In practice the C is chosen by cross-validation and when C is small, we seek to obtain margins that are rarely violated. This gives low bias, but high variance. On the other hand, large C allows more violation of the margin and gives a classifier which fit the data less. A larger margin obtains a classifier that is potentially more biased but may have lower variance. (James et al., 2021)

SVM is an extension of the SVC that introduces the concept of kernels (James et al., 2021). Kernels allow SVM to implicitly map the input features into a higher-dimensional space, making it possible to find a hyperplane that can separate classes more effectively. By enlarging the feature space, we can accommodate a non-linear boundary between the classes. Kernels is just an efficient computational approach, instead of enlarging the feature space in the SVC using polynomial functions of the independent variables. There exist numerous different varieties of the kernel, all who separate the predictor space by different method (James et al., 2021). For this thesis we have selected the best performing model from the three kernels: Linear, Polynomial and Radial, who all separate the predictor space in the manner which they are named.

3.6 Hyperparameter Tuning

The machine learning methods Random Forest and Support Vector Machines can both separate their predictor space and estimate models in several different ways. Since there are many parameters to tune for each problem, manual estimation would be a time-consuming effort. We have utilized a grid search estimator for this problem, by the use of the e1017 package in R (Meyer, 2023). The model takes a range for each parameter and estimates the best performing model on the basis of a 10-fold cross fold validation. Each model was tested on the same ratio split as the in the method, with all defaults included and an oversampling of the bankruptcies at a 20% ratio. For the random forest model, the number of variables selected for each tree estimation in the sequence: *m,* and the number of trees: *ntree*, must be determined for each model estimation. From (James et al., 2021) the square root of number of predictors is usually the optimal solution, and the starting point of our search. For Support Vector Machines there are a number of different kernels and parameters to tune by, and tuning parameters searched is included in appendix (A1).

4. Data

Before introducing the rest of the model, a description of the data selection and filtering is necessary. In this section we will introduce the data set, its sources, and important features. Furthermore, we will explain our approach for selecting the response and independent variables. Most importantly we will justify the filtering of observations for the analysis and the transformation of financial ratios and other predictors.

4.1 Data Introduction

For this thesis we have gathered financial statements for all Norwegian companies from the accounting database created by Centre of Applied Research at the Norwegian School of Economics. The accounting database contains financial and other firm specific information for all Norwegian companies from 1992 to 2020 and is a collection of financial information provided by Dun & Bradstreet, as well as governmental financial institutions: the Brønnøysund Register, Bank of Norway, and Statistics Norway (Mjøs & Mjelde, 2019). The database provides company specific information relating to sector, bankruptcy year, mother organization and more. Additionally, the database provides complete accounting numbers for all companies, as well as mother companies. For this thesis we have limited ourselves to the years after 2005, as certain macroeconomic variables are only available after this year (Clarksons, 2023). We consider this limitation unproblematic, as the number of observations in recent years far outweigh that of the years before the millennium change. In addition, accounting regulations are changed and implemented continuously, and we assume that new accounting standards provide a more accurate representation of the financials of the companies. Secondly, we assume an analysis based on the recent years will more accurately represent the present economic situation and therefore offer better predictive power in the coming years.

4.1.1 Data Processing and Relevant Observations

The purpose of this thesis is to investigate the accuracy in predicting defaults in the shipping sector. To define the sector practically, and select the relevant companies, the variable "*sector*" from the SNF dataset is used a basis of filtering shipping companies. The variable is generated by SNF themselves and represents a common industry group based on industry codes provided from the 2002 and 2007 standards by Statistics Norway, with sector: 2 indicating shipping and offshore. This provides a large range of companies, from domestic ferry traffic to companies engaged in large scale offshore operations. These companies are outside the scope of this thesis and are not all exposed to the same factors or economic environment. To remove irrelevant companies, we select only companies with the main industry description of “International freight” or “shipping business” by the 2002 industry code standard. In cases where a company is further defined by a 2007 industry codes, which is most of the observations, we include companies under the following codes: “International freight with goods”, “Tax oriented investments company. »

Further investigation reveals some problematic characteristics of some companies. Firstly, we have several “dormant” companies, which have a low degree of assets and debt, as well as income. We regard these companies as irrelevant, as the low degree of assets are not compatible with the capital intense nature of international shipping. As a result, only companies with a minimum of 50 million NOK in total assets, and minimum income of 1 million NOK at one point in the dataset are considered for the analysis. Secondly, we observe that some companies experience a high level of debt, and low income in their first years of operation. Assuming that the first years are not representative for the operational nature of a company, only observations after the first two years are considered for further analysis.

When inspecting all bankrupt companies, there is a case of 6 similar companies all going bankrupt in the year of 2016. All these companies have the same mother organization and go by the name of Septem D1 to Septem D6. Several of the companies have almost identical accounting numbers, and they are all created by the same company to purchase and then lease ships (Vanvik, 2019). This might create a bias in the dataset, as will effectively be an oversampling of a single observation, and potentially disturb the effect of time dependent observations in macroeconomic variables. If there is a negative value of GDP growth in this year, it will observe more defaults than what is true, overestimating its effect, and thereby introducing a selection bias towards negative values of GDP growth. If all companies are similar, it also breaks the assumption of random sampling, as one observation is now drawn six times, while other observations can be assumed to be randomly sampled. The assumption of random sampling is important, so the data is representative of the true distribution (Berrar, 2019). If this assumption does not hold, the model accuracy estimates are based on biased data and therefore invalid. This could also lead to overestimation of the accuracy if identical observations are included in both the train and test data. The companies are regarded as one single entity, and accounting numbers are averaged over all 6 companies. After these conditions and transformation of ratios, we are left with 7495 yearly observations from 2005 to 2020, of which there are 889 companies and 19 bankruptcies.

4.1.2 Sectors

The observations deemed relevant still include companies across several subsectors, and companies cannot be separated into further subsector based on company codes. The shipping sector defined by our assumptions, includes companies involved in shipping of several commodities, and it is possible that they are influenced to a different degree by macroeconomic variables introduced in section 5. These effects are considered outside the scope of our thesis. The following chart in figure 4.1 illustrates the distribution of different subsectors, where the sector is identified by its inclusion in the company name. Naturally the most frequent observation is of an unspecified sector, but most of the companies are within general shipping, Offshore, Tankers, Bulk. From Figure 4.1, we can't see a distinct relationship between defaults and sector.

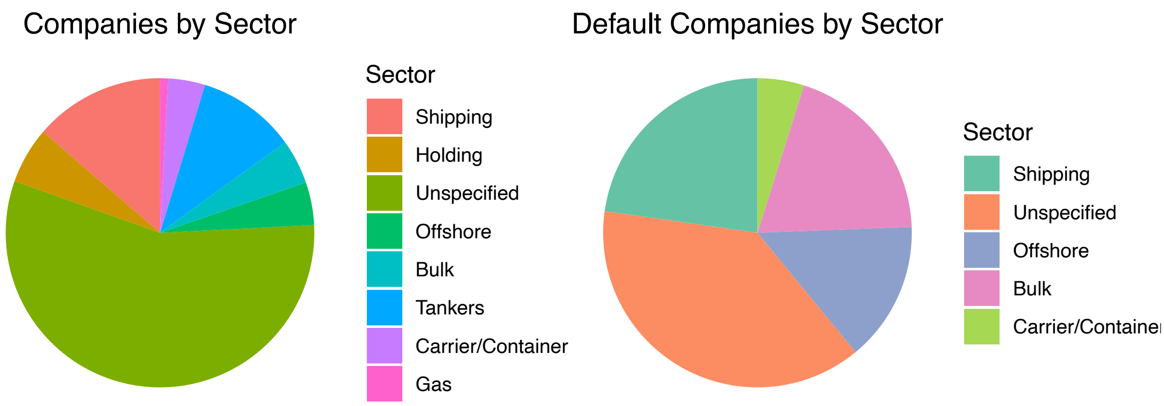

Figure 4.1: Sectors of all Companies, and of default companies.

4.1.3 Financial Data.

The following financial posts were gathered from the dataset to produce the financial ratios: ***Total Income, EBITDA, Yearly Result, Total Debt, Equity, Total Assets, Interest Expense, Cash, Current Assets, Ships and Rigs***. In appendix A4, summary tables have been produced to further investigate differences between healthy observations, and observations where companies default within the next 2 years. Companies which defaults appear to have lower mean Income, negative EBITDA, and results in the years before default. Interestingly they have less average debt and assets than healthy companies, as well as negative equity. This could indicate that default companies are smaller companies that experience heavy losses in years before default, with negative results leading to diminishing equity. Additionally, there are some values of 0 present in financial posts, which could be problematic when calculating ratios.[8]

4.2 Categorization

When computing our dataset, we have a small number of default companies, which are exposed to large variation in the predictors. Some predictor variables may take very large values that have a large effect on the model estimation and as a result become high leverage points (James et al., 2021). High Leverage Points could potentially disturb the estimated effect of other observations of the same variable, and even though they are traditionally more problematic for linear regression models, high leverage points might also have a negative effect on the logistic model estimation. (Nurunnabi, Nasser, & Imon, 2016). It has also been shown that normalizing the dataset is an effective technique for improving the performance of machine learning techniques, as “dominant predictor values introduce significant bias in learning” (Singh & Singh, 2020). Furthermore, “features with greater numeric values dominate those with smaller numeric features values while discriminating patterns from the data.” Data normalization tackle these problems by bringing the predictors values to equal distributions, limiting the dominant effect of the extreme values. It was found that a simple quantile normalization (Peterson & Cavanaugh, 2019) by categorization of the variables improved the prediction accuracy the most. Other methods of removing extreme values were considered, as for example the IQR method, but was found to remove a considerable portion of the data. With a high-class imbalance, and a small number of observations in the class of interest, categorizing variable also retains the most observation out of the methods.

We assume the quantile normalization of the data holds intuitively as well. A shipping company that deviates from the distribution by a high or extreme degree are both at a higher

[8] A value of 0 would result in an infinite value if it is the denominator in a ratio. Infinite values are problematic for regression and machine learning techniques.

risk of bankruptcy. Considering a company in the sample with a leverage ratio in the 10% highest quantile, this will indicate that the firms is experiencing financial distress. The fact that leverage ratio is high compared to its sector competitors might be more informative than its numerical value. Effectively we assume that there is a nonlinear relationship between ratios and the probability of default, with a demeaning effect. A difference between 1 to 3 might increase risk substantially, but a an equally large difference in leverage of 8-11 represents a very little change in risk of default, as it will already be high. Our selected models do not impose a linear relationship between predictor and probability, however normalizing data may still improve predictions (Singh & Singh, 2022). We have therefore chosen to categorize the ratio variables, so we are not at risk of losing important information of the outliers and can stabilize the dataset that will be used to predict bankruptcies.

Predictor variables are in this thesis categorized as: Very Low, Low, Medium, High, Very High, by the quantiles: 10%, 30%, 70%, 90%, 100%, respectively. For a few variables that cannot be separated by these quartiles, they are separate into “High”, “Medium”, “Low”, by other quantiles (see appendix A2). Lastly, certain variables were also shifted by 1 in their denominator and log estimation, to avoid infinite values. The ratios, their categorization[9] and definition[10] can be seen in appendix.

[9] For ratios see appendix A2

[10] For definition see appendix A3

5. Model and Descriptive Statistics

Given that the extended SEBRA-model performs well in predicting defaults on a company level in Norway, we have chosen to use this as our starting point and as a comparison to our own models, which contains shipping specific and macroeconomic variables. In this section we first discuss our response variable. Further, we introduce the ratios in the extended SEBRA-model, followed by the shipping specific variables in our model including descriptive plots. At last, we present the macroeconomic variables used in our prediction model.

