Norwegian School of Economics
Bergen, spring semester 2015

NHH

Factors Affecting the Probability of Bankruptcy

A panel data approach

Håkon Kambestad Sundal & Karoline Hatlestad

Supervisor: Assistant Professor Kai Liu

Master thesis in Economic Analysis and Finance

NORWEGIAN SCHOOL OF ECONOMICS

Abstract

This paper investigates the importance of firm-specific factors in determining or explaining bankruptcy. By studying Norwegian firms from the period 2005-2012, we are able to examine this using binary regression models. First, we identified potential financial measures we believed to be associated with business failure. Then we selected 15 of these that are potentially correlated with the occurrence of bankruptcy along with 3 firm-specific characteristics. These measures were incorporated into different econometric models. During analysis, the 15 financial measures were reduced to 5: Two profitability measures, two solidity measures and one liquidity measure. We conclude that fixed effects are present in the data. Controlling for them enables us to identify the impact of accounting ratios on the probability of a bankruptcy more efficiently. In the logistic regression only two profitability measures remain significant, yet when we construct a prediction model for business failure this model has an overall accuracy of 74 %. Thus, we are also confident that incorporating firm-specific effects in the model enables us to identify good measures of how accounting data affects the probability of bankruptcy.

Preface

This thesis was written as a part of our Master of Science (MSc) degree in Economics and Business Administration at the Norwegian School of Economics, spring 2015. We are majoring in two different profiles, Master in Economic Analysis and Master in Finance respectively. It was therefore important to find a topic that would cover both of our education programs. We also wanted to put our experience with econometrics to use, and that is how we ended up using a dataset on accounting data.

The thesis is based on a dataset of Norwegian accounting figures "Norwegian corporate accounts". The dataset is developed and prepared by Professor Aksel Mjøs, Endre Berner and Marius Olving on yearly numbers provided from the Brønnøysund register, and we gained access to the data material from *Centre for applied research* at NHH.

The particular focus on bankruptcies was somewhat inspired by an engaging lecture courtesy of Professor Kjell Henry Knivsflå about an article written by James Ohlson (1980) on the subject. We knew about the dataset from a previous course held by Assistant Professor Arnt Ove Hopland.

We would sincerely like to thank the people listed above, as well as Kellis Akselsen for helping us getting access to the dataset. Most importantly, we would like to thank our supervisor Assistant Professor Kai Liu for the help and support we have received during this semester. He has been available to us throughout the entire process and his guidance and feedback has been incredibly valuable.

We want to thank our friends and families that have contributed with comments and suggestions to our thesis, and also for being supportive and patient with us all along.

Bergen, June 2015

Håkon Kambestad Sundal

Karoline Hatlestad

Contents

List of figures

List of tables

1. Introduction

Looking into the future, at least doing it with statistical confidence, is very difficult. There are always some elements we cannot foresee. Such is the nature of life. More to the point, this is also the nature of companies. Therefore, accurately predicting corporate bankruptcy is inherently difficult. The literature lists an incredible range of factors that affect the financial state of a company including, but not limited to, factors such as the internal processes (Zhu, 2012), liquidity (Bernanke, 1981), competition in the market (Webb, 1980) and macro-economic conditions (Bhattacharjee, et.al., 2009).

Being able to identify the most important factors in estimating bankruptcies can yield valuable information, especially regarding the population of firms in Norway. What decides the default of a company can differ across cultures. As such, this knowledge can be put to further use upon specifying new models for the actual *forecasting* of bankruptcies within the population.

Different factors explain the variation in different subpopulations. Given all these factors, we are interested in providing some new insight into Norwegian data in particular. The latest authoritative paper on the subject of estimating bankruptcy probabilities in large Norwegian datasets was one by Bernhardsen (2001), which we will discuss in more detail further on. Even so, since then society has been rapidly modernized and we have had periods of financial distress. In addition, software packages and hardware have been developed to accommodate the use of even larger datasets and more demanding estimation techniques. We therefore think a new estimation will add to what we already know about bankruptcies as a phenomenon. The main research questions that we try to shed light on are:

- *How can we best measure the effect of financial factors that determine a bankruptcy?*
- *How does fixed effects influence the estimation approach?*

This thesis is developed on Norwegian data, and is specifically related to the Norwegian spread of bankruptcies. To be able to contribute with new insights we want to take the panel data structure of our data into consideration. Our intuition is that using a good regression based model on a large dataset should yield accurate insight as to how near a firm is to being insolvent if the model is well specified. While we will admit that the latter might be beyond the scope of this thesis, we nevertheless discover evidence that the use of panel data approaches can change the interpretation of what factors determine a bankruptcy. We seek to control for these fixed effects that may be present on a firm specific level. These effects could be anything but say they are managerial capabilities, in which case a panel data approach should be better. This is something we have not found done yet in the literature, probably due to the widespread use of logistic regression, whose fixed effects estimations are very computationally demanding and therefore less used.

Why some businesses are more exposed will be valuable information for both a firm and creditors. A firm would receive input about what measures heightens the risk of a default and be able to improve on these. Any financial institution that provide credit for a company will need input on calculating this risk in order to calculate the interest rates they should charge, so they can maximize their profit and minimize losses. The information could also be useful to decision makers on a macro-level during times of distress when companies will be cutting costs and hoarding equity, whilst unions will underline the importance of employment as firms let go off staff.

Chapter 2 will give an overview and some definitions of the term bankruptcy in relation to Norwegian market conditions, legislation and cyclical variation. Chapter 3 describes what we feel is the most authoritative literature on this subject in regards to our own paper, specifically Altman (1968), Ohlson (1980), Skogsvik (1988) and Bernhardsen (2001). Chapter 4 consists of descriptive statistics and an overview of our dataset. The primary key figures in determining bankruptcy are also derived in this chapter. Furthermore, in chapter 5 the process of selecting and estimating coefficients for our key variables is shown. In order to do so we make use of the ordinary least squares model, random and fixed effect models and logistic regression. The results and analyses from these estimations are also in this chapter. Lastly, in chapter six, we use the most efficient model to predict estimated probabilities and evaluating them.

2. Bankruptcies in Norway and economic cycles

Filing for bankruptcy is something both private persons and companies can do if they are not able to meet their financial obligations. This thesis focuses on bankruptcies in regards to companies. As of late, market conditions in Norway are experiencing changes. The oil sector has been affected by a sudden drop in oil prices. Seeing how the Norwegian economy is tightly connected to how the oil sector performs, this is an interesting time to be able to identify the real drivers of bankruptcy as firms are under financial pressure and some will be more inclined towards failure.

2.1 The Norwegian legislation

The Norwegian law states that if a debtor is insolvent then bankruptcy proceedings are initiated (The Bankruptcy act, 1984, § 60.). Creditors or claimants may petition a bankruptcy proceeding, if the debtor is not able to meet their claims.

The Norwegian law defines insolvent §61:

According to §61 (The Bankruptcy act, 1984, § 61.), the debtor is insolvent when they are not able to meet their financial obligations. The debtor cannot be considered insolvent if the value of their belongings covers the value of their debt. The insolvency must not be considered temporary, and the debtor must struggle to meet their obligations when they are due.

The legislation regarding limited companies includes an act of requirements to the companies' liquidity. One relevant paragraph here is § 3-4 "Requirement of adequate equity" which states that the company shall at all times have an equity and liquidity which is adequate in terms of the risk and scope of the company's business (The Norwegian Public Limited Liability Companies Act, 1997, § 3-4).

2.2 Reasons for bankruptcies

Surviving in a competitive market is a challenge for many companies. In financial recessions, it is observed that more companies experience financial stress that sometimes leads to bankruptcy. Companies may have difficulties adapting to financial stress and not be able to make the necessary adjustments in their costs. Even companies that provide good products and have large earnings struggle if they are not liquid enough. The balancing between investing in the company and having liquidity to meet their obligations is a challenge. Capital-effective industries require more available liquidity than labour-effective industries, so there may be differences between sectors.

Another event that leads to bankruptcies are acts of criminal activity. Although this is not very common, it happens from time to time that companies are exposed to fraud. Being subjected to a scam that largely affects a company's financial state could initiate a down spiralling period. Especially if the company is not able to win potential law suits, and must consider the loss a sunk cost. This could be fraud related to investments, misleading the company in business deals, not delivering products as agreed on etc. Internal controlling is also important to avoid criminal activities within the company. Having trustworthy and loyal employees is crucial for companies, as the employees often have access to finances and reporting systems. Embezzlement is a risk for every company that has disloyal staff, and this could be difficult to discover if the same people are in charge of reporting. Although this is obviously illegal, it still happens.

Distributing the risk of losses is important to companies. If a company base their entire income on a few custumers or projects, they may experience significant percentage-wise losses if something goes wrong. It is like putting all the eggs in one basket, and single cases could destroy companies. If a company makes a mistake in a case like this, they could end up in a lawsuit that they are not able to handle in terms of compensation costs. Depending on what industry the company operates in these risks should be taken into account on different levels.

2.3 Stigma of going bankrupt

The general opinion in the public about bankruptcies differs through cultures. We have the impression that some cultures are more accepting and less negative towards bankruptcies than others are. E.g. the United States is considered more liberal in regards to this than Norway, and it is more common to file for bankruptcy in the US. There are many contributing factors to this, and the fact that bankruptcy proceedings are met with a stigmatization in Norway could be a reason. This stigma towards failing companies has historically been very common. The situation has improved, but a survey done by Deloitte (Helsingeng, 2004) claims that 75 % of managers still find a stigma attached to going bankrupt. In addition to this, other economic perspectives contribute to why the Norwegian market is different from others.

2.4 Consequences of bankruptcy

When a company goes bankrupt, it has many repercussions. The company must file for bankruptcy when they are not able to meet their financial obligations. Creditors are therefore one of the parties that are highly affected. If the company has any debt or other financial obligations, the creditors are at risk of losing this value. Creditors often have models that calculate the probability of bankruptcy times the potential loss, and incorporate this into their strategy of lending. The cost of borrowing capital often is calculated with regards to this risk. Nevertheless, an unexpected failure of a company could provide major losses. Other suppliers that provide the company with goods and products are also in danger of not receiving payment for what they have delivered.

Customers and business partners of a company that goes bankrupt could be affected by a bankruptcy. When a company files for bankruptcy, it terminates the operations. Contracts that the company entered previously may not be fulfilled. This leads to financial and operational consequences for those involved. When the company has to terminate its operations, there are often many employees that lose their jobs. In recessions, this is particularly a problem as the employee can encounter difficulties in finding another job. As Figure 1 and Figure 2 show the bankruptcy rates and the unemployment rates are following a similar pattern. This trend may be due to macroeconomic conditions that are influencing both the bankruptcy rate, and the unemployment rate.

Figure 1: Bankruptcies in Norway

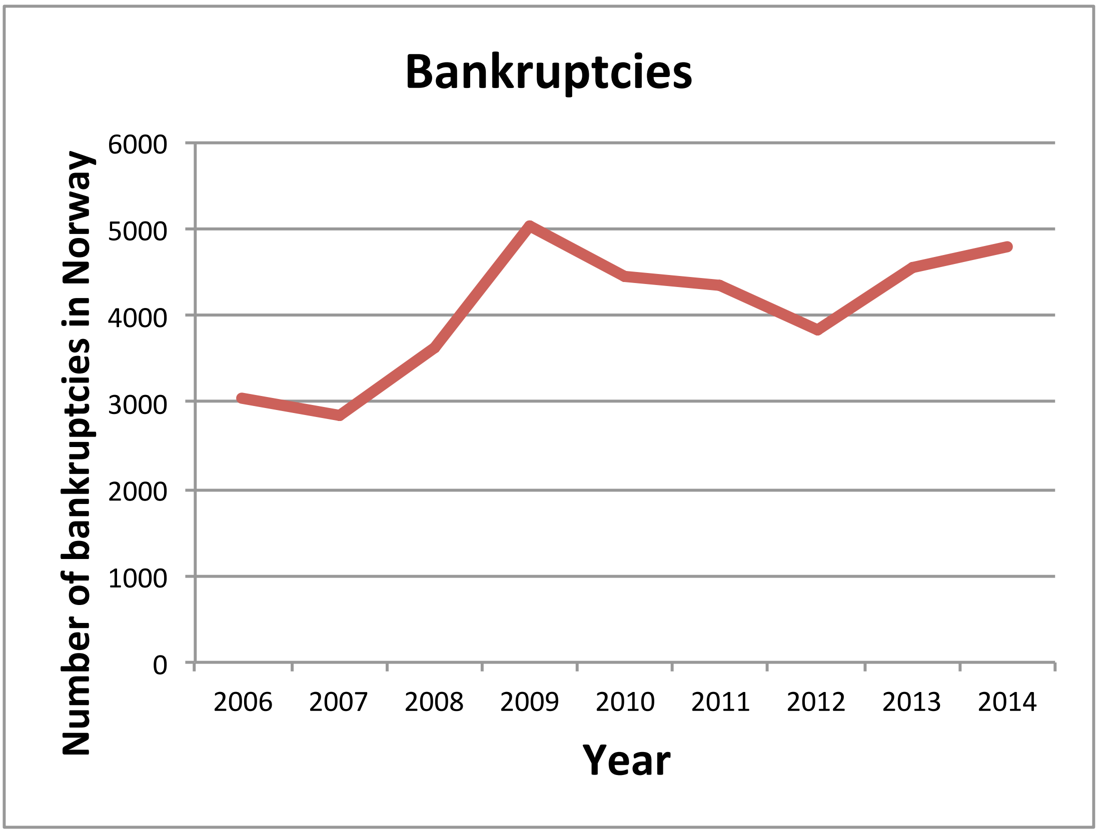
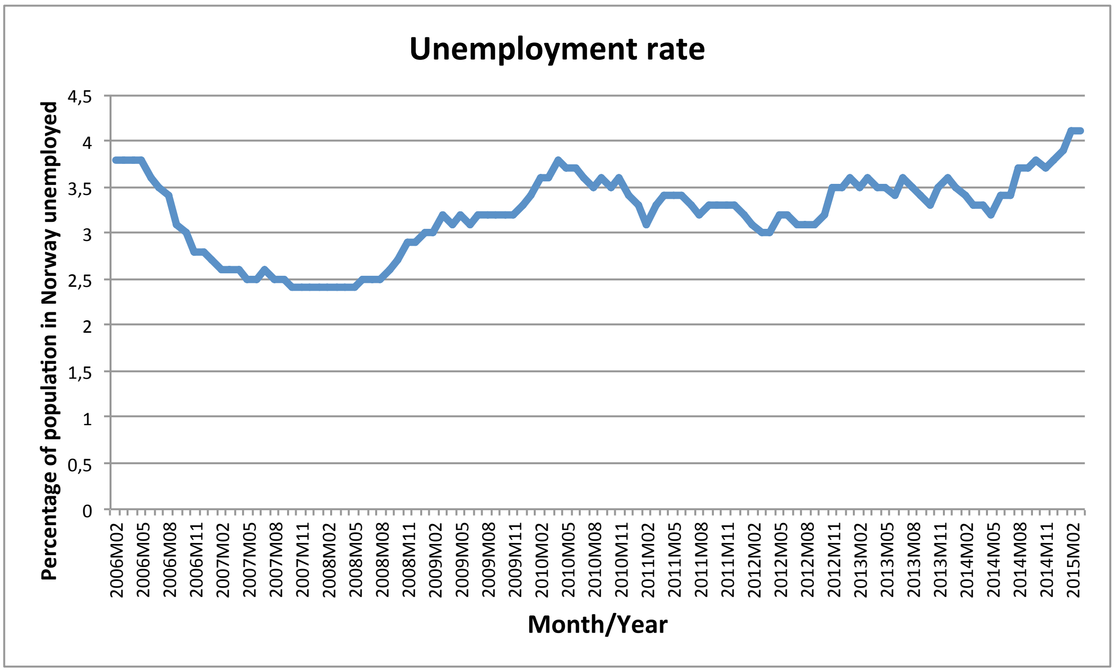


Figure 2: Unemployment rate Norway



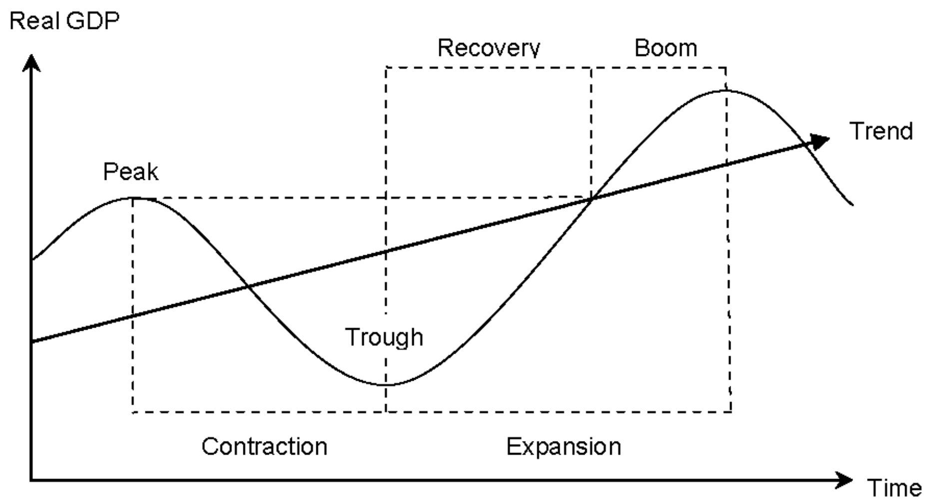
(Statistics Norway 1 & 2 (SSB), 2015)

The social cost of bankruptcy is related to the unemployment rate. If people are losing their jobs and are not able to find a new source of income they rely on the benefits that they get from the state. If many people are unemployed, it becomes a large expenditure on people that are healthy, fit to work and involuntary unemployed.

2.5 Cycles in the economy

There will always be fluctuations in an economy known as business cycles. There are periods of strong growth, and other periods with decreases in the growth. These periods are called recessions and expansion of the economy. The growth is in real terms (inflation adjusted) and can be measured by the GDP and unemployment rates. In recession periods, the general income falls. This is a negative effect for companies, as people have less available capital to spend. The income of businesses will therefore also be decreased in periods like this. It is especially in these difficult circumstances companies have to be alert of the risks of going bankrupt.

