Examining classical capital structure models of debt utilization decisions in Norwegian SME shipping companies

Abstract. This study aims at identifying variables, which have a notable impact on the choice of leverage ratio in 161 Norwegian small- and medium-sized shipping companies during the period from 2008 to 2019. We apply linear-, logistic- and machine learning (ML) regression models. We interpret our findings in terms of the three major theories of companies' capital structure choices: The Trade-off Theory, The Pecking Order Theory and The Market Timing Theory of Capital Structure. Importantly, the ML models support the results of the naive regression models. We find that *tangibility* has the highest impact on leverage ratio and identify a positive relationship between *tangibility* and *leverage*. Also, several models agree on a negative relationship between *profitability* and leverage. The Logistic Regression (logit) model provided full support to The Pecking Order Theory. Generally, the models lend some support to all three theories of capital structure choice.

Key words: Norwegian SME shipping companies, panel regression, logit regression, capital structure choice theories.

1 Introduction

The shipping industry has a reputation of intense use of debt. According to an ABN AMRO Report of 2011 (Gorgels, 2011), the portion of total external funding provided by debt funding is estimated at more than 80%. Further, a recent paper by Drobetz et al. (2013) examines debt utilization in G7 countries across various industries and finds mean leverage ratios in the shipping industry to be almost twice as high as in all other industrial sectors, excluding financials and utilities. When analyzing motivating factors for debt utilization in the shipping industry, empirical studies have so far focused on the largest, globally active and stock exchange-listed companies (MNEs). However, these shipping companies only account for approximately 25% of total global tonnage supply. The remaining portion of about 75% of global tonnage supply comes from small- and medium-sized companies (SMEs). The SMEs remain largely unexplored in terms of use of debt and determinants of capital structure choices. Findings from studies examining capital structure variables' impact on leverage ratios and capital structure choice theories in MNEs may not be appropriate for characterizing SMEs.

Shipping is one of Norway's core export industries consisting of a large number of non-listed shipping firms. Also, we know of no previous studies of debt utilization in Norwegian SME shipping companies.

This study aims at (i) identifying the impact of capital structure variables on leverage ratios in a sample of 161 Norwegian shipping SMEs, and (ii) interpreting the findings with respect to the three major theories of capital structure choice: The Trade-off Theory of Capital Structure, The Pecking Order Theory and The Market Timing Theory of Capital Structure. To this end, we will apply linear- and logistic models which are frequently applied in the literature and also machine learning regression models. We do not know of any previous studies that apply ML regression models on a Shipping SME dataset for this purpose.

Analyzing a company's capital structure based on models presented in this study will be a significant aid in the valuation of small shipping firms. The relative levels of equity and debt affect risk and cash flow and, therefore, the amount an investor would be willing to pay for a share in the company. Capital structure also matters because it influences the cost of capital for these firms. Generally, when valuators use income-based valuation methods — such as discounted cash flow — they convert projected cash flows to present value by applying a discount rate. That rate, which generally reflects the return that a hypothetical investor would require, is derived from the cost of capital, which is commonly based on the weighted average cost of capital (WACC). Many business owners strive to be debt-free, but a reasonable amount of debt can provide some financial benefits. Debt is often cheaper than equity, and interest payments are tax-deductible. So, as the level of debt increases, returns to equity owners also increase — enhancing the company's value. If risk was not a factor, then the more debt a business assumes, the greater its value would be. But at a certain level of indebtedness, the risks associated with higher leverage begin to outweigh the financial advantages. This is especially crucial in the shipping industry, which is more leveraged than other industries. When debt reaches this point, investors may demand higher returns as compensation for taking on greater risk, which has a negative impact on business value. So, the optimal capital structure comprises a sufficient level of debt to maximize investor returns without incurring excessive risk.

The capital structure decision involves a choice by the firm and might be more appropriately captured by a limited dependent variable, i.e., a binary choice variable representing the decision *of whether or not to take on more debt*. The logistic regression model is ideal for modeling binary variables. The fact that the logit model employs a different response variable than the panel regressions and ML models, and broadly agree with the findings of these models, makes our results more robust.

ML models are capable of capturing complex, functional (non-linear) relations between the input and the output. We employ these models to determine whether they agree with the linear models on significant explanatory variables and to find out if ML models are better equipped to distinguish between the three capital choice theories. It turns out that they do agree on significant explanatory variables, but not always on the signs of the coefficients.

Our analysis suggests that tangibility is the most prominent variable in the determination of leverage, with a positive relationship to leverage. Hence, a higher share of fixed assets to total assets might lead to higher leverage. Profitability proved to be an important (significant) variable in several models, which agreed on a negative relationship to leverage. Hence, profitable companies might have lower leverage. The models did not agree on the impact and relation to leverage for the variables *change* in tangibles, operating leverage, company size, and company age. The change in average yearly oil price was considered to have a low impact on leverage by all models.

We further find some limited support for all three theories of companies' capital structure choices. The Logistic Regression model however, provided full support for the Pecking Order Theory. We note that there are only three significant variables in the Logistic Regression model, and two of them concern tangibility. Summarized, the models employed in this study provided some support for all three theories of capital choice, but were unable to label one theory as superior to the others.

The remainder of this paper is organized as follows: Section 2 reviews the capital structure choice theories and some relevant empirical studies, section 3 describes the dataset, section 4 explains the regression models and section 5 presents and interprets our findings. Finally, section 6 provides concluding remarks.

2 Theory

This section briefly reviews the three theories firms' capital structure choice: The *Trade-off Theory* of Capital Structure, *The Pecking Order* Theory, and *The Market Timing* Theory of Capital Structure. In this section we also review some important empirical studies.

2.1 The Trade-off Theory of Capital Structure

The Trade-off Theory of Capital Structure originates from Kraus and Litzenberger's (1973) contribution to the Modigliani-Miller irrelevance proposition (Modigliani and Miller, 1963). The Trade-off Theory of Capital Structure suggests that a company chooses the share of debt and equity financing by balancing the costs and benefits from these sources of financing. Based on existing market imperfections as bankruptcy costs, tax-shield incentives, and agency costs of debt and equity, the theory introduces an optimal level of corporate debt; At the optimal debt level, the cost of the next unit of debt equals its benefit. Companies identify this optimum as their target level for debt utilization, and actively pursue this target.

Later versions of The Trade-off Theory of Capital Structure suggest companies balance agency costs of debt and equity. Examples of agency costs of debt are asset substitution (Jensen and Meckling, 1976) and underinvestment (Myers, 1977). Examples of agency costs of equity are information requirements and the free cash flow problem (Jensen, 1986). The free cash flow problem centers on the assumption that companies with high levels of free cash flow are more likely to undertake value-decreasing investments and takeovers. Diminishing return on equity may cause the stock price to deteriorate, increasing the risk of a hostile take-over. On the other hand, debt may reduce the agency cost of equity. Debt is often supported by collateral, which may force management to make wise investment decisions in order to meet contractual terms. The free cash flow aspect of The Trade-off Theory of Capital Structure has been criticized as it does not account for long-term investments that can be profitable in the long run, such as R&D. Further, leverage exposes the company to interest rate risk beyond the managers' control.

2.2 The Pecking Order Theory

Myers (1984) and Myers and Majluf (1984) offer an alternative framework for explaining corporate capital structure decisions, known as The Pecking Order Theory. According to this theory, a company's funding decisions are based primarily on the asymmetry of information between company insiders (management) and company outsiders (investors) and potential adverse selection problems. The Pecking Order Theory claims that investors pursue any issuance activity in the light of management's superior knowledge of the company's true value and risks. For example, investors may view an equity issuance as a signal that management intends to capitalize on a perceived overvaluation of the company's shares. Investors, being aware of this asymmetry, will react to this signal by assigning a lower valuation to the company, thus causing share prices to drop. This reaction, according to The Pecking Order Theory, makes issuing equity a low-ranking preference for management. The issuance of debt, in comparison, may be viewed as a signal of the management's confidence in the company's value and their belief that the company's equity is undervalued. The theory thus claims that companies in need of funding will first exploit internal funding options through retained earnings. After the depletion of internal funds, companies will turn to the debt markets for further capital needs. Once the issuance of further debt becomes unfavorable or impossible, companies will turn to issuing equity for financing as a last resort. Issuance costs are assumed to increase along this hierarchy, further supporting management's assumed preference for internal financing over debt financing, which in turn is preferred to equity financing.

2.3 The Market Timing Theory of Capital Structure

The Market Timing Theory goes back to the work of Baker and Wurgler (2002), who find firms as generally agnostic to the use of debt and equity financing. According to the market timing theory, companies take their capital structure decisions opportunistically in terms of funding costs, based on relative pricing for each funding source in the capital markets. This theory suggests that a company is more likely to use debt financing in periods of relatively high investor demand for corporate debt and relatively low investor demand for corporate equity. Similarly, companies tend to finance their activities through the issuance of corporate equity in times of relatively high demand for equity, exemplified by high price-to-book ratios for corporate equity. As a result, observable corporate leverage levels are the cumulative product of a company's historical views on favorable financing opportunities in debt and equity markets. The Market Timing Theory of Capital Structure takes the occurrence of mispricing in debt and equity markets as given. Further, the theory assigns corporate treasurers the ability to identify and exploit mispricing through opportunistic issuance activity before the mispricing is corrected in the financial markets. To date, there is only little empirical support for the market timing theory. Baker and Wurgler (2002) manage to relate firm leverage over several subsequent periods back to historical issuance patterns during relatively favorable and less favorable equity market phases. A study by A. Alti (2006) failed to establish a lasting effect of market timing issuance on corporate leverage levels.

