Financial Ratios and Prediction of Corporate Bankruptcy in the Atlantic Salmon Industry

Bård Misund*

University of Stavanger Business School

[Forthcoming in Aquaculture Economics & Management]

*Corresponding author: Bård Misund, Associate Professor, University of Stavanger Business School, N-4036 Stavanger, Norway, phone: +47 51 83 19 08, fax: 51 83 17 54, email: bard.misund@uis.no .

Acknowledgements

The author wish thank to Fiskeri- og havbruksnæringens forskningsfond (FHF) for funding the research, Nordea bank and Kontali analyse for providing data, and Jostein Lillestøl at the Norwegian School of Economics and Administration (NHH) for comments on an earlier version of this manuscript, Frank Asche and Ragnar Tveterås at the University of Stavanger for useful discussions, feedback and suggestions. Thanks are also extended to two anonymous reviewers for providing useful comments and suggestions that have improved this paper.

Financial Ratios and Prediction on Corporate Bankruptcy in the Atlantic Salmon Industry

Abstract

This paper addresses the issue of credit risk in the salmon industry. During the period 2000-2002 the Norwegian salmon industry witnessed a period of low prices leading to a wave of defaults and bankruptcies. The consequences were large monetary losses for both investors and banks, highlighting the importance of early detection of failing firms. Using financial ratios measuring the firms' financial status prior to this event, two credit risk models are developed; one using logit regression and the other Classification and Regression trees. The performance of the two models developed is compared to a cross-industry benchmark model developed by the Norwegian Central Bank. The models estimated on industry data is better at separating between companies with high and low credit risk in the salmon industry compared to the benchmark model.

Keywords: Atlantic salmon production, credit risk, default probability models, logit, Classification and regression trees.

JEL codes: G170, M49, Q22

INTRODUCTION

During 2002-2003 several Norwegian salmon producers went into bankruptcy. This wave of bankruptcies followed three years of low salmon prices, in periods also below the production cost (Figure 1).

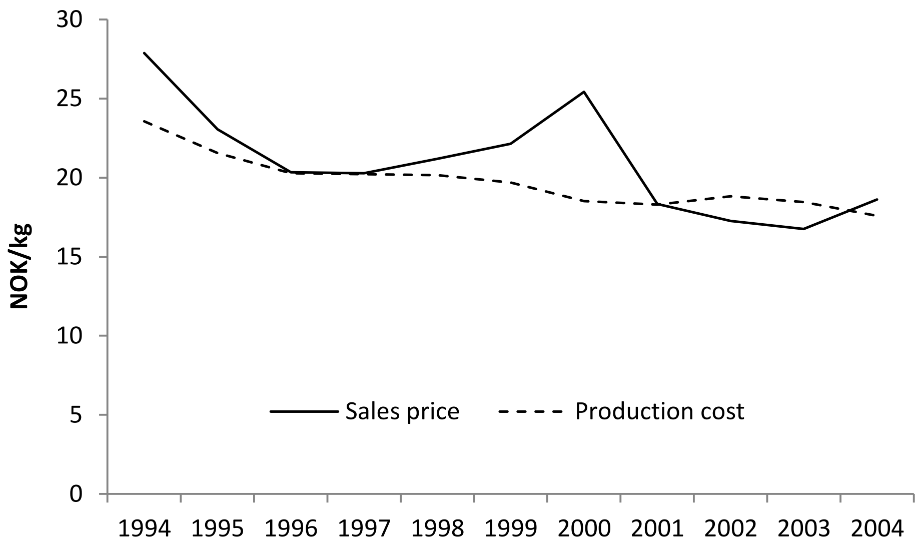


Figure 1. Sales price and production cost for ungutted farmed Norwegian Atlantic salmon (*Salmo salar*). *Source*: The Norwegian Fisheries Directorate (NFD). The dataset used by NFD contains all salmon producers in Norway.

The prolonged period of low prices had a substantial impact on the profitability of the Norwegian salmon farming industry, as shown in Figure 2. In 2001, the average salmon farm broke-even, while in both 2002 and 2003 the profitability measures net profit margin and return on equity were negative.[1]

[1] The net profit margin is calculated as net income after tax divided by total revenues, while return on equity is calculated as net income after tax divided by equity.

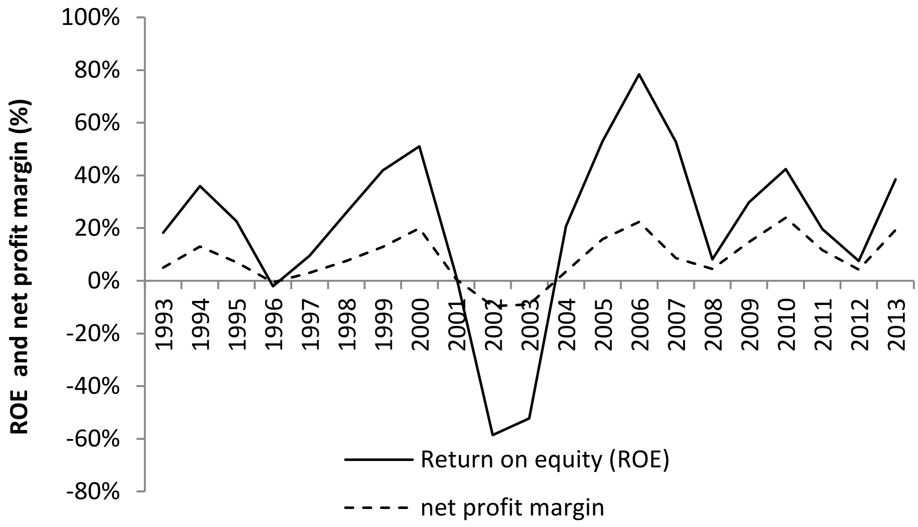


Figure 2. Average profitability for Norwegian salmon farms. Based on data collected from the Norwegian Fisheries Directorate (NFD). The dataset used by NFD contains all salmon producers in Norway.

The losses incurred by many salmon producers also had knock-on effects for investors and for banks operating in the salmon industry. In total, eleven salmon producers went bankrupt, resulting in large losses for both creditors and investors. Arguably, a large proportion of the bankruptcy losses could have been avoided with early detection of failing firms, allowing for measures to be put in place at an early stage. Default probability models are an example of such a pre-emptive measure, and which can easily be adopted by banks and investors, alike.

There is a substantial literature on credit risk modelling and default probability estimation. These studies generally suggest that default probability can successfully be predicted using statistical models based on financial ratios (Beaver, 1967a, 1967b; Altman, 1968, 1984; Ohlson 1980; Zmijewski, 1984).

The aim of this this paper is to investigate if defaults can also be predicted in the seafood industry. A credit analyst building an appropriate default risk model will need to decide between using industry-specific data, or collect data across industries. Hence, an interesting research question emerges. Will an industry specific credit risk model perform better than a more cross-sectional credit risk model? The 2000-2003 crisis in the Norwegian salmon industry serves as a suitable natural experiment in which to answer this research question.