Figure 3: Business cycle



:

(Business cycles, n.d.)

3. Previous research

In this chapter, we want to give an overview of previous literature on the topic of estimating bankruptcies that we find relevant to our thesis. This will give an historical retrospective look at bankruptcy prediction. There have been many studies on this topic over the years, and we will present a selection of these. The literature is meant to be helpful in understanding our research and results. The studies that we have selected were chosen due to their different characteristics in the method used and variable selection. We want to develop our models based on a combination of these features, and supply them with our own ideas. The following literature is used as a base in our model building process.

3.1 Altman's discriminant analysis

In 1968 Altman constructed a model to predict bankruptcy with a multiple discriminant analysis (MDA). Altman found that ratios measuring profitability, liquidity and solvency were the most significant factors in predicting bankruptcy (Altman, 1968). By developing this model, the goal was to find what weights each of these factors should have when predicting bankruptcy, i.e. how much these factors impact the probability of going bankrupt.

The multiple discriminant analysis that was used in this study is a statistical technique for solving a two-class problem (Sundaram, n.d). MDA will find characteristics that are similar within the groups. Altman established two groups with 33 bankrupt firms in one, and 33 non-bankrupt firms in the other. This technique will then derive a linear combination of these characteristics which "best" discriminates between the groups (Altman, 1968 p. 592).

This model includes five financial ratios, and the MDA derives coefficients for each of the ratios. When inputting the financial numbers of a company in the model, it gives a total Z-score that implies the companies' probability of failure. If the models output is a Z-score below 1.81, the company have high probability of bankruptcy. A score between 1.81 and 3.0 is a "grey zone" and a score over 3.0 is a sign of a solid business.

Z is given by:

$$\mathbf{Z} = 1.2X_1 + 1.4X_2 + 3.3X_3 + 0.6X_4 + 1.0X_5$$

X_1 = Working capital/total assets

X_2 = Retained earnings/total assets

X_3 = EBIT/total assets

X_4 = Market value equity/ book value of total debt

X_5 = Sales/total assets

(Altman, 1968)

3.2 Ohlsons logistic regression model

Ohlson contributed to the research in 1980 with his logistic regression model (logit model) for bankruptcy prediction. Ohlson based his model on a larger sample of companies than Altman, with over 2000 individual companies in his dataset. Ohlson explains four different factors he found statistically significant in affecting the probability of failure 1) the size of the company, 2) a measure of financial structure, 3) a measure of performance and 4) a measure of liquidity (Ohlson, 1980).

Using the logit model for prediction, Ohlson was able to find a model that predicts a percentage probability of a firm going bankrupt, since the dependent variable of the logit model is binominal. Ohlson argues that this model gives statistically lower error-rates than previous studies, e.g. Altman (Ohlson, 1980, p. 111). Ohlson chose nine financial ratios that he includes in his model, and these were selected for simplicity reasons. In his study Ohlson pinpoints that for his subset the financial state variables (1-4) are uncorrelated with the performance variables (5-9), which supports the notion that both sets are important for the predictive relationship.

The model is given by:

$$Z = \beta_0 + \beta_1 * SIZE + \beta_2 * TLTA + \beta_3 * WCTA + \beta_4 * CLCA + \beta_5 * OENEG + \beta_6 * NITA + \beta_7 * FUTL + \beta_8 * INTWO + \beta_9 * CHIN + \varepsilon$$

SIZE = log (total assets/GNP price-level index)

TLTA = Total liabilities divided by total assets

WCTA = Working capital divided by total assets

CLCA = Current liabilities divided by current assets

OENEG = One if total liabilities exceeds total assets, zero otherwise

NITA = Net income divided by total assets.

FUTL = Funds provided by operations divided by total liabilities

INTWO = One if net income was negative for the last two years, zero otherwise.

CHIN =$(NI_t - NI_{t-1})/(|NI_t| + |NI_{t-1}|)$ where NI_t is the net income for the most recent period. The denominator acts as a level indicator. The variable intends to measure change in net income (Ohlson, 1980, p. 118).

3.3 Skogsvik- Financial stress

Kenth Skogsvik wrote a research paper in 1988 on predicting when companies are in financial distress. He has based his analysis on the probit model. This is similar to the logit model and it is also a binary prediction model. The output is a percentage probability of financial stress within a company. Included in this work is an analysis of inflation-adjusted figures, and this differs his work from previous research. The aim of his research was to investigate whether the inflation-adjusted numbers gave a better prediction when the accounting data was reported as historical costs. Skogsvik also developed his bankruptcy prediction model based on this.

This study was based on Swedish companies, and the data material included a set of 328 non-bankrupt firms and 51 firms in financial stress. The financial ratios that were included in this model were based on seven categories: 1) Profitability, 2) Costs, 3) Capital turnover, 4) Liquidity, 5) Asset structure, 6) Capital structure, 7) Growth. Within these categories, there are 79 primary variables that are included in his prediction. These are then tested with a component analysis, and non-statistically significant variables are excluded from the model (Skogsvik, 1988).

3.4 Bernhardsen- logistic bankruptcy prediction

Bernhardsen developed a bankruptcy prediction model based on the SEBRA database in 2001, and this is now used by the Norwegian central bank to predict bankruptcies. The SEBRA model predicts bankruptcies in the Norwegian market, and this model is based on key figures that are derived from accounting figures and information about the companies. This model is also derived as a logit model, giving a percentage chance of bankruptcy.

Bernhardsen used data from the years 1990-1996 in his estimation, and small firms with a book value of total assets less than 250.000 NOK were excluded. The sample size was 398.680 non-bankrupt firms, and 8436 bankrupt firms (Bernhardsen, 2001. p. 14). The key ratios included in the analysis are divided in 6 categories. 1) Liquidity, 2) Profitability, 3) Solidity, 4) Age, 5) Size, 6) Industry characteristics. Bernhardsen selected the variables because some were traditionally used in analysis of credit risk, and some based on trial and error (Bernhardsen, 2001. p. 17). His model provided an accuracy of 82-83 %.

3.5 Comparing previous literature

The previous literature provides a useful backdrop for us in developing our model. However, making a model that can control for fixed effects has seemingly not been done before in regards to estimating factors that contribute to bankruptcy. In a way, we are exploring new territories and thus we have very little research to compare our results against. The fixed effects approach will be compared to other estimation techniques that we find relevant, and then we can deduce which of the models will yield the better results.

We have found that researchers do their estimations using different approaches. Still, exploring what explanatory variables and categories they have emphasised is very valuable to us. The validity of the research within this field is considered strong and we rely upon it in selecting our predictors.

Altmans discriminant analysis is not directly comparable to the models we are developing, as he is using an approach that we are not including. However, the categories that he finds significant in terms of companies bankruptcies was relevant to our model. Other than Altman, all of the reviewed papers use a binary dependent variable for their estimations. This approach is used to such an extent, that in order to compare the performance of our own estimators we have also chosen to estimate our models with a binary dependent variable for bankruptcy.

The most important previous research compared to our model, is the one Bernhardsen did in 2001. This model is also based on a Norwegian data set, similar to ours. One difference between his study and our own is that we include companies off all sizes whereas he excluded companies less than a book value of 250.000 NOK. Bernhardsen did a bankruptcy prediction using the logistic regression. We are also including a logistic regression model in this thesis, as an addition to the other models, as we are interested in how our results compare.

4. Method

There are several models for predicting the possible insolvency of a firm. The most important ones according to Jackson & Wood (2013) being: (i) multivariate discriminant analysis (MDA) models (as proposed by Altman ([1968]), (ii) conditional probability models, the most popular being the logit (as used by Ohlson [1980] and Zmijewski [1984]), (iii) more recent models based upon artificial intelligent systems such as neural networks, genetic algorithms, case-based reasoning and recursive partitioning or (iv) models based on pricing theory such as Vassalou and Xing (2004). Jackson & Wood (2013) also assess the efficacy of thirteen selected models using post-IFRS UK data and investigate the distributional properties of model efficacy. They find that the efficacy of the models is generally less than reported in the prior literature. The thirteen different models are of course developed with a particular population in mind, which could explain these results.

Avenhuis (2013) tests the generalizability of some of the most used bankruptcy prediction models, specifically Altman (1968), Ohlson (1980) and Zmijewski (1984). In conclusion, he (Avenhuis) finds that "practitioners should use the bankruptcy prediction models with caution". This due to the fact that (1) the frequency of Type I errors is high (Ohlson [1980] and Zmijewski [1984]) or (2) the accuracy rate is low (Altman [1986]). To use these models in practice, he recommends to "re-estimate the coefficients of the bankruptcy prediction models with a specific and bigger sample to improve the predictive power", which is our intention with this paper. That is, we would like to identify some key variables that can explain the variation in Norwegian data. We also seek to make use of the panel-data structure of our dataset to identify something we think is equally important to the performance of these models, namely the effects of manegerial knowledge and financial management abilities we find specified in a paper by Thornhill & Amit (2003). They underline that newness and smallness are not the only reasons for a company's default. We will differ from the latter mentioned article in the way that we will not have the controls specified as variables (they use survey data to attain information on the knowledge of the firm) but rather control for firm-specific fixed effects to retrieve more unbiased predicted probabilities.

Considering that each firm has their own unique perception about when they default on their payments, simplified as $financial\ state \pm t$, we have that each firm is classified as bankrupt on the date of declaration of bankruptcy ± t depending on if the firm is considered risk averse or risk seeking. If risk averse → -t or if risk seeking → +t. If companies tend enough towards risk aversion they might declare bankruptcy at a point where others would go on and possibly survive. Now, we cannot say that for example boat constructors are particularly risk averse people without any qualitative knowledge about the industry. However, to rule out the uncertainty of when a firm really is bankrupt we still need to control for this when running analyses. Simply running regressions on the mean of this perception would bias the results, while controlling for firm-specific fixed effects intuitively would improve the estimation.

In order to describe the variation in our dataset in the best way possible, we use four different approaches to accommodate our panel data: Ordinary least squares, random effects, the fixed effects approach and lastly the logit estimator. In this chapter we will introduce each model before we later apply them to the data.

4.1 Ordinary least squares (OLS)

As mentioned, our dataset contains cross-sectional units *i* at the time interval *t*. As such, the data has a panel format. In the case that we want to estimate a response y_{it} for any explanatory variables $x_{it1}, x_{it2}, .., x_{itk}$ we need to be aware that we are estimating unobserved factors (error terms) of two types: A component that does not change over time a_i, and a component that does change over time, u_{it}. a_i is the unobserved effect, (also called unobserved hetoregeneity or fixed effect) and varies by each panel unit, i.e. by firm but not by time. This could be capturing characteristics of the firm like ability to run a business. The u_{it} is the idiosyncratic error term and while being specific to unit *i* this will vary over time, thus affecting the outcome y_{it}. Wooldridge (2014) exemplifies this in a regression model:

$$y_{it} = \beta_{t0} + \delta_0 d2_t + \beta_1 x_{it} + v_{it}, t = 1,2,..,T$$

Here one possibility is to use pooled OLS, thus ignoring the panel structure of our data. The composite error term is then: $v_{it} = a_i + u_{it}$. Because of a_i, the error terms v_{it} will be correlated. This is a serial correlation within panels, also called cluster correlation (each unit is a cluster), which causes bias in our OLS. This can easily be solved by clustering the standard errors. If OLS is to be consistent, we also require that x_{it} and the composite error v_{it} are uncorrelated.

Because $v_{it} = a_i + u_{it}$ we need

$$Cov(x_{it}, a_i) = 0$$

$$Cov(x_{it}, u_{it}) = 0$$

The first of these is violated if x_{it} is determined based on systematic differences in units. When $Cov(x_{it}, a_{it})$ is not equal to zero it is often said that pooled OLS suffers from heterogeneity bias (omitted variable bias) due to not incorporating the fixed effects. Endogeneity, i.e. $Cov(x_{it}, u_{it})$ different from zero, occurs when there is a correlation between an independent variable x_i and the error term u in a regression (Wooldridge, 2014). If the dependent variable is correlated with the error term, the OLS estimation is biased. A consequence of not eliminating endogeneity is that the model could lead to spurious results.

Next, we comment on the logistic regression, before we describe different ways of correcting for the omitted variable bias.

4.2 Logistic regression (logit)

The logistic regression is a technique that is often used when the explanatory variable of a model is binary. The binary dependent variable takes on two values, e.g. 0 or 1.

$$y = \begin{cases} 1 \; if \; yes \\ 0 \; if \; no \end{cases}$$

This will model the probability of the outcome being 1. The probability estimation of this model will be between 0 and 1.

$$P(y = 1|x) = \; P(y = 1|x_1, x_2, \ldots, x_k)$$

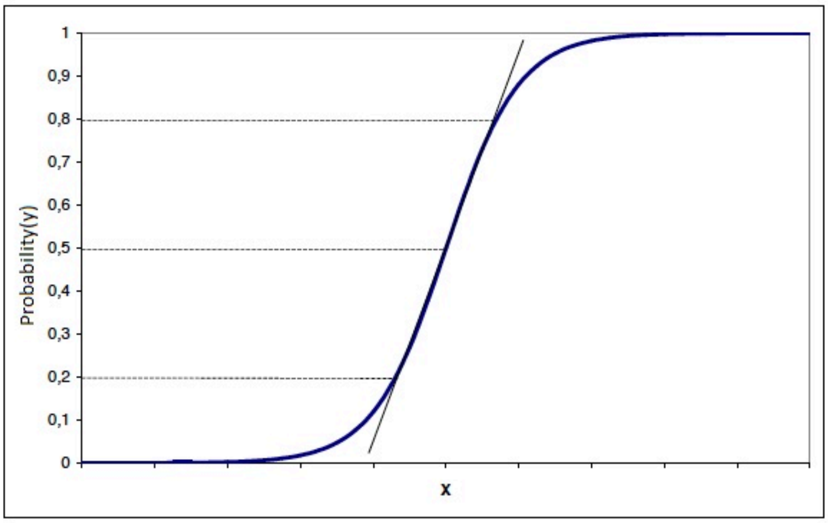
(Wooldridge, 2014).

The probit model is a similar form of model, which also produces very similar results to the logit model. The difference between the logit and the probit model is the assumptions we make on the standard errors. In the logit model, the cumulative distribution function is of the standard logistic distribution. For the probit model, we assume the standard normal cumulative distribution (Wooldridge, 2014). In this thesis we are not reporting probit results and the theory described will hence be related to the logit model.

The logit model parameters are estimated using the *maximum likelihood method.* The logit model will solve a function to maximize the probability of the observed y-values (0 and 1) (Tufte, 2000). This maximizing problem is developed to find the coefficients that give the highest probability of a correctly estimated dependent variable.

The logit-model is a non-linear model. The relationship of a dichotomous outcome variable plotted against a continuous independent variable will show the shape of the function in a logit model. The S-shaped curve is portraying this relationship.

Figure 4: The S-shaped curve



(Tufte, 2000)

The probabilities in a logit model will never be below 0 or above 1 (Tufte, 2000), as they will in a linear probability model. In order to transform the logistic regression so that it is able to take on all values, we have to convert the variables into odd ratios and the log of odd ratios. The odd ratios will remove the upper limit of one, and the log of the odd ratios will remove the lower limit of 0 (Tufte, 2000).

Tufte (2000), explains the *odds ratio* in a logistic model in the formula:

$$Odds = \frac{p}{(1-p)}$$

The odds ratio measure how large the probability of y=1 is relative to the probability of y=0. It measures the relative risk of the logit model. The logistic regression coefficient output is the natural logarithm of the odds ratio (Peng, Lee & Ingersoll, 2002).

The log of the odds ratios is called the logit, and is expressed by the following formula (Tufte, 2000)

$$L = \ln\left(\frac{p}{1-p}\right)$$

The interpretation of the logit model is that an increase in the x variable will make the probability of y more or less likely. We are able to interpret the sign of the coefficient, telling us whether the probability will increase or decrease, but we cannot directly interpret the magnitude of the estimated coefficients. To correct for any fixed effects we need to introduce different estimators, starting with random effects.

4.3 Random effects (RE)

We start with the equation written for a unit *i*:

$$y_{it} = \delta_t + x_{it}\beta + a_i + u_{it}, t = 1,2,..,T$$

Where $x_{it}\beta = \beta_1 x_{it1} + \beta_2 x_{it2} + .. + \beta_k x_{itk}$. The δ_t *he*re represents different time intercepts.

Random effects assumes that the unobserved effect is random. In other words; that the composite term $v_{it} = a_i + u_{it}$ is uncorrelated with *x*, thus that a_i is uncorrelated with *x*. The other side to this is that RE is consistent only under this assumption. If this assumption holds, RE is the most efficient way of correcting for the omitted variable bias (Wooldridge, 2014). We will later show that a fixed effects transformation (FE) removes the a_i from the estimating equation in order to get consistent estimates of β in spite of correlation between a_i and *x*,.

RE can include time-constant observed controls (these are differenced or time-demeaned away with FE). With good time-constant controls, RE may be convincing as more is taken out of a_i as we add time-constant variables (Wooldridge, 2014). The RE estimators also assume no serial correlation and no heteroscedasticity in the error term u_{it}.

Heteroscedasticity arises when the error term u has a not consistent variance given any value of the explanatory variable: $Var(u|x) \neq \sigma^2$. Heteroskedasticity does not cause bias or inconsistency in the estimation. However, it will cause the standard errors to be invalid for constructing confidence intervals and t-statistics (Wooldridge 2014, p. 213).

If the assumptions about serial correlation and heteroscedasticity fail we use cluster-robust inference as for the other models.

The RE estimator is a feasible generalized least squares (FGLS) procedure. It uses the fact that under the RE assumption the composite error term is serially correlated in a particular way, giving rise to the following correlation between errors in period t and s (Wooldridge, 2014 p.396):

$$Corr(v_{it}, v_{is}) = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2} = \rho$$

The RE transformation similarly removes the serial correlation in the composite error terms;

$$\theta = 1 - \left[\frac{1}{1 + T(\sigma_a^2/\sigma_u^2}\right]^{1/2}$$

Where $\sigma_a \; and \; \sigma_u$ can be found from FE-estimates (see chapter 4.3).

