LUND UNIVERSITY
School of Economics and Management

MASTER THESIS

Bankruptcy Prediction: Static Logit Model versus Discrete Hazard Models Incorporating Macroeconomic Dependencies

Master Programme in Finance

Lund University

Spring 2010

Authors:
Matej Duda
Henning Schmidt

Tutor:
Dr. Hossein Asgharian

Abstract

The purpose of this master thesis is to (i) compare the out-of-sample prediction power of one static logit model and two hazard models, (ii) explore whether incorporating macroeconomic patterns improves forecasting results, and (iii) examine the determinants of corporate failures from the pool of accounting and market driven variables.

We perform our study on 102 US listed manufacturing firms that defaulted between 2000 and 2009. The out-of-sample period spans over the crisis period 2007-2009.

We find that the static logit outperforms both hazard model specifications in out-of-sample accuracy. Next, we find that incorporating macroeconomic patterns can improve forecasting results of hazard models. The most important firm-specific determinants of bankruptcy are profitability, stock return, short-term solvency, further significant variables are cash holdings, relative market size, and leverage.

Keywords: bankruptcy prediction, discrete-time hazard model, static logit, multi-period logit, macro dependent baseline function

Acknowledgement

We wish to thank our supervisor and programme director Professor Hossein Asgharian for his helpful comments, suggestions and invaluable guidance. We also would like to convey thanks to Wolfgang Hess for providing his insights into hazard models.

Finally, we thank Professor Lynn M. LoPucki (UCLA School of Law, USA) for providing us with his Bankruptcy Research Database.

Table of Contents

Tables

Figures

I. Introduction

This chapter gives background information about default probability prediction as well as our motivation to contribute to the ongoing discussion on credit risk assessment. We also state our purpose and comment on delimitations. Finally, we outline the rest of the thesis.

A. *Background and Motivation*

The global financial crisis and the increased number of corporate defaults emphasize the importance of credit risk management. In finance terms, credit risk management refers to methods and processes of assessing credit risk which can be defined as "the potential that a borrower or counterparty will fail to meet its obligations in accordance with the terms of an obligation's loan agreement." (Sobehart, Keenan, and Stein (2001)) It has been widely acknowledged that the lack of understanding of credit exposures may lead to incorrect use of risk management tools, which in turn might cause major losses by financial institutions triggering their bankruptcies. As experienced in 2008 and 2009, losses suffered as a result of credit defaults may lead to a shrinking or rather a freezing of credit markets. This may result in an overall decrease in economic productivity. For this reason, correct estimates of probability of bankruptcy are of tremendous importance to banks and financial institutions. Consequently, the field of bankruptcy forecasting has gained significant attention in economic research and among practitioners alike.

Bankruptcy prediction refers to the process of calculating the probability of bankruptcy or financial distress of public companies. Default probabilities can be quantified in several ways, but one of most common approaches is the use of statistical models for default prediction building upon so-called credit scores. Credit scores are numerical expressions that characterize the creditworthiness of borrowers based on the statistical analysis of their files. This method is well established and such credit-scoring models are commonly used in banks' loan approval processes[1]. Credit scoring models employ market data as well as the data from the accounting of public firms to calculate various ratios, which indicate the risk of failure.

[1] Mester (1997) provides evidence showing that 70 percent of banks use credit scoring in their business lending.

In this thesis, we examine three of the most discussed ways of calculating credit scores by estimating the out-of-sample probabilities of default for the US manufacturing industry. Our three models are (i) a simple static logit model, (ii) a discrete hazard model allowing for time-varying firm-specific characteristics, and (iii) a hazard model allowing for time-varying firm-specific characteristics and macroeconomic dependencies. The models are fitted to the data using the in-sample period from 2000 to 2006 (estimation window) and evaluated on the out-of-sample period (validation window) consisting of the years between 2007 and 2009.

A number of studies have addressed the bankruptcy prediction using simple logistic regressions. Papers by Ohlson (1980), Wiginton (1980), Zmijewski (1984), Lenox (1999), and Westgaard and van der Wijst (2001) are some of the most influential ones. A common finding is that the probability of default is triggered by ratios within the following categories: (i) cash flow, (ii) profitability, (iii) leverage, (iv) size, (v) liquid asset, (vi) short term solvency, and (vii) activity.

Simple logit models have been criticized for imposing the assumption that default data is composed of two distinct and separate populations. (Nam et al. (2008)) In this case, simple logit might not produce accurate results in out-of-sample forecasting. Another deficiency of static models is the requirement of stable failure rates. In order to work properly static models need a failure process, which is stable over a considerable period of time. (Nam et al. (2008)) Additionally, it has been argued that these models ignore the dynamic character of firms' financial structures. This leads to biased and inconsistent estimates of default probabilities. (Shumway (2001))

Hazard models address the issues mentioned by Shumway (2001) and Nam et al. (2008) by drawing samples from two identical populations, accounting for time, as well as by the possibility of capturing the impact of general macroeconomic conditions on corporate defaults. Consequently, hazard models yield generally consistent estimates (Shumway (2001)).