5.1 Response Variable

Bankruptcy or default can be qualified as a rare event among the population of active companies (Calarese et al., 2016). In our dataset, only 0,2% of all observations are categorized as defaulted, indicating a very low proportion. This is an important factor that we must consider in our analysis and modeling as this results in an imbalanced dataset. There are several ways to define when a company is considered default. In our study, we use the variable *konkaar* from the original dataset. This item represents the year that liquidation proceedings started in the company for the first time (Mjøs & Selle, 2022). Companies often stop submitting annual accounts before going into liquidation, and thus it is not uncommon for *konkaar* to be recorded after the final statement. In such cases, we consider the year of default as the same year as submitting their final statement. The logic behind this is that companies in the liquidation process have stopped operating normally and their last normal year is the year the company filed their final statement.

Companies classified as default receive a value of 1, while non-defaulted companies are assigned a value of 0. Most of the companies get assigned the correct default status, but we have identified potential noise in our response variable. One factor contributing to noise is the inclusion of observations where companies fail to submit their accounts, or not having a board or an auditor, leading to a classification of default. In these instance, false data may occur. To deal with this we have looked further into the companies in the Brønnøysund Register to verify their default status.

5.2 Variables Included in the Extended SEBRA-Model

In this section we will describe the variables as given in Bernardsen and Larsen (2007) which is used to create our replicated SEBRA-model. For predictions one year prior to default, we use the variables from the preceding year. Predictions two years before default are calculated based on the variables observed two years before the default. Considering a default observed in 2013, the model is trained and predictions made with 2012 and 2011 financials, in two separate predictions.

Profitability

Profitability is a measure of a company's income relatively to its expenses (Gartner, n.d.). For sustainable financial health in the long term, firms must generate positive profits from its operations to service their debt. In the short run, firms with negative profits will quickly drain the liquidity of the firm. Moreover, a firm's profitability influences its capacity to secure external financing (Bernhardsen, 2001). To quantify a firm's profitability, the variable used is "earnings over total debt".

Liquidity

In numerous cases, the primary cause of bankruptcy is deficiency in liquidity (Bernhardsen, 2001). Liquidity measures a company's ability to meet their financial obligations, with the liquid assets available to them (Hayes, 2023). Failure to meet financial obligations will impact the trust from lenders and customers (Torgersen , 2018). The SEBRA-model uses the variables "cash minus short term debt to revenue", "unpaid indirect tax as a percentage of total assets" and "trade accounts payable as a percentage of total assets" to measure the firm's liquidity (Bernhardsen, 2001).

Solidity

The greater the share of shareholders equity relatively to debt, the lower are the financial risk and the more likely is the firm to obtain external finance. A higher equity ratio allows a company to survive longer periods of weak earnings. The variables used to measure the equity ratio is simply "equity in percentage of total assets" and a dummy variable for weakened equity, "invested equity less than book value of equity" (Bernhardsen, 2001)

Age:

The SEBRA-model includes an indicator variable for the company's age from 1 to 8. The result by Eklund, Larsen and Bernhardsen (2001) and other international studies shows that bankruptcy is more frequent in newly established firms than wall established firms. It takes time to develop relevant expertise, gain access to the capital markets and establish business ties to suppliers and customers. From Figure 5.1 it is obvious that newly established companies default at a much higher rate than older companies. We also see that both new and old companies are in risk of defaulting.

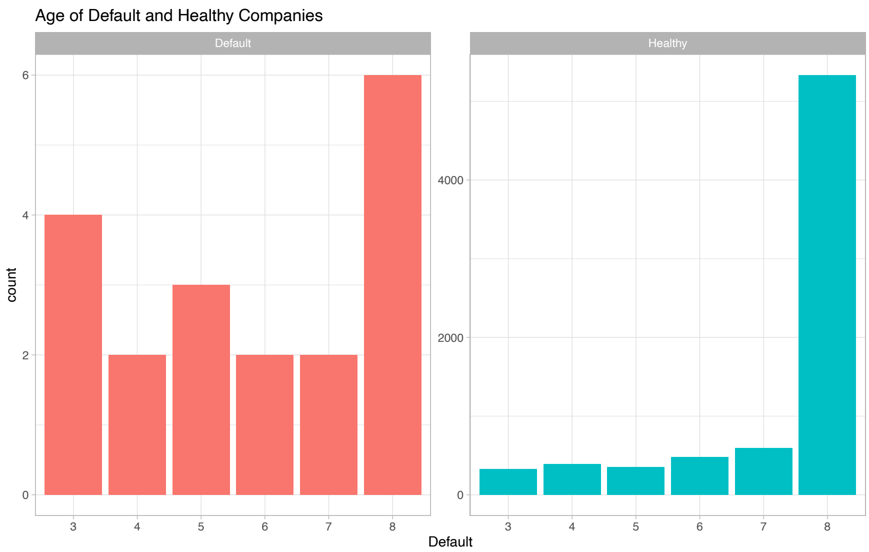

Figure 5.1: Distribution of age for both default and healthy companies

Size

Bankruptcy frequencies tend to be higher for small firms than for bigger firms (Eklund et al., 2001). Small firms are more vulnerable to individual events, due to a limited product range and operate in a limited geographical area. The small firms are also often new established and face many of the same problems as mentioned for the variable age. The firm size is represented as the logarithm of total assets in our analysis. Figure 5.2 demonstrates that smaller companies have a higher default rate compared to larger companies.

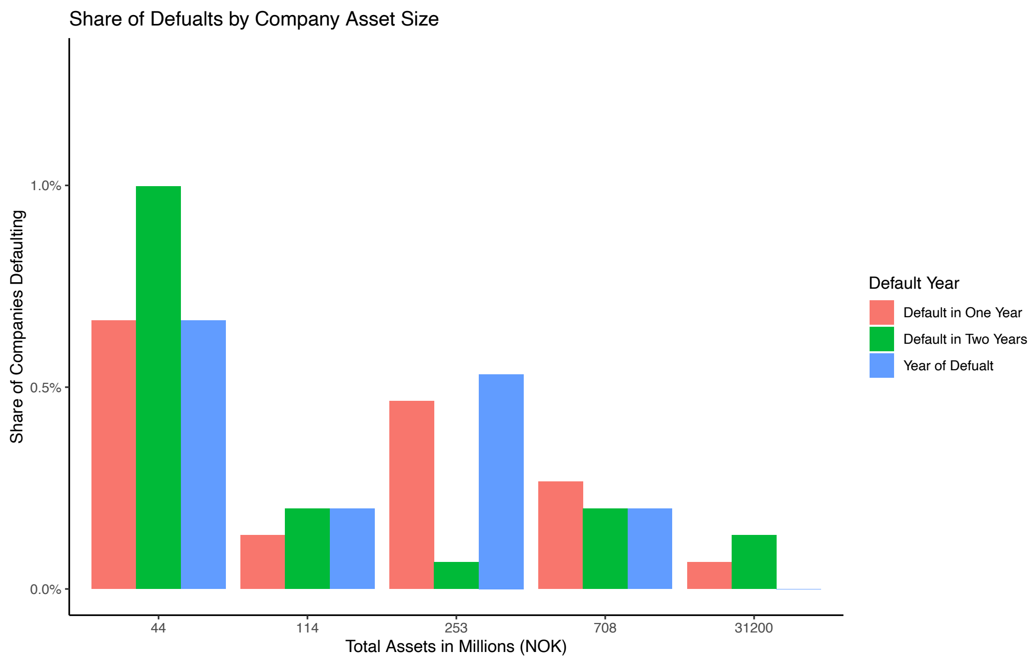

Figure 5.2: Share of defaulting companies by size.

5.3 Financial and Non-financial Variables Included In Our Model

Instead of relying on the variables included in the extended SEBRA-model, we have chosen to include our own set of financial variables, which we consider as better predictors of defaults in the shipping industry. These variables are also found in earlier studies on shipping defaults such as those by Kauvossanos & Tsouknidis (2016) and Lozinskaia et al. (2017). Our model introduces three new financial ratios. Additionally, we incorporate leverage and the non-financial variables age and size, which are part of the SEBRA-model. For our model, we predict defaults using observations from one and two years prior to the actual default event, same approach as for the SEBRA variables.

Current Ratio:

Instead of the liquidity measures included in the extended SEBRA-model we have chosen to use the current ratio. The current ratio measures how many times the current assets of the company cover its current liabilities and are calculated as "current assets over current liabilities" (Kavussanos & Tsouknidis, 2016). For shipping companies, the current ratio is relevant due to significant upfront costs and variable revenue streams and give a good picture of the company's ability to meet their short-term obligations. A high current ratio is suggesting a strong liquidity and is often seen as positive[11]. Figure 5.3 illustrates the difference in current ratio for companies about to default, with a lower current ratio being more common for defaulting companies.

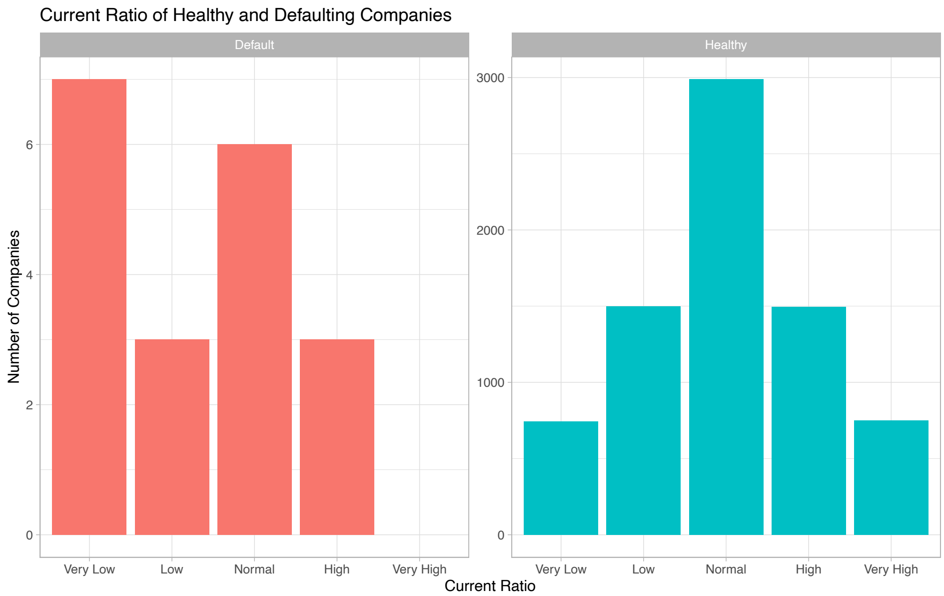

Figure 5.3: Difference in observed current between companies that are default or healthy the following year.

Return on Assets

Return on debt (ROD) is an uncommon metric in financial analysis (Kenton, 2021). Instead of the ROD ratio used in the extended SEBRA-model we use return on assets (ROA) as a profitability ratio. ROA is not a perfect measure, but it is the most effective, broadly available financial measure to assess company performance (Hagel & Brown, 2013). In the shipping industry companies' primary assets is their ships and ROA evaluates the effectiveness of managing and operating the fleet to generate returns. ROA is also included in the study by Lozinskaia et al. (2017) and is calculated as "annual return over total assets". Due to a highly volatile market, shipping companies may experience a great deal of uncertainty in income. In certain cases, they have a significant negative income a few years before default. To account for this volatility, we use the average return of past three years when calculating Return on

[11] A current ratio which is too high may not be positive as it may indicate that the company leave workable assets on the sideline, which could had been used to expand operations or improve equipment. (Ross, 2021)

Assets. Figure 5.4 illustrate the difference in ROA between companies that default the following year in contrast to companies that are healthy the following year.