A training sample of data comprising both salmon farming and fisheries companies for the time period 1994-1999 (henceforth: Data sample 1) are used to estimate two credit risk models, a parametric logit model and a non-parametric Classification and Regression Tree (CART) model. The performance of the credit risk models is evaluated using a holdout sample from salmon farmers over the period 2000-2002 (henceforth: Data sample 2). Moreover, the performance of the models is compared to a benchmark model developed by the Norwegian Central Bank for all industry sectors.

The results suggest that both models, estimated on a larger sample aquaculture and fisheries data, are able to separate between companies with high and low credit risk in the salmon industry. Moreover, the two models outperform the benchmark model.

The results contribute to the literature in two ways. First, the results suggest that it is important to use industry specific data when estimating the credit risk model. Second, the comparison of the two types of models, logit versus CART, shows that the CART methodology is better than logit regression on the training-sample, but performs worse for the holdout sample. This suggests that the CART approach tends to overfit the data, even after pruning.

The results should be of interest to investors and banks in the fisheries and salmon industry.

The remainder of the paper is organized as follows. The next section provides an overview of the literature, followed by a description of the research design, econometric specification and hypothesis development. The fourth and fifth sections describes the data and presents the results and discussion, respectively. The last section concludes the findings.

THE SALMON INDUSTRY

The production of farmed salmon production has been rapidly growing for more than 30 years, surpassing 2.5 million tonnes in 2013. High productivity growth has been highlighted as crucial for supporting the rapid production growth in both in the Norwegian salmon industry and globally (Nilsen, 2010; Vassdal & Holst, 2011; Asche & Roll, 2013; Roll, 2013). A considerable demand growth has coincided with this productivity growth (Asche et al., 2011; Brækkan & Thyholdt, 2014; Brækkan, 2014). However, the demand growth has been uneven, and the price of farmed salmon has a limited effect on supplied quantity (Asheim et al., 2011); a combination which can result in volatility, both in terms of prices and company profitability. Prior studies have demonstrated that the salmon price is volatile and contributes substantially to a salmon farmer's risk (Guttormsen, 2008; Oglend & Sikveland, 2008; Solibakke, 2012; Oglend, 2013), also compared to other seafood prices (Dahl & Oglend, 2014; Asche et al., 2015a). A recent study determines that salmon prices are the main driver for salmon farming profitability (Asche & Sikveland, 2015). Consequently, market volatility should lead to earnings volatility, and ultimately also impact on the financial status of the salmon producers.

Market risk is not the only risk that salmon farmers face. The production cycle for farmed Atlantic salmon is a lengthy process covering several years (Asche & Bjørndal, 2011). First, juvenile salmon (smolts) are produced from roe in fresh water (Sandvold & Tveteras, 2014). Next, after completing a metamorphosis allowing the fish to survive in salt water (smoltification), they are transferred to the sea. Over the following 16 to 24 months they are reared in sea-based pens until they reach approx. 3-8 kilos, whereby they are processed and marketed. The quantity of salmon which the farmers are able to sell is dependent on key factors such as growth and survival. The growth rate is dependent on size, feed conversion rate, feed quantities, seawater temperatures, season, artificial lighting and fish health. Hence, these biophysical factors represent a significant production risk. Consequently, the literature on salmon economics has focused on the production stage (Forsberg & Guttormsen, 2006; Asche et al., 2009; Vassdal & Holst, 2011; Roll, 2013; Asche et al., 2013; Sandvoll & Tveterås, 2014; Thyholdt, 2014). Finally, fish health and the impact of fish diseases has also been highlighted as a key production risk (Asche & Tveteras, 1999; Tveteras, 1999; Torrisen et al., 2011).

A strand of the literature also addresses the potential mitigation of market and production risks through horizontal and vertical integration (Kvaløy & Tveteras, 2008; Olson & Criddle, 2008; Oglend & Tveteras, 2013; Asche et al., 2013), use of bilateral contracts (Kvaløy & Tveteras, 2008; Larsen & Asche, 2011; Asche et al., 2014; Straume, 2014), and the use of financial derivatives (Asche et al., 2015b, 2015c; Asche & Misund, 2015).

Although the effect of market risk and production risk on profitability seems to be clear, the impact of these risk factors on the financial health of salmon producers, and thereby also the risk of bankruptcy, has not been investigated. This is important since there has been an intensification and concentration of the salmon industry (Asche et al., 2013). The effects of future bankruptcies could be bigger simply because the companies are becoming larger.

CREDIT RISK MODELS

Banks are subject to many sources of risk, comprising both financial and non-financial risks (e.g. operational risk). The most important financial risks are credit risk, liquidity risk, interest rate risk, market risk, exchange rate risk and solvency risk (Bessis, 2002). Of these, credit risk is considered to be one of the major risk exposures for banks. Credit risk is defined as the risk that the counterparty cannot meet their obligations (default) and the creditor consequently suffers a loss. A default on a loan agreement does not necessarily lead to bankruptcy. Often, the default will instigate a negotiation between the debtor and the creditor. Bankruptcy will typically require both *insufficiency* (i.e. the value of the assets are lower than the value of the debt) and *illiquidity* (i.e. the company does not possess sufficient liquid funds to cover day-to-day commitments). Hence, credit risk embodies both default risk and bankruptcy risk. Prior to a default the company will often encounter a shorter or longer period of deteriorating financial situation. This element of credit risk is called credit migration and transition risk, and a negative credit migration will increase the risk of default.

Default risk is typically quantified as an expected probability of default (PD). The relationship between default risk and a bank's expected loss is given as

$$EL = PD \times LGD \times EAD \qquad (1)$$

where *EL* is expected loss, *LGD* is loss given default and *EAD* is exposure at default.

Another important issue is concentration risk, which occurs when the *PD* for company A is correlated with the *PD* for company B (Ong, 1999). This is particularly relevant for the salmon industry. Asche & Sikveland (2015) find that salmon prices are the major driver for profitability for salmon farmers. Consequently, a prolonged salmon price drop will increase the overall default risk in the salmon industry as a whole.

In order to avoid costly defaults and bankruptcies it is imperative to be able to identify failing firms. For nearly 50 years, finding models which accurately classify firms according to default likelihood has been the topic of numerous academic studies. The seminal work of Beaver (1967a) is commonly considered to be the point of departure. Beaver (1967a) developed a univariate discriminant analysis model. This approach involves using a dichotomous classification procedure based on a set of financial ratios selected to be able separate between firms of varying financial standing.

Shortly after, Altman (1968) developed the multiple discriminant analysis method which applies multivariate analysis to estimate a composite z-score. The latter measure is a weighted average of financial ratios which can be used to discriminate between failing and non-failing firms. Consider that firm *i* is characterized by a vector of predictors (e.g. financial ratios) $\boldsymbol{X_i} = \{x_1, x_2, x_3, \dots, x_n\}$, Altman's z-score can be calculated for each firm as

$$z_i = \boldsymbol{AX_i} \tag{1}$$

where z_i is the z-score for firm i, $\boldsymbol{A}$ is a coefficient vector $(a_1, a_2, a_3, \dots, a_n)$, and a_0 is a constant term. The calculated z-score is then compared to a critical z-value, z*. If the z-score is higher than z* then the model predicts that the firm will default, compared to no default if z<z*. Discriminant analysis is used to select the financial ratios (and also determine their weights) that best classify companies that are expected to default from the ones that are not.