As such, the RE estimates can be obtained from the pooled OLS regression:

$$y_{it} - \hat{\theta}\bar{y}_i \; on \; x_{it} - \hat{\theta}\bar{x}_i, t = 1,..,T; i = 1,..,N.$$

Where $y_{it} - \hat{\theta}\bar{y}_i$ is called a «partially-time-demeaned» variable. Because θ lies between zero and one, only a fraction of the mean is in fact removed (Wooldridge, 2014);

$$\hat{\theta} \approx 0 \; \leftrightarrow \; \hat{\beta}_{RE} \approx \hat{\beta}_{POLS}$$

$$\hat{\theta} \approx 1 \; \leftrightarrow \; \hat{\beta}_{RE} \approx \hat{\beta}_{FE}$$

We note that θ is close to one when σ_a^2/σ_u^2 is large (the variance of the unobserved part of the composite error term is "dominating") or when T is large. With large T, there should be small or no difference between the FE- and the RE-estimator.[1]

[1] Stata does this transformation in a similar way to the fixed effects transformation, but with a different subscript.

4.4 Fixed effects estimator (FE)

The fixed effects estimator, also called “within-transformation” removes the within i time averages. Another name for it is “time-demeaning” of the variable, which is fitting, seeing how the transformation demeans away any observation in our regression that is constant (Wooldridge, 2014 p.387)

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it}$$

We average this equation across t for each unit i to get

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_i + a_i + \bar{u}_{it}$$

Which we then subtract from other time periods:

$$y_{it} - \bar{y}_i = \beta_1(x_{it} - \bar{x}_i) + (u_{it} - \bar{u}_i)$$

Where $\bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it}$ is a “time average” for unit i.

We then use OLS on the deviations from time averages to estimate β_1, the FE estimator.[2] The u_{it} might have serial correlation (and heteroscedasticity), and so we use cluster robust inference as with the RE estimator. The FE estimator requires strict exogeneity, i.e. that the x_{it} are uncorrelated with u_{is} for all s (Wooldridge, 2014).

Computing pooled OLS and FE estimators can be informative and we will use both in applications. If OLS is different from FE, it indicates explanatory variables correlated with a_i.

[2] We do not do the within transformation manually: Stata does this for us with the xtreg command.

4.5 Discussion of the estimators

In this introduction of our models, we have described four estimators that we will apply to our dataset. Firstly, we explain the pooled OLS estimator (OLS), which we do on the levels. Next, we comment on the logit as a possible solution to the linear probability constraint. Then, we elaborate on the random effects estimator (RE) which essentially is OLS on the partially time-demeaned variables. Lastly, we elaborate on the fixed effects estimator (FE), which is a OLS run on the time-demeaned variables.

OLS on the levels is usually deficient, unless we include things like lagged y which is not allowed in the other methods. With good controls and lags of y in OLS, we might be able to make a convincing analysis. On the other hand, if we believe an unobserved fixed effect is important, we do prefer the other estimators.

Due to the above-mentioned fact that FE does not consider time-constant variables RE is always more efficient than FE provided that the RE assumption $Cov(x_{it}, a_i) = 0$ holds. Given this, RE and FE will be similar if T is large. If RE and FE estimates are very different, we want to know if we can use the more efficient RE, or whether we must reject RE in favor of FE.

As is the case for us later on, we can test for RE versus FE using the Hausman-test. The basic intuition of this test is that under the main assumption of the RE-estimator ($Cov(x_{it}, a_i) = 0$), both RE and FE are consistent estimators. However, if only RE is efficient the null hypothesis in the Hausman-test is that the above assumption actually holds, i.e. that there is no systematic within i variation. As a consequence, if the p-value of this test is small, we reject the null and prefer the FE over the RE-estimator.

5. Data material

5.1 Sample selection

This thesis is an empirical study based on a dataset of Norwegian accounting figures. The data material was originally separated by years and divided in to two categories, based on accounting data and general information about each company. We decided to merge the years and the two categories so that we are able to perform our analysis on the entire range of companies in Norway. As a result, we were left with panel-data, each panel having information for both the separate sets, distributed across as yearly observations. This way we are able to gather a large amount of information for each firm. Before we were able to start with our analysis, we had to clean the dataset so that the remaining data were relevant for our model.

The dataset includes data spanning from 1992-2012. However, in 1998 changes were made in the Norwegian legislation. The accounting act of 1998 presented new tax regulations for group contributions between companies within the same group. It also changed the requirements for which companies that had to submit consolidating accounts (Berner, Mjøs & Olving, 2014 p.1). This radically changed the company structure for many businesses in Norway, making it difficult to compile data across panels before this point.

Further, IFRS (International Financial Reporting Standards) rules were implemented from 2005 and onwards, making all listed companies in Norway required to report after this standard. Reporting for the period between 1998 and 2005 therefore differs from the ones after in regards to listed companies, enough to make us focus on the period 2005-2012.

In our dataset, we have a variable that takes on the value of the year a company goes bankrupt. Specifically it lists the year the liquidation proceedings starts (Berner, Mjøs & Olving p. 26). In transforming this to a binary variable we first had to change the year format, starting by taking into consideration that every panel contained this information. As such, we had to make a rule that only allowed one observation of the actual bankruptcy per firm. There were also observations of firms that had filed for bankruptcy more than once.

There is variation between companies as to whether the bankruptcy date is set one year or two years after their last active year of operations. In defining our new variable, we made the bankruptcy occur one year after the last active year of operations for all companies. Since the Norwegian legislation states that a bankruptcy proceeding starts when a company is insolvent, we found it appropriate to define the variable like this. As a result, we could now generate a binary variable across all panels for the bankruptcy observations, such that y_{t+1}.

5.2 Descriptive statistics

Our dataset consists of almost 2 million separate panel observations on the accounting data of Norwegian firms in the period 2005-2012. Because of the bankruptcy definition used, we end up with some 18194 firms that have gone bankrupt during this period, but distributed on the years 2006-2013 as shown below.

Table 1: Number of bankrupt and non-bankrupt firms in the data set

Year	Non-Bankrupt	Bankrupt	Total	Bankruptcy rate
2006	169,087	1,637	170,724	0,96 %
2007	199,666	1,501	201,167	0,75 %
2008	218,315	3,196	221,511	1,44 %
2009	230,938	2,662	233,600	1,14 %
2010	234,861	2,545	237,406	1,07 %
2011	238,247	2,148	240,395	0,89 %
2012	246,317	2,342	248,659	0,94 %
2013	261,086	2,163	263,249	0,82 %

During the merging of accounting and company data we lost 1418 observations, mostly due to lack of company information available, as some companies were not listed in this registry. It is assumed however, that these companies were too small to have had any implication on the analysis since they either have not started up properly, or refrained from financial activity and none of them ever declared bankruptcy.

Whilst describing the data, we sensed that younger firms seemed more susceptible to financial distress. There are also many authoritative articles describing this “liability of adolescence” (Freeman et al. 1983), (Carroll, 1983) and (Sorensen and Stuart, 2000). Carroll (1983) describes exit rates that decline monotonically to a positive asymptote, which we can see from appendix A.1 is the case for our distributions as well. As we take a closer look on bankruptcies distributed by age, we observe that almost 30% of the bankrupt firms in our study go out of business during their first year, declining in a positive asymptote.

Further, a regression on age categories (Table 2) also confirms this distribution and the way it affects the risk of a company going bankrupt. For a firm between 5 and 9 years there is a negative and significant effect on going bankrupt (in regards to the 0-4 category). This effect increases in magnitude as a company ages, and it seems being an older firm means you are less likely to fail.

Table 2: Regression on age categories

	(1)
	Bankrupt
Age 5-9 years	-0.00633***
	(0.000198)
Age 10-25 years	-0.00942***
	(0.000176)
Age 26+ years	-0.0118***
	(0.000296)
Constant	0.0154***
	(0.000120)
R^2	0.002
Observations	1816540

Standard errors in parentheses
$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

However, as previously discussed newness and smallness are not the only reasons for bankruptcy. Even older firms can default on their obligations given a pattern that exists prior to the bankruptcy itself, which is what we seek to identify.

The bankruptcies in our dataset does follow this even distribution, except for the years 2008-2010 during which the financial crisis and its repercussions created a spike in firms that could not fulfill their financial obligations. We can discover this by looking at the more uneven histogram in appendix A.2 and comparing it to the one in A.1. We spot that older firms are more exposed during the financial crisis, which implies that some firms are more likely to fail regardless of age given enough exposure. However, well over 20 % of the firms are still in their first year of business. As we later analyse the data, these patterns should add to our explanation albeit with the risk of overstating the effects. The simple fact remains that there exits characteristics within firms that can make them more exposed to a default, and these factors should thus be evaluated.

Amongst the other points of interest as to what can explain why a company cannot fulfill their financial obligations is the particular industry it operates in and whether this has some special characteristic that makes businesses within it different or more risk seeking. Appendix A.3 shows a spread of bankruptcies by industry, where we can clearly see that some industries are more exposed.

Interestingly, category 30 is the category for office machinery and computers, while 55 is the category for hotels and restaurants. Office machinery and computers has the largest percentage of bankruptcies by far, which initially sounds strange, as there were other categories we expected had a higher rates, such as the production of boats. However, especially the wholesale and manufacture of computers can be very resource intensive, so it is possible to imagine. In either case, there could be an added risk that is industry dependent, unless there is culture within the industry for sloppy accounting/management. In any event, this gives raise to the intuition that industry specific effects should be taken into account during the analysis.

Appendix A.4 is a bar chart to illustrate the difference between big and small firms. This difference seems to be consistent, i.e. larger firms are less likely to go bankrupt. This might be due to larger savings, better bank relations or continuous demand for their product, the fact remains that size should matter.

All of the above certainly suggests that newness and smallness does have large impact. Keeping this in mind, we seek to identify the remaining controls for what determines a bankruptcy.

5.3 Choosing variables and key figures

When deciding which variables to use as input in our model, we base our choices on extensive analysis of previous research, combined with own ideas. The variables are key figures and financial ratios that we have developed from the data material. We are presenting most of our variables as ratios that are comparable across companies. Starting by choosing key variables from a variety of literature, mainly Skogsvik (1988), Altman (1968) and Ohlson (1980), many financial measures were first identified. These were then reduced to 15 by identifying those relevant to the economic framework a Norwegian firm must relate to. We also chose 5 firm-specific characteristics that all have been described as determining factors, size, age and industry being the most noteworthy.

To ensure that we include variables from different categories of accounting and company data, we are using Bernhardsen's main categories from his research paper (2001) as a framework. The categories also follow a similar classification in the previous research that we have described.

In addition to this, we include a category related to growth similar to Skogsvik (1988). This way we are able to examine if expanding a business leads to a higher or lower risk of bankruptcy. Both Skogsvik (1988) and Bernhardsen (2001) find that growth ratios significantly predicts corporate failure.

Table 3 presents an overview of the categories and the key figures that are included in each of them, then we make an in depth review of each variable.

Table 3: Overview of key-ratios and dummy variables

Liquidity

Liquidity ratio 1

Liquidity ratio 2

Liquidity ratio 3

Profitability

Profit margin 1

Profit margin 2

Return on equity (ROE)

Return on assets (ROA)

Net loss dummy (Y)

Solidity

Solidity 1

Solidity 2

Debt ratio 1

Debt ratio 2

Solidity dummy (X)

Age

Dummy variables on the age of the company

Size

Dummy variables on size of the company

Industry characteristics

Dummy variables on each industry

Growth

Growth revenue

Growth assets

Growth equity

Growth current liabilities

5.3.1 Liquidity

A company is insolvent if they are not able to meet their obligations in time. Having enough liquidity is therefore important to avoid insolvency and possibly bankruptcy. Liquidity ratios measure how much cash and liquid assets a company holds. The need for liquidity in a business will depend on the business type, however we find liquidity ratios relevant to include in our model. Appendix A.5 contains the means of the liquidity measures. These show that bankrupt firms have consistently lower ratios of liquidity as defined in Liquidity ratio 2 and 3. Surprisingly, bankrupt firms have higher ratios of L1.

Liquidity ratio 1

$$\boldsymbol{L1 = \frac{Cash + current\ assets}{Total\ assets}}$$

Liquidity ratio 1 is a measure of how large share of the companies' assets is liquid. This is an important ratio to get an indication of how liquid the business is. If this ratio proves to be significant in the model, this is a ratio the companies can change themselves. A company with a small percentage of their assets as liquid may consider changing the structure of their assets if they are at risk of failure.

Liquidity ratio 2

$$\boldsymbol{L2 = \frac{Cash + current\ assets}{Short - term\ debt}}$$

Meeting short-term liabilities is crucial for a company. Liquidity 2 gives information on the ratio between liquid assets and debt that is due within one year. A low L2 may be an indicator that the company struggles to meet their liabilities.

Liquidity ratio 3 (Altman, 1968)(Ohlson, 1980)

$$\boldsymbol{L3 = \frac{Working\ capital}{Total\ assets}}$$

The working capital is defined as current assets - current liabilities. Liquidity ratio 3 is also a measure of how liquid a business is, and it is incorporated the current liabilities. These liabilities will impact the companies' financial state, and the capacity to meet short-term debt.

5.3.2 Profitability

In the long run, a company has to be profitable to be an attractive business. A profitable business will attract more investors. Having investors interested in the company could lead to a larger market value if shares are sold, and hence raising more capital. More capital can be used to expand the company, or to meet current obligations. If a company is leveraged, they need a certain profitability to be able to handle these obligations. The ratios used in this category are commonly known profitability ratios that are often used to measure the performance and to compare companies. A.6 contains a spread of the Means of the profitability measures. Bankrupt firms have higher values of P1, which we later discuss in chapter 6. Otherwise, non-bankrupt firms have higher profitability margins.

Profit margin 1 (Altman,1968)

$$\boldsymbol{P1} = \frac{\boldsymbol{Sales\ income}}{\boldsymbol{Total\ assets}}$$

Profit margin 1 is a standard measure for total asset turnover that varies greatly from industry to industry.

Profit margin 2 (Altman, 1968)

$$\boldsymbol{P2} = \frac{\boldsymbol{EBIT}}{\boldsymbol{Total\ assets}}$$

Profit margin 2 measures the profitability of the company compared to the amount of assets they own. EBIT is earnings before interest and tax and equals operating result + other income. This ratio is similar to one used in Skogsvik from 1988, except he only included the operating results/total asset. This ratio includes “other income“ and is somewhat more informative as it encapsulates more about the total profitability of the company.

Return on equity (ROE)

$$\boldsymbol{ROE} = \frac{\boldsymbol{Net\ income}}{\boldsymbol{Equity}}$$

This ratio indicates how much income the company generates compared to how much equity the shareholders have in the company. A high ROE means that the company is profitable in terms of creating value from the amount of equity available. This ratio does not give information on how much debt the company has. This may impact the net income, as debt gives possibilities for larger investments. We have included this feature in the next variable, ROA.

Return on assets (ROA)

$$\boldsymbol{ROA = \frac{Net\ income}{Total\ assets}}$$

ROA is a standard measure of how much earnings the company is able to get compared to how much assets they hold. Compared to the profit margin variable, the net income also includes interest expenses/income and taxes. This ratio was also used in Ohlson's model from 1980, but with a different variable name. This variable measures the profitability over both equity and debt capital.

Net loss dummy (Y)

$$\boldsymbol{Y = \begin{cases} 1\ if\ net\ loss\ for\ two\ years \\ 0\ otherwise \end{cases}}$$

A dummy to indicate if a firm runs deficits for two consecutive years.

5.3.3 Solidity

The intention of this category is to include a measure of how solid the companies are. The solidity of a business is an indication of how well they are able to handle changes in external circumstances in e.g. cyclical downturns in the economy. We are including variables that describe the leverage situation in the companies in this category. A change in e.g. interest rates will have a higher impact on companies that are more financed with debt. A.8 in the appendix describes the mean of these solidity measures, bankrupt firms being consistently less solid and perform worse in terms of leverage.

Solidity ratio 1 (Skogsvik, 1988)

$$\boldsymbol{S1 = \frac{Equity}{Total\ assets}}$$

Solidity ratio 1 is a variable that gives us information on how much of the total capital in the business is owned by the shareholders. The leveraged part of the capital may impose a risk if the macro economical situation change.

Solidity 2

$$S2 = \frac{mortgaged\ assets}{total\ tangible\ fixed\ assets}$$

A ratio that describes how much of a firm's assets that are considered as collateral by the creditors.

Debt ratio 1

$$D1 = \frac{Total\ debt}{Total\ assets}$$

Debt ratio 1 is the counterpart of Solidity 1. Total assets are a combination of equity and debt. The debt ratio is relevant in a bankruptcy prediction model, as the debt imposes a risk on the business. The debt is also a capital gain that the company could use to pay off their liability.

Debt ratio 2

$$D2 = \frac{Net\ income}{Total\ debt}$$

Debt ratio 2 provides information on how much income the business generates compared to the amount of debt they have.

Solidity dummy (X)

$$X = \begin{cases} 1\ if\ Total\ debt > Total\ Assets \\ 0\ otherwise \end{cases}$$

Typically zero, this dummy indicates whether or not a company is "solid". A value of 1 means the company is technically bankrupt and should in theory default on their obligations.

5.3.4 Age

We find that there is a relationship between the age of the company and when it goes bankrupt. It seems like younger companies more often fail than older companies. Due to this we want to include a variable in the model that can capture this.

Dummy variable of the age of a company

Based on our variable in the dataset of the companies starting year, we have developed age intervals as dummy variables. We have divided the age into four intervals in a similar manner to Olsen & Øien (2009):

- 0-4 years
- 5-9 years
- 10-25 years
- 26 → years

5.3.5 Size

Our hypothesis is that the size of the company may have an impact on the probability of going bankrupt. To investigate this, we are including a size variable in our model.

Dummy variable of the size of a company

The European commission is basing their definition of small companies on three factors: 1) Number of employees in the company, 2) Turnover, 3) Balance sheet total (European commission, 2003).