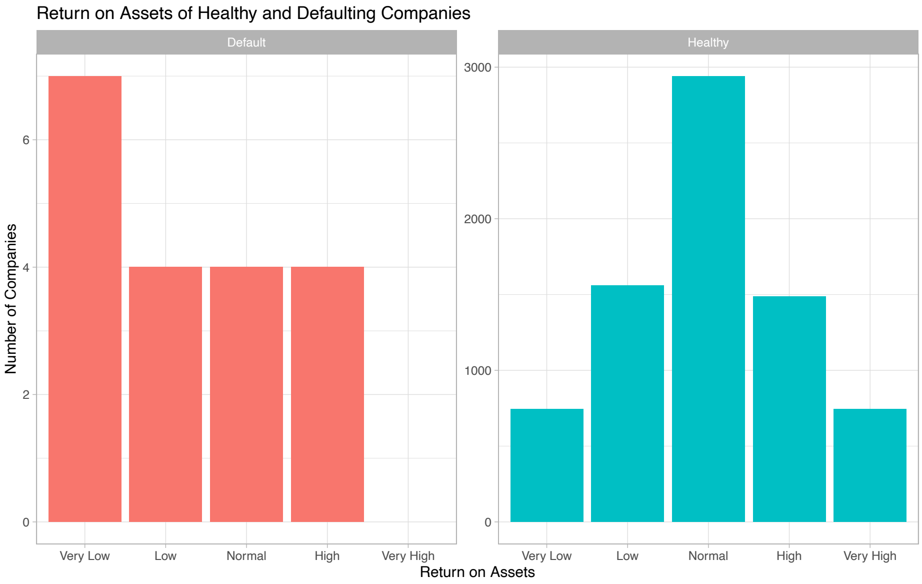

Figure 5.4: Difference in observed ROA between companies that are default or healthy the following year.

Interest Expense Ratio:

The interest expense ratio measures the interest expenses of the company in percentage of its EBITDA and is included in the study by Kauvossanos & Tsouknidis (2016). In the shipping industry, where there is high capital intensity and a reliance on debt financing, the interest expense ratio holds importance due to substantial capital expenditures. A lower ratio is generally favourable, indicating stronger interest coverage through operational earnings. The ratio is calculated as “interest expenses over EBITDA”. In figure 5.5 we find no apparent relationship between interest expense in percentage of EBITDA.

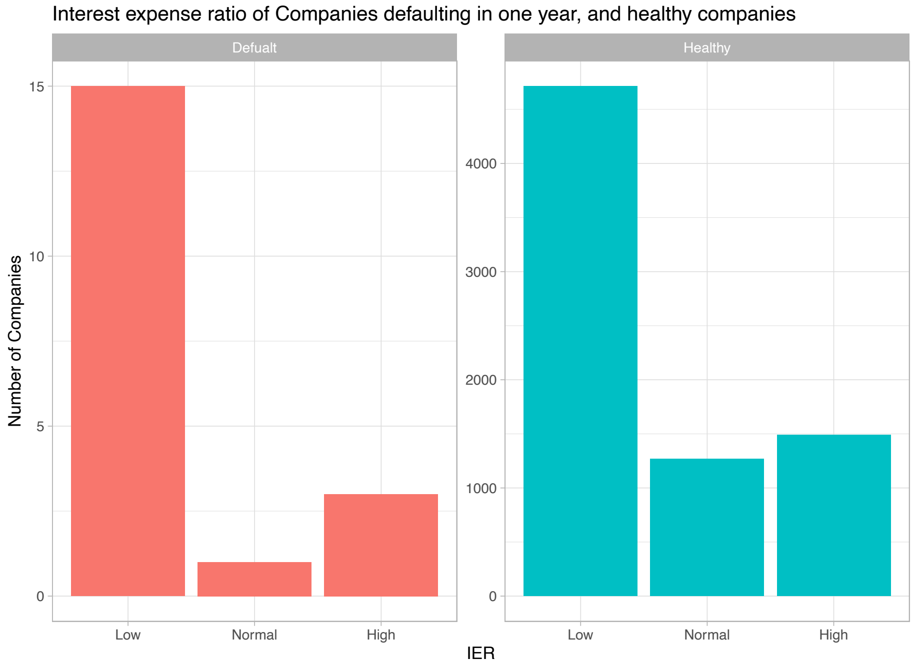

Figure 5.5: Difference in observed Interest Expense Ratios between companies that are default or healthy the following year.

Leverage

We have chosen to also include the financial variable leverage in our prediction model for the same reason as in the SEBRA-model. Shipping companies often have high amounts of debt and the leverage ratio capture the financial risk associated with this. In the study by Kavussanos and Tsouknidis (2016) a positive relationship between financial leverage and the probability of default in the shipping industry was found. In this study we have used "equity over total assets" as a measure of leverage[12], based on the ratio from the SEBRA-model. Figure 5.6 show that companies defaulting in one year have a lower degree of equity ratio, and therefore higher degree of leverage.

[12] Leverage = $\frac{Debt}{Total\ Asset}$ and Total Asset = Debt + Equity,

Therefore: Leverage = $\frac{Total\ Assets - ek}{Total\ Assets} = 1 - Equity\ Ratio$

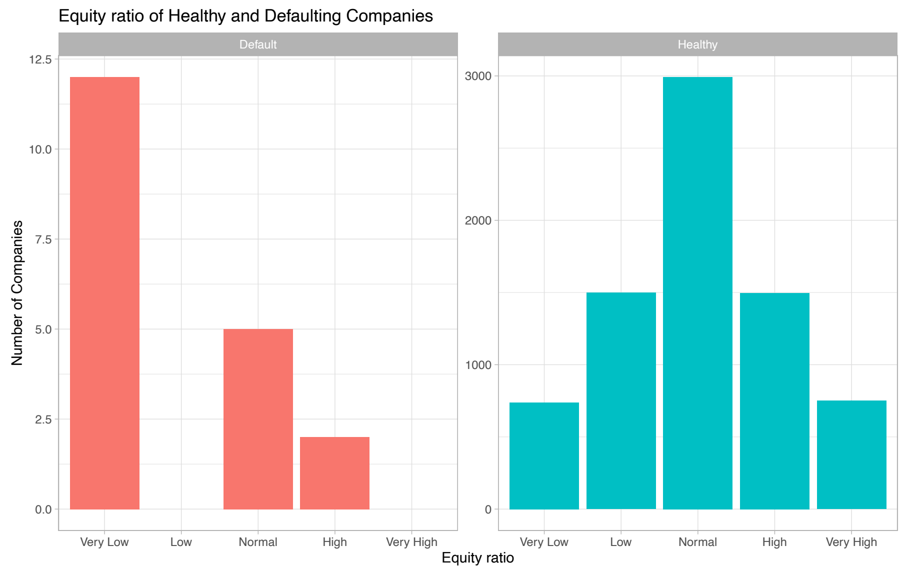

Figure 5.6: Difference in observed Equity Ratio between companies that are default or healthy the following year.

Size and Age

The non-financial variables size and age is also included in our model. We expect that younger shipping firms has less experience in the shipping industry and therefore higher probability of default. For companies with more total assets, we expect a lower probability of defaults and vice versa. These non-financial variables are also found in the studies by Kauvossanos & Tsouknidis (2016) and Lozinskaia et al. (2017), as well as in the SEBRA-model.

5.4 Macroeconomic Variables Included In Our Model

In addition to the financial and non-financial variables we have also included macroeconomic variables as an attempt to capture the dynamics of the shipping industry as a global industry and to better predict defaults in the Norwegian shipping industry.

ClarkSea Index:

The ClarkSea-index approximates the profitability of the shipping sector, by computing the daily earnings for the primary ship types and weighting them based on their share in the total fleet (OECD, 2018). This index is a valuable tool for investors, analysts, and industry experts, offering insights into the present condition and developments within the shipping market. A decline in the index signal reduced freight rates, resulting in diminished cash flow for companies and a potentially weakened debt capacity (CFI, n.d.). We source the ClarkSea index from Clarksons (2023) and use the annual observations. When predicting defaults one year prior, we use the ClarkSea index for the current year. For instance, if predicting default for a company that defaulted in 2015, we use the observed value for the ClarkSea index in 2014.

Orderbook:

Shipbuilding is a long-cycle industry, with the process of ordering a ship until its actual delivery typically spanning between 1 to 4 years. During favorable market conditions, companies want to increase their earnings by expanding their fleet. These decisions to invest are only based on expected future demand. When the orderbook is then delivered the earnings may drop due to higher supply, and the companies who bought ships on top of the cycle may end up losing money on their investment (Stopford, 2009). We include a variable of orderbook over total fleet captured annually, obtained from Clarksons (2023), which serves as forward-looking indicator. In our prediction we use the observation one and two year prior to actual default. A high orderbook may indicate lowering freight rates and income in the years to come and higher probability of defaults.

Oil Price:

Oil is used as fuel for most of the ships. Out of total voyage costs, fuel costs can constitute to as much as 47%, and a change in the oil price will have a large impact on the shipping costs (Stopford, 2009). Crude oil is also the largest individual commodity traded by sea, and a price change which affect the worldwide demand for oil, will impact the shipping companies (Stopford, 2009). The oil price is also obtained from Clarksons (2023) and contain yearly oil prices. Such as for the other variables we use the observed oil price for one and two years prior to predict default.

GDP:

The most important single influence on shipping demand is the world economy. During periods of economic prosperity, there is an increase in the import of raw materials for manufacturing and the trade in manufactures products, leading to increased demand for shipping services (Stopford, 2009). To assess the world economy, we have used the world GDP from the World Bank (2023). In our prediction we use the GDP observed one and two years prior to default. Lower GDP may indicate less demand and higher probability of default for shipping companies.

From Figure 5.7 we see that defaults are concentrated within the years 2009 to 2017, with a low degree of bankruptcies in the recent years. From our graph in Figure 1.1, the largest shipping companies went bankrupt around 2016, as we saw a great deal of value lost to default in these years. In figure 5.7, we see that the orderbook declines gradually after a sharp decline in the ClarkSea-index, representing the saturation of supply in the market, and the following decline in shipping rates. As the orderbook has stayed low in several years, with positive GDP growth, the ClarkSea-index is rising again, as the degree of supply is following relative to the market demand.

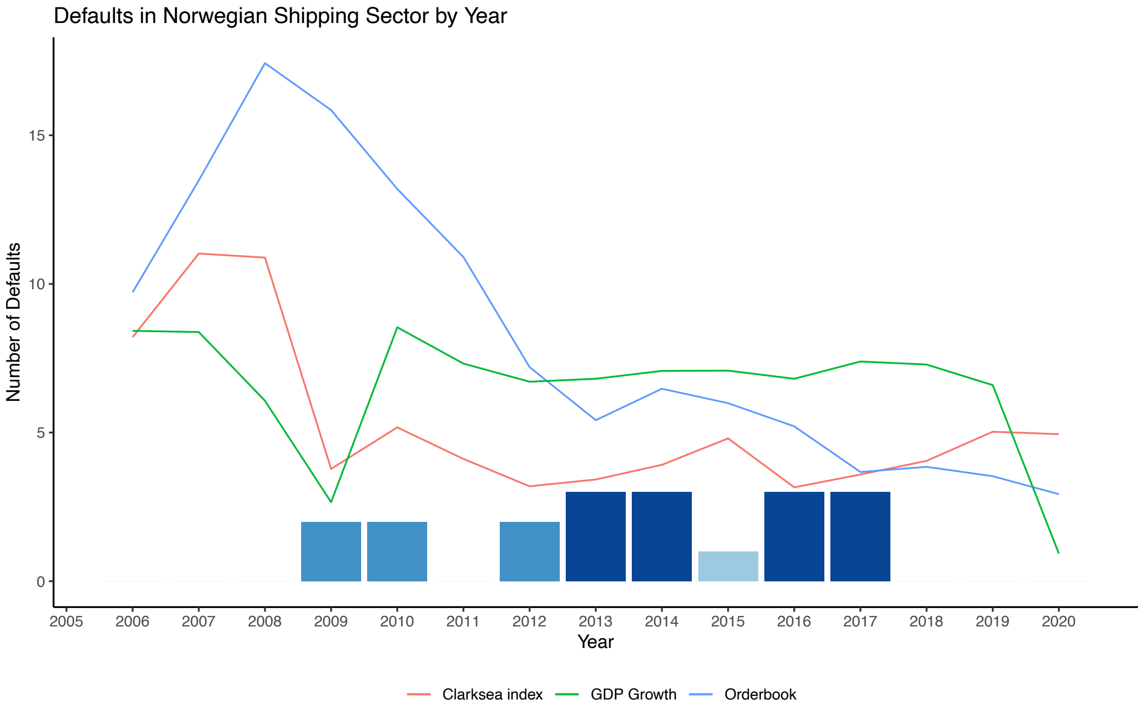

Figure 5.7: Yearly observations of defaults, compared to the change in Clarksea-Index, oil price, and GDP- growth.