The limitation of Altman's approach is that it does not result in a default probability. For this reason, alternative models which are capable of explicitly calculating PDs were developed. Categorical regression models are a popular family of models for the purpose of estimating linear probability models. Ohlson (1980) developed a logit model for predicting company failure, which can be specified as

$$z_i = \beta_0 + \boldsymbol{\beta X_i} + u \tag{2}$$

where $\boldsymbol{\beta}$ is a vector of unknown parameters, β_0 and u are the constant and error terms, respectively. The parameters are obtained by maximum likelihood estimation. Let $P(\boldsymbol{X_i}, \boldsymbol{\beta})$ denote the probability of bankruptcy for any given $\boldsymbol{X_i}$ and

$\boldsymbol{\beta}$, and P is some probability function, $0 \leq P \leq 1$. According to this model, the probability to go bankrupt, given $\boldsymbol{X_i}$, is $P(\boldsymbol{X_i}, \boldsymbol{\beta}) = F(z_i)$, where $F(z_i)$ is the cumulative logistic function:

$$F(z_i) = \frac{1}{1 + e^{-z_i}} \tag{3}$$

Based on the estimated probability, the firm can be classified as bankrupt or non-bankrupt using a cut-off probability.[2]

Zmijewski (1984) developed a probit model for financial distress prediction. Probit models are similar to logit models, except that the probability is calculated using the cumulative standard normal distribution function:

$$F(z_i) = \int_{-\infty}^{z_i} \frac{1}{(2\pi)^{1/2}} e^{-z^2/2} dz. \tag{4}$$

In addition, several artificial intelligence and machine learning methods have been developed, such as neural networks (Odom & Sharda, 1990), genetic algorithm (Back et al., 1996); Anandarajan et al., 2001) and support vector machines (Shin et al., 2005; Min & Lee, 2005; Sun & Li, 2012).

[2] See Balcaen and Ooghe (2006) for a review of classical statistical methodologies.

Academic studies suggests that decision trees could be superior predictors of business failure as compared to discriminant analysis (Gepp et al., 2010). In fact, Frydman et al. (1985) find that the Recursive Partitioning Algorithm (RPA) outperform discriminant analysis in most original sample and holdout comparisons. The RPA algorithm was later succeeded by Classification and regression trees (CART) models (Breiman et al., 1984). CART is a type of machine-learning method, and is developed by recursively partitioning the data and fitting a simple prediction model within each partition. If the outcome is continuous, CART will produce a regression tree, while if the outcome is categorical, a classification tree is developed. In the case of business failure prediction the outcome is continuous in the range [0,1] and a regression tree is the result.

The next section describes the CART methodology in more detail and relies heavily on Breiman et al. (1984) and Yohannes & Webb (1999). In order to build a regression tree one needs to create a precise set of classification rules. Building a regression tree consists of three parts: 1) building a large tree, 2) pruning the tree, and 3) optimal choice of pruned tree.

A CART tree is built by creating nodes at different levels in the tree. From the start node, a binary decision is made to go either left or right. This process is based on three elements:

1) A set of questions of the type: Is $x_i \leq d$? where x_i is a variable and d is a constant. The answer to this question is Yes or No. For bankruptcy prediction

models, typical questions could be: *Is the equity-to-assets ratio below a certain level?*

2) A ‘goodness-of-split’ criteria for the choice of the best split rule for a particular variable, and

3) Generation of a summary statistic for the end nodes (only relevant for regression trees).

The process starts in a start node where the algorithm searches over several possible splits for each decision variable. For instance, this would mean a search over many possible levels of the equity-to-asset ratio. The splits result in different degrees of ‘impurities’ (i.e. variation). A relevant measure of impurities is the least squares criterion (LS) which calculates the sum of squares, *SS(t)*

$$SS(t) = \sum_{i=1}^{N} \left(y_{i(t)} - \bar{y}_t\right)^2 \tag{5}$$

where $y_{i(t)}$ are the values for the dependent variable at node *t*, and $\bar{y}_t$ is the resulting mean at node *t*. Let *s* be the split decision determining if the next step is to move to the next left node (t_L) or the right node (t_R) in the tree. The goodness-of-split is then measured as

$$\phi(s,t) = SS(t) - SS(t_R) - SS(t_L) \tag{6}$$

where $SS(t_R)$ and $SS(t_L)$ are the sum of squares for the right and left daughter nodes, respectively. The best split is found by maximizing $\phi(s, t)$, which gives the largest reduction in impurities. This would for instance result in a specific cut-off value for the equity-to-assets ratio. The data are subsequently separated and the process is continued separately to each subgroup recursively until the end nodes are reached. Along the way, more decision points are generated with attached cut-off values. Unfortunately, the resultant tree is too complex, and in the extreme, the process can lead to each observation becoming an end node. To avoid overfitting it is necessary to 'prune' the tree (Steinberg & Colla, 1995).

Pruning entails reducing the original tree such that a sequence of subtrees is created by combining some of the branches in the original tree and removing sections of the tree that provide little classification power. This is done by applying a pruning algorithm (see Breiman et al. (1984) and Steinberg & Colla (1995) for details on this topic). In essence, the LS approach uses the mean squared error (MSE) to measure predictor accuracy and to be able to rank the sequence of trees generated by pruning. The minimal cost tree is identified as the tree with the lowest MSE. The optimal tree is chosen by applying a rule versus the minimal cost tree, not necessarily selecting the minimal cost tree (which could be too simplified). In essence this is the result of a trade-off between complexity (lower fit of the data with decreasing complexity) and cost (increasing the relative re-substitution cost). Avoiding misclassification costs represents a trade-off between accuracy and tree complexity. The relative re-substitution cost is a measure of the cost related to misclassification when choosing a reduced model. An example of a rule that can be applied to find the optimal pruned tree is the one-standard-error rule, which selects the optimal tree as the tree within one standard error of the minimum cost tree. Finally, under the LS approach, the end-node values are calculated by computing mean and standard deviations of the dependent variable. The mean becomes the predicted value for a particular end node.

The resulting partitions can be represented as a decision tree in the stylized example in Figure 3. Starting with a complete dataset, the CART algorithm result in splitting dataset initially according to the equity-to-asset ratio, then according whether the firm's book equity was positive or negative. The algorithm also provides the optimal cut-off value for the variables. For instance, an equity-to-asset ratio of 20% is optimal in terms of distinguishing between firms that will default and those that do not.