2.4 Empirical Studies

The body of empirical studies attempting to support the theories is huge. However, we cannot decisively claim that one of these theories is more suitable than the others. According to empirical evidence so far, the main capital structure models appear incomprehensive and may provide contradicting estimations. Nevertheless, certain variables appear to explain variation in a more reliable manner. Among them are asset tangibility, profitability and industry median levels of leverage (Murray and V. Goyal, 2007) (Lemmon et al., 2008).

Fama and French (2002) point out shortcomings in explanatory power for The Trade-off Theory of Capital Structure. Correspondingly, Graham and Harvey (2001) and Mugoša (2015) find mixed support for The Trade-off Theory of Capital Structure. However, Drobetz et al. (2013) conclude that there exist target leverage ratios and relatively high adjustment speed against these target ratios.

Empirical studies of The Pecking Order Theory have delivered ambiguous results, and there is only slight support for a dominating role of a pecking order. Helwege and Liang (1996) estimate a *logit* model to study the Pecking Order Theory in the context of a set of US firms, which had recently been listed on the stock market. They hypothesize that newly listed firms are more likely to require external funding to finance growth than more mature firms. Their core objective is to determine factors which affect the *probability* to incur external debt. The dependent variable is thus binary. They find some limited support for the Pecking Order Theory. While Frank and V. K. Goyal (2003) reject The Pecking Order Theory for a sample of small companies where information asymmetry should be more prevalent than with large companies, Shyam-Sunder and Myers (1999) find a higher explanatory power in The Pecking Order Theory when compared to the static The Tradeoff Theory of Capital Structure. In a study of privately held companies in Brazil, Zeidan et al. (2018) propose that owners of these companies follow the financing hierarchy proposed by The Pecking Order Theory. Equal to The Trade-off Theory of Capital Structure, Fama and French (2002) point out shortcomings in explanatory power for The Pecking Order Theory.

There is marginal empirical support for The Market Timing Theory of Capital Structure. Baker and Wurgler (2002) manage to relate company leverage over several subsequent periods back to historical issuance patterns during relatively favorable and less favorable equity market phases. A study by Alti (2006) failed to establish a lasting effect of market timing issuance on corporate leverage levels.

We have only found one study which specifically examine debt utilization decisions of *SME* shipping firms. Kotcharin and Maneenop (2018) study capital structure decisions in 71 non-listed SME shipping companies, employed in sea and coastal freight water transport in Thailand. Broadly in line with our findings, they discover that tangibility, size, and growth are positively associated with capital structure whereas profitability is negatively related. They further find that the impact of being in the shipping industry is greater than being SMEs, either family owned or not.

Several studies however examine capital structure decisions of larger listed shipping firms. Arvanitis et al. (2012) identify factors that affect the capital structure of European oceanic shipping firms and search for the existence of an ideal-target capital structure ratio. They use static fixed effect and dynamic (GMM) econometric models, employing data from the financial statements of 32 *listed* European shipping companies for the period 2005-2010. Their results lend support to the pecking order theory and identify a positive relationship between tangible assets and tax benefits (which we do not consider) against leverage. Generally, they observe a negative relationship between size or profitability and debt. Their results are broadly in line with our findings for the pecking order theory.

Paun and Topan (2016) study capital structure decisions of shipping companies listed on four major stock exhanges, covering 246 firm-year observations for the crises period 2009 to 2011. Dependent variables in their study representing debt utilization are book value of total debt and book value of total liabilities. They are only able find support for the trade-off theory, and confirming our results they discover that tangibility, profitability, size are important variables explaining debt utilization decisions of shipping firms.

Kotcharin and Maneenop (2020) study the impact of geopolitical risk (GPR) on 118 listed shipping firms' capital structure decisions. Their main explanatory variable of interest is a GPR index, but they also employ macroeconomic and firm specific variables and like us they discover that tangibility and profitability (among others) are significant explanatory variables of debt utilization.

Tsatsaronis (2018) study capital structure determinants for 50 listed shipping companies during the period 2006 to 2016, covering sub-periods of high and low risk, using a dynamic panel data approach. In line with our results, he discovers that tangibility is positively related to the use of debt in all market environments, whereas profitability can be both positively and negatively related to debt depending on the risk environment.

3 Data Set

This section describes the sample of shipping SMEs, the definition of the variables, and provides an analysis of the dataset.

3.1 Sample of Shipping SMEs

We collected and organized financial data from the company database of Bureau van Dijk, Orbis. The sample consists of annual financial data for 161 Norwegian shipping SMEs. The time horizon analyzed spans the period from 2008 to 2019, resulting in 1,224 company-year observations for financial leverage. Hence, the time period includes the latest Global Financial Crisis (GFC), the subsequent sovereign debt crisis and the oil price crisis in 2014.

The sample of this study concerns companies with sector code ”Shipping within transportation”, excluding shipyards, passenger water transport and inland freight water transport. Further, companies are required to engage in sea and coastal freight water transport, and to be classified as small- or medium-sized. Thus, the companies have operating revenue below 10 million EUR, total assets below 20 million EUR, and less than 150 employees. Furthermore, exclusive holding companies with no employees are also excluded from the sample. In addition, more than one observed year of financial data is required per company, otherwise the company is dropped from the sample.

The 161 Norwegian shipping SMEs provide transportation services globally, and other offshore shipping services mainly in the North Sea.

3.2 Definition of Variables

In this study, the panel data analysis uses the companies' leverage ratios, LEVR, as the dependent variable. Leverage ratio is defined as the ratio of non-current liabilities to total assets. The independent variables can be grouped further in company level and macroeconomic variables. Company level variables are asset tangibility, profitability, operating leverage, company size and company age. Oil price can be classified as a macroeconomic variable.

A company's asset tangibility, TANG, is defined as the ratio of fixed assets to total assets, while CH TANG is the change in tangible assets over a year. The company's profitability, PROF, is the ratio of operating revenue to total assets. The operating leverage, OPLEV, is the ratio of operating expenses to total assets. The company size, SIZE, is defined as the logarithm of total assets, and a company's age, AGE, is the time span from its inception to the respective date of observed leverage, measured in years and based on 365 days per year. Finally, OIL is the change in the average yearly price of Brent Crude Oil from one year to the next.

3.3 Data Analysis

Table 1 provides descriptive statistics of the variables employed in the panel regressions of this study. The mean leverage ratio, LEVR, is approximately 44%. Hence, the mean leverage ratio of our dataset is lower than the approximated leverage ratio in the shipping market of 55%–65% according to the ABN AMRO Report of 2011 (Gorgels, 2011). Further, the distribution of LEVR and the independent variables are illustrated in section B.1 in the appendix, which may be assessed by clicking the following link: https://www.ntnu.no/documents/1265701259/1281473463/Appendices+to+Examining+classic+capital+structure+models+of+debt+utilization+in+Norwegian+shipping+SME.%28001%29.pdf/73f34151-0a5b-dd77-4055-aaf02dbba64c?t=1655705644113

Variables	Mean	Median	Max	Min	Std.Dev.	Skewness	Kurtosis
LEVR	0.4372	0.3700	5.3400	0.0100	0.4366	33.7307	4.2128
TANG	0.5760	0.6200	1.0000	0.0100	0.2813	-0.3567	-1.1407
CH TANG	0.1329	0.0000	26.6666	- 0.9687	1.4533	13.7902	218.3042
PROF	0.0337	0.0400	0.9800	- 0.9300	0.1824	-0.7227	5.1996
OPLEV	1.1604	0.8400	0.0100	9.2000	1.1636	2.1405	7.1446
SIZE	10.1539	9.955	15.5300	7.4000	1.2300	0.7278	0.9856
AGE	12.7003	12.7003	24.8600	0.0600	6.9265	-0.0977	-1.1938
OIL	0.0183	-0.0100	0.3600	- 0.4700	0.2591	-0.9260	-0.3792

Table 1: Analysis of the dataset variables from the company database of Bureau van Dijk, Orbis. The sample consists of annual financial data for 161 Norwegian shipping SMEs. The time horizon analyzed spans the period from 2008 to 2019, resulting in 1,224 company-year observations for financial leverage. LEVR: Leverage, TANG: Tangibility, CH TANG: Change Tangibility, PROF: Profitability, OPLEV: Operating Leverage, SIZE: Total assets of the company, Age: Age of the company, OIL: Oil price (relative changes).

Table 2 provides a correlation matrix of the dataset variables. The variable with the highest correlation with leverage is TANG, equal to 0.34. This relationship is illustrated in figure 17 in section B.2 in the appendix. Additionally, there is a moderate negative correlation between TANG and OPLEV, and SIZE and OPLEV, illustrated in Section B.2 in the appendix. The correlation between the remaining variables is low if we assume low correlation corresponds to correlation values lower than ± 0.30.