5.4.2 Multicollinearity

Although not necessary for prediction accuracy, multicollinearity between the variables in the data can be problematic in the estimate the variable coefficients, and therefore the empirically interpretation of financial ratios and macroeconomic variables (James et al., 2019). We estimate the degree of covariance between the financial ratios and macroeconomic variables using the Variable Inflation factor. Given that no VIF estimates of the predictors are near the problematic threshold of 5, we consider multicollinearity unproblematic for further prediction

(James et al., 2021) (see Appendix A5). Additionally, Random Forest estimates trees one a subset of predictors, and the effect of potential multicollinearity is therefore reduced. Predictors that would be correlating are included in estimations separate from each other, and their effect more accurately estimated as the variables they correlated with are not present in the estimation.

6. Results

In this section we will first introduce the results from the hyperparameter tuning, before subsequently presenting and discussing results for each model on one and two years ahead predictions. Lastly, we will investigate the variable importance of the Random Forest model.

Hyper Parameter Tuning

The tuning grids used in the search of optimal parameters are presented in appendix A2. For Random Forest 3 predictors per split were found to be optimal, and no improvement were found in increasing number of tree estimations above 500. For the SVM tuning, a linear kernel, with a cost of 0.01 was found to produce the best result, considering it had similar average AUC as other kernels, but lower variance.

Models

Table 6.1 shows the variables used in each model and their classification method. The left-hand side of the table present the variables included in the replication of the SEBRA-model, named Logistic SEBRA, which is used for comparison. The right-hand side presents the variables included our new model with shipping specific ratios and macroeconomic variables. A logistic regression is used for both models, where we have named our model Logistic Macro. The added machine learning methods are only applied on our new model.

Logistic SEBRA – Model	Logistic Macro / Random Forest / SVM
• Return on Debt • Accounts Payable over Total Assets • Leverage (Equity Ratio) • Liquidity • Public charges owed over Total Assets • Log Total Assets • Age • Equity Loss	• Average Return on Assets last 3 years • Leverage (Equity Ratio) • Interest Expense over EBITDA • Current Ratio • Orderbook • Clarksea-index • GDP growth • Oil Price • Age • Log Total Assets

Table 6.1: Table of each predictor variable included in the model estimation. The SVM and Random Forest are both based on the new variables.

6.1 Results for Prediction One Year in Advance

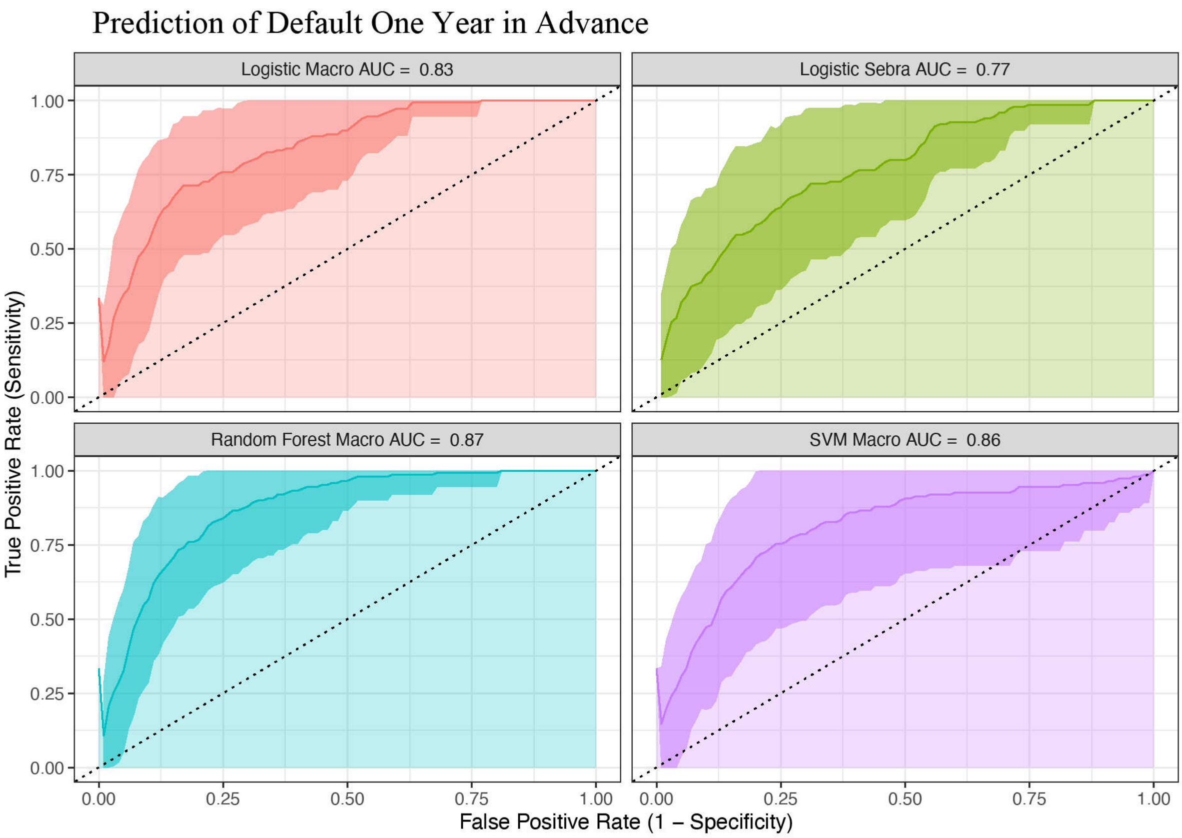

Figure 6.1: Mean TPR and FPR for each threshold, observed across 50 samples, by a one year before default model.

Through the results presented in Figure 6.1 we find that the Random Forest model performs better than the other models, achieving the highest mean AUC[13] of 87%. Mean AUC is quite similar for all the models based on the extended model with macroeconomic variables, with 86% for SVM and 83% for Logistic Macro. The replication of the SEBRA-model has an estimated mean AUC of 77%, much lower than the original model by Bernardsen and Larsen (2007) which has an AUC of 89%. We assume this difference is due to a different dataset and response variable, as well as implementation of the variables, possibly through categorization. More importantly, the confidence interval estimate of the Random Forest model is narrower than that of the other models, indicating that it is has a more stable performance on the given dataset, and therefore produce more reliable predictions. A possible reason for this is its use

[13] The AUC in each model heading, is computed by the mean AUC for each curve, and is therefore not accurate for the area under curve in the plot.

of the bagging method, which is an effective method of reducing variance in smaller samples. Through Figure 6.2 we show that the Random Forest mean ROC curve is better than that of the other models, with the curve indicating mean TPR of approximately 83%, at a 25% FPR, higher than that of the other models at 75% TPR for SVM and Logit Macro, and 63% TPR for the SEBRA replication. Considering a hypothetical new dataset with identical characteristics of size and class imbalance, these ratios would result in around 15 of 19 correctly predicted defaults, predicting around 1850 false defaults of 7486 healthy observations, further illustrating the difficulty of predicting default under this class imbalance. These predictions would also result in a model accuracy of around 75%.

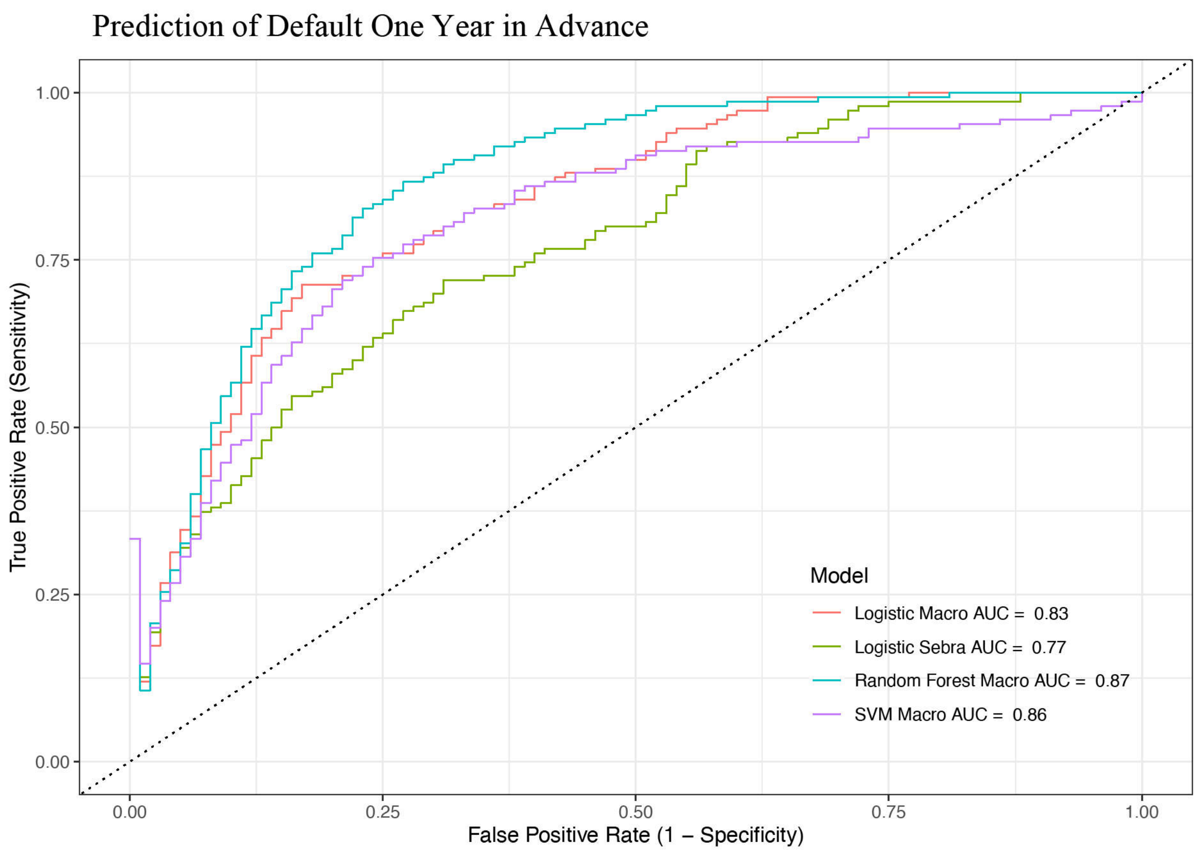

Figure 6.2: Mean ROC – curves from Figure 6.1, plotted against each other.

With the Random Forest being the most accurate, both the Logistic Macro and the SVM method perform very similar in terms of ROC-curve, and all three models based on the new ratios and the macroeconomic variables perform better than the replication of the SEBRA-model. Still, all four models manage to separate the default and healthy class at reasonable levels. This means that variables included in the SEBRA-model hold predictive power of default in a 1 – year period. Due to the increase in accuracy from the logistic model with the replicated SEBRA variables, to the logistic model macroeconomic variables, we can assume that there is additional predictive power in either the new added ratios or in the macroeconomic variables. Next it is interesting to see the performance of the models two years prior to bankruptcy, as the earlier it is detected, the greater the chance of preventing it.