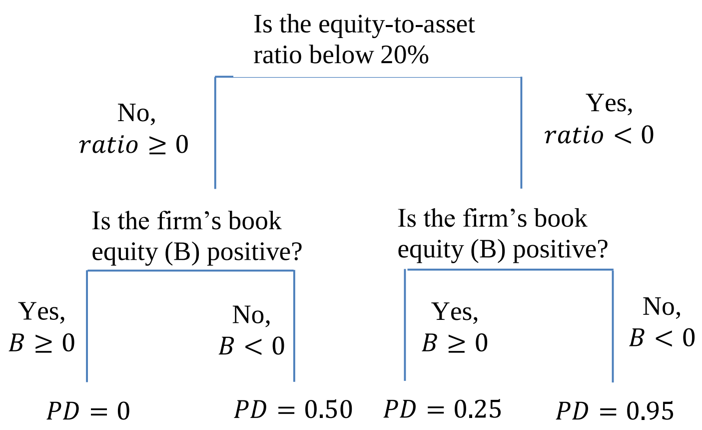


Figure 3. Stylized example of the use of CART for calculating the probability of default (PD). The numbers in the figure are arbitrary.

The tree can then be applied as follows to estimate *PD*. The first step is to start at the top of the tree in a start node. In the simplified example presented in Figure 3 this means that the firm is categorized based on its equity-to-asset ratio. If the ratio is above 20% then one moves to down and to the right in the tree. At the next decision node, the firm is categorized based on the sign of its equity. If the book equity (B) is negative, then one moves to the right and to the end node at the far right. In this simplified model, the *PD* = 0.95, meaning the model assigns a 95% probability of default.

Yohannes & Hoddinott (1999) point out some benefits of CART: 1) the model makes no distributional assumptions, 2) the explanatory variables can be a mixture of categorical, interval and continuous, 3) can deal with missing values, and 4) not affected by outliers, collinearity, or heteroskedasticity. This latter element is relevant since financial ratios are known to create outliers (e.g. when the denominator is a low value).

In this paper the CART methodology is applied in addition to the ‘standard’ logit model in order to estimate the probabilities of default.

METHOD

The method section comprises i) choice of accounting measures (financial ratios) ii) fitting empirical models and iii) validation of models.

Six groups of financial ratios as predictors of financial distress are considered. These measure traits which are thought to be relevant for evaluating the financial situation for firms. The ratios are based on financial accounting information, found in the income statement, balance sheet and cash flow statements disclosed by the annual accounts for salmon farmers. One of the benefits of using financial ratios is they are independent of size (White et al., 1997). In the credit modelling literature ratios are selected according to the following criteria: 1) economic interpretation, 2) contain information that can be used to predict defaults, 3) based on easily available accounting information and 4) cover several aspects of the firm's economics. In order to meet the last criteria, the metrics are therefore selected from several groups such as turnover/activity, liquidity, profitability, solidity/leverage, interest rate coverage and size.

Several metrics can be calculated for each of these groups. For instance, profitability can be measured as gross profit margin, net profit margin, return on equity, return on assets, return on capital employed etc. Several ratios within a group were considered, and the best ratio in each group were selected by making plots of financial ratio versus default frequency (not shown). The aim of these graphs is to ascertain if there is a correlation between the level of the ratio and default frequency. Ideally, the graphs should suggest either a positive or a negative relationship between the metric and default frequency. For instance, if the graph suggested that high (low) values of the ratio were associated with a high (low) default frequency, and vice versa, then that particular ratio was selected to be included in empirical models.

The resulting set of variables are listed below

1. *TFA*: Fixed assets turnover is a measure of activity, and is calculated as the ratio of earnings before interest and taxes (EBIT), measured from time *t-1* to time *t*, to fixed assets measured at time *t*. This ratio provides information on how much capital (investments) is needed to generate profits.
2. *NPM*: Net profit margin is a measure of profitability, and is calculated as the ratio of earnings after tax divided by revenues, measured from time *t-1* to time *t*. A high *NPM* signals high net income or low costs relative to revenues.
3. *CR*: The current ratio is a measure of liquidity, and is calculated as the ratio of current assets to short-term debt, measured at time *t*. This ratio provides information on the firm's ability to service short-term debt using funds which can be raised in the short-term.
4. *EQ*: The equity to asset ratio is a measure of leverage and solidity, and is calculated as the ratio of book equity to total assets at time *t*. A low *EQ* signifies a high debt level and a low solidity.
5. *IC*: The EBIT/interest ratio is a measure of interest rate coverage, and is calculated as the ratio of EBIT to net interest expenses, measured from time

t-1 to time *t*. The *IC* ratio provides information on the firm's ability to service interest expenses from current pre-tax profits.

6. *DC*: The EBITDA/NIBD ratio is a measure of debt coverage, and is calculated as the ratio of earnings before interest, taxes, depreciation and amortization (EBITDA), measured from time *t-1* to time *t*, to net interest bearing debt (NIBD), measured at time *t*. This ratio provides information on the firm's ability to service debt from operating cash flows.[3]

A priori, the expectation is that *PD* will increase with lower *TFA* and *EQ*, and decrease with higher *NPM*, *CR*, *IC* and *DC*.

Using the above set of explanatory variables the following logit model is estimated

$$z_i = \beta_0 + \beta_1 TFA + \beta_2 NPM + \beta_3 CR + \beta_4 EQ + \beta_5 IC + \beta_6 DC + \varepsilon \qquad (7)$$

where ε is the error term.

The CART model was estimate by applying the same explanatory variables in the CART model developed by Salford Systems (www.salford-systems.com/product/cart).

[3] EBITDA is often used by financial analysts as a proxy for pre-tax cash flows.

Serving as a benchmark for the two models estimated in this paper is a logit model developed by the Norwegian Central Bank (Bernhardsen, 2001; Eklund et al., 2001). The latter model is estimated on a cross-industry sample of approximately 400.000 observations for Norwegian firms during 1990-1996. The model is a logit model including financial ratios for profitability, solidity (leverage) and liquidity, and therefore comparable to the logit model estimated in this study.

A forecasting model is seldom perfect, and will result in some degree of misclassification (Table 1). The strength of a model lies in its ability to separate good companies from bad companies. This is done by analysing degree of misclassification as a function of a certain cut-off probability, α. If the *PD* is higher than α, the model predicts a default. No default is predicted for *PD* lower than α.

Table 1. Misclassification types

		Actual	
		Low credit quality	High credit quality
Model	Low credit quality	Correct prediction	Type II error
	High credit quality	Type I error	Correct prediction

Since choice of cut-off probabilities is often subjective Moody's applied a so-called Cumulative Accuracy Profiles (CAP), also known as power curves, for visual model performance evaluation (Soberhart et al., 2000; Kealhofer, 2003). This approach is useful since it measures both Type 1 and Type 2 errors at the same time.