	LEVR	TANG	CH TANG	PROF	OPLEV	SIZE	AGE	OIL
LEVR	1.00							
TANG	0.34	1.00						
CH TANG	-0.06	0.01	1.00					
PROF	-0.15	-0.09	-0.01	1.00				
OPLEV	-0.22	-0.44	0.00	0.01	1.00			
SIZE	0.22	0.29	0.00	-0.11	-0.43	1.00		
AGE	-0.12	-0.14	-0.02	-0.08	-0.01	0.04	1.00	
OIL	-0.01	-0.01	-0.04	-0.03	0.01	-0.02	-0.01	1.00

Table 2: Correlation matrix of the variables used from the company database of Bureau van Dijk, Orbis. The sample consists of annual financial data for 161 Norwegian shipping SMEs. The time horizon analyzed spans the period from 2008 to 2019, resulting in 1,224 company-year observations for financial leverage. LEVR: Leverage, TANG: Tangibility, CH TANG: Change Tangibility, PROF: Profitability,

OPLEV: Operating Leverage, SIZE: Total assets of the company, Age: Age of the company, OIL: Oil price (relative changes).

3.3.1 Machine Learning Data Set

The machine learning data set is split into a training set and a test set, and the purpose of the split is described in section 4.3. In this study, 70% of the dataset constitutes the training set, and the remaining 30% constitutes the test set. Table 3 provides the mean and median of the training- and test sets and can be used to control if the two sets follow the same pattern. By comparison of the means and medians, we conclude that the two sets follow the same patterns due to minor offsets. The OIL variable proved to have a low impact on leverage in an early phase of the construction of the machine learning models. Hence, the variable is excluded from the machine learning dataset.

Variables	$\text{Mean}_{training}$	Mean_{test}	$\text{Median}_{training}$	Median_{test}
LEVR	0.44	0.44	0.35	0.38
TANG	0.57	0.57	0.62	0.63
CH TANG	0.13	0.13	0.00	0.00
PROF	0.03	0.03	0.04	0.03
OPLEV	1.19	1.08	0.86	0.80
SIZE	10.13	10.20	9.94	10.09
AGE	12.62	12.88	13.51	13.90

Table 3: Display mean and median of the training- and test sets. In this study, 70% of the dataset constitutes the training set, and the remaining 30% constitutes the test set. Data used from the company database of Bureau van Dijk, Orbis. The sample consists of annual financial data for 161 Norwegian shipping SMEs. The time horizon analyzed spans the period from 2008 to 2019, resulting in 1,224 company-year observations for financial leverage. LEVR: Leverage, TANG: Tangibility, CH TANG: Change Tangibility, PROF: Profitability, OPLEV: Operating Leverage, SIZE: Total assets of the company, Age: Age of the company, OIL: Oil price (relative changes).

4 Method

In this study, linear-, logistic- and machine learning regression models are applied to identify variables with notable impact on companies' debt ratio. Additionally, the regression output forms the basis for interpreting the capital structure choice theories. This section explains the more advanced regression models. The standard models are explained in appendix 1.

4.1 Logistic Regression Model (XTLogit)

It is interesting to examine whether a non-linear logit model, which is estimated using maximum likelihood techniques, evaluate the data differently than the linear panel regression (OLS) models. The Logistic Regression (logit) model can be applied when the dependent variable is binary, equal to zero or one (Gujarati, 2009). As mentioned in section 2.4, Helwege and Liang (1996) used a logit model to determine the factors which affect a firm's probability of raising external financing. Obviously, leverage is not binary, and mostly varies between zero and one. Some companies even have negative equity with leverage exceeding one. Thus, we construct a binary variable associated with the *choice* of employing more debt. We define a binary variable y_i (*BIN_LEVERAGE*$_i$) equal to 1 if a company increases leverage and 0 otherwise. The binary variable y_i should be viewed as the realization of a random variable, Y_i, which assumes the values 1 or 0 with probabilities π_i and $1-\pi_i$ respectively. The random variable is following a Bernoulli distribution with parameter π_i (Bertsekas and Tsitsiklis, 2008). For $y_i = 0$ or $y_i = 1$, the distribution of Y_i can be written:

$$P\{Y_i = y_i\} = \pi_i^{y_i}(1-\pi_i)^{(1-y_i)} \tag{1}$$

If $y_i = 1$ we obtain π_i, and if $y_i = 0$ we obtain $1-\pi_i$. Suppose we study all the companies in our sample and view the event of an increase in leverage as independent among companies and with the same probability π for each company. To the extent that companies make decisions independently, the independence assumption seems reasonable. However, in the context of the market timing theory, the independence assumption might be violated. Still this assumption is required by the logit model. Also, assuming the same probability for each company to increase the use of leverage is an analytical simplification which may, or may not, be reasonable.

We let Y represent the number of successes (increase in leverage) in n independent trials. Then, Y is a binomial distributed variable with parameters (n, π). The probability mass function of a binomial random variable with parameters (n, π) is

$$P\{Y_i = i\} = \binom{n}{i}\pi^i(1-\pi)^{(n-i)} \tag{2}$$

For $i = 0, 1, \dots, n$.

The mean of this function is π, and the variance is $n\pi(1-\pi)$. The mean and the variance of this variable depend on the underlying probability π. Hence, any factor which affects the probability π will alter the variance of this variable. Accordingly, a linear regression model that assumes constant variance is not suitable for analyzing binary data.

In a suitable model for the structure of our data, the probabilities of an increase in leverage, π, should depend on observed, explanatory variables, *x*. In the linear model $\pi = x^T\beta$, the term on the right-hand side may take any real number, while the probability on the left-hand side must lie between zero and one. The Logistic Regression we have employed in this study is given in equation 3 (Brooks, 2014).

$$\pi_i = \frac{1}{1+e^{x^T\beta}} = \frac{1}{1+e^{-(\beta_1+\beta_2 x_{2i}+\cdots+\beta_k x_{ki}+e_{i,t})}} \tag{3}$$

for $i = 1, \dots, n$, where π_i is the probability that y_i equals 1. Obviously, this model is non-linear both in the x's and the β's. In our notation the model is written:

$$\pi_i = \frac{1}{1+e^{-(\beta_1+\beta_2 CH_{TANG}+\beta_3 TANG+\beta_4 PROF+\beta_5 OPLEV+\beta_6 SIZE+\beta_7 AGE+\beta_8 OIL+e_{i,t})}} \tag{4}$$

This model can compute the effect on the *probability* that a company will incur more debt in response to changes in the explanatory variables. It is therefore well suited for examining validity of the three capital structure models.

4.2 Machine Learning Regression Models

In the capital structure literature, there is no consensus on which statistical model estimates fits best and more accurately describe capital structure choices. The most common statistical models used are linear regression modes, panel data methods and methods for discrete choice (logit and probit models) or combinations of these. These models have the advantage of being relatively transparent and thus understandable for most researchers and practitioners. However, these statistical models are bound to certain assumptions (e.g., linearity and a limited number of variables) that might restrict their prediction accuracy. Machine learning methods have yielded promising results in numerous fields, including the related field of bankruptcy prediction. New types of data (with more variables) require new analytical approaches. Such techniques have already evolved in fields with a long tradition in crunching big data (e.g., default data for private and corporate bank customers). The objective of the present paper is to apply ML-techniques and assess if and how such models can confirm results form more classical models, and possibly propose new variables. ML covers random forest, neural networks, support vector machines, and boosting.

We will here review the fundamentals and more details of random forest and boosting techniques that we apply is this study. As pointed out in the introduction, the rationale for employing these models is to determine whether they agree with the linear models on significant explanatory variables and to find out if ML models are better equipped to distinguish between the three capital choice theories. We present the machine learning algorithms *Random Forest*, *Gradient Boosting* and *Extreme Gradient Boosting*. These three algorithms are based on the same method; they construct a series of decision trees in order to determine the target variable.

To explain the fundamentals of machine learning regression models, sub-section ”Feature Engineering” describes how the features of the dataset are prepared. Sub-section ”Training-Validation- and Test set” explains how and why the dataset is split into subsets, and how these subsets can be used to improve the performance of machine learning models. Finally, sub-section ”Model Evaluation”, explains how to evaluate the models.

Feature Engineering

Features are characteristics, properties, and attributes of raw data. A feature can be explained as numeric representation of data. (Zheng and Casari, 2018). Feature engineering is the process of using domain knowledge to extract features from raw data, given the employed model and the modelling purpose. In this study, we apply tree-based machine learning algorithms. As tree-based models are not sensitive to the scale of the input features, the features of our dataset will not be transformed into normalized features. In our panel dataset, all rows consisting of missing feature values are excluded. Further, rows consisting of extreme outliers are excluded.

Training-, Validation- and Test set

For large datasets, the data can be split into a training-, a validation and a test set. The training set is used for training the machine learning model, the validation set is used to evaluate the model while training, and the test set is used to evaluate a trained model (Friedman et al., 2018). If the amount of data is limited, the data set can be split into a training- and test set. To replace the need of an independent validation set, we use can use the method of Cross Validation.