Since the EU area is very different in size compared to the Norwegian market, we have decided to use lower rates on each of the factors than the EU. We have defined our variable "small" as a dummy variable that takes on the value 1 if:

1. The company have ≤ 50 employees and turnover is ≤ than 70.000.000 NOK per year and/or
2. Balance sheet total ≤ 35.000.000 NOK

5.3.6 Industry characteristics

The industry a company operates in is affecting many aspects of the business. Industries are related to the market, and political and economical changes may affect them differently. Regulations and competition that are directly connected to specific industries will also affect companies in the given industry. Our descriptive statistics gives us an indication of which industry that is most prone to bankruptcy. A variable that includes which industry a company is operating in is therefore included in our model.

Dummy variable of industry

Our dataset contains a numeric variable for SIC2002 industry codes (A.3) where the numbers represent different industries. We developed dummy variables for each of the industries and included them in our model. This way we are able to see if being in a certain industry affects the probability of going bankrupt.

5.3.7 Growth

A business in a growing state is often vulnerable due to large investments and unsecure future results. Businesses in their early state are often exposed to these risks of expanding. A category of variables that capture this is therefore included in our model to explore if the growth leads to a higher or lower probability of failure. A.9 in the appendix describes mean ratios for bankrupt and non-bankrupt firms. The former growing less all over, except for the measure on growth in revenue, where bankrupt firms show a spike in growth for later years.

Growth revenue (Skogsvik, 1988)

$$\boldsymbol{G1} = \frac{\boldsymbol{\Delta Income}}{\boldsymbol{Income\ previous\ year}}$$

Growth assets (Skogsvik, 1988)

$$\boldsymbol{G2} = \frac{\boldsymbol{\Delta Assets}}{\boldsymbol{Assets\ previous\ year}}$$

Growth equity (Skogsvik, 1988)

$$\boldsymbol{G3} = \frac{\boldsymbol{\Delta Equity}}{\boldsymbol{Equity\ previous\ year}}$$

Growth current liabilities (Skogsvik, 1988)

$$\boldsymbol{G4} = \frac{\boldsymbol{\Delta Current\ liabilities}}{\boldsymbol{Current\ liabilities\ previous\ year}}$$

Table 4: Descriptive statistics of key ratios

	mean	p50	p10	p90	min	max
Liquidity ratio 1	.9451953	.9952105	.0577957	1.929078	-3117	1444
Liquidity ratio 2	43.68965	2.309154	.4078883	25.54545	-383260.7	1467595
Liquidity ratio 3	-.9228768	.436553	-.1505945	1.528292	-202398	19945
Profit margin 1	1.528903	.1619335	0	3.452365	-841	19585
Profit margin 2	-.3507728	.0290608	-.2393423	.3850686	-19282	19497.4
Return on equity	.4186543	.126995	-.3609467	1.141949	-59266	33751.25
Return on assets	-.3700188	.0239502	-.2203978	.3112033	-19282	19497.4
Solidity ratio 1	-2.431669	.3286214	-.1219008	.9703704	-202397	19946
Solidity ratio 2	.0000395	0	0	0	0	18.5
Debt ratio 1	3.431526	.6713062	.0295567	1.121844	-19945	202398
Debt ratio 2	2.657801	.0445691	-.3574661	1.014	-736359	1424964
Growth revenue	1.936119	.0292398	-.516129	.7811745	-14475.71	129291
Growth assets	3.721158	.0075361	-.3181818	.6717325	-10372.83	194136
Growth equity	1.633112	.0309278	-.6801872	1.063588	-138484	188140
Growth c.liabilities	12.29643	.0179196	-.69967	1.684685	-70047.5	1299999
2-year deficit	.1827458	0	0	1	0	1
Debt > Total assets	.1441042	0	0	1	0	1
Observations	1816540					

6. Results

This part of or thesis is assigned to present our results, and our analysis of the results.

The very first step was to test all the primary variables that we selected. Eliminating a number of the primary variables was necessary because we wanted the model to be manageable, with not to many variables. It was also important to have variables that did not correlate strongly with each other. If we were to have multicollinearity present the overall estimation would be correct, but the partial effects are difficult to estimate (Wooldridge, 2014, p. 262). Since we wish to estimate the impact of these partial effects on the probability of failure, and in particular if the estimations differ across the different approaches, we had to address this issue. The process of downsizing the number of independent financial measures has made the model more solid in respect to this and trial and error testing of the primary variables enabled us to identify ratios and key figures that determine a bankruptcy.

Estimations of the OLS, logit, RE and FE models are reported. Effects of the control variables are listed in the appendix. For consistency, effects are reported and discussed in percentage points.

6.1 OLS

The OLS results are shown in appendix A.10. First, we include the right hand variables stepwise to see if there are any relationships we should be aware of as we expand the model from:

$$(1)\ y_{it+1} = \beta_0 + \beta_1 * Profitmargin1_{it} + \beta_2 * Liquidityratio1_{it} + \beta_3 * Solidityratio1_{it} + \beta_4 * Solidityratio2_{it} + a_i + \mu_{it}$$

$\vdots$

$$\begin{aligned}(5) y_{it+1} = \beta_0 &+ \beta_1 * Profitmargin1_{it} + \beta_2 * Liquidityratio1_{it} + \beta_3 * Solidityratio1_{it} \\ &+ \beta_4 * Solidityratio2_{it} + \beta_5 * 2year\ deficit_{it} + \beta_6 * Size(1\ if\ small)_{it} \\ &+ \beta_7 * Age(5{\sim}9)_{it} + \beta_8 * Age(10{\sim}25)_{it} + \beta_9 * Age(26{\sim})_{it} + \beta_{10} \\ &* Growth\ revenue_{it} + \beta_{11} * Growth\ assets_{it} + \beta_{12} * Growth\ equity_{it} \\ &+ \beta_{13} * Growth\ current\ liabilites_{it} + a_i + \mu_{it}\end{aligned}$$

Secondly, we control for year effects in column (6-8), industry effects (7-8) and year×industry in (8) which is is exemplified as:

$$y_{it+1} = \beta_0 + \beta_1 x_{1,it} + \cdots + \beta_k + x_{k,it} + e_2 E_2 + \cdots + e_n E_n + \delta_2 T_2 + \cdots + \delta_t T_t + a_i + \mu_{it},$$

where y_{it+1} is the dependent variable for bankruptcy (set one year after the last registered year of operations). We have that I = entity and t = time. $x_{k,it}$ represents the independent variables that are named in regression (5), β_k is the coefficients for the independent variables, $a_i + \mu_{it}$ is the error term, E is the entity n. Since the industry variables are binary we have n-1 entities included in the model. e_n is the coefficient of the binary regressors (entities). T_t is the binary variable for time, such that we have t-1 time periods. δ_t is the coefficient for the binary time regressors. The intercept in the first (base) period is β_0 and that for the next period is $\beta_0 + \delta_2$.

The year dummies allows the intercept to change over time accounting for any general shifts (i.e. trends) in our data. The industry dummy captures this effect for industries while the interaction term allows the slopes to change over time.

In appendix A.10 we report the results of linear regressions on our bankruptcy variable. As well as being used for forecasting, the simple linear regression model is valuable in evaluating the historical effects of the predictors. To do this we use normal statistical inference methods. We clarify if there is any identifiable effect of x on y, i.e. if there is enough evidence to suggest a relationship to, in our case, the probability of going bankrupt. The test is a simple test of the slope parameters, for us this is equivalent to saying:

$$\mathrm{H_0}: \beta_{\mathrm{it}} = 0$$

Then preferring the alternative $\mathrm{H_A}: \beta_{\mathrm{it}} \neq 0$ wherever the coefficients are significantly different from zero. For lower p-values the observed data are extremely unlikely to have come to be if the null hypothesis is true and so we choose our initial variables based on the existence of a likely relationship. A requirement of doing this "pooling" on data is to fix the serial correlation of $v_{i1} and\ v_{i2}$ etc, in the composite error term: $v_{it} = a_i + u_{it}$ (Torres-Reyna, 2007). This problem is solved by using robust standard errors from column (2) and onwards.

Initially, in column (1) we identify the significant effects of *Profit margin 1, Liquidity ratio 1*, *Solidity ratio 1* and *Solidity ratio 2*. The values of the coefficients are altered once we control for a 2-year deficit in column two (2). The profitability and liquidity ratios both do not change significantly in columns (3) through (7) which shows that the figures are independent from each other. However, when we include the industry effects the profit margin decreases in magnitude and the sign on liquidity ratio now has an impact of -0.00134 percentage points. A change like this upon including the industry effects suggests that there are industry specific effects and that these are also creating biased results, underlining the intuition that we should control for them. Intuitively, increasing either the profit margin or the liquidity of a firm should not increase the probability of going bankrupt. As to why *Profit margin 1* never changes sign it might make sense to increase sales, but probably not to scale down business while trying to avoid bankruptcy. After all, the ratio is sales/total assets and scaling down the assets alongside the sales of a firm is captured in this ratio. And with fewer assets a firm will also be more exposed to failure.

Solidity ratio 1 has a low effect, and is never significant throughout this analysis. This ratio is an important measure of the firm's equity, and we are reluctant to drop it from the model. We are keeping it to see if the same results are also the case for any of the other estimators.

Solidity ratio 2 has an effect of -0.00165 percentage points (2) increasing in magnitude for each step ending up at a -0.00205 percentage point change in the bankruptcy probability for a unit change in the ratio.

Column (2) adds the dummy *2-year* deficit. At the beginning of the variable inclusion, we were also working with another dummy variable controlling for cases where debt was larger than total assets. At first glance these dummies, *2-year deficit* and *Debt > Total assets,* were somewhat correlated, but the model was able to separate each effect without the remainder of the model changing. However, controlling for a technical bankruptcy when predicting bankruptcies seemed superfluous, so *Debt > Total assets* was dropped from further analysis also removing any bias from including both dummies. Both variables suggested evidence that their coefficients are significantly different from zero and we conclude that the null is not preferable in their case, thus we kept the dummy that controlled for successive deficits for later analysis, and that result is included in column (2) and onwards. Its impact on the probability of going bankrupt is 0.0218 percentage points for a change in the dummy for two consecutive years of running a deficit. This value increases somewhat when adding our growth figures (5) before it decreases again when controlling for year effects (6), industry effects (7) and year specific industry shocks (8) ending up at around 0.0227 percentage points. This variable thus has a very large impact on the probability and a positive as such. I.e. for any given firm that runs two consecutive deficits the probability of bankruptcy goes up by a whole lot. Comparing it to the mean of going bankrupt in our dataset, 0.01, any such firm would have its probability of default more than tripled.

Further, in column (3) we include our measure of size, *Size (1 if small)*. This is at first of some impact to the probability of failure, about a 0.0065 percentage point increase for being small company. It makes sense that being a smaller firm actually would increase your probability of going bankrupt. This effect interestingly becomes smaller when we include growth, then it becomes larger again as we include the dummies for the years, staying significantly different from zero throughout and in (8) has a value of 0.0045 percentage points. Smaller firms are clearly subject to a higher probability of bankruptcy. Should anything occur, they have less of a safety net to cover their losses.

The addition of the age categories in column (4) provides the insight that younger firms are more likely to go bankrupt. The values of the age categories in our table is referenced to the omitted category of the youngest firms (0-4 years), thus the probability of going bankrupt is increasingly negative the older a firm gets, in thread with our initial testing and the initial idea of younger firms being more exposed. Probably they have less experience and competence to draw upon in times of financial pressure.

In column (5) we add the growth measures. We decided to incorporate these in our model in order to capture the effects of expanding a company on bankruptcy probability. Surprisingly, these variables turned out to be not significant. Counter-intuitive, as one would expect that these variables were to have some effect on the probability of a default. In addition the literature describes these measures as being important predictors for bankruptcy. However, none of the growth variables show any sign of significance.

From the year dummies we observe that simply being a firm in 2008 or 2009 clearly exposed firms to more risk as all the other years has negative impacts compared to these two years, 2009 being somewhat "worse" than 2008.

Industries that we can say with confidence impact the probability of going bankrupt reported by industry code (SIC):

Increasing probability	**Reducing probability**
2 Forestry and logging	**18** Wearing apparel., fur
15 Food products and beverages	**75** Public administration and defence
20 Wood and wood products	**91** Membership organizations n.e.c.
28 Fabricated metal products	
32 Radio, TV sets, communication equip.	
45 Construction	
52 Retail trade, repair personal goods	

We note that most of the industries that have a reduced probability of going bankrupt are also ones that receive protection or subsidies in some way.

Looking at the interaction term, we identify some year specific industry shocks that increased probability for bankruptcy. In 2007, as a consequence of the global recession the industries: **50** Motor vehicle services **55** Hotels and restaurants and **67** Auxiliary financial intermediation experience a higher probability of going bankrupt. The industry **55** Hotels and restaurants receives another shock like this in 2011.

Conversely, some industries also experienced positive shocks, i.e. a decrease in probability of going bankrupt: **45** Construction and **62** Air transport experience such shocks in 2008. In 2009 **14** Other mining and quarrying and **45** Construction have lower probabilities of going bankrupt, while in 2012 **2** Forestry and logging, **36** Furniture, manufacturing n.e.c. and **45** Construction receive these shocks. To some extent, this confirms the validity of our data as we are able to identify the impact of the financial crisis, and what we believe to be results of government policy. Just to take some examples, in 2008, the monopoly on air shuttle traffic ended in Norway. In 2009, the Norwegian government renewed promises of keeping the coal industry on Svalbard alive (Ministry of Justice, 2009, p.1). In regards to construction, we can only make educated guesses but it could be attributed to public funding of large projects and/or an ever increasing demand for housing.

6.2 Logit model

The logit model is a model that uses a binary dependent variable. The results we get from this binary model are interpreted as the percentage likelihood of a business going bankrupt. We think that this percentage likelihood of bankruptcy is a good addition to our thesis, as it is a model that transforms the data in a different way and thus serves as a robustness check.

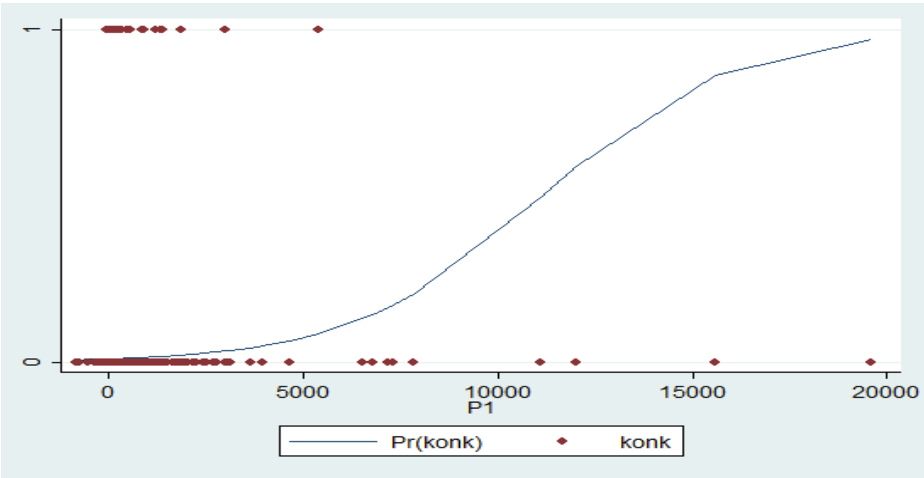
Since this is a binary model, we needed a dummy variable for bankruptcy. The dummy variable takes the value 0 if the company never has gone bankrupt, and the value 1 if the company has gone bankrupt.

$$y_{t+1} = \begin{cases} 1 \; if \; the \; company \; have \; gone \; bankrupt \\ 0 \; otherwise \end{cases}$$

This binary dummy variable is the same dependent variable we have used in all of the models we have estimated. The OLS dependent variable, often a continuous variable, is in our case binary. Therefore we find it interesting to compare this with a logistic model that is specifically designed to handle a dichotomous outcome. We cannot directly compare the coefficients from the logit model with the other models, but we are interested to se if there are any additions to our analysis we can draw from the logit model (A.13, appendix). We have also modelled a logistic regression that gives odd ratios as output, for the intention of making the analysis more informative (A.11, appendix).

When we plot our binary dependent variable against the profitability margin, we find the S-shaped curve that the logistic regression provides.

Figure 5: Plot of the S-kurve



The sign of the coefficients in the logit model is what we are interested in analysing. They are interpreted together with the odds ratios for the variables.

The net loss dummy for a 2-year deficit that are describing a net loss in the two previous years, get an odds ratio value of 4,7. In addition to this the sign of the coefficient is positive, indicating a higher probability of default. The odds ratios tell us that if a company have a two year period with a deficit, the probability of bankruptcy is 4, 7 times higher than if they do not have a net loss in two consecutive years.

The size variable is defined as a dummy variable that takes on the value of 1 if the company is small, and 0 if the company is large. When we estimate the model with odd ratios as output, we find that the odd ratio for size is approximately 2.2. This suggest that companies that are defined as small are 2,2 times more likely to go bankrupt than the larger companies. The sign of the coefficient in the logit regression is also positive, which indicates a higher probability of bankruptcy if the company is small. The size of the company will impact its probability of default.

For the age dummies that we have included, we se that the sign of the coefficients are negative. As they are measured against the reference category of age 0-4 years, this is a sign that as the company gets older, the probability of bankruptcy falls. The odd ratios of the age variables have a lower value for each age interval the company go up, also proof of a lower probability of bankruptcy as the company ages. This is consistent with what we assumed.

The financial crisis have been up for debate and analysis to a large extent the last years, and it have been a understanding that the Norwegian market was less affected by the financial crisis compared to other countries. Nevertheless, the signs of the coefficients in the year dummies are positive in the years 2007, 2008 and 2009, meaning that the probability of bankruptcy is higher in this period. This is interesting, because even though the Norwegian marked did not get hit as hard as others, it implies that the probability of bankruptcy in Norway was affected by the financial crisis. The world is globalized, and the national Norwegian marked has been affected by the recession in other parts of the world.