To plot CAP, the companies are first ranked according to PD from lowest to largest. For a given percentage of the total number of companies, the CAP curve is constructed as the *y(x)%* of the default companies which has a *PD* which is lower or equal to x%. For instance, for x = 20%, y(20%) will be the proportion of the bankruptcies predicted by the model for the 20% lowest *PD*'s. Typically, the performance of estimated credit risk models are compared to that of two theoretical models; the ideal and the naïve models. The ideal model is a perfect model that predicts all bankruptcies, while the naïve model mimics the result of a coin-toss. For instance, the latter model will predict 50% of the bankruptcies for 50% of the population. In the stylized example below (Figure 4), the naïve model will, for 20% of the population capture 20% of the bankruptcies, compared to 85% for Model 1 and 95% for Model 2. The resulting conclusion is that both estimated credit risk models are better than the naïve model but worse than the ideal model. Moreover, the figure suggests that Model 1 is better than Model 2.

To be able to measure the degree of difference between the different models it is possible to measure the area under the curves. The Accuracy ratio (AR) is calculated as the ratio of the area under the curve for the model under consideration compared to that of the ideal and naïve models. Low AR values mean that the model is not much better than a toss of a coin (naïve model), while models with high AR demonstrate good ability for default prediction.

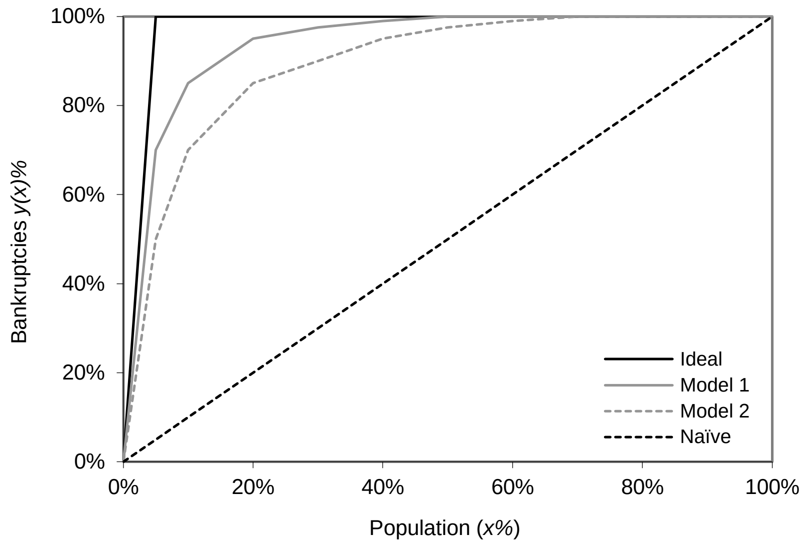


Figure 4. CAP for a stylized example. The numbers in the figure are arbitrary.

DATA SAMPLES

In order to reduce the possibility of overfitting, two non-overlapping data samples are used for estimating the empirical credit risk models, and for the subsequent performance evaluation. The composition of the datasets are different. Data sample 1 consists of data for Norwegian fisheries and salmon industry prior to the wave of salmon producer bankruptcies for the years 1994-1999.[4] These observations are used to estimate the logit and CART models. The second data set, Date sample 2 consists of only Norwegian fish farming firms for the years 2000-2002, and is used to evaluate the model performance. The reason for using two samples is the twofold. First, the aim of the paper is to examine models that can be used to calculate bankruptcy risk in the Norwegian aquaculture industry. Hence, Data sample 2 only

[4] This dataset also included observation for part of 2000, but these were excluded in the final data sample.

contains aquaculture firms. Second, number of defaults in the Norwegian aquaculture has been small historically. The 2001-2003 period was characterised by an exceptional increase in the number of defaults/bankruptcies. In order to have a sufficient amount of data allowing us to estimate an empirical model reliably, it is preferable to include more observations. This was accomplished by collecting data from companies in the fisheries industry in addition to aquaculture firms. The rationale is that fisheries companies share many similarities with aquaculture companies. Table 2 describes the composition of the training sample. The original sample consists of 2878 companies. After removing observations due to missing data for some of the accounting variables the sample decreases to 1551. Truncating the data by removing the highest and lowest 2 percentiles of values, results in the final training sample of 1464 observations. Truncation is not necessary for the CART estimation, but is done for ease of comparison with the logit model.

Table 2. Composition of original sample

Code	Description	N	N^b	N^b(%)
50000	Services for fisheries and aquaculture	25	1	4.00
50100	Fisheries	341	16	4.69
50110	Ocean and coastal fisheries	405	31	7.65
50200	Hatcheries	997	27	2.71
50210	Production of fish and shellfish	150	2	1.33
50220	Smolt production	49	0	0
152000	Fish and seafood processing	546	57	10.44
152001	Trading and packaging	137	13	9.49
152010	Production of salted and dried fish	50	4	8.00
152020	Freezing of fish and shellfish	27	1	3.70
152030	Canning	15	1	6.67
152090	Production of other conserved seafood products	80	8	10.00
154110	Production of fish oil	18	1	5.56
157000	Production of fish feed	38	3	7.89
	Total	2878	165	5.70

Note: N is the number of observations and N^b is the number of bankruptcies. The data consists of fisheries and aquaculture companies for the years 1994-1999.

The number of bankruptcies per year in the original sample is presented in Table 3. It seems that on average around 20 firms go into bankruptcy per year, based on the observations in the data set.

Table 3. The number of bankruptcies in the original sample by year (N = 2878).

Year	1994	1995	1996	1997	1998	1999
N^b	12	17	35	29	25	41

Note: N^b is the number of bankruptcies. The data consists of fisheries and aquaculture companies for the years 1994-1999.

Table 4 presents the descriptive statistics for the variables for the training sample. Even after truncation it seems that there is a substantial variation in the ratios as measured by the standard deviation. For instance, the mean profit margin is -0.361, while the median is 0.018, suggesting that there is still some skewness left originating from a few observations with very low values.

Table 4. Descriptive statistics for Data sample 1 (N = 1464)

Variable	Mean	St.dev.	25 percentile	Median	75 percentile
TFA	11.231	223.159	0.180	0.500	1.241
NPM	-0.361	12.628	-0.079	0.018	0.100
CR	1.849	3.750	0.715	1.139	1.760
EQ	-0.208	4.606	0.027	0.168	0.330
IC	9.314	107.412	0.694	2.452	5.114
DC	0.747	17.478	0.037	0.214	0.462

Note: TFA = Fixed asset turnover, NPM = net profit margin, CR = current ratio, EQ = equity-to-assets ratio, IC = interest coverage ratio, and DC = debt coverage ratio. Data sample: Data sample 1: Model calibration sample (1464 fisheries and aquaculture companies 1994-1999.

Table 5 presents the correlations between the variables for the training sample. The correlations are smaller in magnitude compared to what would normally cause concerns about multicollinearity.

Table 5. Correlations for the variables (N = 1464)

	TFA	*NPM*	*CR*	*EQ*	*IC*	*DC*
TFA	1					
NPM	-0.489	1				
CR	0.090	-0.022	1			
EQ	0.013	0.183	0.181	1		
IC	-0.072	0.216	0.123	0.152	1	
DC	-0.142	0.198	0.057	0.147	0.481	1

Note: TFA = Fixed asset turnover, NPM = net profit margin, CR = current ratio, EQ = equity-to-assets ratio, IC = interest coverage ratio, and DC = debt coverage ratio. Data sample: Data sample 1: Model calibration sample (1464 fisheries and aquaculture companies 1994-1999.