6.2 Results for Prediction Two Years in Advance

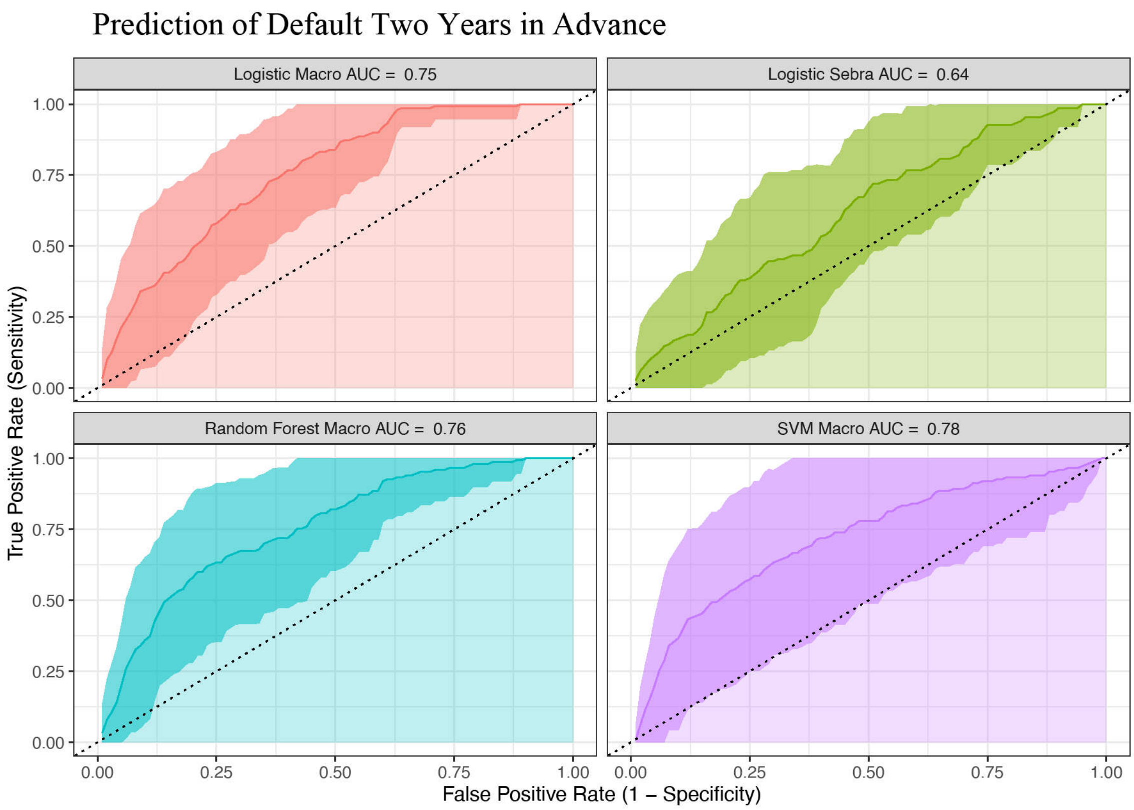

Figure 6.3: Mean TPR and FPR for each threshold, observed across 50 samples, in two years before default model.

For predictions of default two years ahead, we find that the SVM model is the one achieving the highest AUC at 78%, with a 76% AUC for the Random Forest and 75% AUC for the Logistic Macro. The AUC of the replication of the SEBRA-model has now decreased to 64%. Overall, the mean of the AUC has decreased for all models, and the estimated variance in the ROC curves has increased. Even though the SVM achieves the highest AUC, its ROC curve confidence estimate overlaps the diagonal line and is in risk of performing no better than a random guess in certain cases. For this reason, it is difficult to say with certainty that the SVM model performs well. The Random Forest and Logistic Macro models are also subject to increased variance, but the confidence estimate of their ROC curves is better than that of a random guess. Given its reliability, we consider the Random Forest model as the best performing model in two years ahead predictions as well. We still find that the Random Forest model holds predictive power, but at a lower level than before. Compared to the last result, it now has a TPR estimate around 60% for a 25% FPR, resulting in at around 11 of 19 correctly predicted defaults, with 1850 false positives. Prediction of default 2 years in advance is therefore proving to be harder with the given dataset and definitions, which further highlights the difficulty of default prediction. In this result the new models also outperform the replication of the SEBRA-model, and like the SVM model its confidence estimate overlap the diagonal line.

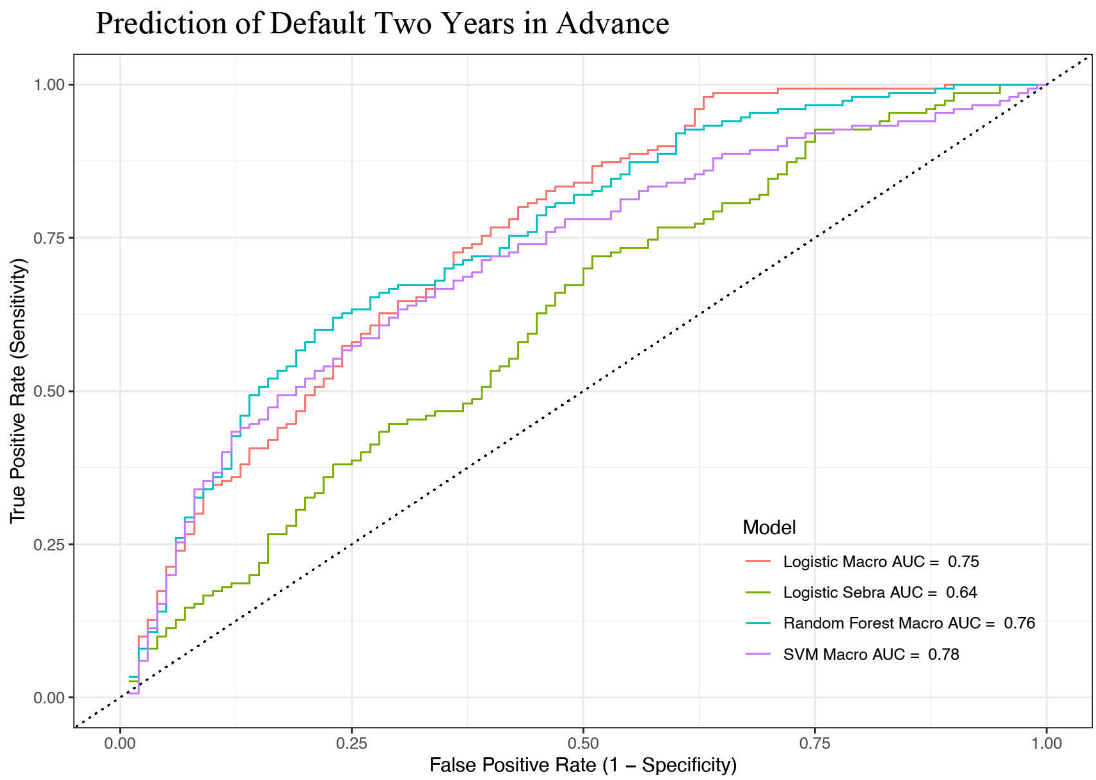

Figure 6.4: Mean ROC – curves from Figure 6.3, plotted against each other. As in Figure 6.3, the mean ROC curve is not representative for the mean AUC.

Conclusively, we can say that the models containing new variables perform better than the replication of the SEBRA-model, most likely due to a relationship between the new variables and shipping defaults. Considering the results over both estimations, a Random Forest model performs best in terms of the mean AUC and confidence interval of the true positive and false positive rates. Given the lower variance, the Random Forest model produce more stable results and is therefore the most reliable model for predictions, although having less accuracy in predicting two years ahead of default. Given the assumptions of relevant observations, transformations of variables and resampling of training set, Random Forest with extended variables is the best performing model.

6.3 Variable Importance

Based on the Random Forest model, we created a variable importance plot. As mentioned, the plot estimates the most important variables by their Mean Decrease in Gini by each tree estimation where the variable is included. The plot in Figure 6.5 is the mean estimate of each predictor across all samples. We find that Log Assets and company's Age are the most important predictors of default in the model, and that the macroeconomic variables contribute additional predictive power. Interestingly we see that apart from the average return on assets, other financial ratios contribute very little to the prediction of default. The variable importance will be discussed further in the following sections.

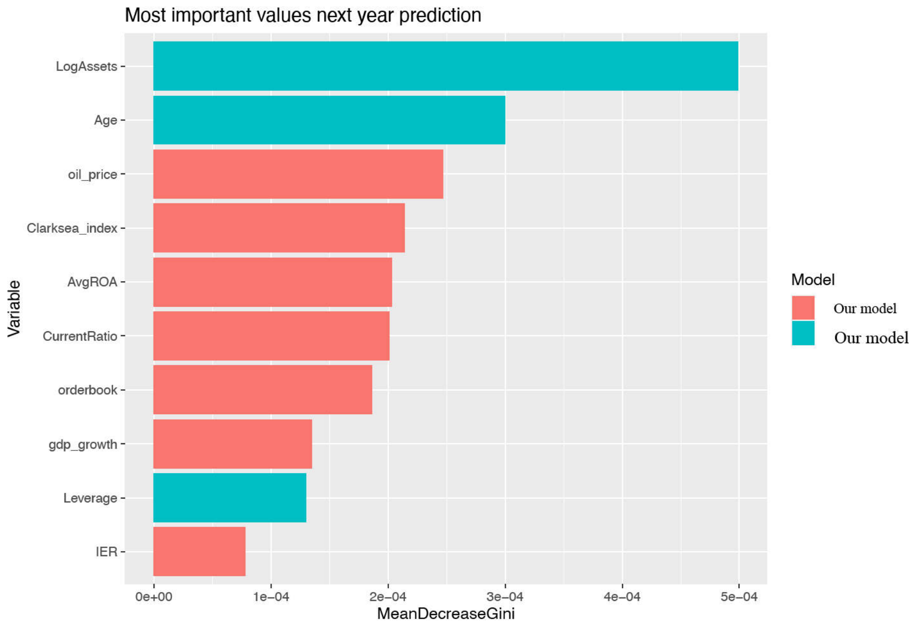

Figure 6.5: Mean decrease in GINI index for each variable. The values are calculated by their mean across the folds.

7. Illustration of Our Model

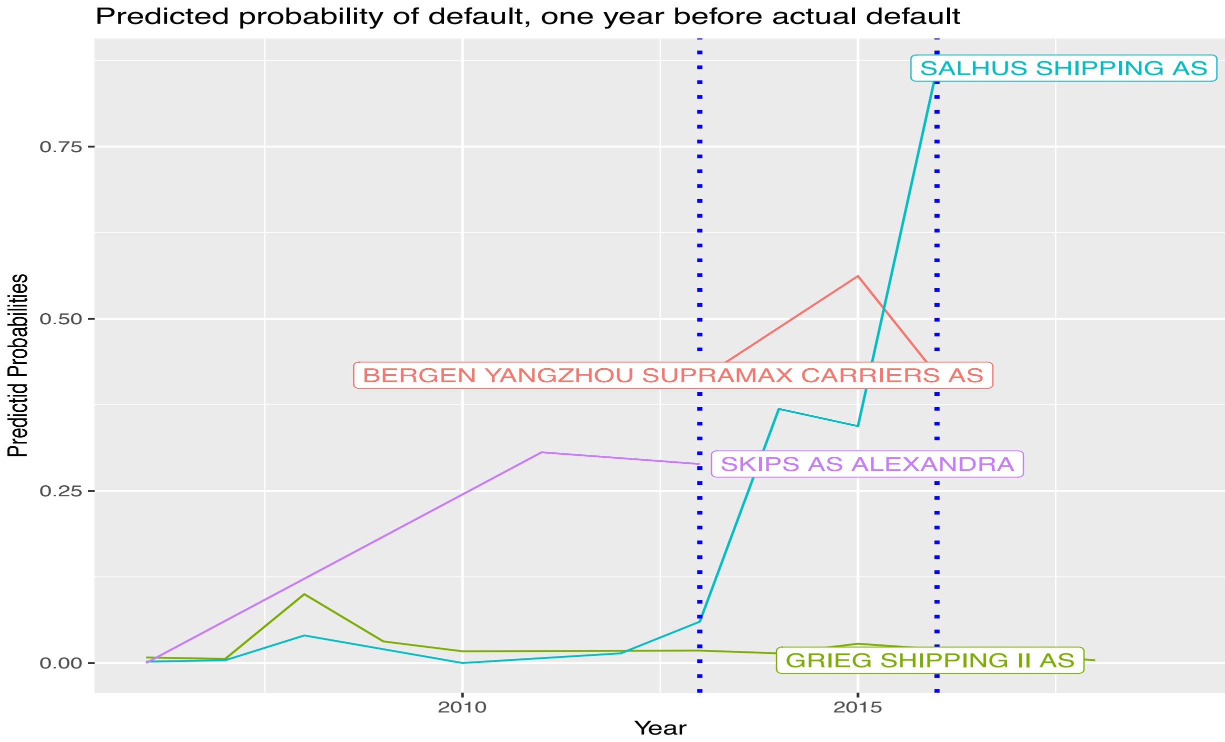

Figure 7.1: Predicted probabilities of default for four companies, by Random Forest model estimating the probability of default next year.