6.3 Random effects

In the random effects model we assume that the variation across companies is random, i.e. that the difference across the units is uncorrelated with the explanatory variables. Comparing differences between companies, the goal of using the RE approach is to have the most efficient estimate. RE is preferred over FE because it can be better assuming there are no fixed effects present. We later test for this assumption. When the variables change across time and between companies with one unit, the coefficient will show the average effect of the variables over the bankruptcy dependent variable. Thus we are able to interpret the time invariant variables like the variables for size or age of the company. The independent variables will include both within-entity effects and between-entity effects.

The liquidity ratio in this model gives us an estimation that indicates some signs of trouble. As liquidity increases (increased by a unit), the probability of default also goes up. This is a result that makes no sense, and contradicts results from the other estimators. The 2-year deficit variable has a significant coefficient of 0.0163 percentage points, which is lower than in the OLS estimation. The probability of bankruptcy will in this estimation increase with this 0.0163 percentage points if the company have a 2-year net loss, so this factor still makes sense.

The size variable in this model has a coefficient that is significant and positive. Age categories are negative measured against the youngest age interval of 0-4 years. This also makes sense as the probability of default is reduced as the company becomes older.

By performing a Breusch and Pagan Lagrange- multiplier test for random effects we find that the null hypothesis of var(u)=0 is preffered due to a non-significant test value. This indicated that there are not sufficient random effects in the model. Due to this, OLS with good controls, or FE, might be a better approach than RE.

6.4 Fixed effects

In the fixed effects model we are analysing the impact of the variables that change over time. The effects that are fixed within each firm is controlled for in this model. Intuitively we argue that this will be the better approach compared to the OLS and RE. The RE model is more efficient (Wooldridge, 2014) only if all the data within each entity is random, which we have established that it is not. After estimating the FE model we can perform a Hausman test to investigate if fixed effects are present.

Appendix A.12 displays results from running the FE model. We have incorporated the variables stepwise; to be able to look at the change each variable contributes. The coefficients in this model are interpreted as the percentage point change in the probability of $y_{t+1} = 1$, i.e. the change in the probability of going bankrupt. This is measured against the mean value of the sample, to get an idea of how much impact on the probabilities the variables have. The mean value for the probability of going bankrupt is in our sample is approximately 0.01.

In order to take into consideration the time- and industry-specific fixed effects that we assume exists within our dataset we make use of the fixed effects transformation described in chapter 6.4, thus:

$$(y_{it} - \bar{y}_i) = \beta_{it}(X_{it} - \bar{X}_i) + (\mu_{it} - \bar{\mu}_i)$$

And we can apply OLS to this equation and estimate the β_i from our OLS as $\hat{\beta}_i$,

$$(1)\ y_{it+1} = \beta_0 + \beta_1 * Profitmargin1 + \beta_2 * Liquidityratio1 + \mu_{it}$$

$$\vdots$$

$$\begin{aligned}(6) y_{it+1} = \beta_0 &+ \beta_1 * Profitmargin1 + \beta_2 * Liquidityratio1 + \beta_3 * Solidityratio1 \\ &+ \beta_4 * Solidityratio2 + \beta_5 * 2year\ deficit + \beta_6 * Size(1\ if\ small) + \beta_7 \\ &* Age(5{\sim}9) + \beta_8 * Age(10{\sim}25) + \beta_9 * Age(26{\sim}) + \beta_{10} \\ &* Growth\ revenue + \beta_{11} * Growth\ assets + \beta_{12} * Growth\ equity + \beta_{13} \\ &* Growth\ current\ liabilites + e_2E_2 + \cdots + e_nE_n + \delta_2T_2 + \cdots + \delta_tT_t + a_i \\ &+ \mu_{it}\end{aligned}$$

Again y_{it+1} is the dependent variable for bankruptcy where I = entity and t = time. $x_{k,it}$ represents the independent variables as seen in (6), β_k is the coefficients for the the independent variables, $a_i + \mu_{it}$ is the error term, E is the entity n. Since the industry variables are binary we have n-1 entities included in the model. e_n is the coefficient of the binary regressors (entities). T_t is the binary variable for time, such that we have t-1 time periods. δ_t is the coefficient for the binary time regressors. Again, the intercept in the first (base) period is β_0 and that for the next period is $\beta_0 + \delta_2$.

We are then controlling for the average differences across industries in any observable *or unobservable* predictors, such as difference in the quality of firms, sophistication etc. The new coefficients soak up all the across group action variation. The effects we are able to isolate from the FE-model are how internal rates are actually affecting the probability of a company going bankrupt, holding everything else constant.

The last column (6) in the model is enhanced from the FE estimation one previous (5). Both of these models are included in the overview (appendix A.13). We have included an interaction term:

$$\boldsymbol{Interaction\ dummy = Year\ x\ Industry}$$

This dummy captures time specific events within each industry. The dummy takes on the value of 1 if the company is in a certain industry *and* in a certain year.

In the first column (1) in our FE table the results are reported for our key accounting figures following the discussion in the section on OLS. *Profit margin 1* is significant at the 1 % level, showing a 0.001 percentage point increase in the probability of failure for a one unit change in the ratio for any given firm. This result is persistent throughout the introduction of more controls except for a small stepwise decrease in magnitude ending at 0.0005 in column (6). This indicates that the profit margin will increase the probability of bankruptcy, as the ratio becomes larger. The profit margin is measured as sales/total assets, and the interpretation of this is the same as in OLS.

Liquidity ratio 1 shows a decrease of -0.002 percentage points for a given unit increase in the ratio. This is in thread with the OLS results after introducing the industry specific effects that are now also controlled for by the FE model. The liquidity of a company is necessary to meet obligations. This is important for short-term obligations, as they are due often within 1 year. To cover the longer obligations, companies have the option of turning illiquid assets into liquidity. These results prove that the more liquid a company is, the less of a probability of bankruptcy. The company must consider the trade-off between increasing liquidity and decrease the probability of bankruptcy, as they determine their strategy. This is a part of the operations that the management are able to influence.

Solidity ratio 1 has a decrease of -0.0002 percentage points for a given unit increase in the ratio, which is to say that increasing solidity is a good thing. The value is persistent throughout the model but never significant on anything else than the 10% level except when we include the interaction in (6), where it is significant on the 5% level. The FE model leaves decidedly more noise out of the error terms, as this figure was never significant in our initial OLS models although common sense says it should be. The percentage of equity the company has is impacting the ability to handle their debt claims. Banks often have a minimum equity percentage of total assets that the company needs to stay above to keep receiving credit. If a company is less leveraged, they have capital buffer in the form of equity in case of unforeseen events.

Solidity ratio 2 is at first not significant (1) but takes on a value of -0.02 percentage points in the last model (6). This is an effective decrease in probability for bankruptcy for a unit increase in the ratio. This value increases upon adding the different controls when we are expanding the fixed effect model. The more assets that are a collateral to the creditors, the more solid the business. The ratio is mortgaged assets divided by collateral fixed assets. An increase in this will strengthen the banks' security of getting the loans repaid when they provide capital. The interest rate on the loan may be lower due to a lower risk, and the banks are willing to provide higher loans.

In column (2) we add the dummy for a 2-year deficit. This has an impact of a 0.02 percentage point increase in the probability of going bankrupt for a change in status, at first then decreasing as we add more controls, ending up at 0.009 percentage points. This variable infers that if a company has a net loss in two consecutive years, the probability of bankruptcy increases. The net loss is a product of both income and the costs of the operation. There are many aspects that can be altered to improve the bottom line results.

The variables that are not likely to change from one category to another, does not give us any additional information to this analysis. A potential limitation of FE is that we cannot assess the effect of variables that have little within-group variation, which applies to some of our variables. E.g., not many businesses change their status from small to large (or the other way around) within an industry. The interpretation of *Size (1 if small)* in column (3) becomes negative at first, then impossible to determine as we add the interaction variable. It is highly unlikely that being a small firm should decrease the probability of going bankrupt and we must forgo the FE model if we are to isolate this effect in particular. However, as shown, that would expose us to potential omitted variables and for the rest of the variables FE seems more efficient. Furthermore, we do not have to worry about unobservable factors that are correlated with the variables included in our regression. If we instead had developed a size variable based on e.g. the logarithm of sales income, we would be able to interpret the size effect in the FE-model.

Upon adding the age categories in (4) we see that it might be the case that not sufficiently many firms change their age status either. Seeing how being an older firm now increases the risk of going bankrupt in reference to the youngest firms it would probably be a good idea to estimate the model with smaller intervals for age, or a continuous measure.

Column (4) also introduces our growth variables, out of which only *Growth revenue* seem to be significantly different from zero (at the 10% level), having a -0.0000017 percentage point impact on the probability of going bankrupt for a unit change in the ratio. The lack of explanatory power in the growth variables is what surprised us most. Intuitively, we assumed that expanding and growing was associated with some kind of extra risk. However, it does not look like the growth has any significant impact on the probability of bankruptcy.

Hausman test

The Hausman test is a way to formally test for statistically significant differences in the coefficient of the explanatory variables (Wooldridge, 2014). The test proves if there is fixed effects in a panel data or not. Fixed effects are unobservable characteristics or unique errors u_i that differ within each entity in the panel. If fixed effects are present in the data material that is modelled, then the fixed effects model (FE) is a better approach to get solid results. The results are more solid since the FE model captures these individual traits.

H_0: The regressors are not correlated with the unobservable characteristics and the preferred model is random effects model (RE).

H_A: The regressors are correlated with the unobservable characteristics, and the preferred model is fixed effects model (FE).

The models that we have estimated include a fixed effects model and a random effects model. Our sample consists of data sorted by different organization numbers, and we have assumed that there is fixed effects within each company that affect the result. Each company may have individual features that could impact on their probability of default. This could be due to employees, internal procedures, management and so on. To test this hypothesis, we have used the Hausman test which gives us a chi-squared value of 0.000. This indicates that our hypothesis is correct. There are fixed effects within the panels that have to be controlled for. The FE model in our results is therefore likely to be more efficient.

6.5 Comparison of the models

We find that many of the results as to what factors determine bankruptcy are similar across the different models, which in itself is a validation of each measure. The fixed effects model incorporates the causes of changes within a firm. Comparing it to the OLS, we see that the interpretation of some key figures change, which is a sign that FE is a better fit. Intuitively the FE results also outperforms the ones from the RE, which leads us to prefer the FE in this case too. The Hausman test confirms this. Together with the fixed effects model, we believe the results from the logistic regression provide us with a good insight. By controlling for external circumstances, the model is able to estimate the financial measures, and measure how large impact they have on the probability of bankruptcy.

Across all models, neither of the growth variables are significant, and we cannot say how company growth has impact on its probability of going bankrupt, if indeed at all.

The profit margin is significant throughout all of the models. One should be careful about increasing the ratio on the wrong terms. The fraction is part of two different aspects related to bankruptcy, hence upping sales is ok but scaling down assets is not.

We have a liquidity ratio that changes the sign in front of the coefficient as we include it in different estimation models. The sign in RE is positive, but in the rest of the models it is negative, i.e. an increase in liquidity is lowering the probability of bankruptcies. This makes us suspicious the RE model cannot be trusted.

Solidity ratio 1 is not significant until we control for fixed effects. This is a ratio that incorporates a lot of information about a company's financial state. Solidity ratio 2 has a very low magnitude of the coefficient, but also this ratio increases the significance level when we control for fixed effects.

To avoid bankruptcy, companies should avoid running consecutive deficits. This ratio shows an increase in the probability in all of the models we have estimated.

7. Predictions and evaluation

7.1 Predicting bankruptcy

In order to evaluate the accuracy of our model we use categorical regression analysis on a variable Z that indicates bankruptcy. In the dataset we have previously ran regressions on the variable already defined as:

$$Bankrupt \begin{cases} 0\ for\ not\ going\ bankrupt \\ 1\ for\ going\ bankrupt \end{cases}$$

The model that we want to estimate sets Bankrupt as Z and looks like the specification by Ohlson (1980):

$$Z = \beta_0 + \beta_1 * F_1 + \ \beta_2 * F_2 + \beta_n * F_n + \ \varepsilon$$

Where $\beta_0, \beta_1,\ldots, \beta_n$ are regression coefficients and ε denotes the error term.

In the logit model, Z then denotes the probability of bankruptcy:

$$p = p(Bankrupt = 1|Z) = F(Z)$$

Where F is a cumulative distribution (increasing between 0 and 1), which is to say that we have a probability for bankrupcy:

$$p = F(\beta_0 + \beta_1 * F_1 + \ \beta_2 * F_2 + \beta_n * F_n)$$

In order to find p, we assume this cumulative distribution is logistically distributed, i.e.:

$$F(Z) = \frac{e^z}{1 + e}$$

Such that our probability p can be allowed to be:

$$p = \frac{1}{1 + e^{-z}}$$

(while Z is still equal to: $\beta_0 + \beta_1 * F_1 + \ \beta_2 * F_2 + \beta_n * F_n$). Z is given by the logistical regression. A higher Z means probability p is higher too.

The two categories can be grouped from the critical value $p^* \; if \; p \leq p^*$ this means a firm is categorized as credit worthy and grouped as 0. Whereas in the other case, where $p \geq p^*$ it is grouped as 1 and described as a probable bankrupcy. Alternatively we can categorize based on Z^* which is given by:

$$\mathrm{p}^* = \frac{1}{1 + \mathrm{e}^{-\mathrm{Z}^*}}$$

If Z ≤ Z^* then a firm is credit worthy and grouped as 0. If Z > Z^* this means it is grouped as 1 and described as a probable bankrupcy. Here high Z is a sign of danger. Z^* is estimated by our model. [3]

This table is an overview of the models predictions and the correct values. We want the model to estimate as low number of "false" as possible.

Table 5: Predicting errors

	Predicted= 1	Predicted= 0
Correct =1	True	False
Correct =0	False	True

H_0: A company is likely to go bankrupt, =1

H_A: A company is not likely to go bankrupt, =0

Type 1 errors occur when the null hypothesis is rejected, but it is in fact true (Smith, 2011). In our case this would be predicting that a company are not likely to go bankrupt, when in reality they are.

Type 2 errors occur when we are not rejecting the null hypothesis, when the alternative hypothesis is true (Smith, 2011). In our case this would be predicting that a company are likely to go bankrupt, when in the reality they are not.

[3] Each categorization can be done in one estimation using Stata's estat classification after running the logit

7.2 Evaluation and prediction using logit

The logistic model gives the percentage probability of a company going bankrupt. This is only meaningful information if the model gives a valid number of correct estimations. We are testing our model against the data we have to evaluate if we can predict bankruptcies accurately for our dataset. We want to find the percentage of prediction that is correct.

To evaluate the competency of our logit model, we needed to find a probability cut-off as to how we wanted to classify companies as bankrupt or not. To find this cut-off point, we estimate the specificity/sensitivity trade-off (A.14 in appendix).[4] The two functions are crossing at the point where misclassifications are in a steady state.

The prediction fares rather well (A.15), and we can safely assume that our fixed effects approach has enabled us to identify good measures that determine bankruptcy. There is a total percentage of correct classifications in the logistic model of 74,22 %. Nevertheless, as the literature warns, anyone using the regression approaches of this model must keep in mind the 25,75% that the model will predict falsely. This logistic regression does not incorporate fixed effects in estimating the probabilities. It may be possible to improve on these results using a logistic regression that controls for this.

[4] Stata's lsens command does this for us

8. Discussion and summary

8.1 Criticism of the estimated models

The models we have formulated must be used with caution. The results can give a company insight as to which factors that are contributing to a higher risk of failure, but the complexity of this issue makes these types of models to some degree uncertain. It is impossible to control for every factor that may contribute to a higher risk of bankruptcy.

Due to differences in reporting the data, we had to prepare it by making some assumptions. Bankruptcies were reported either on the same year as the companies last active year, or in one of the previous years. To be able to do the analysis, we had to set the bankruptcy date to one year after the last active year for all companies. In our estimation, these kinds of assumptions may have influenced the results to some extent. The dependent variable was a binary dummy variable, which considered if the company was bankrupt/not bankrupt. In regards to our OLS, RE and FE approach it is evident that linear estimations suffer from logical constraints, such as predicting probabilities outside the probability range, something the logit does not do.

The fixed effects models also show some limitations when it comes to determining the impact of size on the probability of a company going bankrupt. Regardless, we have found that small companies have a higher probability than large. This seems to be a solid result. However, the size variable is only divided in two categories and could give a more nuanced view if the model was estimated with more fragmented size categories. The same goes for the variable age. We have estimated the model with age categories, but if the model included a continuous measure for age we would be able to get the "per year change" effect.

8.2 Suggestions for further research

The estimations applied controls for other factors than pure financial measure. It would be interesting to expand on this, so that it includes additional perspectives that could affect the probability of bankruptcy. Ideally, a fixed effects model that takes the binary nature of our data into consideration.

Some research based on industries has been done in this thesis. An elaboration on industry specific estimation models could provide more information about the differences. Some industries are densely clustered in the same geographical areas in Norway. E.g. the oil sector has a large percentage of their operations in Stavanger, the Shipping industry in Bergen, the finance sector is clustered in the larger cities, the aluminium production is concentrated in Sogn & Fjordane, and the fishing industry is located at the coast etc. It is possible to analyse whether industry or geographical aspects have an impact in determining bankruptcy probabilities.

The period our data material is based on, is from the years 2005-2012. Developing a model using longer panels would make it less biased by the years of the financial crisis. In our model these years are a large part of the total sample, and even though we are controlling for year specific effects as well as year specific shocks the companies were in this period largely impacted by the stress of the crisis. Many employees were let go, and skills and knowledge may have been lost during this period.

Given that we find different attributes that are beneficial in both the logit and the FE, a suggestion for further research is a combination of these models. There exists a logistic regression model that is specified for panel data.[5] The logistic regression will then take into consideration the fixed effects present in the panel. This estimation process will be computationally demanding, especially with such a large dataset. We believe it will identify much of the same factors, but may yield more accurate predictions.

[5] The command for this model is xtlogit in Stata

8.3 Conclusion

The purpose of this thesis was to identify the impact of the factors that are present in determining the probability of corporate bankruptcy. The emphasis was laid on how fixed effects influence the estimation approach. We uncover differences as to what internal factors explains the probability of bankruptcy depending on which model we use in estimating. We were able to find some general results, as we measured many of the variables as ratios, giving us comparable output.