Cross Validation can be applied to evaluate the model during the training phase. In this method, a fraction of the training data is kept out and used to estimate a test error (Friedman et al., 2018). In this study, we use *K-fold Cross Validation*. When employing K-fold Cross Validation, the dataset is randomly split into k non-overlapping groups of equal size *n*. In our analysis, k = 10. By iterating over the dataset k times, each group will be used as a validation set once, while the other k − 1 groups constitute the training set. For each iteration, the test error of the validation set is computed as a Mean Square Error, MSE. Finally, the K-fold Cross Validation estimate is computed by averaging the MSE values of all iterations, given by equation 5.

$$CV_k = \frac{1}{k}\sum_{j=1}^{k} MSE_j \tag{5}$$

The MSE_i of equation 5 is computed by equation 6:

$$MSE_j = \frac{1}{n}\sum_{i=1+(j-1)n}^{jn} \left(y_i - \hat{f}(x_i)\right)^2 \tag{6}$$

In equation 6, the size of the training set equals $(k-1)n$ and the size of the validation set equals n for each iteration. The target variable which the model aims to predict is y_i and $\hat{f}(x_i)$ denotes the predicted variable of the i-th observation. The purpose of the K-fold Cross Validation estimate is to identify the method providing the lowest CV_k to tune the hyperparameters. (James et al., 2013)

In this context, hyperparameters refer to the configuration of a machine learning model. Hyperparameter tuning can improve the performance of the model, because different hyperparameter configurations suit different datasets (Hutter et al., 2019). In this study, we have applied a grid search to tune the hyperparameters. In a grid search, a finite set of hyperparameters will be explored. We apply the hyperparameters resulting in the most suitable model evaluated by the components.

Model Evaluation

In general, a model's performance can be determined by examining how well a model performs on an independent test set (Friedman et al., 2018). Model evaluation is used for two main purposes, model selection and model assessment. Model selection refers to the process of evaluating different models in order to choose the best suited model. Model assessment refers to estimation of the chosen model's prediction error on an independent dataset.

In order to measure the error rate between the target variable and the predicted variable, we apply a loss function, $L\big(y, \hat{f}(x)\big)$. The Squared Error (SE) or Absolute Error (AE) are common measures in the computation of loss functions. The loss function is given by equation 7, where y denotes the target variable and $\hat{f}(x)$ denotes the predicted target variable. The predicted variable is computed by applying the model *f* to a dataset where *x* denotes the features.

$$L\big(y, \hat{f}(x)\big) = \begin{cases} \big(y - \hat{f}(x)\big)^2 & Squared\ Error, SE \\ \big|y - \hat{f}(x)\big| & Absolute\ Error, AE \end{cases} \tag{7}$$

When evaluating the model, the loss function is applied to both the training set and the test set taking the squared error and the absolute error respectively as arguments. The training set error, ERR_z, is given by equation 8 and provides the average loss over the specific training set. The training set is denoted by z, the number of observations of the training set is given by N, and *i* refers to the *i*-th observation.

$$ERR_z = \frac{1}{N}\sum_{i=1}^{N} L\big(y_{zi}, \hat{f}(x_{zi})\big) \tag{8}$$

The test set error given by equation 9 measures the average model loss over an independent test dataset. The test set is denoted by τ, the number of observations of the test set is given by *N*, and *i* refers to the *i*-th observation.

$$ERR_\tau = \frac{1}{N}\sum_{i=1}^{N} L\big(y_{\tau i}, \hat{f}(x_{\tau i})\big) \tag{9}$$

The expected test set error, $E[Err_\tau]$, equals the training set error and is given by equation 10. However, the test set error will typically be greater than the training set error as the model is fitted to the training set.

$$E[ERR_\tau] = ERR_z \tag{10}$$

An additional measure to evaluate the model is the model accuracy. In this paper, R^2 is used as a measure for the model accuracy (scikit-learn, 2021b). The model accuracy is given by equation 11. The target variable is given by y, the mean target variable is given by $\bar{y}$ and the target variable predicted by the model is given by $\hat{f}(x)$. Further *x* denotes the features of the model, *N* is the total number of observations and *i* refers to the *i*-th variable in the range from 1...*N*.

$$\begin{gathered} R^2 = 1 - \frac{u}{v} \\ v = \sum\nolimits_{i=1}^{N} \big(y_i - \hat{f}(x_i)\big)^2 \\ u = \sum\nolimits_{i=1}^{N} (y_i - \bar{y})^2 \end{gathered} \tag{11}$$

The model accuracy can be computed for both the training set and the test set. The model accuracy of the training set is R_z^2, and the model accuracy of the test set is R_τ^2. The expected model accuracy of the test set equals the model accuracy of the training set. However, the accuracy of the training set is typically higher for the training set, because the model can adapt to its specific patterns while training.

When the complexity of a model increases, the model increases its adaptability to the underlying structures of the training set. (Friedman et al., 2018) However, if the model complexity increases too much, the training set error decreases while the test set error increases, which results in an overfitted model. An overfitted model follows the errors of the training set too closely and cannot be generalized. A symptom of an overfitted model, is when the model performs well on the training set, but not on the test set. (Amazon Web Services, 2021) In contrast, an underfitted model is a model which lacks complexity. A symptom of an underfitted model is when the model performs poorly on the training set.

4.2.1 Random Forest (RF)

A Random Forest consists of a large number of independent individual decision trees, operating as an ensemble. Each individual tree in the forest produces a class predictor characterizing a *feature* of the tree. The class predictor that appears in most of the trees gets the most votes and becomes the model's class predictor. No correlation between individual trees is a key property of this model, since this assures that individual trees protect against individual errors, allowing the group of trees to move in the right direction. Hence, the Random Forest algorithm builds a large collection of de-correlated decision trees and averages them (Friedman et al., 2018). In order to grow new trees in the collection, random vectors are generated (Breiman, 2001). A new tree is generated so that the *k*th tree consists of a random vector which is independent of the previous trees or vectors, but follows the same distribution. Hence, the result is a collection of many unbiased models with a high level of noise (Friedman et al., 2018). Since each tree has a high degree of noise, there is great benefit from averaging them, since this reduces the variance.

Assuming we have B trees where each tree has a variance of σ^2, then the average variance is $\frac{\sigma^2}{B}$. If the trees follow the same distribution, but are not completely independent, their correlation coefficient is given by ρ, and the average variance is computed by equation 12. An increasing number of trees causes the second term to approach zero. Therefore, the correlation coefficient from the first term limits the benefit of averaging. Consequently, we aim at generating de-correlated trees to limit the variance.

$$\rho\sigma^2 + \frac{1-\rho}{B}\sigma^2 \tag{12}$$

The datasets used by the Random Forest algorithm are generated using a *bootstrapping* technique, where random subsets of the same dimension as the original dataset are selected from the training data. This procedure is referred to as *bagging*. To grow new trees, we randomly select $m \leq p$ features from the bagged dataset, where p represents all the features. A reduction of m leads to reduced correlation between any pair of trees, and hence reduced variance. Among the m features, we choose a variable to be split into two daughter nodes. This process is repeated until the minimum node size is reached. T_b refers to a tree b among the total ensemble of B trees. The ensemble of all trees is given by $\{T(x;\theta_b)\}_1^B$. A *prediction* at a point x is given by equation 13, where θ_b provides the characteristics of the bth tree.

$$\hat{f}_{rf}^B(x) = \frac{1}{B}\sum_{b=1}^{B} T(x,\theta_b) \tag{13}$$

4.2.2 Gradient Boosting (GB)

Gradient Boosting combines many weak learners to improve the predicting performance of a model. The goal is to teach a model $\hat{f}$ to predict values of the form $y = \hat{f}(x)$, for instance by minimizing the mean squared error, $\frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{f}(x_i)\right)^2$. A weak learner is a learning algorithm which performs slightly better than random guessing. Despite low performance of one weak learner, it is possible to combine many weak learners to produce a model with high accuracy.

Gradient Boosting is based on an additive model. (Scikit-learn, 2021a) The computation of an estimate of the output variable y_i is given by equation 14. Assuming the Gradient Boosting algorithm has M stages, some imperfect model $\hat{f}_M$ is improved at each stage of the boosting algorithm. The model aims to predict y_i by summing the M output of the weak learners, h_m, taking the features, x_i as an argument.

$$\hat{f}_M(x_i) = \sum_{m=1}^{M} h_m(x_i) \tag{14}$$

Gradient Boosting produces a sequence of weak learners and will at each step choose the optimal h_m to minimize a loss function given the previous estimation of $\hat{f}_m(x_i)$.

$$\hat{f}_m(x_i) = \hat{f}_{m-1}(x_i) + h_m(x_i) \tag{15}$$

Equation 16 provides the computation of a new weak learner, h_m. The new weak learner is computed by minimizing the loss function, L_m.

$$h_m = \underset{h}{\operatorname{argmin}}\, L_m = \underset{h}{\operatorname{argmin}} \sum_{i=1}^{n} l\left(y_i, \hat{f}_{m-1}(x_i) + h(x_i)\right) \tag{16}$$

Equation 17 provides the loss function of the Gradient Boosting algorithm, which is computed using a first-order Taylor approximation. A Taylor approximation can be applied to approximate a function f at a point a, given that the function is differentiable at point a, and a is a real or complex number. The first-order approximation of a Taylor function is given by $l(d) \approx l(a) + (d-a)\frac{\partial l(a)}{\partial(a)}$

where $d = \hat{f}_{m-1}(x_i) + h_m(x_i)$ and $a = \hat{f}_{m-1}(x_i)$. In our notation the Taylor approximation becomes:

$$l\left(y_i, \hat{f}_{m-1}(x_i) + h_m(x_i)\right) \approx l\left(y_i, \hat{f}_{m-1}(x_i)\right) + h_m(x_i)\left[\frac{\partial l\left(y_i, \hat{f}(x_i)\right)}{\partial \hat{f}(x_i)}\right]_{\hat{f}=\hat{f}_{m-1}} \tag{17}$$

The final term in equation 17, $\left[\frac{\partial l\left(y_i, \hat{f}(x_i)\right)}{\partial \hat{f}(x_i)}\right]_{\hat{f}=\hat{f}_{m-1}}$, corresponds to the derivative of the loss function with respect to the second parameter at point $\hat{f}_{m-1}(x_i)$. Because the loss function is differentiable at point $\hat{f}_{m-1}(x_i)$, the final term of equation 17 can easily be computed in a closed form, denoted by g_i. By removing the constant terms, we can rewrite the function of a weak learner from equation 16 in equation 18.

$$h_m \approx \underset{h}{\operatorname{argmin}} \sum_{i=1}^{n} h(x_i) g_i \tag{18}$$

By fitting $h(x_i)$ to predict a value proportional to $-g_i$, the expression in equation 18 will be minimized. When initiating the Gradient Boosting model, $\hat{f}_0$ is chosen as the constant minimizing the loss function. When the loss function is chosen as least-squares, the constant $\hat{f}_0$ equals the mean of y_i.