To illustrate the intended function of the model, we have chosen four companies from our dataset to use in a case. With the Random Forest model predicting default one year ahead, estimated probabilities of default are made in each year. The estimated probabilities can all be seen increasing for the companies which defaults in the years in advance. Skips AS Alexandra is defaulting in 2013, with Bergen Yangzhou Supramax Carriers AS and Salhus Shipping defaulting in 2016, which is indicated by the blue lines. Grieg Shipping does not go bankrupt and does not have an increase in predicted probability. The effect of age is also apparent, as Bergen Supramax have a higher probability of default from their first years.

For Skips AA Alexandra, bankruptcy proceedings opened in 2016 by the data retrieved from SNF (2023), but the real year of default is considered to be 2013 by our assumptions, as the last year of accounting information are submitted in 2013. In the case, the financial information in the year of default is also used to make a prediction of default in the next year. For Skips AS Alexandra this means that a prediction of default is made for 2014, with financial information from 2013. This is interesting as the predicted probabilities are equal in both years.

This might indicate that the company is either effectively default one year before the real default is considered, or that the company is under similar risk of default both years without being default. This further illustrates that uncertainty in response variable might produce inaccurate predictions. This case must not be seen as a validation of the model, as the companies selected are not sampled randomly, but selected to illustrate the functionality of the model.

8. Discussion:

In this section we will first discuss the estimated models and look closer at the most interesting variables. Further we compare our results to earlier shipping defaults studies before we investigate the tendencies of defaulted companies. Lastly, we discuss the use of the model and limitation and further research related to our study.

8.1 Discussion of the Estimated Models

The extended SEBRA model by Bernhardsen and Larsen (2007) achieved an AUC of 89%, while our replicated SEBRA model without macroeconomic and shipping financial ratios achieve an AUC of 77% one year prior to default. This indicates that the extended SEBRA model is superior to our replicated version. However, our replicated version of the SEBRA-model is not directly comparable to the original model. Our dataset exclusively comprises companies in the shipping sector, whereas the original model was initially estimated on companies across all sectors. Consequently, our dataset possesses unique characteristics that may not be captured by the variables in the SEBRA-model. We also have very few defaulted companies relative to the SEBRA dataset, which contains over 20 000 bankrupt companies. Additionally, our definition of the response variable differs slightly from the extended SEBRA-model. We define our response variable as defaulted companies rather than exclusively bankrupt ones. In the SEBRA model the company is considered bankrupt when "the company stops submitting accounts next year" and "bankruptcy is filed" coincide, while we define companies as defaulted when the liquidation process starts, or the last year of submitting accountings.

Our models with shipping specific ratios and macroeconomic variables, achieves higher AUC than our replicated SEBRA-model. Interestingly, when employing the Random Forest approach one year before default, we obtain an AUC that is approximately the same as the original extended SEBRA-model used by the Norwegian Central Bank. This implies that we may have identified variables that better capture the dynamics of default within the Norwegian shipping companies compared to the variables included in the original SEBRA-model. All our models also achieve a higher AUC than the replicated SEBRA-model when predicting default two years before. This shows that our variables also have good predictive power two years prior to the actual default. For our own model we achieve a higher AUC for the models estimated by machine learning methods compared to the logit model. Although the SVM achieves the highest AUC for two-year prediction it has a much higher variance then the Random Forest and it seems that Random Forest is the approach which performs best overall.

There may be several reasons for why the Random Forest outperform the other models. One reason may be that Random Forest is better than the logit model to handle noisy data and outliers (Singh, 2020), which we also find in our dataset. The Random Forest prediction will therefore be less likely to overfit the data and lead to more accurate predictions. Another reason may be that we have transformed some of the variables into categories and Random Forest perform well on both categorical and continuous variables (Singh, 2020). Lastly the Random Forest can handle linear and non-linear relationships well (Singh, 2020), which may exist between our response varaible and predictors. Although the Random Forest is the model with highest AUC, the model can be hard to interpret, and it can be difficult to understand how the variables are included in the model. For this reason, we introduced the variable importance measure.

8.2 Variable Importance

As shown in Figure 6.5 the non-financial variables size and age are identified as the most influential factors in predicting default. One explanation to the predicting power of the non-financial variables, is that older and larger firms may exhibit distinct characteristics which has contributed to their existence and growth. The survival of these companies over several years suggests an ability to adapt to dynamic market conditions, implement innovative strategies and maintain competitiveness within the industry. Larger and older firms may also have better access to financing due to their reputation and credibility, which can be critical in challenging economics times.

An additional note is that some of the defaulted companies are quite young in relation to the rest of the companies in the dataset, as well as being subsidiaries for larger companies. We suspect the reason for this is that new companies are created to undertake new projects, and that projects that fail to be profitable after a few years are declared bankrupt. There might exist a survivorship bias, in that older companies are based on profitable projects over time, and that default is therefore caused by their company specific project, rather than the effect of macroeconomic variables and ratios. This is just however a hypothesis formed on subjective observations.

For the macroeconomic variables we find that oil price is identified as the most important predictor. When we predict defaults in the Norwegian shipping industry the oil price may have higher importance than in other countries as developments in oil prices greatly affect the Norwegian shipping companies, as well as the Norwegian economy (Norwegian Shipowners' Association, 2022). Our dataset also contains observations around the oil crisis in 2014 (IMF, 2017), which also may contribute to the importance of the oil price variable. Although the oil price is an important predictor, the effect of change in oil price are ambiguous. Shipping companies with a greater focus on the oil industry experience increased earnings with a rise in oil prices, whereas those not involved in the oil sector experience increased costs when the price of oil increases, due to rising fuel costs (Norwegian Shipowners' Association, 2022). In this thesis we refrain from delving deeper into the impact of the oil price on defaults. However, further investigation should be done, before definitive conclusions can be drawn regarding the effect of the oil price on defaults.

Regarding the remaining variables, we observed that GDP is identified as the least important among the macroeconomic variables. This is surprising, as mentioned, Stopford (2009) state that the world economy is the most important variable affecting the shipping demand. Although we have tested for multicollinearity between the macroeconomic variables, we believe that some of the effect of GDP may be captured by the oil price and explain the lower importance of GDP in our prediction. A study by Jiménez-Rodríguez & Sánchez (2004) supports this idea revealing evidence of a non-linear relationship between oil price and real GDP. Further we found that both ClarkSea index and orderbook have similar importance in our prediction model. For our financial variables we discuss these further in the next section.

8.3 Compared to Earlier Shipping Studies

Previous research on shipping defaults frequently bases default estimations on datasets comprising loans or bonds such as the studies by Grammenos et al. (2008), Mitroussi et al. (2016) and Kavussanos and Tsouknidis (2016) rather than directly on shipping companies such the study by Lozinskaia et al. (2017) and our study. Although we consider the study by Lozinskaia et al. (2017) as the most pertinent, we will examine the results and variables of earlier studies and draw comparisons to our own findings.

The only shipping study employing ROC as a performance metric of the studies we have reviewed is the study by Kavussanos and Tsouknidis (2016). Their most effective model attains an AUC of 92,93%. In comparison, our top-performing model achieves an AUC of 87%. Grammenos et al. (2008) and Lozinskaia et al (2017) evaluate prediction performance using the percentage of correctly predicted observations. Grammenos et al. (2008) correctly predict 87,83% of the observations, while Lozinskaia et al. (2017) achieve a slightly lower percentage of correctly predicted observations at 69,61%. Our study predicts correctly across classes at around 75% for a one year ahead prediction, whilst having been more focused on the accuracy and reliability in prediction of true defaults.

Several variables in our analysis have shown to have a significant effect on default in the earlier studies. Lozinskaia et al. (2017) highlighted the significance of total assets, with larger firms expected to be more solvent. The study by Grammenos et al. (2008) linked shipowner's experience, potentially aligned with a company's age, to have a negative relationship with the likelihood of bond default. Bellovary et al. (2007) noted that current ratio and ROA, both featured in our model, are the ratios most used in previous default studies. Despite being identified as important financial variables in our prediction, they lack significance in earlier shipping default studies. Our least important variable IER is not found significant in the study by Kavussanos and Tsouknidis (2016). Conversely, leverage, which is less important in our study, are found significant in the study by Kavussanos and Tsouknidis (2016). Lozinskaia et al. (2017) found a negative association between GDP and default, while the oil price lacked statistical significance in predicting default probability. These differences in variable importance may suggest unique dynamics in the Norwegian shipping industry or differences in modeling approaches, however it is essential to recognize that a variable's importance in our prediction doesn't necessarily imply its statistical significance.

8.4 Tendencies of Bankrupt Companies

For many of the companies classified as default, we observe a trend where their total income appears to rise in the period leading to their eventual bankruptcy, as presented in appendix A5. Simultaneously, there is an observed decline in total assets, yearly result, and equity, while debt remain almost constant. In our case in section 7, we note a similar pattern where the probability of default slightly decreases the year before default for two of the companies. This suggest that companies, when facing financial distress, may attempt to undergo restructuring. This restructuring process could involve selling vessels, decreasing debt ratio through negotiations with debt holders or through equity infusion, and engaging in discussion with stakeholders, such as suppliers, customers, and employees. In the context of our predictive analysis, restructuring efforts by these companies might introduce complexities or changes that impact the accuracy of our predictions.

When it comes to the companies undergoing restructuring, there will be variations among them. Companies may initiate restructuring at different times or in different ways, making it challenging to synchronize these efforts for each company. Additionally, for companies undergoing restructuring there will also be different outcomes. Some companies will experience a successful restructuring, while others will not succeed and proceed into a defaulted status. If companies also use financial ratios as preventive measures or to decide when to undergo a restructuring, this could potentially reduce the predictive power of these ratios, as companies may strategically adjust their financial ratios to project a healthier financial position. Unfortunately, we lack data to further investigate how this affects our default predictions.

8.5 Use of the Model

The model is intended to be used by banks, investors, shipping companies, suppliers and other stakeholders which are involved in the Norwegian shipping industry to predict which companies that may default. In a stakeholder position it would be possible to choose a threshold for the model which corresponds to a desired level. For debtholders it is crucial to predict defaulted companies correct as defaulted companies could be extremely costly. It will therefore be in the interest of debtholders to choose a low threshold. More companies would then be categorized as defaulted, both defaulted and non-defaulted companies. On the other hand, for stakeholders where a default is not that costly, such as suppliers, a higher threshold may be appropriate.

While our model achieves a relatively strong AUC, it tends to predict a significant number of non-defaulted companies as defaulted and misclassifying some defaulted companies as non-defaulted. Accurate prediction of defaulted companies is crucial due to the potential financial losses investors may face if they invest in companies incorrectly classified as non-defaulted. There is also a cost associated with misclassifying non-defaulted companies, representing missed opportunities for stakeholders who might avoid mistakenly predicted companies. Consequently, users of the model should supplement it with additional indicators such as credit ratings or their own analyses before making decisions.

We have predicted the probability of default for both one and two years preceding the actual event. This affords stakeholders an opportunity to respond to the predictions and implement preventive actions before it becomes too late. While a longer lead time for predicting default is advantageous, the model's accuracy tends to decrease the further into the future it attempts to predict. The trade-off between lead time and accuracy may be different for different stakeholders. For instance, investors may be more interested in early detection of potential defaults to adjust their investment portfolios and manage risks. Financial institutions on the other side might prioritize high accuracy to minimize financial risks associated with lending.