The data we uncover on company characteristics appear to be in line with previous findings. The models prove that the probability of bankruptcy falls as the company becomes larger and older. The FE model was not the best to describe the latter, as it only estimates coefficients for firms that change their status within the time period. It is not very often a company changes their size and age. The logit model was a better approach to analyse these factors. We found a diminishing probability as the company grew older, and that a small company is approximately twice as likely to go bankrupt as a big company according to our odds ratio output.

The accounting data also provided us with insights. Initially we had many primary variables that we expected to be relevant. After the selection process we were left with the final accounting variables to be analysed. When analysing the profit margin we expected the variable to reduce the probability of bankruptcy as the ratio got higher. However, the results are opposite. The profit margin is a fraction, and a product of both sales income and total assets. Even if sales income goes up, it does not weigh up for when assets are decreasing, and hence the ratio goes up. The probability actually becomes higher when the ratio increases.

The liquidity ratio is a measure of how much liquidity a company has. Our initial beliefs about this ratio were that it should decrease the probability of bankruptcy. It should be beneficial to be more liquid, as the company has better chance of covering their obligations. The sign of this coefficient is negative in both the logit model and in the fixed effects model, and thus confirming this belief. The coefficient in the logit model is not significant though. The same applies to the equity ratio *Solidity 1*. It is negative in both the logit model (not significant) and in the FE. The reason for this is that in the FE model we are also controlling for firm-specific attributes. This means that even controlling for things like management, operational systems and so on, the liquidity ratio and *Solidity 1* still makes the probability of default decrease. Incorporating the fixed effects is better than using the logit when analysing the accounting ratios.

The net loss dummy that measures a 2-year deficit proved to be significant and positive, i.e. gives a higher probability of a default. The result is logical, since the company is losing money. The odd ratios gave us a numeric value on how much the impact is, and the company is almost 4.7 times more likely to go bankrupt if they run a deficit for two years.

To summarize, we find the logistic model appropriate when we are looking at how the age and the size of a company affect the probability of bankruptcy. However, the fixed effects model is superior when it comes to estimating the factors that affect the probability of bankruptcy based on accounting ratios.

References

Accounting act (1998). Act relating to annual accounts. [online] Available at: https://lovdata.no/dokument/NL/lov/1998-07-17-56#KAPITTEL_1 [Accessed 14 Mar. 2015].

Altman, E. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4), p.589.

Avenhuis, J (2013) Testing the generalizability of the bankruptcy prediction models of Altman, Ohlson and Zmijewski for Dutch listed companies. Netherlands: University of Twente

Bernanke, Ben (1981) Bankruptcy, Liquidity and Recession. *The American Economic Review, 71,* p. 155-159

Berner, E. Mjøs, A. Olving, M. (2014). Norwegian Corporate Accounts – Documentation and quality assurance of SNF's and NHH's database of accounting and company information for Norwegian companies. Bergen: Centre for Applied Research at NHH.

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

Bharath, S. and Shumway, T. (2008). Forecasting Default with the Merton Distance to Default Model. *Review of Financial Studies*, 21(3), pp.1339-1369.

Bhattacharjee, A., Higson, C., Holly, S., Kattuman, P. (2009) Macroeconomic Instability and Corporate Failure. London: Cambridge

Business cycles (n.d.). [online] Available at: http://financialplanningbodyofknowledge.com/wiki/Business_cycles [Accessed 22. April. 2015].

Carroll, G. (1983). A stochastic model of organizational mortality: Review and reanalysis. *Social Science Research*, 12(4), pp.303-329.

European commission (2003). Commision recommendation concerning the definition of micro, small and medium-sized enterprises. Official Journal of the European Union [online]

Available at: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:124:0036:0041:EN:PDF [Accessed 10. April. 2015].

Freeman, J., Carroll, G. and Hannan, M. (1983). The Liability of Newness: Age Dependence in Organizational Death Rates. *American Sociological Review*, 48(5), p.692.

Helsingeng, A. (2004). *Konkurs og en ny start: Holdninger til konkurser og konsekvenser av konkurs,* Deloitte.

Hillegeist SA, Keating EK, Cram DP, Lundstedt KG (2004) Assessing the probability of bankruptcy. Rev Account Stud 9:5-34

Jackson, R. and Wood, A. (2013). The performance of insolvency prediction and credit risk models in the UK: A comparative study. *The British Accounting Review*, 45(3), pp.183-202.

Konkursrådet (2000). *Konkurs i praksis.* Sandefjord: Konkursrådet

Ministry of justice (2009). St.meld. nr. 22 (2008-2009) Bankruptcies [online] Available at: https://www.regjeringen.no/no/dokumenter/stmeld-nr-22-2008-2009-/id554877/[Accessed 15. June 2015].

Norwegian Public Limited Liability Companies Act (1997). Norwegian Public Limited Liability Companies Act [online] Available at: https://lovdata.no/dokument/NL/lov/1997-06-13-44?q=aksje [Accessed 22. Mar. 2015].

Ohlson, J. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1), p.109

Ohlson, J. A. (1980). “Financial Ratios as Predictors of Failure”, *Journal of Accounting Research* 4, 71-102

Olsen, H. Øien, M. (2009) Finnes det en sammenheng mellom bedrifters resultat og utbetaling av styrehonorar? *NHH*

Peng, C., Lee, K. and Ingersoll, G. (2002). An Introduction to Logistic Regression Analysis and Reporting. *The Journal of Educational Research*, 96(1), pp.3-14.

Philippe & Partners (2002). *Bankruptcy and a fresh start: Stigma on failure and legal consequences of bankruptcy.* Brussel: Deloitte & Touche Corporate finance.

Pohlman, John T.; Leitner, Dennis W. (2003). *A Comparison of Ordinary Least Squares and Logistic Regression*. Ohio Journal of Science: Volume 103, Issue 5. The Ohio State University Knowledge bank.

Skogsvik, K. (1988). *Prognos av finansiell kris med redovisningsmått, En jämförelse mellan traditionell ochinflationsjusterad redovisning.* Stockholm: Ekonomiska Forskningsinstitutet vid Handelshögskolan i Stockholm

Skogsvik, K. (1990). Current cost accounting ratios as predictors of business failure; the Swedish case*. Journal of Business Finance & Account.*

Smith, M. (2011). Common mistakes in using statistics [online] Available at: http://www.ma.utexas.edu/users/mks/statmistakes/errortypes.html[Accessed 10. June 2015].

Sorensen, J. and Stuart, T. (2000). Aging, Obsolescence, and Organizational Innovation. *Administrative Science Quarterly*, 45(1), p.81.

Statistics Norway 1 (2015). Bankruptcies [online] Available at: https://www.ssb.no/statistikkbanken/selecttable/hovedtabellHjem.asp?KortNavnWeb=konkurs&CMSSubjectArea=virksomheter-foretak-og-regnskap&checked=true [Accessed 05. Feb. 2015].

Statistics Norway 2 (2015). Unemployment rate [online] Available at: https://www.ssb.no/statistikkbanken/selectvarval/Define.asp?subjectcode=&ProductId=&MainTable=SyssArbledSes02&nvl=&PLanguage=0&nyTmpVar=true&CMSSubjectArea=arbeid-og-lonn&KortNavnWeb=akumnd&StatVariant=&checked=true [Accessed 05. Feb. 2015].

Sundaram, R (n.d). *Multiple discriminant analysis* [online] Available at: http://www.isip.piconepress.com/courses/msstate/ece_8443/papers/2001_spring/mmda/p02_paper_v0.pdf[Accessed 01. Feb. 2015].

The Bankruptcy act (1984). The Bankruptcy act. [online] Available at: https://lovdata.no/dokument/NL/lov/1984-06-08-58?q=konkurs [Accessed 02. Mar. 2015].

Thornhill, S and Amit, R (2003) Bankruptcy, Firm Age and the Resource-Based View. Canada: Richard Ivey School of Business

Torres-Reyna, O. (2007). *Panel Data Analysis Fixed and Random Effects using Stata.*[online] Available at: http://www.princeton.edu/~otorres/Panel101.pdf [Accessed 05. Mar. 2015].

Tufte, P. (2000). En intuitiv innføring i logistisk regresjon. *Statens institutt for forbruksforskning.*

Vassalou, M. and Xing, Y. (2004). Default Risk in Equity Returns. *J Finance*, 59(2), pp.831-868.

Webb, G (1980) Competition and Bankruptcy in the Australian Road Transport Industry. *ATRF, 25,* p. 427-435

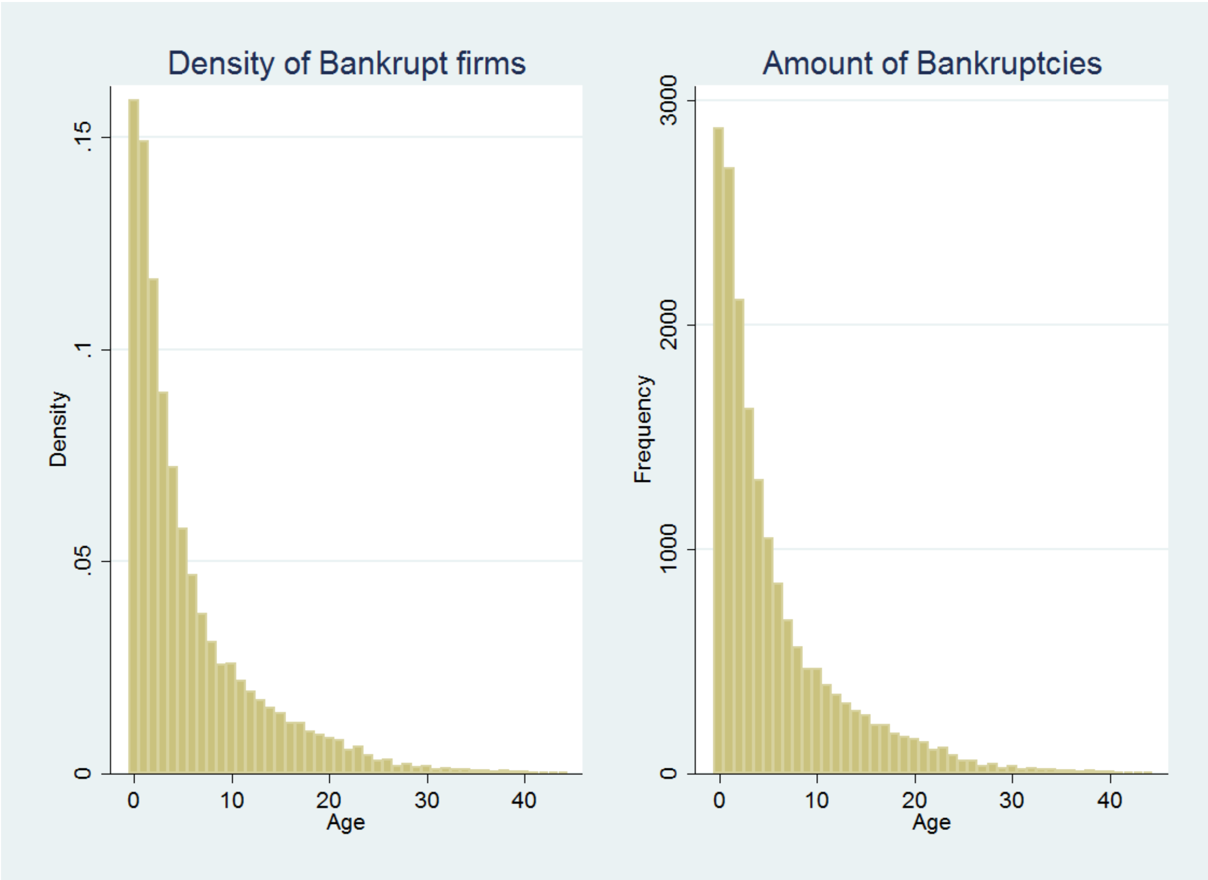
Wooldridge, J. (2014). *Introduction to econometrics*. Europe, Middle east and Africa edition. United Kingdom: Cengage learning.

Zhu, Xianping (2012) Internal Factors Associated with Business Failures of Public Companies in China. *International Journal of Business and Social Science, 3,* p. 49

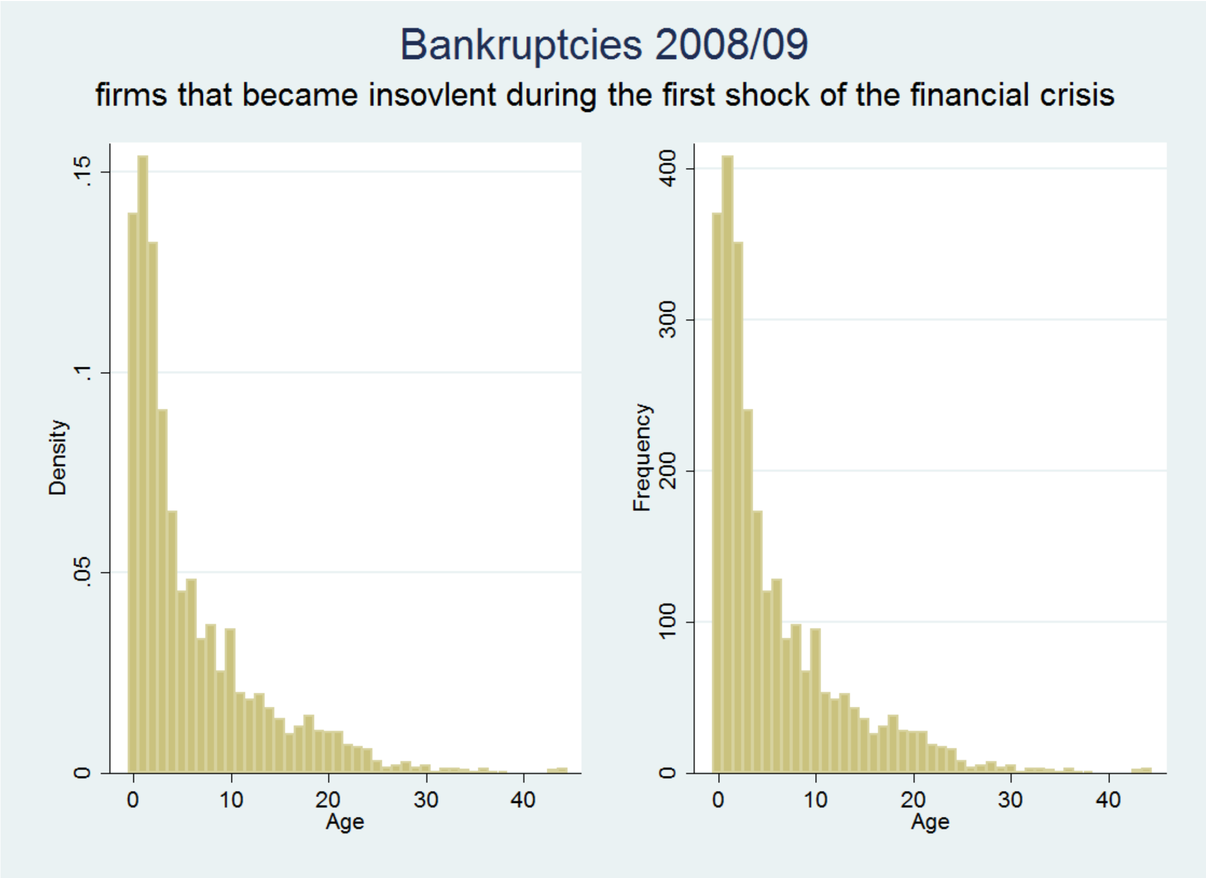
Zmijewski, M. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, 22, p.59.