4.2.3 Extreme Gradient Boosting (XGB)

Like Gradient Boosting, Extreme Gradient Boosting is a machine learning system for tree boosting (Chen and Guestrin, 2016). Extreme Gradient Boosting is known as a scalable algorithm, which runs faster than most common machine learning algorithms whose memory is limited.

XGB uses an additive model to predict the output variable, provided by equation 19. M denotes the number of additive functions to predict y_i. Further, h_m corresponds to a weak learner expressed as a decision tree.

$$\hat{f}(x_i) = \sum_{m=1}^{M} h_m(x_i) \tag{19}$$

Equation 20 provides the loss function, L, where y_i is the target variable, $\hat{f}(x_i)$ is the predicted output variable, and l is a loss function computing the difference between the predicted variable and the target variable. Further, Ω is a regularization term penalizing model complexity. The model takes advantage of the regularization term, as it smooths the computed weights and avoids overfitting.

$$\mathcal{L} = \sum_i l\left(y_i, \hat{f}(x_i)\right) + \sum_m \Omega(h_m) \tag{20}$$

To optimize the loss function in equation 21, we apply the loss function in an additive manner by adding the h_m that improves our model at iteration m.

$$\mathcal{L}_m = \sum_{i=1}^{n} l\left(y_i, \hat{f}_{m-1}(x_i) + h_m(x_i)\right) + \Omega(h_m) \tag{21}$$

We apply a second-order Taylor approximation to provide an approximation of the loss function. The second-order Taylor approximation of the loss function is given by equation 22.

$$\mathcal{L}_m \cong \sum_{i=1}^{n} [l\left(y_i, \hat{f}_{m-1}(x_i) + d_i h_m(x_i)\right) + \frac{1}{2}\alpha_i h_m^2(x_i)] + \Omega(h_m) \tag{22}$$

where $d_i = \partial_{\hat{f}_{m-1(x_i)}} l\left(y_i, \hat{f}_{m-1}(x_i)\right)$ and $\alpha_i = \partial^2_{m-1(x_i)} l\left(y_i, \hat{f}_{m-1}(x_i)\right)$. By removing the constant terms, we get equation 23.

$$\tilde{\mathcal{L}}_m \cong \sum_{i=1}^{n} [d_i h_m(x_i) + \frac{1}{2}\alpha_i h_m^2(x_i)] + \Omega(h_m) \tag{23}$$

The optimal scoring function of a tree structure q can be obtained by rewriting equation 23. The total number of leaves in a tree structure is J. As the score of the prediction model is computed by summing the corresponding weighted leaf nodes for each regression tree, we define I_j as the instance set of leaf node j. To obtain the optimal scoring function of a tree structure, we expand the final term $\Omega(h_m)$, where $\Omega(h_m) = \gamma J + \frac{1}{2}\alpha \sum_{j=1}^{J} w_j^2$ and w_j corresponds to the weight of a leaf node j. Finally, the scoring function measuring the quality of a tree structure is given by equation 24.

$$\tilde{\mathcal{L}}_m(q) = -\frac{1}{2}\sum_{j=1}^{J}\frac{\left(\sum_{i\in I_j} d_i\right)^2}{\sum_{i\in I_j}\alpha_i + \lambda} + \gamma J \tag{24}$$

5 Results

In this section we identify variables with notable impact on leverage and interpret our findings in the context of the three theories of capital structure choice reviewed in section 2.

5.1 Linear Regression Models

5.1.1 The Random Effects Model

Initially, we run the static panel regression with Random Effects of equation A.1, section A.1.1. The output is given in column RE in Table 6. Three of the explanatory variables, TANG, CH TANG, and SIZE, are statistically significant at the 5% level.

We run a Hausman test on the Random Effects panel regression model in order to evaluate if this model is a good fit for our data. The output of the Hausman test is given in table 9, section A.2 in the appendix. As the p-value is less than 5%, the null hypothesis of random effects must be rejected. Since the Hausman test leads to a rejection of the null hypothesis of random effects, we do not interpret the regression output from this model. A Fixed Effects model or a Pooled OLS, might more appropriately capture the dynamics of our dataset.

5.1.2 The Fixed Effects Model

We run the Fixed Effects model of equation A.4 in section A.1.2. The regression output is given in column FE in table 6. Three of the explanatory variables, TANG, SIZE, and AGE, are statistically significant at a 5% level.

To determine whether the Fixed Effects model is preferable to a Pooled OLS, we examine redundant fixed effects presented in section A.3 in the appendix. Considering the zero-probabilities in the last column of table 10, we will not accept the null hypothesis of redundant cross section effects, i.e., that the intercepts α_i in equation A.4 are uncorrelated with the explanatory variables x_i. We conclude that a Fixed Effects model better captures the characteristics of our dataset.

Accordingly, we interpret the regression output of the static Fixed Effects model in column FE of table 6. As noted in section A.1.2, we have removed the time-invariant company specific characteristics in the fixed effects specification to assess the net effects of changes to the explanatory variables on the dependent variable. These net effects apply to all companies in the dataset. Hence, we can assess the impact on the use of leverage in the SME shipping industry “as a whole” from changes to the explanatory variables. We only consider the statistically significant variables, TANG, SIZE, and AGE. The findings will be discussed with respect to the theories: The Trade-off Theory of Capital Structure, The Pecking Order Theory and The Market Timing Theory of Capital Structure. An overview of the findings is given in table 4.

Variables	**Trade-Off Exp.**	**Pecking Order Exp.**	**Market Timing Exp.**	**FE Verif.**
TANG	Positive	Positive	Negative	Positive
SIZE	Positive	Negative	Positive	Negative
AGE	Positive	Negative	–	Positive

Table 4: Dependent variable LEVR. Overview of significant variables TANG, SIZE, and AGE in the four different models; The Trade-off Theory of Capital Structure, The Pecking Order Theory, The Market Timing Theory of Capital Structure, and the fixed effect model. LEVR: Leverage, TANG: Tangibility, SIZE: Total assets of the company, Age: Age of a company.

The Trade-off Theory of Capital Structure

The Trade-off Theory of Capital Structure suggests that there is an optimal mix of equity and debt financing. The optimal mix is dependent on the costs and benefits of employing a marginal unit of either equity or debt financing.

The TANG variable, measuring the share of fixed assets to total assets, is positively related to the use of debt. The positive coefficient could imply that the higher the share of fixed assets, the lower the credit risk, because the fixed assets serve as collateral for the banks. Further, the lower the credit risk, the cheaper it will be to obtain debt. Consequently, a company with a higher share of fixed assets will use more debt to arrive at the optimal mix. Hence, the positive coefficient of the TANG variable lends support to The Trade-off Theory of Capital Structure.

The negative SIZE variable coefficient indicates that debt decreases when total assets increase. This may contradict The Trade-off Theory of Capital Structure as more total assets serve as collateral for debt and lowers credit risk. Hence, it would be cheaper to obtain debt, and according to The Trade-off Theory of Capital Structure, more debt could be beneficial to the company.

Finally, the positive coefficient of the AGE variable implies that a mature company is more likely to obtain more debt. This could imply that the company has proved to be resistant to market changes over time and may be associated with lower credit risk. Hence, it will be cheaper to obtain debt, which provides support to The Trade-off Theory of Capital Structure.

The Pecking Order Theory

The Pecking Order Theory suggests that companies will typically select the cheapest form of financing available. When the cheapest capital sources are exhausted, companies will utilize more expensive capital, i.e., capital that require payment of higher rates of return to investors. First, companies will utilize internal funds, and thereafter external funds if further financing is required.

The positive coefficient of the TANG variable indicates that a company with a high level of fixed assets to total assets is more likely to obtain more debt. As the SME shipping industry is capital intensive, an increase in tangibles may require larger investments. As we argued above, the positive coefficient on the TANG-variable could further suggest that the higher the share of fixed assets, the less risky investors perceive its balance sheet and the cheaper it will be to obtain debt. If we assume internal funds are exhausted when obtaining new tangibles, the positive coefficient could provide support to The Pecking Order Theory, as it suggests debt increases when the investment exceeds internal funds.

The negative coefficient of the SIZE variable indicates that a company with more total assets is less likely to obtain more debt. If we assume larger companies have larger internal funds, and are less dependent on external financing, The Pecking Order Theory can be supported as the theory suggests companies utilize internal funds rather than external financing. Also, larger companies may have less need to grow their balance sheet and may use more internal funds.

Finally, the positive coefficient of the AGE variable indicates that a mature company is more likely to obtain more debt. However, this might contradict The Pecking Order Theory if we assume mature companies have built up internal funds over time. According to The Pecking Order Theory, access to internal funds would lower the demand for debt financing.