8.6 Limitations & Further Research

Although our study has shown that it is possible to make a prediction model which suits the Norwegian shipping industry, there are several limitations to consider. Firstly, we made assumptions considering which companies were relevant for analysis based on their registered company code and age. Furthermore, we made assumptions on whether size of the company made it exposed to macroeconomic variables of the international shipping market. If too many observations were included, the results could be disturbed by noise in the observations. If too few observations were included, this thesis is biased towards the small number defaults in our dataset, and not representative for the shipping industry. Companies are also possibly affected differently or oppositely by the macroeconomic variables. An example of this is the oil price, which could possibly affect some of the companies oppositely, depending on offshore or freight operations. This would be problematic in the estimation of the effect of the variable and create inaccurate predictions. A natural proposition for further research is therefore to perform predictions on a subsector individually. It might also be interesting to further investigate the effect of oil price on defaults, both on a new dataset, but also for another country, to see if oil price remains important in the new prediction.

Another limitation of the model is the categorization of the variables. Categorization was performed before the cross-fold validation split for computational reasons. The estimation of the ratio category was therefore based on observed variables outside of its training and test data. We considered the effect to be minimal, under the assumption that the distribution of the ratios would not change significantly given many observations in each set. This is however a potential limitation should this assumption not hold, and a proposition for further research. Under the section of tendencies of default companies, we observed a decline in income when financials were scaled within the companies (see appendix A5). A proposition for further research is to scale variables within companies to include this effect in a prediction model.

The rare nature of defaults is also a limitation. Our dataset contains relatively few defaults and makes it harder to make a reliable prediction model due to the imbalanced data. In this thesis, we performed a 5-fold cross validation split, with 10 repetitions to estimate the variance of the models relative to each other. This might create a false representation of the model's performance when used on new data in coming years. The reason for this is that a low number of defaults are to be expected, and a prediction of defaults in for example 2022, will be subject to a high degree of uncertainty. This uncertainty could possibly exceed that of the variance implied in the results. The accuracy of predictions is in risk of deviating significantly from the mean accuracy rates observed in the results.

Lastly there is significant uncertainty in whether the response variable of default accurately describes the actual year of default. As mentioned previously, default is defined in the SNF database as the year that liquidation or bankruptcy proceedings starts (Mjøs & Selle, 2022), even though the last year of submitted accounting information often occurs years before. Furthermore, we observe that companies possibly use the last years before default to restructure the company or sell important assets. In these cases, one could assume that the impact of ratios and macroeconomic factors that lead to default, falls outside of the estimation window. The effect of the observations is then assigned to healthy companies, and the observed values in the last years are the values of an already defaulted company. The time incongruence in the response variable is therefore a considerable limitation that contributes to inaccuracy in the model estimation and variable importance understanding in this thesis. Conclusively we can say that due to difficulty in separating different operations of the companies and the uncertainty of the response variable, the estimation in this thesis is subject to a great deal of noise, and the result should be interpreted in relation to these effects.

9. Conclusion

The primary intention of this thesis has been to create a model for predicting defaults in the Norwegian shipping sector. The shipping sector is important to the Norwegian economy, and from the introduction we saw that defaults in this sector incurs heavy losses not only for shareholders, but for customers and connected industries as well. Combining prior research of defaults outside and in the field of shipping, we identified financial ratios and macroeconomic variables useful for default prediction. We further compared this with the local adaption of the SEBRA – model, used by the Norwegian Central Bank for predicting default amongst Norwegian companies (Eklund, Larsen, & Bernhardsen, 2001).

Through assumptions of size, age and company, a selection of relevant shipping companies was made, of which we had 889 companies and 19 defaults. Furthermore, data was categorized for enhanced model performance. The specific assumptions made regarding sector and size, offer a range of alternative assumptions possible for further research.

Predictions were made with the three classifications models: Logistic regression, Random Forest, and Support Vector Machines, under the hypothesis that machine learning methods might better handle potential non-linear relationships between the predictors. Our findings indicate that the Random Forest model, with our specific indicators perform better than the other models, both in terms of expected accuracy, but also in reliability of the results. Due to a low number of defaults, we deem it possible that some of the model's performance comes from its ability to better handle categorized data, as well as variance.

Despite achieving a high degree of accuracy in our chosen metrics of model evaluation, we found that predicting defaults remains a difficult task. To a degree it comes from a large class imbalance, making default prediction subject to noise. Furthermore, due to the complexity of the market, there exist much uncertainty in the whether the year of a true default could be assigned accurately. Some companies liquidate assets years before default, effectively ending their operations, years before going default. Through a variable importance estimate of the Random Forest model, it was found that the most important indicators of default, were size and age. Subsequently, shipping financial variables also holds predictive powers, but to a lesser degree.

Bibliography

Alexandridis, G., Kavussanos, M. G., Kim, C. Y., Tsouknidis, D. A., & Visvikis, I. D. (2018). A survey of shipping finance research: Setting the future research agenda. *Transportation Research Part E: Logistics and Transportation Review, Volume 115,*, 164-212.

Altman, E. I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance, Vol. 23*, 589-609.

Beaver, W. H. (1966). Financial Ratios As Predictors of Failure. *Journal of Accounting Research, Vol. 4*, 71-111.

Bellovary, J. L., Giacomino, D. E., & Akers, M. D. (2007). A Review of Bankruptcy Prediction Studies: 1930-Present. *Journal of Financial Education*, 42.

Berg, D. (2007). Bankruptcy prediction by generalized additive models. *Applied Stochastic Models in Business and Industry,*, 129-143.

Bernhardsen, E. (2001). *A Model of Bankruptcy Prediction.* Oslo: Norges Bank.

Bernhardsen, E., & Kai, L. (2007). Modelling credit risk in the enterprise sector – further development of the SEBRA model. *Economic Bulletin vol. 78*, 102-108.

Berrar, D. (2019). *reaserchgate.net.* Retrieved from reaserchgate.net: https://www.researchgate.net/profile/Daniel-Berrar/publication/324701535_Cross-Validation/links/5cb4209c92851c8d22ec4349/Cross-Validation.pdf

Breiman, L. (2001). Random Forests. *Machine Learning 45*, 5-32.

Calarese, R., Marra, G., & Osmetti, S. A. (2016). Bankruptcy prediction of small and medium enterprises using a flexible binary generalized extreme value model. *The Journal of the Operational Research Society, Vol. 67*, 604-615.

Careratings. (2018, 08). *Rating Methodology - Shipping Companies*. Retrieved from Careratings: https://www.careratings.com/upload/NewsFiles/GetRated/RatingMethodology-ShippingCompanies_Aug2019.pdf

CFI. (n.d.). *Debt Capacity*. Retrieved from Corporate Finance Institute: https://corporatefinanceinstitute.com/resources/commercial-lending/assessing-debt-capacity/

Clarksons. (2023). *Shipping Intelligence Network*. Retrieved from Clarksons: https://www.clarksons.net/n/#/sin/timeseries/browse

Eklund, T., Larsen, K., & Bernhardsen, E. (2001). Model for analysing credit risk in the enterprise sector. *Economic Bulletin Q3*, 99-106.

GAO. (2020, 01). *Ocean Carrier Bankruptcies: Federal Agencies Have a Limited Role in Addressing Effects on Cargo*. Retrieved from GAO: https://www.gao.gov/products/gao-20-353r

Gartner. (n.d.). *Profitability*. Retrieved from Gartner: https://www.gartner.com/en/finance/glossary/profitability

Gissel, J. L., Giacomino, D., & Akers, M. D. (2007). A Review of Bankruptcy Prediction Studies: 1930-Present. *Journal of Financial Education*, 42.

Graham, L. (2016, 09). *CBNC*. Retrieved from All at sea? The ripple effect of Hanjin's crisis: https://www.cnbc.com/2016/09/23/all-at-sea-the-ripple-effect-of-hanjins-crisis.html

Grammenos, C. T., Nomikos, N. K., & Papapostolou, N. C. (2008). Estimating the probability of default for shipping high yield bond issues. *Transportation Research Part E: Logistics and Transportation Review, Volume 44*, 1123-1138.

Hagel , J., & Brown, J. S. (2013, 10). *Success or struggle: ROA as a true measure of business performance* . Retrieved from Deloitte: https://www2.deloitte.com/xe/en/insights/topics/operations/success-or-struggle-roa-as-a-true-measure-of-business-performance.html

Hayes, A. (2023, 09). *Understanding Liquidity and How to Measure It*. Retrieved from Investopedia: https://www.investopedia.com/terms/l/liquidity.asp

Illmer, A. (2017, 02). *Hanjin: Final curtain falls on shipping saga*. Retrieved from BBC: https://www.bbc.com/news/business-38953144

IMF. (2017). *A Crude Shock: Explaining the Impact of the 2014-16 Oil Price Decline Across Exporters*. Retrieved from International Monetary Fund: https://www.imf.org/en/Publications/WP/Issues/2017/07/18/A-Crude-Shock-Explaining-the-Impact-of-the-2014-16-Oil-Price-Decline-Across-Exporters-44966

IMO. (2015, 11). *"Shipping: indispensable to the world" selected as World Maritime Day theme for 2016*. Retrieved from International maritime organization: https://www.imo.org/en/MediaCentre/PressBriefings/Pages/47-WMD-theme-2016-.aspx

James, G., Witten , D., Hastie, T., & Tibshirani, R. (2022). *An Introduction to Statistical Learning with Applications in R (2nd edition).* Springer-Verlag New York Inc.

Jiménez-Rodríguez, R., & Sánchez, M. (2004). *Oil price shocks and real gdp growth.* Frankfurt: European Central Bank.

Katsomitros, A. (n.d.). *The shipping industry must adapt if it is to survive in the modern world*. Retrieved from World Finance: https://www.worldfinance.com/markets/the-shipping-industry-must-adapt-if-it-to-survive-in-the-modern-world

Kavussanos, M. G., & Tsouknidis, D. A. (2016). Default risk drivers in shipping bank loans. *Transportation Research Part E: Logistics and Transportation Review, Volume 94,*, 71-94.

Kenton, W. (2021, 06). *Return On Debt: What it is, How it works*. Retrieved from Investopedia.com: https://www.investopedia.com/terms/r/return-on-debt.asp

Kuhn , M. (2023, 08 23). *tune: Tidy Tuning Tools* . Retrieved from r-project: https://cran.r-project.org/web/packages/tune/index.html

Lozinskaia, A., Merika, A., Merikas, A., & Penikas, H. (2017). Determinants of the Probability of Default: The Case of the Internationally Listed Shipping Corporations. *Maritime Policy & Management*.

Mitroussi, K., Abouarghoub, W., Haider, J. J., Pettit, S. J., & Tigka, N. (2016). Performance drivers of shipping loans: An empirical investigation. *International Journal of Production Economics, Volume 171, Part 3*, 438-452.

Mjøs, A., & Selle, S. F. (2022). *Regnskapsdatabasen - Norwegian Corporate Accounts. Documentation and quality assurance of SNF and NHH's database of accounting and corporate information on Norwegian companies.* Bergen: SNF.

NHO. (n.d.). *Fremtidsmuligheter i maritime næringer*. Retrieved from NHO: https://www.nho.no/publikasjoner/veikart/fremtidsmuligheter-i-maritime-naringer/

Norwegian Shipowners' Association . (2022). *Maritime Outlook 2022.* Oslo: Rederi.no.

Norwegian Shipowners' Association. (2021). *Maritime outlook 2021.* Rederi.no.

Nurunnabi, A., Nasser, M., & Imon, A. (2016). Identification and classification of multiple outliers, high leverage points and influential observations in linear regression. *Journal of Applied Statistics*, 509-525.

OECD. (2018). *COUNCIL WORKING PARTY ON SHIPBUILDING.* OECD.