The holdout sample consists of Norwegian salmon farms over the period 2000-2002. This sample contains 162 salmon farming companies. Default does not always mean that a company goes bankrupt. Bankruptcy is in this paper defined as the case where the company has either gone bankrupt or has defaulted on its bank loans with the consequence that the banks has taken over the control of the company. According to this definition results in 11 bankruptcies, representing 6.8 percent of the total sample.

RESULTS AND DISCUSSION

The results from the logit regression are presented in Table 6.

Table 6. Logit regression

$$z_i = \beta_0 + \beta_1 TFA + \beta_2 NPM + \beta_3 CR + \beta_4 EQ + \beta_5 IC + \beta_6 DC + \varepsilon$$

Variable	coefficient	Wald	p-value
Constant	-0.922	15.23	<0.001
TFA	0.011	0.429	0.512
NPM	-0.007	0.011	0.916
CR	-1.376	31.877	<0.001
EQ	-0.042	0.201	0.654
IC	-0.084	9.333	<0.001
DC	-0.636	3.353	0.067
Regression summary			
Log-likelihood			609.510
Goodness-of-fit (pseudo R2)			
Cox-Snell			0.115
Nagelkerke			0.272

Note: TFA = Fixed asset turnover, NPM = net profit margin, CR = current ratio, EQ = equity-to-assets ratio, IC = interest coverage ratio, and DC = debt coverage ratio. The regression goodness-of-fit is measured with Cox-Snell (Cox and Snell, 1989) and Nagelkerke (Nagelkerke, 1991) pseudo R2. Data sample: Data sample 1: Model calibration sample (1464 fisheries and aquaculture companies 1994-1999. N = 1464.

In contrast to expectations, the coefficient on the leverage variable is not significant.

Since low or negative equity is typically associated with default risk, the equity-to-assets ratio is therefore replaced with a dummy variable which takes the value 1 if book equity is negative, and 0 otherwise. The reason is that the equity-to-asset ratio does not fully capture the negative effects of negative equity. The latter typically being the result of large losses or write-downs of assets, with is often associated with business failure. Moreover, *TFA* and *NPM* are omitted in the final model since these are insignificant, and the logit model is re-estimated. This procedure improves the results (Table 7). The coefficients in the improved model are now all significant at the 10% level, and except for the debt coverage ratio, also significant at the 5% level.

Table 7. Logit regression: Simplified and including a dummy for negative equity

$$z_i = \theta_0 + \theta_1 CR + \theta_2 IC + \theta_3 DC + \theta_4 NEGEQ + \epsilon$$

Variable	coefficient	Wald	p-value
Constant	-1.678	39.402	<0.001
CR	-1.063	20.827	<0.001
IC	-0.062	5.665	0.017
DC	-0.629	3.216	0.073
NEGEQ	0.904	12.567	<0.001

Note: CR = current ratio, IC = interest coverage ratio, DC = debt coverage ratio, and NEGEQ is a dummy variable for negative equity. The regression goodness-of-fit is measured with Cox-Snell (Cox and Snell, 1989) and Nagelkerke (Nagelkerke, 1991) pseudo R2. Data sample: Data sample 1: Model calibration sample (1464 fisheries and aquaculture companies 1994-1999.

Next, the regression tree is estimated. The unpruned tree is very large, comprising a total number of leaves of 235. This model is clearly too complex and a pruning algorithm is applied, resulting in the model in Figure 5.

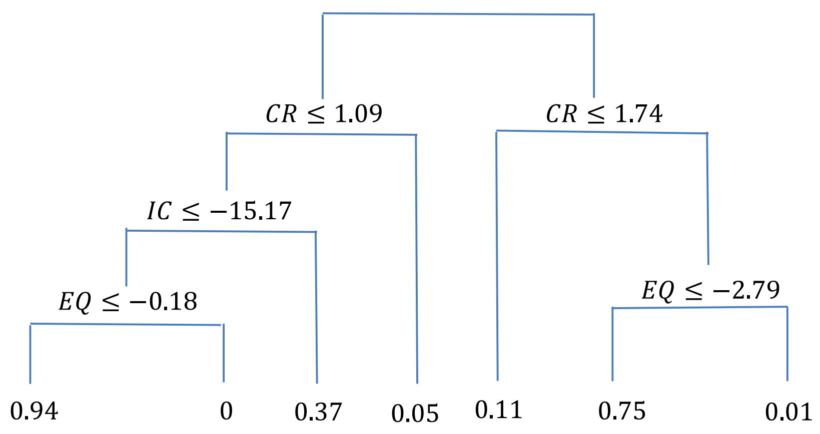
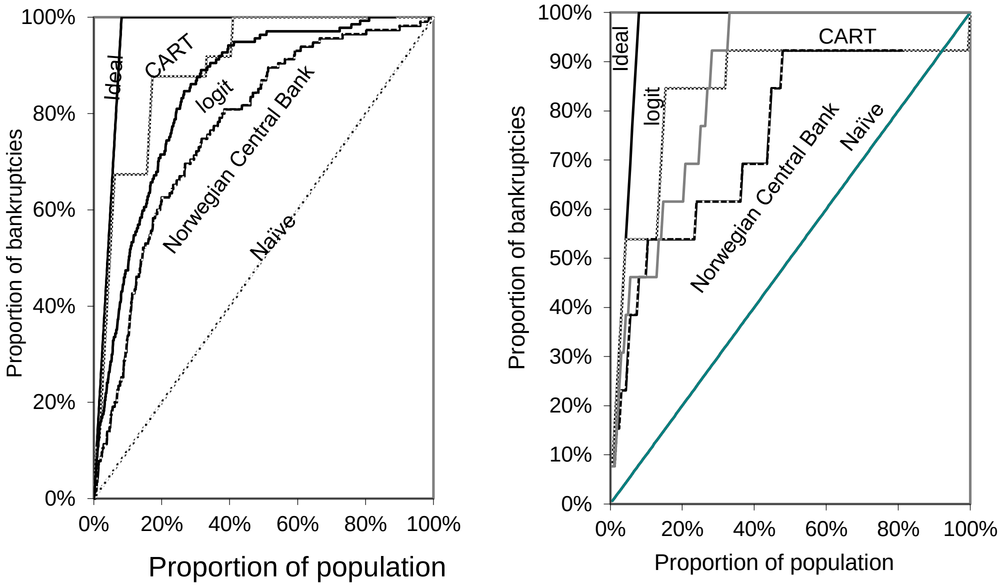


Figure 5. Pruned CART tree. Data sample 2: Model performance evaluation sample (162 aquaculture companies 2000-2002).