The Market Timing Theory of Capital Structure

According to The Market Timing Theory of Capital Structure, when a company's profitability is low, and presumably market conditions are poor, investors tend to become more risk averse. Consequently, investor demand for equity decreases compared to the demand for investment grade credit bonds. Typically, accommodative monetary policy during economic recessions, pull interest rates (and eventually spreads) further down. Under these conditions, The Market Timing Theory of Capital Structure suggests that a company's capital structure is highly debt intensive.

The positive coefficient of the TANG variable indicates that an increase in the share of tangible assets to total assets implies an increase in debt. If we assume an economic recession forces companies to sell off tangibles to avoid bankruptcy, we would expect debt to decrease. However, this might contradict The Market Timing Theory of Capital Structure, as the theory suggests debt increases during recessions.

Further, the negative coefficient of the SIZE variable indicates an increase in total assets leads to a decrease in debt. If we assume an economic boom leads to higher profitability and an increase in total assets, The Market Timing Theory of Capital Structure can be supported by the negative SIZE coefficient, as it suggests debt decreases under an economic boom.

Finally, we cannot evaluate AGE with respect to The Market Timing Theory of Capital Structure, as the variable is consistent during various market conditions unless the company is closed down.

5.1.3 Dynamic Panel Regression

The Dynamic Panel Regression output is given in the DYN column of Table 6. Three of the explanatory variables, TANG, SIZE, and AGE, are statistically significant at a 5% level, and we note that the Dynamic Panel Regression and the Fixed Effects model agree on the significance and slope of these explanatory variables. Further, with a lag of one year the variables TANG, PROF, SIZE and LEVR become significant in the Dynamic Panel Regression model. As the Dynamic Panel Regression model is the only model of this study with lagged variables, we do not compare this model to the other models, nor interpret the results with respect to the capital structure choice theories. However, as the variables TANG, PROF, SIZE, and LEVR becomes significant with a lag of one year, we note that this study indicates that these variables may be relevant in the companies' assessment of the leverage ratio one year ahead.

5.1.4 Pooled OLS

The output from the Pooled OLS is given in the OLS column in table 6. Four explanatory variables: TANG, PROF, SIZE, and AGE, are statistically significant at a 5% level. The coefficients of the significant variables are different for SIZE and AGE in the Pooled OLS compared to the Fixed Effects regression. However, it is well known that a Pooled OLS can bias the slope estimate (Gujarati, 2009). As we concluded in section 5.1.2, the Fixed Effects model is preferable to a Pooled OLS on our dataset. Hence, we do not interpret the regression output of the Pooled OLS.

5.2 Logistic Regression Model

The output from the Logistic Regression model of equation 4 in section 4.1 is given in the XTlogit column of table 6. The dependent variable, y_i is binary, and in our case $y_i \in \{0,1\}$. Three explanatory variables, TANG, CH TANG, and PROF, are statistically significant at a 5% level. Further, a summary of the interpretation output with respect to the theories are given in Table 5.

Variables	Trade-Off Exp.	Pecking Order Exp.	Market Timing Exp.	XTLogit Verif.
TANG	Positive	Positive	Negative	Positive
CH TANG	Positive	Positive	Negative	Positive
PROF	Positive	Negative	Negative	Negative

Table 5: Expected and Verified Relationships of significant variables. Dependent variable LEVR. Overview of significant variables TANG, CH TANG, and PROF in the four different models; The Trade-off Theory of Capital Structure, The Pecking Order Theory, The Market Timing Theory of Capital Structure, and the XTLogit model. LEVR: Leverage, TANG: Tangibility, CH TANG: Change Tangibility, PROF: Profitability.

Taking table 5 at face value, we see that the regression output from logistic model offers full support for the Pecking order theory, i.e., agree with this theory on the sign of all three significant variables. It also offers some support for the Trade-Off theory, but is hard to interpret in favor of the Market Timing theory of capital structure.

	RE	**FE**	**OLS**	**DYN**	**XTLogit**
TANG	0.2908***	0.2498***	0.3996***	0.2694***	0.8139**
	(5.0610)	(3.8047)	(6.0662)	(4.0323)	(2.9600)

TANG(-1)				-0.1662* (-2.4556)	
CH _TANG	-0.0131*	-0.0101	-0.0212	-0.0035	0.1326*
	(-2.0769)	(-1.6247)	(-4.5289)	(-0.6150)	(2.2898)
CH_TANG(-1)				0.0034 (0.6358)	
PROF	-0.0878	0.0036	-0.2905*	-0.0666	-3.5943***
	(-1.5148)	(0.0611)	(-2.3377)	(-1.3196)	(-7.9248)
PROF(-1)				-0.1444** (-2.9257)	
OPLEV	-0.0059	0.0251	-0.0218	-0.0042	0.1209
	(-0.4255)	(1.6143)	(-1.5496)	(-0.2791)	(1.6957)
OPLEV(-1)				0.0206 (1.3679)	
SIZE	-0.0361*	-0.1188***	0.0387*	-0.1161***	0.0379
	(-2.4593)	(-6.6148)	(2.0874)	(-5.2083)	(0.6207)
SIZE(-1)				0.0870*** (3.6129)	
AGE	0.0006	0.0096**	-0.0066*	0.0282*	0.0115
	(0.2592)	(3.2450)	(-2.1684)	(2.4207)	(1.1338)
AGE(-1)				-0.0221 (-1.8618)	
OIL	-0.0323	-0.0121	-0.0280	-0.0355	0.0115
	(-0.9566)	(-0.3660)	(-0.9982)	(-1.2639)	(1.2216)
OIL(-1)				0.0019 (0.0712)	
LEVR(-1)				0.6910*** (24.200)	
c	0.6471***	1.3504***	-0.0644		-1.5512*
	(4.1575)	(7.4137)	(-0.3634)		(-2.2477)
Adj. R^2	0.1546	0.5664	0.1546	0.4379	
AIC					1324.2639
BIC					1364.0148

Table 6: Regression results for random effect model (RE), fixed effect model (FE), dynamic panel regression (DYN), and logit model (XTlogit). t-statistics in parenthesis (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$)

5.3 Machine Learning Regression Models

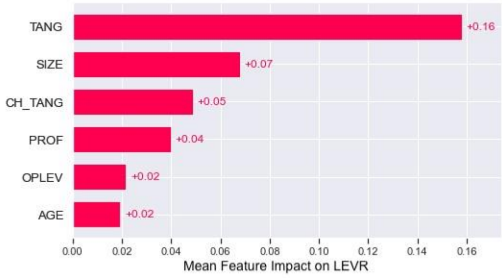
This section presents the results of each machine learning (ML) model. The features' relationship to leverage is presented in table 7. As all three models agree on the relationships to leverage, we will not distinguish between the three algorithms when we interpret ML results with respect to the capital structure choice theories in section 5.3.4. The main difference between the model results is the estimated impact of each feature on leverage. In the following, we will only interpret features with a mean impact on leverage greater than 0.1. Consequently, we will not consider the model results of CH TANG, PROF, and AGE in the Gradient Boosting model.

Variables	RF	GB	XGB
TANG	Positive	Positive	Positive
CH_TANG	Negative	-	Negative
PROF	Negative	-	Negative
OPLEV	Ambiguous	Ambiguous	Ambiguous
SIZE	Positive	Positive	Positive
AGE	Negative	-	Negative

Table 7: Expected and Verified Relationships of significant variables. Dependent variable LEVR. Overview of significant variables TANG, CH TANG, PROF, OPLEV, SIZE, and AGE in three four different machine learning models; Random forrest (RF), Gradient Boosting (GB), and X Gradient Boosting (XGB). LEVR: Leverage, TANG: Tangibility, CH TANG: Change Tangibility, PROF: Profitability, OPLEV: Operating Leverage, SIZE: Total assets of the company, Age: Age of the company.

5.3.1 Random Forest

Figures 1 and 2 provide the output of the Random Forest model. According to figure 1, TANG has the greatest mean impact on LEVR. Further, the features SIZE, CH TANG, PROF, OPLEV, and AGE provide a mean impact greater than 0.01, and hence, all features will be used for interpretation. Figure 2 illustrates how varying input feature values have different impact on predicted LEVR. Blue color indicates that the input feature value is low, while red color indicates high input feature value. According to the figure, the relationship between OPLEV and LEVR is ambiguous because low OPLEV leads to both positive and negative impact on LEVR. Further, high OPLEV leads to none or a slightly negative impact on LEVR.



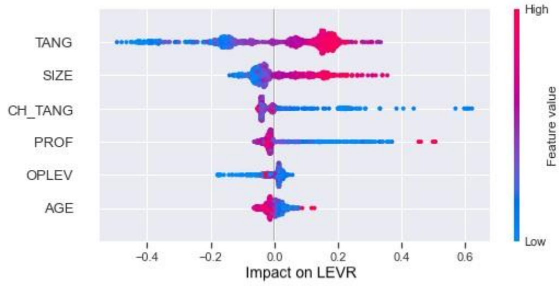
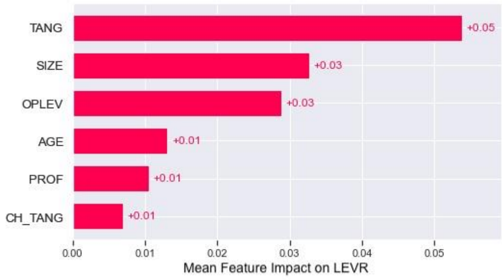


Figure 1: RF. Feature Importance.