OECD. (n.d.). *Ocean shipping and shipbuilding*. Retrieved from OECD: https://www.oecd.org/ocean/topics/ocean-shipping/

Ohlson, J. A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research, Vol. 18*, 109-131.

Peterson, R. A., & Cavanaugh, J. E. (2019). Ordered quantile normalization: a semiparametric transformation bilt for the cross-validation era. *Journal of Applied Statistics*, Voolume 47.

Ross, S. (2021, 07). *Is there a downside to having a high liquidity ratio?* Retrieved from Investopedia.com: https://www.investopedia.com/ask/answers/122414/there-downside-having-high-liquidity-ratio.asp

Shrestha, N. (2020). Detecting multicollinearity in regression analysis. *American Journal of Applied Mathematics and Statistics Vol.8*, 39-42.

Singh, D., & Singh, B. (2020). Investigating the impact of data normalization on classification performance. . *Applied Soft Computing*, Volume 97.

Singh, K., & Xie, M. (2010). *Bootstrap Method.* International Encyclopedia of Education (Third Edition).

Stautland, K. (2021, 03). *Norway is the world's fourth largest shipping nation*. Retrieved from Norwegian Maritime Autorithy: https://www.sdir.no/en/news/news-from-the-nma/norway-is-the-worlds-fourth-largest-shipping-nation/

Stopford, M. (2009). *Maritime Economics (3rd edition)* . Routledge.

Syvertsen, B. D. (2004). *Hvor godt treffer kredittrisikomodeller for norske foretak?* Oslo: Norges Bank.

The Guardian. (2016, 09). *Hanjin Shipping bankruptcy causes turmoil in global sea freight*. Retrieved from The Guardian: https://www.theguardian.com/business/2016/sep/02/hanjin-shipping-bankruptcy-causes-turmoil-in-global-sea-freight

The World Bank. (2023). *GDP growth (annual %)*. Retrieved from World Bank: https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG

Torgersen , O. (2018, 05). *Likviditet*. Retrieved from NDLA: https://ndla.no/nb/subject:1:792414c5-896f-470a-9558-6101d7266237/topic:1:04afff6e-bea4-48b3-82fa-a4f7b21527aa/topic:1:461389e9-2cc3-4874-99d1-b5469823f088/resource:1:116000

Vanvik, H. (2019, 04 03). *Britisk bank mener den ble ført bak lyset av Pareto – krever 350 millioner av meglerhuset*. Retrieved from Dagens Nærinsgliv: https://www.dn.no/marked/pareto-securities/semptem-offshore/varada-marine/britisk-bank-mener-den-ble-fort-bak-lyset-av-pareto-krever-350-millioner-av-meglerhuset/2-1-555653

Citations of R packages used for model Estimation:

A. Liaw and M. Wiener (2002). Classification and Regression by randomForest. R News 2(3), 18--22.

Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F (2023). _e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien_. R package version 1.7-13, <https://CRAN.R-project.org/package=e1071>.

R Core Team (2023). _R: A Language and Environment for Statistical Computing_. RFoundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.

Appendix

A1 Hyperparameter Tuning

Random Forest:

Predictors		Trees	
Searched	Selected	Searched	Selected
3,4	3	300,400,500	500

Table A1.1: Search grid for Random Forest

Method	kernel	Parameter	Range
Support Vector Machines	Linear	Cost	0.01, 0.1, 1 5, 10,
	Polynomial	Cost	0.1, 1, 5, 10,
		Gamma	1, 2, 3,
		Degree	1, 1.5, 2, 2.5, 3
		Coef	0.1, 1, 1.5, 2
	Radial	Cost	0.1, **1**, 5, 10
		Gamma	0.1, **1,** 2

Table A1.2: Search grid for SVM

Selected: Linear with cost = 0.01

A2 Variable Categorization

Quantiles	Predictors
0%, 10%, 30%, 70%, 90%, 100%,	ROD, Leverage, APTA, AvgROA, Current-ratio,
0%, 20%, 80% 100%	Liquidity, IER.
0%, 10%, 90%, 100%	PBR

Table A3 Categories for each of the variables, by their quantiles.

A3 Ratio Calculation

Predictor	Formula	Database notation
Variables included in the SEBRA-Model		
Profitability	$\frac{EBITDA}{Total\ Debt}$	$\frac{EBITDA}{Total\ debt + 1}$
Solidity	$\frac{Equity}{Total\ Assets}$	$Leverage = \frac{Ek}{Sumeiend + 1}$
Loss of Equity Indicator	Equity < Invested Equity	$Ek < Inv$_ek (0,1)
Liquidity	$\frac{Current\ Assets - Short\ term\ Debt}{Total\ Income}$	$Liquitidy = \frac{oml - kgjeld}{totinn + 1}$
Age Indicator	Age of company	(1,2,3,4,5,6,7,8)
Asset Size scaled by currency	Log (Total Assets)	Log (Sumeiend + 1)
Accounts payable over Total Assets	$\frac{Accounts\ Payable}{Total\ Assets}$	$\frac{kgjeld}{\text{sumeiend} + 1}$
Public charges owed over Total Assets	$\frac{Public\ Charges\ Outstanding}{Total\ Assets}$	$PBR = \frac{offavg}{Sumeiend + 1}$
Variables included in our model		
Current Ratio	$\frac{Current\ Assets}{Short\ Term\ Debt}$	$Current\ Ratio = \frac{oml}{kgjeld + 1}$
Return on assets	$\frac{Annual\ result}{Total\ assets}$	$ROA = \frac{aarsrs}{sumeiend + 1}$
Interest expense ratio	$\frac{Interest\ Expense}{EBITDA}$	$\frac{rentekost}{EBITDA + 1}$
Leverage	$\frac{Equity}{Total\ Assets}$	$Leverage = \frac{Ek}{Sumeiend + 1}$
Age	Age of company	(1,2,3,4,5,6,7,8)
Size	Log (Total Assets)	Log (Sumeiend +1)

Table A3 Variables computed, with their name in the SNF database.

A4 Descriptive Statistics Financial Information

Summary statistics for healthy companies

	Variable	Median	Mean	Min	Max	1st Quantile	3rd Quantile
1	Total Income	27484	167311	-499341	15135071	1164	104750
2	EBITDA	2157	19856	-5676193	1588502	-628	24051
3	Result	94	10047	-4283185	11272141	-3660	12169
4	Total Debt	93778	365703	-3	15496122	27374	285293
5	Equity	44342	334889	-1092450	25617035	1876	206595
6	total Assets	171821	700695	0	31181378	60649	565126
7	Interest Expense	22	7056	-68639	551392	0	3371
8	Cash	6667	41935	-452	4974638	1039	26003
9	Current Assets	26336	124440	0	7556294	6918	86084
10	Ships and Rigs	0	234709	0	11466373	0	163774

Table A4.1 Summary statistics of healthy companies.

Summary statistics for companies going default within two years

	Variable	Median	Mean	Min	Max	1st Quantile	3rd Quantile
1	Total Income	15533	65324	-951	943570	4409	34440
2	EBITDA	-64	-3524	-317104	111194	-9547	6556
3	Result	-4278	-38654	-1070529	142000	-23296	0
4	Total Debt	89777	186464	0	862000	28616	256168
5	Equity	24	-15794	-693961	716000	-25386	26960
6	total Assets	50018	170670	0	1578000	13396	177663
7	Interest Expense	1679	8185	0	97002	4	4659
8	Cash	2479	9163	-384	137777	183	8427
9	Current Assets	9438	27020	-378	440000	3110	25012
10	Ships and Rigs	502	82358	0	598849	0	115248

Table A4.2 Summary statistics of default companies.

A5 Variable Inflation Factor test.

Variable Inflation Factor SEBRA Model	GVIF	Df	GVIFDf))	Variable Inflation Factor Extended	GVIF	Df	GVIFDf))
ROD	1.317	4	1.035	AvgROA	2.134	4	1.099
Leverage	2.100	4	1.097	Leverage	1.587	4	1.059
APTA	1.470	4	1.049	CurrentRatio	1.368	4	1.040
LogAssets	2.044	1	1.430	Age	1.148	1	1.071
Age	1.074	1	1.037	LogAssets	1.385	1	1.177
EqLoss	1.905	1	1.380	IER	1.319	2	1.072
Liquidity	1.792	2	1.157	orderbook	1.825	1	1.351
PBR	1.197	2	1.046	gdp_growth	1.573	1	1.254
				Clarksea_index	1.655	1	1.287
				oil_price	1.437	1	1.199

Table A5: Results of Variable Inflation Factor test.

A6 Tendency of defaulted companies

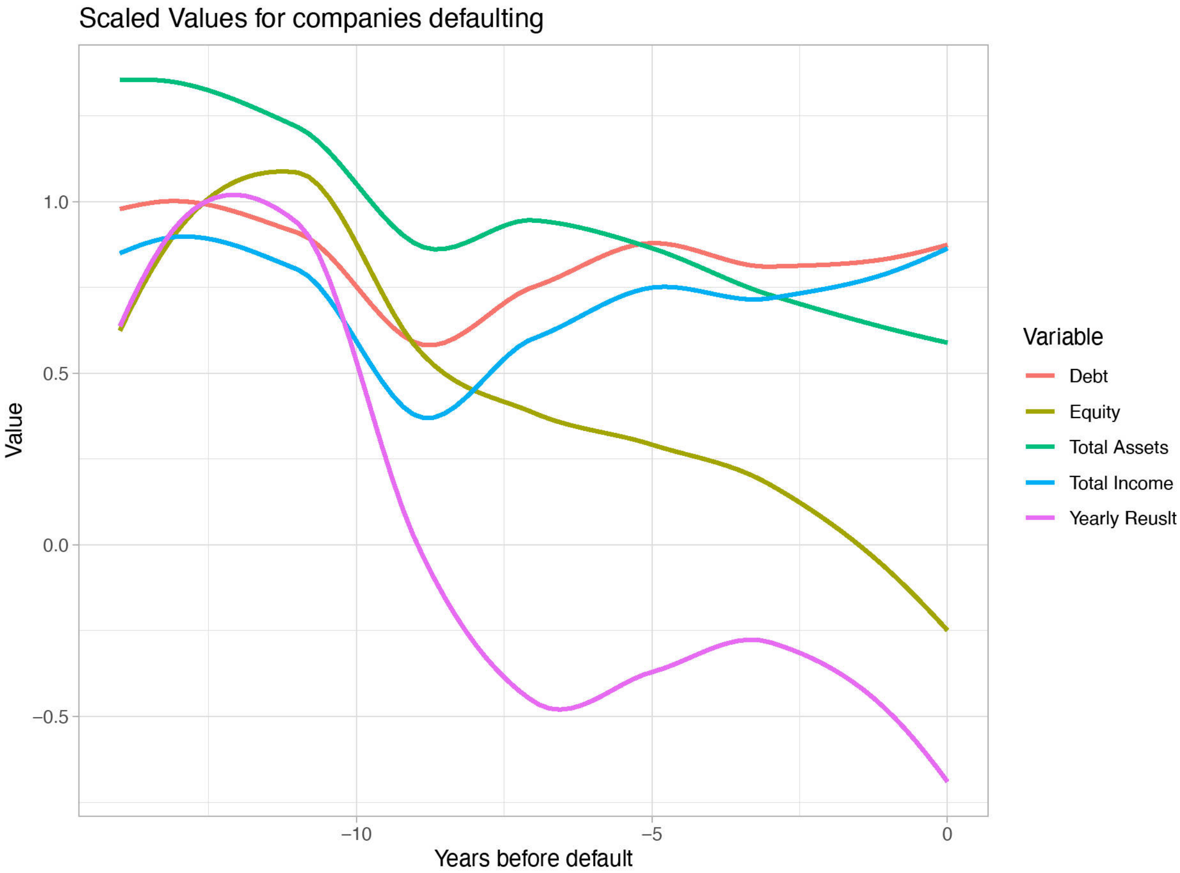

Figure A6: Scaled values for default companies