The benefit of CART, compared to linear models such as logit, is demonstrated by the leverage/solidity variable. The algorithms in CART are able to find relevant cut-off values. For instance, for firms with an interest rate coverage of less than 0.14, plus a current ratio of less than 1.09, combined with an interest rate coverage of less than 15.17, the model attributes a *PD* of 94% if the equity is less than minus 18%. In the logit model, it was necessary to replace the leverage variable with a negative equity dummy variable, while the CART model was able to select a more exact cut-off point for the equity variable. This is an example of the benefits of regression trees.

The next step is to perform a model evaluation using power curves (Figure 6) and accuracy ratios (Table 8) for both datasets.



(a) Data sample 1 (b) Data sample 2

Figure 6. Power curves (CAP) for the logit and CART models compared to the ideal, and naïve and benchmark models. Data samples: Data sample 1: Model calibration sample (1464 fisheries and aquaculture companies 1994-1999; Data sample 2: Model performance evaluation sample (162 aquaculture companies 2000-2002).

Table 8. Accuracy ratios (%) for the logit and CART models compared to the ideal, and naïve and benchmark models

Model	Data sample 1	Data sample 2
Ideal	100%	100%
Logit	71.8%	79.2%
CART	86.7%	75.2%
Benchmark	57.7%	53.5%

Data samples: Data sample 1: Model calibration sample (1464 fisheries and aquaculture companies 1994-1999; Data sample 2: Model performance evaluation sample (162 aquaculture companies 2000-2002).

The results show that both the logit and CART models perform better than the naïve model. They are able to predict 75-80% of the bankruptcies. The results also show that they are much better than the Central Bank model suggesting that one should use industry data for estimating credit risk models. In fact, the logit model has an accuracy ratio of nearly 80% compared to only approximately 50% for the benchmark model. Hence, the model developed by the Norwegian Central Bank would not have accurately predicted many of the bankruptcies which occurred in the Norwegian salmon farming industry during the early 2000s.

Finally, the results also show that the simple logit model performs better than the CART model. An explanation is that the CART model tends to overfit the data. The CART was the best model in the training sample, but was second best on the holdout sample. The reduction in the accuracy ratio was approximately ten percentage points.

CONCLUSIONS

During 2000-2003 the Norwegian salmon industry witnessed a marked increase in bankruptcies following several years of low wholesale prices, even below production costs. This time period serves as a natural experiment to test statistical default probability prediction models. Bankruptcies are very costly, both for the owners and creditors to the affected firm. These costs could be substantially reduced, even avoided all together if the failure of these firms could have been predicted at an earlier stage. In this paper, two default probability prediction models are estimated. A logit and a CART model, are calibrated from a set of fisheries and aquaculture data. The models are subsequently evaluated on a different on-overlapping data sample consisting of salmon producers prior to the bankruptcy wave. The aim is to ascertain if the empirical credit risk models can reliably predict the defaults. For comparison, a benchmark model developed by the Norwegian Central Bank is also applied. The results suggest that the model estimated on fisheries and aquaculture data outperforms the benchmark model. Moreover, the results suggest that the logit model performs better than the CART model, possibly due to an overfitting of the latter model.

The results demonstrate the importance of using an industry sample to develop relevant default probability models. Like many other industries, the firms in the seafood sector are characterized by certain traits, resulting in many similarities, which ultimately might be manifested in similar values of financial ratios. For instance, the salmon industry is characterized by relatively high leverage due to the lengthy production process and the tendency to rely on bank loans and supplier credit. A low leverage in itself might not be a predictor of business failure. What is important is to be able to separate firms that fare worse, or substantially worse, than what is considered normal in that particular industry. Hence, the implication of the study is that commercial banks calculating default probabilities or central banks trying to assess overall credit risk, need to consider industry specificities in their credit assessment models.

The results of the study should be of interest to wide audience including salmon producers, investors and creditors, both commercial banks and also other suppliers who also extend credit to the farmers. A default probability model, such as the logit model, are fairly simple to implement and to use in practice.

REFERENCES

Altman, E.I. (1968) Financial ratios, discriminant analysis, and the prediction of corporate bankruptcy. Journal of Finance, 23(4), 589-609.

Altman, E.I. (1984) The success of business failure prediction models – An international survey. Journal of Banking and Finance, 8, 171-198.

Anandarajan, M., P. Lee & A. Anandarajan (2001) Bankruptcy prediction of financially stressed firms: An examination of the predictive accuracy of artificial neural networks. International Journal of Intelligent Systems in Accounting, Finance and Management, 10, 69-81.

Asche, F. & B. Misund (2015) Hedging efficiency of Atlantic salmon futures. Working Paper. University of Stavanger, Stavanger, Norway.

Asche, F. & K.H. Roll (2013) Determinants of inefficiency in Norwegian salmon aquaculture. Aquaculture Economics and Management, 17(3), 300-321.

Asche, F. & M. Sikveland (2015) The Behavior of Operating Earnings in the Norwegian Salmon Farming Industry. Forthcoming in Aquaculture Economics and Management.

Asche, F. & R. Tveterås (1999) Modeling production risk with a two-step procedure. Journal of Agricultural and Resource Economics, 24, 424-439.

Asche, F. & T. Bjørndal (2011) The Economics of Salmon Aquaculture, Wiley-Blackwell: Chichester.

Asche, F., R.E. Dahl & M. Steen (2015a) Price volatility in seafood markets: Farmed vs. wild fish. Forthcoming in Aquaculture Economics and Management.

Asche, F., R.E. Dahl, D.V. Gordon, T. Trollvik & P. Aandal (2011) Demand growth for salmon in the European market. Marine Resource Economics, 26(4), 255-265.

Asche, F., R.E. Dahl, D. Valderrama & D. Zhang (2014) Price transmission in new supply chains – The case of salmon in France. Aquaculture Economics and Management, 18, 205-219.

Asche, F., A.G. Guttormsen & R. Nielsen (2013) Future challenges for the maturing Norwegian salmon aquaculture industry: An analysis of total factor productivity change from 1996 to 2008. Aquaculture, 1, 396-399.

Asche, F., B. Misund & A. Oglend (2015a) The spot-forward relationship in the Atlantic salmon market. Working Paper. University of Stavanger, Stavanger, Norway.

Asche, F., B. Misund & A. Oglend (2015b) Production risk and the futures price risk premium. Working Paper. University of Stavanger, Stavanger, Norway.

Asche, F., K.H. Roll & R. Tveteras (2009) Economic inefficiency and environmental impact: An application to aquaculture production. Journal of Environmental Economics and Management, 58(1), 93-105.

Asche, F., K.H. Roll, H.N. Sandvold & A. Sørvig (2013) Salmon aquaculture: Larger companies and increased production. Aquaculture Economics and Management, 17(3), 322-339.

Asheim, L.J., R.E. Dahl, S.C. Kumbhakar, A. Oglend & R. Tveteras (2011) Are prices or biology driving the short-term supply of farmed salmon? Marine Resource Economics 26(4), 343-357.