Figure 2: RF - Feature Impact on LEVR

5.3.2 Gradient Boosting

Figures 3, and 4 provide the output of the Gradient Boosting Model. According to figure 3, TANG has the greatest mean impact on LEVR, followed by SIZE, and OPLEV. Further, AGE, PROF, and CH TANG provide a marginal impact on LEVR. Hence these features will not be used for interpretation.



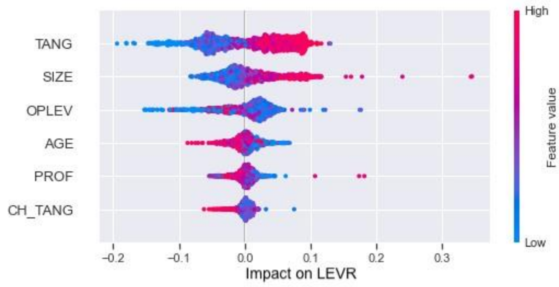
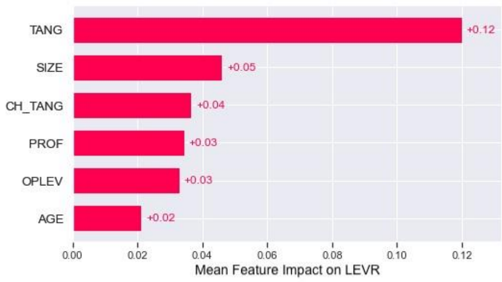


Figure 3: GB - Feature Importance

Figure 4: GB - Feature Impact on LEVR

5.3.3 Extreme Gradient Boosting

Figures 5 and 6 provide the output of the Gradient Boosting Model. According to figure 5, TANG has the greatest mean impact on LEVR. Thereafter follows SIZE, CH TANG, PROF, OPLEV, and AGE. According to figure 6, OPLEV has an ambiguous relationship to LEVR due to the same reasoning as for the Random Forest model.



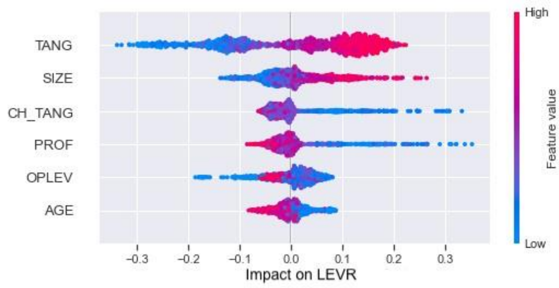


Figure 5: XGB - Feature Importance

Figure 6: XGB - Feature Impact on LEVR

5.3.4 Combined Machine Learning Results

As noted above, since the ML models suggest the same relationships between the features and leverage, we will interpret the ML results with respect to the three capital structure choice theories without distinguishing between the algorithms. While Random Forest and Extreme Gradient Boosting take advantage of all features in the determination of leverage, Gradient Boosting suggests CH TANG, PROF, and AGE have marginal impact on leverage. Hence, we conclude that Random Forest and Extreme Gradient Boosting are best suited for our dataset, and we will not interpret the results of Gradient Boosting. In the following when we refer to the ML models, we are only referring to Random Forest and Extreme Gradient Boosting (table 8). Furthermore, Random Forest and Extreme Gradient Boosting agree on the order of the features' mean impact on leverage.

Variables	Trade-Off Exp.	Pecking Order Exp.	Market Timing Exp.	ML Verif.
TANG	Positive	Positive	Negative	Positive
CH _TANG	Positive	Positive	Negative	Negative
PROF	Positive	Negative	Negative	Negative
OPLEV	Negative	Negative	Positive	Ambiguous
SIZE	Positive	Negative	Positive	Positive
AGE	Positive	Negative	-	Negative

Table 8: ML - Expected and Verified Relationships between Features and LEVR for four different models; The Trade-off Theory of Capital Structure, The Pecking Order Theory, The Market Timing Theory of Capital Structure, and the Machine Learning Model. LEVR: Leverage, TANG: Tangibility, CH TANG: Change Tangibility, PROF: Profitability, OPLEV: Operating Leverage, SIZE: Total assets of the company, Age: Age of the company.

The Trade-off Theory of Capital Structure

TANG is positively related to debt. The Trade-off Theory of Capital Structure can be supported by this variable due to the same reasoning as for the Fixed Effects model.

CH TANG is negatively related to debt, which indicates that companies with an increase in tangibility are less likely to obtain debt. This might not support The Trade-off Theory of Capital Structure, assuming lower debt financing costs are associated with more tangibles.

PROF is negatively related to debt. The Trade-off Theory of Capital Structure cannot be supported by the variable due to the same reasoning as for the Logistic Regression model.

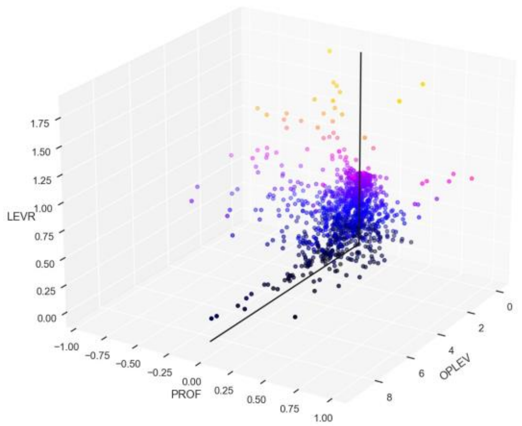


Figure 7: RF - Relationship of OPLEV, PROF and LEVR

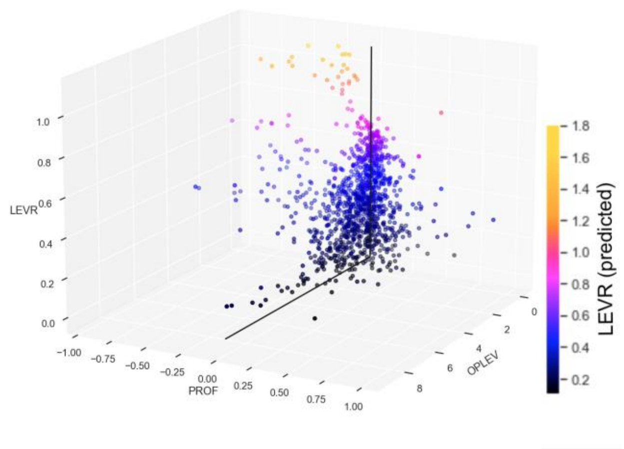


Figure 8: XGB - Relationship of OPLEV, PROF and LEVR

According to the ML regression output, OPLEV is not intuitively related to debt. Figures 7 and 8 better capture this relationship, with PROF as extra variable. When profitability and operating leverage are low, companies have higher probability of incurring debt. When operating leverage is low and profitability high, a company has a slightly lower probability of incurring more debt. This might contradict The Trade-off Theory of Capital Structure, assuming profitable companies can obtain lower debt financing costs. Further, the theory suggests a company requires more debt to arrive at the optimal mix. Further, when operating leverage increases, a company is less likely to obtain debt. High operating leverage can be associated with increased credit risk and debt financing costs. This might support the theory, as the theory suggests increased debt financing costs require less debt to arrive at the optimal mix.

SIZE is positively related to debt. We assume these companies are perceived as less risky with cheaper debt financing. The Trade-off Theory of Capital Structure can be supported as the theory suggests cheaper debt financing costs requires more debt to arrive at the optimal mix.

The AGE variable implies a mature company is more likely to obtain debt. We assume mature companies prove to be resistant to market changes and are associated with lower credit risk. Hence, it can be cheaper to obtain debt, which provides support to The Trade-off Theory of Capital Structure.

The Pecking Order Theory

TANG is positively related to debt. The Pecking Order Theory can be supported by this variable by the same reasoning as for the Fixed Effects model.

CH TANG is negatively related to debt. Assuming internal funds are exhausted when obtaining tangibles, the negative relationship contradicts the The Pecking Order Theory.

PROF is negatively related to debt. As profitable companies produce more internal funds, The Pecking Order Theory can be supported by the same reasoning as for the Logistic Regression model.

OPLEV is not intuitively correlated with debt. According to figures 7 and 8, low operating leverage and profitability supports the theory, as negative profitability reduces internal funds and may result in need for external funding. A company with low operating leverage and high profitability is slightly less likely to obtain debt. This might support the theory, as internal funds increase reducing the need of external financing. According to figures 7 and 8, when operating leverage increases, the profitability converges to zero and the probability of incurring debt is low. Hence, we cannot interpret the results with respect to the theory, as internal funds are not affected.

SIZE is positively related to debt. If we assume larger companies have larger internal funds and are less dependent on external financing, The Pecking Order Theory cannot be supported.

According to the AGE variable, a mature company is less likely to obtain more debt. This can provide support to The Pecking Order Theory if we assume mature companies have built up internal funds over time.

The Market Timing Theory of Capital Structure

TANG is positively related to debt. The Market Timing Theory of Capital Structure cannot be supported by the variable by the same reasoning as for the Fixed Effects model.

CH TANG is negatively related to debt. Assuming an economic recession forces companies to sell tangible assets, the variable provides support to The Market Timing Theory of Capital Structure.

PROF is negatively related to debt. The Market Timing Theory of Capital Structure can be supported by the variable due to the same reasoning as for the Logistic Regression model.