Appendix

A.1 Bankrupt firms distributed by age

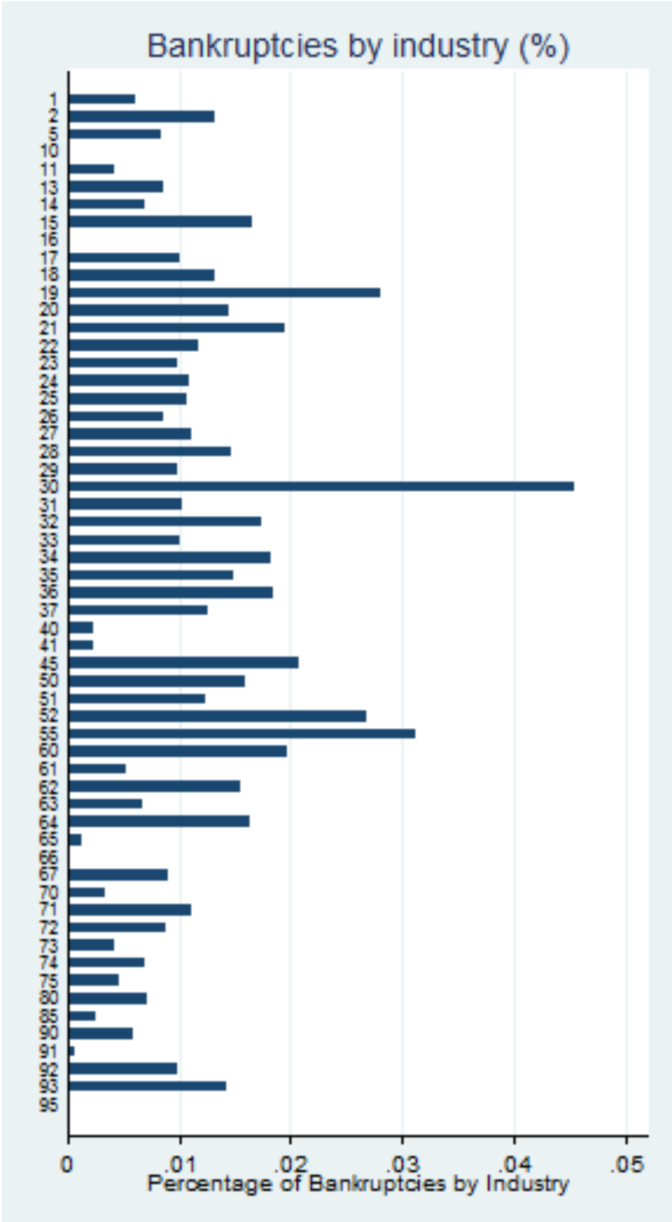


A.2 Bankruptcies 2008/2009

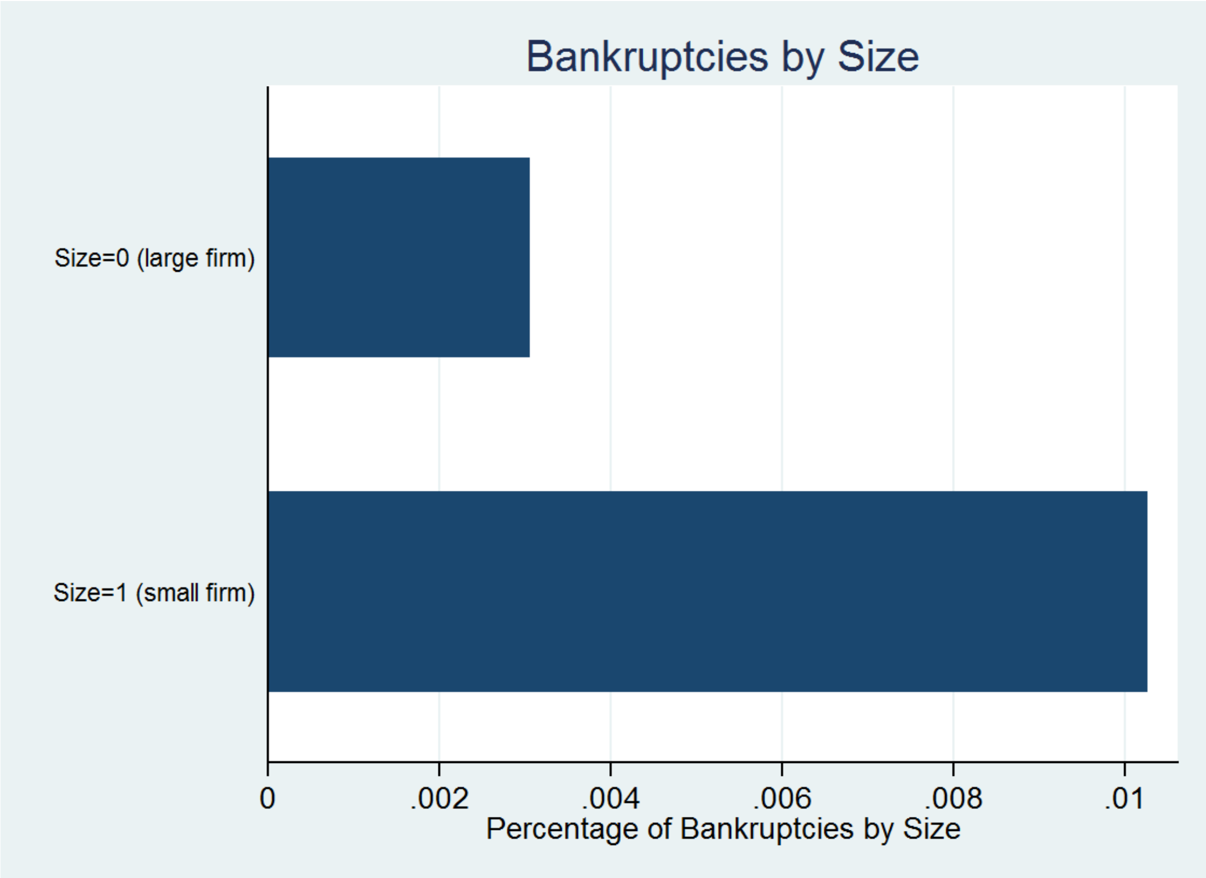


A.3 Bankruptcies by industry, **Standard Industrial Classification** (**SIC**)

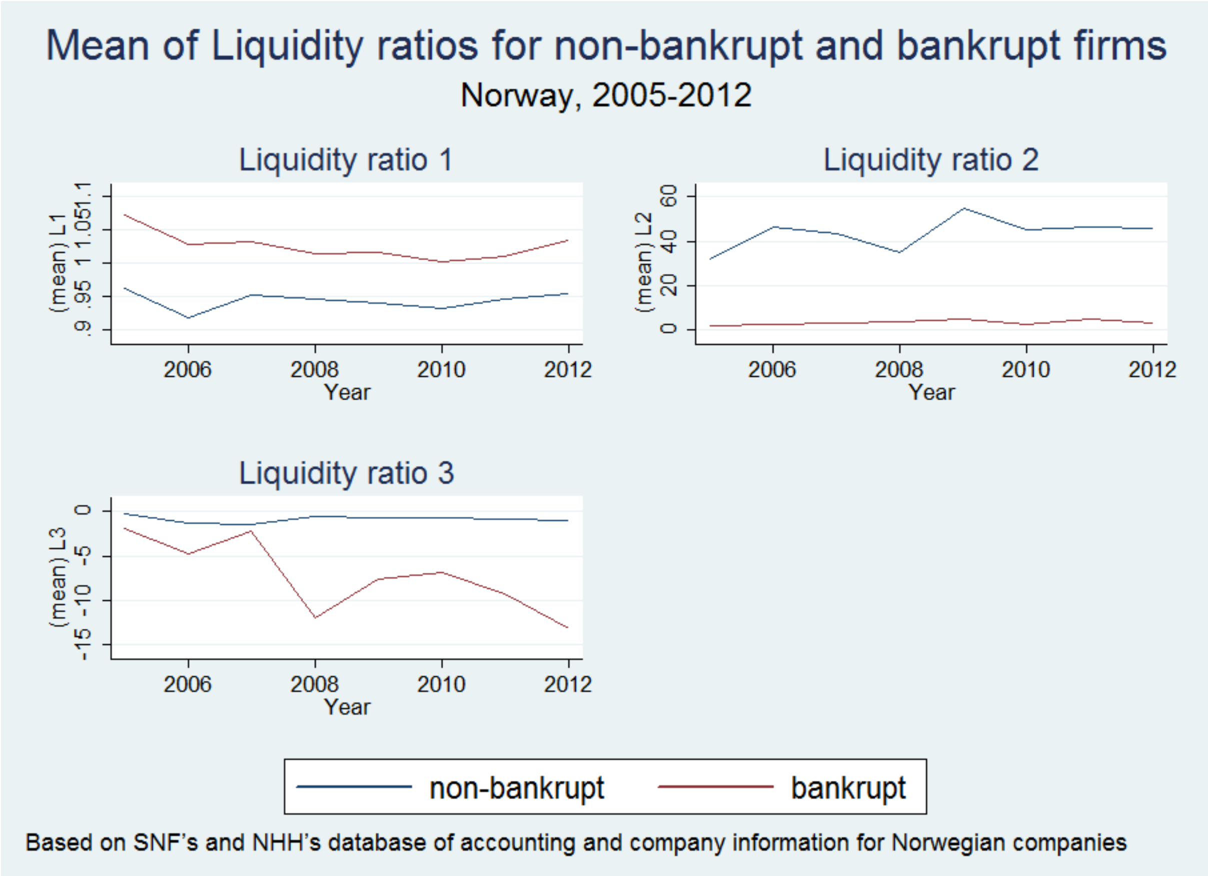
01 Agriculture and hunting
02 Forestry and logging
05 Fishing, fish farming, incl.services
10 Coal mining and peat extraction
11 Oil and gas extraction, incl.serv.
12 Mining of uranium and thorium ores
13 Mining of metal ores
14 Other mining and quarrying
15 Food products and beverages
16 Tobacco products
17 Textile products
18 Wearing apparel., fur
19 Footwear and leather products
20 Wood and wood products
21 Pulp, paper and paper products
22 Publishing, printing, reproduction
23 Refined petroleum products
24 Chemicals and chemical products
25 Rubber and plastic products
26 Other non-metallic mineral products
27 Basic metals
28 Fabricated metal products
29 Machinery and equipment n.e.c.
30 Office machinery and computers
31 Electrical machinery and apparatus
32 Radio, TV sets, communication equip
33 Instruments, watches and clocks
34 Motor vehicles, trailers, semi-tr.
35 Other transport equipment
36 Furniture, manufacturing n.e.c.
37 Recycling
40 Electricity, gas and steam supply
41 Water supply
45 Construction
50 Motor vehicle services
51 Wholesale trade, commision trade
52 Retail trade, repair personal goods
55 Hotels and restaurants
60 Land transport, pipeline transport
61 Water transport
62 Air transport
63 Supporting transport activities
64 Post and telecommunications
65 Financial intermediation, less ins.
66 Insurance and pension funding
67 Auxiliary financial intermediation
70 Real estate activities
71 Renting of machinery and equipment
72 Computers and related activities
73 Research and development
74 Other business activities
75 Public administration and defence
80 Education
85 Health and social work
90 Sewage, refuse disposal activities
92 Cultural and sporting activities
93 Other service activities
95 Domestic services