Back, B., T. Laitinen & K. Sere (1996) Neural networks and genetic algorithms for bankruptcy predictions. Expert Systems with Applications, 11, 407-413.

Balcaen, S. & H. Ooghe (2006) 35 years on business failure: an overview of the classical statistical methodologies and their related problems. The British Accounting Review 38, 63-93.

Beaver, W.H. (1967a) Financial ratios as predictors of failure. Empirical research in accounting: Selected studies 1966. Journal of Accounting Research, 4, 71-111.

Beaver, W.H. (1967b) Alternative accounting measures as predictors of failure. American Accounting Association Manuscript Contest for 1967, 113-121.

Bernhardsen, E. & K. Larsen (2002) Bankenes prising av risiko ved utlån til foretaksektoren (In Norwegian). Penger og Kreditt 4, 185-194.

Bessis, J. (2002) Risk Management in Banking. John Wiley & Sons.

Brækkan, E. & S.B. Thyholdt (2014) The bumpy road of demand growth. Forthcoming in Marine Resource Economics, 29, 339-350.

Brækkan, E. (2014) Disentangling demand and supply shifts: The impact on world salmon price. Applied Economics, 46, 3942-3953.

Breiman, L., J.H: Friedman & C.J. Stone (1984) Classification and Regression Trees. Chapman and Hall.

Cox, D.R. and E.J. Snell (1989) Analysis of binary data. Second edition. Chapman and Hall.

Dahl, R.E. & A. Oglend (2015) Fish price volatility. Marine Resource Economics, 29, 305-322.

Eklund, T., K. Larsen & E. Bernhardsen (2001). Modell for analyse av kredittrisiko i foretakssektoren (In Norwegian). Penger og kreditt, 2, 109-115.

Forsberg, O.I. & A.G. Guttormsen (2006) The value of information in salmon farming. Harvesting the right fish at the right time. Aquaculture Economics and Management, 10(3), 183-200.

Frydman, H., E.I. Altman & D.-L. Kao (1985) Introducing recursive partitioning for financial classification: The case of financial distress. Journal of Finance, 40, 269-291.

Gepp, A., K. Kumar & S. Bhattacharya (2010) Business failure prediction using decision trees. Journal of Forecasting, 29(6), 536-555.

Guttormsen, A.G. (2008) Forecasting weekly salmon prices: Risk management in fish farming. Aquaculture Economics and Management, 3(2), 159-166.

Kealhofer, S. (2003) Quantifying credit risk I: Default prediction. Financial Analysts Journal, 59(1), 30-44.

Kvaløy, O. & R. Tveteras (2008) Cost structure and vertical integration between farming and processing. Journal of Agricultural Economics, 59(2), 296-311.

Larsen, T.A. & F. Asche (2011) Contracts in the salmon industry: An analysis of Norwegian salmon exports. Marine Resource Economics, 26(2), 141-150.

Min, J.H. & Y.-C. Lee (2005) Bankruptcy prediction using support vector with optimal choice of kernel function parameters. Expert Systems with Applications, 28, 128-134.

Nagelkerke, N.J.D. (1991) A not on a general definition of the coefficient of determination. Biometrika 78, 691-692.

Nilsen, O.B. (2010) Learning-by-doing or technological leapfrogging: Production frontiers and efficiency measurement in Norwegian salmon aquaculture. Aquaculture Economics and Management, 14, 97-119.

Odom, M. & R. Sharda (1990) A neural networks model for bankruptcy prediction. Proceedings of the IEEE international conference on Neural Network, 2, 163-168.

Oglend, A. & M. Sikveland (2008) The behavior of salmon price volatility. Marine Resource Economics, 23(4), 507-526.

Oglend, A. & R. Tveteras (2009) Spatial diversification in Norwegian aquaculture. Aquaculture Economics and Management, 13(2), 94-111.

Oglend, A. (2013) Recent trends in salmon price volatility. Aquaculture Economics and Management, 17, 281-299.

Ohlson, J.A. (1980) Financial ratios and the probabilistic prediction of bankruptcy. Journal of Accounting Research, , 109-131.

Olson, T.K. & K.R. Criddle (2008) Industrial evoluation: A case study of Chilean salmon aquaculture. Aquaculture Economics and Management, 12(2), 89-106.

Ong, M.K. (1999) Internal Credit Risk Models. Risk Books.

Roll, K.H. (2013) Measuring performance, development and growth when restricting flexibility. Journal of Productivity Analysis, 39(1), 15-25.

Sandvoll, H.N. & R. Tveterås (2014) Innovation and productivity growth in Norwegian production of juvenile salmonids. Aquaculture Economics and Management, 18(2), 149-168.

Shin, K.-S., T.S. Lee & H.-J. Kim (2005) An application of support vector machines in bankruptcy prediction models. Expert Systems with Applications, 28, 127-135.

Soberhart, J.R., S.C. Keenan & R.M. Stein (2000) Benchmarking quantitative default risk models: A validation methodology. Moody's White Paper (www.moodys.com).

Solibakke, P.J. (2012) Scientific stochastic volatility models for the salmon forward market: Forecasting (un-)conditional moments. Aquaculture Economics and Management, 16(3), 222-249.

Steinberg, D. & P. Colla (1995) CART. Tree-Structured Nonparametric Data Analysis. San Diego, California, U.S.A.: Salford Systems.

Straume, H.M. (2014) Currency invoicing in Norwegian salmon export. Marine Resource Economics, 29, 391-409.

Sun, J. & H. Li (2012) Financial distress prediction using support vector machines: Ensemble vs. individual. Applied Soft Computing, 12, 2254-2265.

Torrisen, O., R.E. Olsen, R. Toresen, G.I. Hemre, A.G.J. Tacon, F. Asche, R.W: Hardy & S.P. Lall (2011) Atlantic salmon (Salmo salar) – The super-chicken of the sea? Reviews in Fisheries Science, 19(3), 257-278.

Tveteras, R. (1999) Production risk and productivity growth: Some findings for Norwegian salmon aquaculture. Journal of Productivity Analysis, 12(2), 161-179.

Thyholdt, S.B. (2014) The importance of temperature in farmed salmon growth: Regional growth functions for Norwegian farmed salmon. Aquaculture Economics and Management, 18(2), 189-204.

Vassdal, T. & H.M.S. Holst (2011) Technical progress and regress in Norwegian salmon farming: A Malmquist index approach. Marine Resource Economics, 26, 329-342.

White, G.I., A.C. Sondhi & D. Fried (1997) The analysis of financial statements 2nd edition. John Wiley & sons.

Yohannes, Y. & J. Hoddinott (1999) Classification and regression trees: An introduction. Food Policy Research Institute.

Yohannes, Y. & P. Webb (1999) Classification and Regression Trees, CART. A User Manual for Identifying Indicators of Vulnerability to Famine and Chronic Food Insecurity. Food Policy Research Institute.

Zmijewski, M.E. (1984) Methodological issues related to the estimation of financial distress prediction models. Journal of Accounting Research (Supplement), 59-82.