According to OPLEV in figures 7 and 8, a company with low operating leverage and high profitability is slightly more likely to reduce debt. If we assume operating leverage decrease and profitability increase during an economic boom, The Market Timing Theory of Capital Structure can be supported, as it suggests debt decrease during an economic boom. However, the theory cannot be supported by higher operating leverage, as the theory suggests debt increases during an economic recession. According to the results, debt converges to zero.

The SIZE variable implies that a company with a higher share of fixed assets to total assets is more likely to obtain debt. We assume an economic boom leads to higher profitability and an increase in total assets. The Market Timing Theory of Capital Structure cannot be supported as this theory suggests that debt decreases during an economic boom.

Finally, we cannot evaluate AGE with respect to The Market Timing Theory of Capital Structure, as the variable is consistent during various market conditions, unless the company is discontinued.

6 Concluding Remarks

This study aims at revealing underlying considerations in the choice of leverage in shipping SMEs. We have (i) identified the impact of explanatory capital structure variables on leverage ratios in a sample of 161 Norwegian shipping SMEs, and (ii) interpreted the findings with respect to the three major theories of capital structure choice; The Trade-off Theory of Capital Structure, The Pecking Order Theory and The Market Timing Theory of Capital Structure.

To this end, we have implemented and evaluated four linear regression models; *Random Effects*, *Fixed Effects*, *Dynamic Panel Regression*, and *Pooled OLS*, one logistic regression model, *Logistic Regression*, and three machine learning regression models; *Random Forest*, *Gradient Boosting*, and *Extreme Gradient Boosting*. Of these models, the most suitable models for our dataset are: Fixed Effects, Logistic Regression, Random Forest, and Extreme Gradient Boosting.

The principal findings of our study are:

Significant variables in explaining the level of debt utilization of Norwegian SME shipping firms are:

- Tangibility (positive in all models)
- Profitability (positive in logit- and ML models)
- Change in tangibility (positive in logit model, negative in all ML models)
- Size (positive in all ML models, negative in FE model)

- Age (positive in FE model, negative in all ML models)
- Operating leverage (ambiguous)

Other studies of debt utilization in various industries also point to these variables as significant explanatory variables of capital structure. As such, the Norwegian SME shipping industry does not stand out from the crowd.

Also, in line with findings in the studies cited in this paper, we find some support for all theories of debt utilization, but we are not able to point to one of the three capital structure models as superior to the others for Norwegian SME shipping firms.

Bibliography

Alti, Aydo˘gan (2006). How Persistent Is the Impact of Market Timing on Capital Structure? *The Journal of Finance*, Vol. LVII, No. 1.

Amazon Web Services, AWS (2021). *Model Fit: Underfitting vs. Overfitting*. url: https://docs.aws.amazon.com/machine-learning/latest/dg/model-fit-underfitting-vs-overfitting.html (visited on 21st Apr. 2021).

Arvanitis, S. H., Tzigkounaki, I. S., Stamatopoulos, T. V., Thalassinos, E. I. (2012). Dynamic Approach of Capital Structure of European Shipping Companies, *International Journal of Economic Sciences and Applied Research*, ISSN 1791-3373, Eastern Macedonia and Thrace Institute of Technology, Kavala, Vol. 5, Iss. 3, pp. 33-63

Baker, Malcolm and Jeffrey Wurgler (2002). 'Market Timing and Capital Structure'. *The Journal of Finance*, Vol. LXI, No. 4.

Bertsekas, Dimitri P. and John N. Tsitsiklis (2008). *Introduction to probability*. 2rd. Athena Scientific. Chap. 1,5.

Breiman, Leo (2001). 'Random Forests' In: Machine Learning, Kluwer Academic Publishers 45, pp. 5–32.

Brooks, Chris (2014). *Introductory Econometrics for Finance*. 3rd. Cambridge University Press. Chap. 5,7,11,12.

Chen, T., and Estrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Drobetz, Wolfgang et al. (2013). 'Capital structure decisions of globally listed shipping firms. *Transportation Research: Part E. Logistics and Transportation Review*, Vol. 52, 49-76.52, pp. 49–76.

EViews (2021). *Panel Equation Testing, Fixed Effects Testing*. url: http://www.eviews.com/help/helpintro.html#page/content/panel-Panel Equation Testing.html (visited on 21st Apr. 2021).

Fama, Eugene F. and Kenneth R. French (2002). 'Testing Trade-Off and Pecking Order Predictions about Dividends and Debt'. *Review of Financial Studies*, Vol 15, No. 1, 1-33.

Frank, Murray Z. and Vidhan K. Goyal (2003). 'Testing the pecking order theory of capital structure'. *Journal of Financial Economics,* Vol. 67, 217–248 I

Friedman, Jerome, Trevor Hastie and Robert Tibshirani (2018). *The Elements of Statistical Learning*. Vol. 697. Springer.

Gorgels, Joep (2011). 'Shipping Finance and Investment: Current trend in ship finance'. In: url: http://old.mareforum.com/ISTANBUL III PRESENTATIONS/GORGELS.pdf.

Graham, John and Campbell Harvey (2001). 'The Theory and Practice of Corporate Finance: Evidence from the Field'. *Journal of Financial Economics,* Vol. 60, 187-243.

Gujarati, D N. (2009). *Basic Econometrics*. Tata McGraw-Hill Education. Chap. 15.

Helwege, J and Liang N. (1996). Is there a Pecking Order? Evidence from a Panesl og IPO Firms, *Journal of Financial Economics*, 40, 429-58

Hutter, Frank, Lars Kotthoff and Joaquin Vanschoren (2019). *Automated Machine Learning*. Springer.

James, Gareth et al. (2013). *An introduction to statistical learning*. 1st. Springer.

Jensen, Michael C. (1986). 'Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers'. *American Economic Review*, Vol. 76 No. 2, 323–329.

Jensen, Michael C. and William H. Meckling (1976). 'Theory of the firm: Managerial behavior, agency costs and ownership structure'. *Journal of Financial Economics*, Vol 3, 305-360.

Kotcharin S. and Maneenop S. (2018). All in the Family: The Role of Related Shareholders in Capital Structure Decisions for the Shipping Industry. Working paper, Thammasat University, https://scholar.google.co.th/citations?view_op=list_works&hl=en&hl=en&user=9SEyR-8AAAAJ

Kotcharin S. and Maneenop S. (2020). Geopolitical risk and shipping firms' capital structure decisions in Belt and Road Initiative countries, *International Journal of Logistics Research and Applications*, 23:6, 544-560, DOI: 10.1080/13675567.2020.1766003 https://doi.org/10.1080/13675567.2020.1766003

Kraus, Alan and Robert H. Litzenberger (1973). 'A State-Preference Model of Optimal Financial Leverage'. *The Journal of Finance*, Vol. 28, 911–922.

Lemmon, Michael, Michael Roberts and Jaime Zender (2008). 'Back to the Beginning: Persistence and the Cross-Section of Corporate Capital Structure'. *The Journal of Finance*, Vol. 63, No. 4, 1575-1608.

Modigliani, Franco and Merton Miller (1963). 'Corporate income taxes and the cost of capital: a correction'. *American Economic Review*, Vol. 53, No. 3, 433–443.

Mugoša, Ana (2015). 'The determinants of capital structure choice: Evidence from Western Europe. *Business and Economic Horizons*, Vol. 11, No. 2, 76–95.

Murray F. Z., and Goyal, V. K. (2007) Capital Structure Decisions: Which Factors are Reliably Important?

SSRN: https://ssrn.com/abstract=567650 or http://dx.doi.org/10.2139/ssrn.567650

Myers, Stewart C. (1977). 'Determinants of corporate borrowing'. *Journal of Financial Economics,* Vol. 5, 147– 175.

— (1984). 'The Capital Structure Puzzle'. *The Journal of Finance*, Vol. 39, No. 3, Papers and Proceedings, Forty-Second Annual Meeting, American Finance Association, July, 575-592.

Myers, Stewart C. and Nicholas S. Majluf (1984). 'Corporate financing and investment decisions when firms have information that investors do not have'. *Journal of Financial Economics*, Vol. 13, No. 2, 187–221.

Paun C. and Topan V. (2016). Capital Structure in the Global Shipping Industry, *PANOECONOMICUS*, 2016, Vol. 63, Issue 3, pp. 359-384

scikit-learn (2021a). *Gradient Tree Boosting*. url: https://scikit-learn.org/stable/modules/ensemble. html#gradient-boosting(visited on 21st Apr. 2021).

— (2021b). *sklearn.ensemble. RandomForestRegressor, Score*. url:https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html#sklearn.ensemble.RandomForestRegressor. score (visited on 21st Apr. 2021).

Shyam-Sunder, Lakshmi and Stewart C. Myers (1999). 'Testing static tradeoff against pecking order models of capital structure'. *Journal of Financial Economics*, Vol. 51, No. 2, 219-244

Tsatsaronis M. (2018). Capital Structure Determinants According to Risk Periods in Shipping Companies, *International Conference on Business &amp*; Economics of the Hellenic Open University

Zeidan, Rodrigo, Koresh Galilc and Offer Moshe Shapir (2018). 'Do ultimate owners follow the pecking order theory?' *The Quarterly Review of Economics and Finance*, Vol. 67, Issue C, 45-50.

Zheng, Alice and Amanda Casari (2018). *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientist*. O'Reilly Media, Inc. ISBN: 9781491953242.