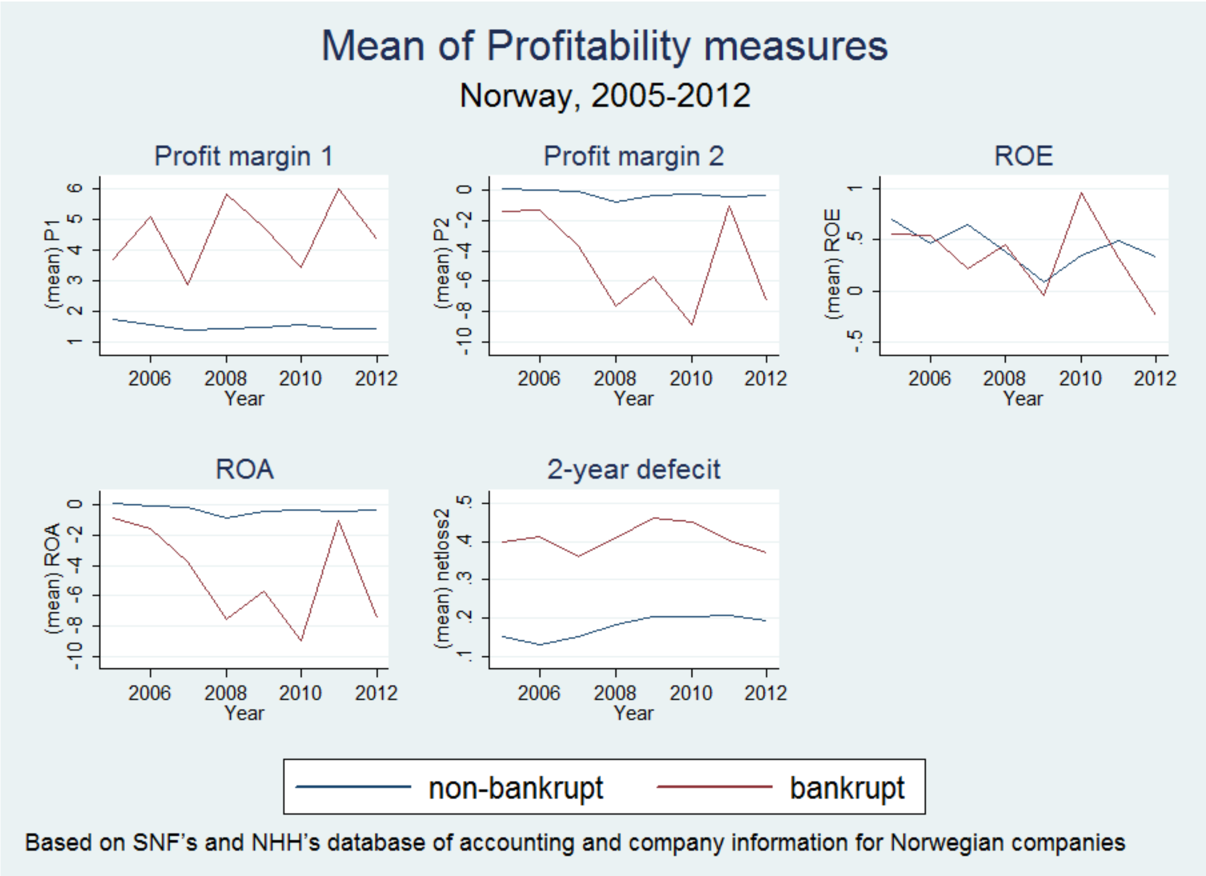
A.4 Bankruptcies by Size



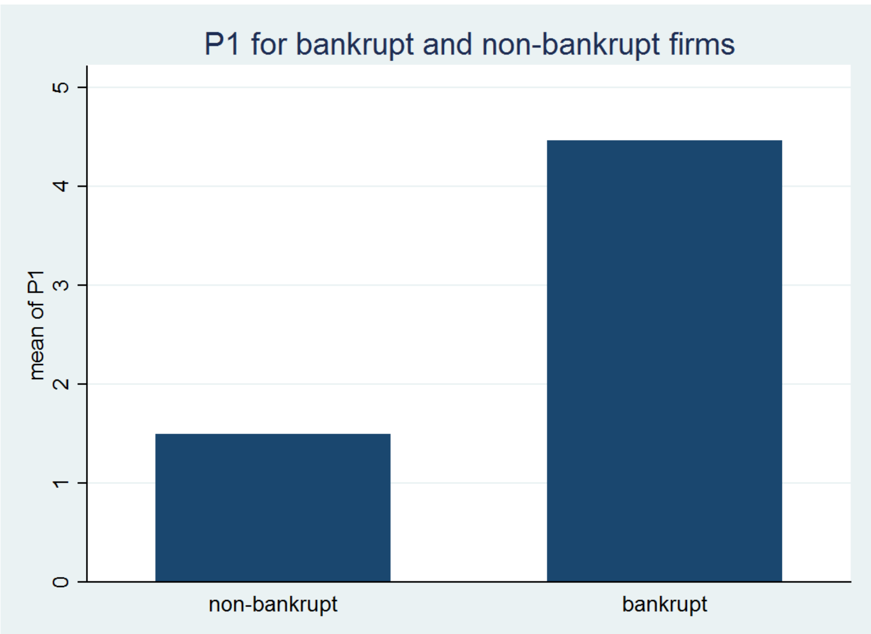
A.5 Mean of Liquidity ratios



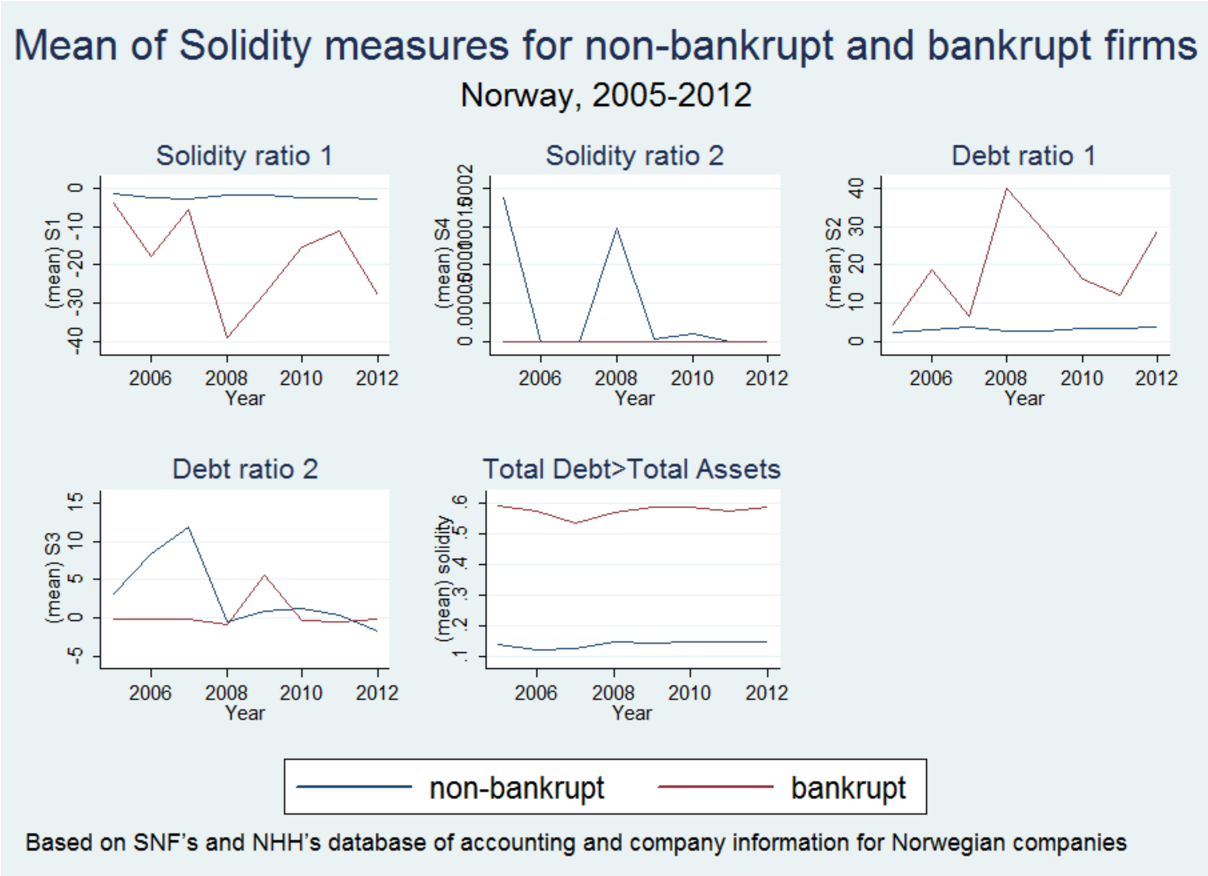
A.6 Mean of Profitability measures



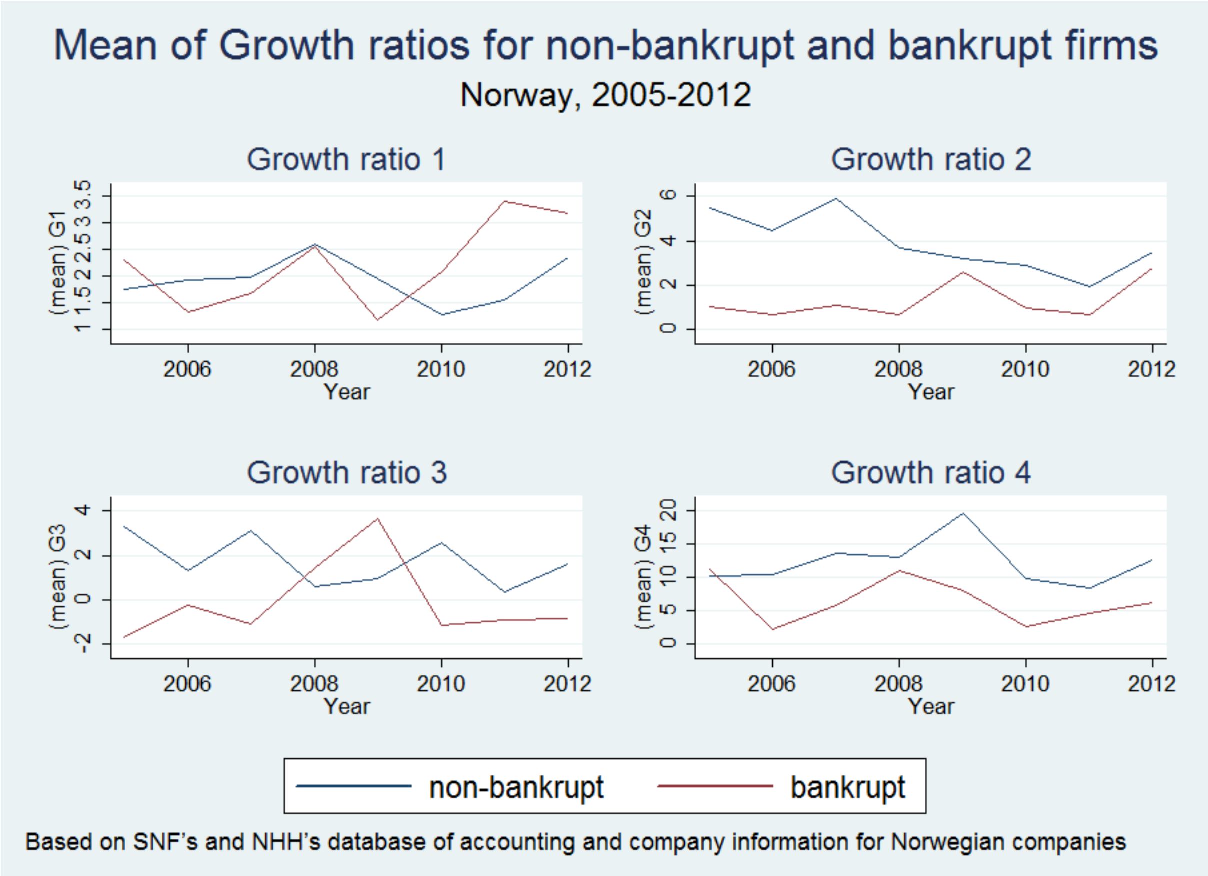
A.7 Mean of Profitability 1



A.8 Mean of Solidity measures



A.9 Mean of Growth ratios



A.10 Regression estimates (OLS)

Bankrupt	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Profit margin 1	0.00202***	0.00197***	0.00197***	0.00190***	0.00197***	0.00197***	0.00135***	0.00135***
	(0.000284)	(0.000277)	(0.000277)	(0.000269)	(0.000269)	(0.000270)	(0.000224)	(0.000225)
Liquidity ratio 1	0.00105***	0.00280***	0.00274***	0.00307***	0.00204***	0.00203***	-0.00134***	-0.00139***
	(0.000396)	(0.000389)	(0.000389)	(0.000382)	(0.000429)	(0.000429)	(0.000305)	(0.000306)
Solidity ratio 1	-0.0000889*	-0.0000563	-0.0000557	-0.0000586	-0.0000108	-0.0000105	-0.0000288	-0.0000283
	(0.0000533)	(0.0000479)	(0.0000478)	(0.0000469)	(0.0000403)	(0.0000404)	(0.0000324)	(0.0000323)
Solidity ratio 2	-0.00108***	-0.00165***	-0.00165***	-0.00177***	-0.00180***	-0.00181***	-0.00203***	-0.00205***
	(0.000380)	(0.000184)	(0.000179)	(0.000266)	(0.000175)	(0.000147)	(0.000305)	(0.000419)
2-year deficit		0.0218***	0.0217***	0.0205***	0.0246***	0.0246***	0.0227***	0.0227***
		(0.000422)	(0.000422)	(0.000420)	(0.000474)	(0.000476)	(0.000463)	(0.000463)
Size(1 if small)			0.00658***	0.00489***	0.00382***	0.00411***	0.00455***	0.00456***
			(0.000276)	(0.000276)	(0.000294)	(0.000296)	(0.000317)	(0.000319)
Age 5-9 years				-0.0105***	-0.00796***	-0.00771***	-0.00723***	-0.00734***
				(0.000323)	(0.000361)	(0.000363)	(0.000358)	(0.000360)
Age 10-25 years				-0.0137***	-0.0112***	-0.0111***	-0.0108***	-0.0108***
				(0.000274)	(0.000318)	(0.000318)	(0.000312)	(0.000313)
Age 26+ years				-0.0150***	-0.0124***	-0.0122***	-0.0116***	-0.0117***
				(0.000330)	(0.000379)	(0.000380)	(0.000354)	(0.000355)
Growth Revenue					-0.000000184	-0.000000190	-0.000000119	-0.000000116
					(0.000000333)	(0.000000330)	(0.000000358)	(0.000000355)
Growth assets					0.000000606	0.000000470	0.000000901	0.000000736
					(0.00000235)	(0.00000233)	(0.00000242)	(0.00000241)
Growth equity					-0.000000221	-0.000000212	-0.000000171	-0.000000172
					(0.000000201)	(0.000000201)	(0.000000166)	(0.000000166)
Growth current liabilities					0.000000314	0.000000291	0.000000513	0.000000516
					(0.000000676)	(0.000000678)	(0.000000684)	(0.000000682)
Constant	0.00742***	0.00263***	-0.00358***	0.00696***	0.00562***	0.00379***	0.00391***	0.00651
	(0.000209)	(0.000210)	(0.000325)	(0.000392)	(0.000443)	(0.000556)	(0.00124)	(0.00421)
Year effects	No	No	No	No	No	Yes	Yes	Yes
Industry effects	No	No	No	No	No	No	Yes	Yes
YearXindustry	No	No	No	No	No	No	No	Yes
R^2	0.003	0.009	0.009	0.013	0.015	0.015	0.018	0.019
Observations	1074192	1074192	1074192	1074192	918804	918804	901981	901981

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.11 Logistic regression with odds ratios

Bankrupt	(1)
Profit margin 1	1.028***
	(10.84)
Liquidity ratio 1	0.989
	(-0.56)
Solidity ratio 1	0.999
	(-0.61)
Solidity ratio 2	1
	(.)
2-year deficit	4.700***
	(67.27)
Size(1 if small)	2.204***
	(9.17)
Age 5-9 years	0.611***
	(-17.84)
Age 10-25 years	0.355***
	(-37.08)
Age 26+ years	0.229***
	(-23.98)
Growth Revenue	1.000
	(-0.24)
Growth assets	1.000
	(0.18)
Growth equity	1.000
	(-1.37)
Growth current liabilities	1.000
	(1.12)
Year effects	Yes
Industry effects	Yes
Observations	901140

Exponentiated coefficients; t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.12 Regression estimates (FE)

Bankrupt	(1)	(2)	(3)	(4)	(5)	(6)
Profit margin 1	0.00104***	0.000991***	0.000991***	0.000600***	0.000612***	0.000539***
	(0.000259)	(0.000246)	(0.000246)	(0.000196)	(0.000197)	(0.000180)
Liquidity ratio 1	-0.00155***	0.000412	0.000409	-0.00109**	-0.00206***	-0.00229***
	(0.000436)	(0.000405)	(0.000405)	(0.000491)	(0.000526)	(0.000537)
Solidity ratio 1	-0.000203*	-0.000165*	-0.000165*	-0.000303*	-0.000292*	-0.000305**
	(0.000111)	(0.0000982)	(0.0000982)	(0.000166)	(0.000165)	(0.000155)
Solidity ratio 2	0.0000466	-0.000169*	-0.000160*	-0.000250***	-0.000466***	-0.00171**
	(0.0000643)	(0.0000916)	(0.0000876)	(0.0000483)	(0.0000533)	(0.000795)
2-year deficit		0.0229***	0.0229***	0.0109***	0.00991***	0.00889***
		(0.000473)	(0.000473)	(0.000521)	(0.000517)	(0.000504)
Size(1 if small)			-0.000851*	-0.00158**	-0.00381***	0.000461
			(0.000500)	(0.000633)	(0.000645)	(0.000640)
Age 5-9 years				0.0133***	0.00598***	0.00574***
				(0.000416)	(0.000454)	(0.000447)
Age 10-25 years				0.0221***	0.00535***	0.00508***
				(0.000585)	(0.000685)	(0.000673)
Age 26+ years				0.0269***	0.00124	0.000552
				(0.000862)	(0.00102)	(0.00100)
Growth Revenue				-0.00000170**	-0.00000140*	-0.00000132*
				(0.000000860)	(0.000000779)	(0.000000746)
Growth assets				-0.0000110	-0.00000947*	-0.00000764
				(0.00000673)	(0.00000549)	(0.00000503)
Growth equity				-2.00e-08	-3.08e-08	-2.25e-08
				(8.00e-08)	(7.32e-08)	(6.59e-08)
Growth current liabilities				-0.000000155	-3.29e-08	-0.000000108
				(0.000000150)	(0.000000135)	(0.000000123)
Constant	0.0109***	0.00608***	0.00689***	-0.00519***	-0.00540***	0.0150***
	(0.000520)	(0.000498)	(0.000697)	(0.000867)	(0.000914)	(0.00295)
Year effects	No	No	No	No	Yes	Yes
YearXindustry	No	No	No	No	No	Yes
R^2	0.001	0.008	0.008	0.005	0.010	0.014
Observations	1074081	1074081	1074081	825318	825318	808837

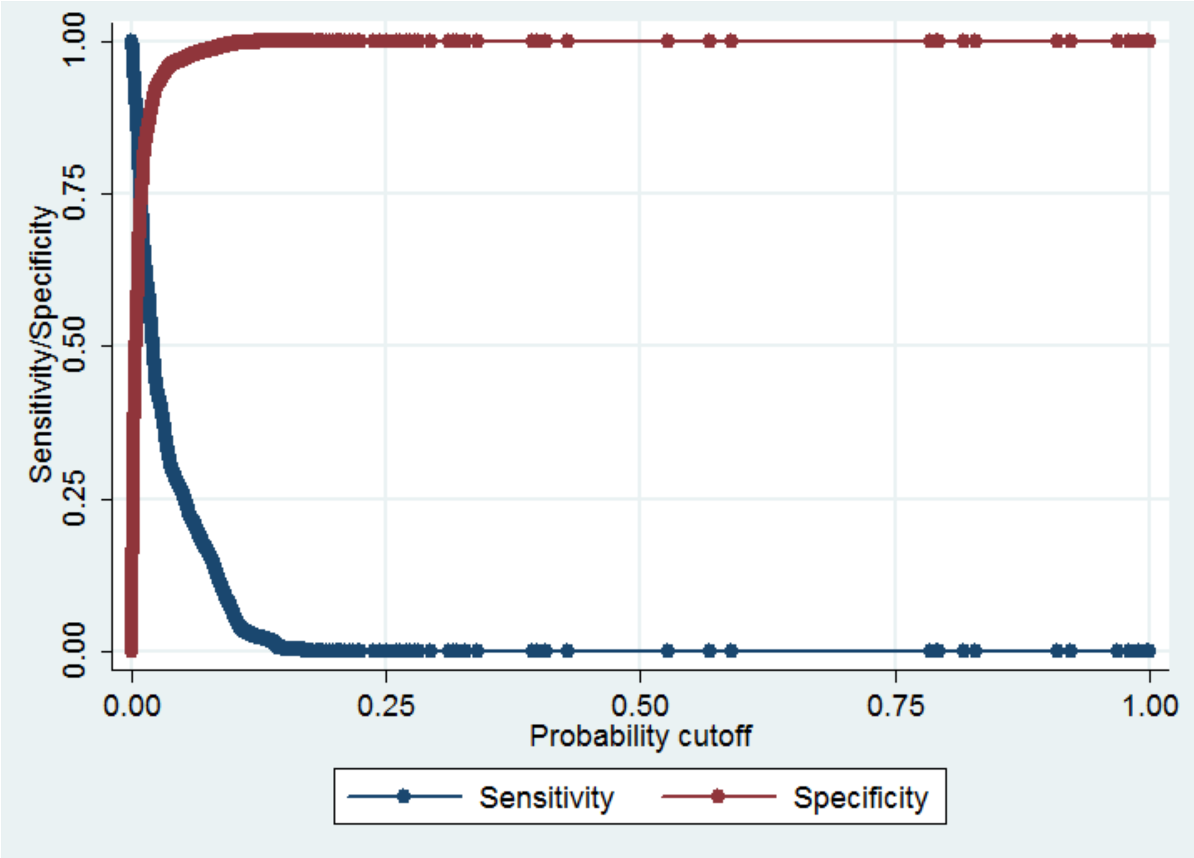
Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.13 Regression estimates of OLS, logit, RE and FE

	(1) OLS	(2) OLS w. growth	(3) OLS w. interaction	(4) logit	(5) RE	(6) FE	(7) FE w. interaction
Profit margin 1	0.00113***	0.00135***	0.00135***	0.0277***	0.00129***	0.000612***	0.000539***
	(0.000194)	(0.000224)	(0.000225)	(0.00255)	(0.000286)	(0.000197)	(0.000180)
Liquidity ratio 1	-0.00107***	-0.00134***	-0.00139***	-0.0106	0.00147***	-0.00206***	-0.00229***
	(0.000261)	(0.000305)	(0.000306)	(0.0190)	(0.000362)	(0.000526)	(0.000537)
Solidity ratio 1	-0.0000777*	-0.0000288	-0.0000283	-0.000507	-0.000191	-0.000292*	-0.000305**
	(0.0000409)	(0.0000324)	(0.0000323)	(0.000834)	(0.000143)	(0.000165)	(0.000155)
Solidity ratio 2	-0.00205***	-0.00203***	-0.00205***	0	0.0000935	-0.000466***	-0.00171**
	(0.000208)	(0.000305)	(0.000419)	(.)	(0.000468)	(0.000533)	(0.000795)
2-year deficit	0.0189***	0.0227***	0.0227***	1.548***	0.0163***	0.00991***	0.00889***
	(0.000414)	(0.000463)	(0.000463)	(0.0230)	(0.000484)	(0.000517)	(0.000504)
Size(1 if small)	0.00547***	0.00455***	0.00456***	0.790***	0.00222***	-0.00381***	0.000461
	(0.000297)	(0.000317)	(0.000319)	(0.0862)	(0.000526)	(0.000645)	(0.000640)
Age 5-9 years	-0.00961***	-0.00723***	-0.00734***	-0.492***	0.00102***	0.00598***	0.00574***
	(0.000321)	(0.000358)	(0.000360)	(0.0276)	(0.000364)	(0.000454)	(0.000447)
Age 10-25 years	-0.0132***	-0.0108***	-0.0108***	-1.037***	-0.00464***	0.00535***	0.00508***
	(0.000272)	(0.000312)	(0.000313)	(0.0280)	(0.000442)	(0.000685)	(0.000673)
Age 26+ years	-0.0140***	-0.0116***	-0.0117***	-1.474***	-0.0115***	0.00124	0.000552
	(0.000310)	(0.000354)	(0.000355)	(0.0615)	(0.000611)	(0.000102)	(0.00100)
Growth Revenue		-0.000000119	-0.000000116	-0.0000209	-0.000000862**	-0.00000140*	-0.00000132*
		(0.000000358)	(0.000000355)	(0.0000873)	(0.000000404)	(0.000000779)	(0.000000746)
Growth assets		0.000000901	0.000000736	0.000113	-0.00000561*	-0.00000947*	-0.00000764
		(0.00000242)	(0.00000241)	(0.000638)	(0.00000292)	(0.00000549)	(0.00000503)
Growth equity		-0.000000171	-0.000000172	-0.000182	-9.69e-08	-3.08e-08	-2.25e-08
		(0.000000166)	(0.000000166)	(0.000133)	(0.000000106)	(7.32e-08)	(6.59e-08)
Growth current liabilities		0.000000513	0.000000516	0.0000630	2.28e-08	-3.29e-08	-0.000000108
		(0.000000684)	(0.000000682)	(0.0000563)	(0.000000277)	(0.000000135)	(0.000000123)
Constant	0.00493***	0.00391***	0.00651	-5.828***	0.00346***	-0.00540***	-0.0150***
	(0.00113)	(0.00124)	(0.00421)	(0.170)	(0.000824)	(0.000914)	(0.00295)
Year effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry effects	Yes	Yes	Yes	Yes	No	No	Yes
YearXindustry	No	No	Yes	No	No	No	Yes
R^2	0.017	0.018	0.019			0.010	0.014
Observations	1051403	901981	901981	901140	918804	825318	808837

Standard errors in parentheses * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

A.14 lsens output



A.15 estat classification output

```
Logistic model for konk

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |      6315        230054  |     236369
     -     |      2254        662517  |     664771
-----------+--------------------------+-----------
   Total   |      8569        892571  |     901140

Classified + if predicted Pr(D) >= .0095106
True D defined as konk != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   73.70%
Specificity                     Pr( -|~D)   74.23%
Positive predictive value       Pr( D| +)    2.67%
Negative predictive value       Pr(~D| -)   99.66%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   25.77%
False - rate for true D         Pr( -| D)   26.30%
False + rate for classified +   Pr(~D| +)   97.33%
False - rate for classified -   Pr( D| -)    0.34%
--------------------------------------------------
Correctly classified                        74.22%
--------------------------------------------------
```