Nevertheless, the choice between static logit model and hazard model is not clear particularly in terms of the out-of-sample performance. Whilst aforementioned researchers proclaim superiority of complex models over single period logit models, there is also recent scientific evidence provided by Fuertes, A.-M., and Kalotychou (2006), Rodriguez and Rodriguez (2006) and Fantazzini and Silvia (2009) showing that parsimonious models produce better out-of-sample forecasts.

Hence, further investigation of the out-of-sample performance of bankruptcy predicting models is necessary in order to indentify the best model and to solve the existing controversy. Our investigation is performed on the newest bankruptcy data, which makes this study a considerable contribution to the field.

B. *Purpose*

The purpose of this thesis is threefold.

First, we compare a static logit model and two hazard models allowing for multi-period data in terms of bankruptcy prediction accuracy during the recent financial crisis. To date, there has been little empirical research conducted on bankruptcy data of recent financial crisis. Additionally, previous comparisons of these models during non-crisis times show controversy. While some studies document the superiority of more complex models (hazard models), other authors find parsimonious models (simple logit) to be more accurate (s. Section II).

Second, we investigate potential performance improvements reached by incorporating macroeconomic patterns into hazard models, as suggested by previous research.

Third, one common critique of default probability models based on credit scores is the incorporation of historical accounting based information. In contrast, predictive models should include forward-looking information, such as market-data[2], rather than historical information. Therefore, the third purpose is to find an optimal combination of accounting and market driven variables able to predict bankruptcy.

C. *Delimitations*

We are aware of certain boundaries of our study such as the fact that employed estimation models are based on historical, accounting based data pulled from the companies' financial statements and do not reflect all information possible embedded in the market value

[2] Market data expresses today's value of future expectations.

of the firm. We partly mitigate this problem by incorporating market variables into our models. However, the models might still be sensitive to possible manipulations of book earnings, which are the common practices in order to give more positive view of the company, particularly when on the verge of bankruptcy.

Additionally, employing yearly data from financial statements ignores the fact that companies' financial position at the time of filing might be substantially deteriorated relative to the last reporting date. In such cases, the data drawn does not reflect the true economic reality at the time of bankruptcy and the conclusions based on such data may be dubious. Availability of e.g. quarterly data would improve this limitation.

Our dataset consists of US manufacturing companies between 2000-2009. Considering only US manufacturing companies does not allow us to investigate the effects of industry or country. Since the number of defaulted companies is very low relative to healthy companies in the population, the sample is constructed by so-called one-to-one matching. This procedure violates the random selection principle and might lead to biased results.

D. *Outline of the Thesis*

The rest of the thesis is structured as follows. Section II presents previous research on corporate default prediction. Section III specifies applied theoretical models, tests for misspecification as well as evaluation and comparison methods. Section IV describes the data and explains the sample selection process. In Section V, the estimation results are described and discussed. In Section VI, we compare the explanatory power of the models and rank our models. The final Section concludes.

II. Previous Research

The literature review is divided into three sub-sections. In the first sub-section, we present the previous research on discriminant analysis (DA) and static logit models followed by hazard models discussed in the second sub-section. Third sub-section discusses some controversies in empirical works that compares the predictive performance of static logit and hazard model classes.

A. *Default Prediction with Discriminant Analysis (DA) and Static Logit*

There have been various attempts to design models for predicting the probability of default of corporations. Altman (1968) is to our best knowledge the first author who employs a statistical model in order to predict bankruptcy. He uses DA, on a grouped dataset of failed and non-failed manufacturing companies, to construct a financial-ratio based model for predicting corporate defaults. The resulting Z-score gained global prominence. (Peresetsky, Karminsky, and Golovan (2004))

The DA has been heavily criticized for restrictive distributional assumptions. This technique is based on the assumptions that the explanatory variables follow a multivariate normal distribution, and that a sample is randomly selected from the populations of defaulted and surviving companies. While the former assumption has been proved to be violated regularly by real world data, the latter assumption may be violated by the matching technique of sample selection. (Lenox (1999)) This turned the attention of researchers to logit models, which address the restriction of normality.

One of the first authors employing logit models for bankruptcy prediction is Ohlson (1980). As determinants of bankruptcy of US industrial companies, he identifies four statistically significant basic factors: (i) size of the company, (ii) financial structure, (iii) performance, and (iv) current liquidity. Additionally, he finds that error rates of the models can be improved by employing timely financial data. This means that a short time span between the disclosure of the relevant annual report and the date of default adds to accurate failure risk assessment.

Westgaard and van der Wijst (2001) employ the logit approach in order to predict bankruptcies in the Norwegian business sector and find that the model is able to predict defaults sufficiently well. As determinants of default Westgaard and van der Wijst (2001) identify the ratios of cash flow to debt, liquidity, solidity3, financial coverage as well as size and age to be statistically significant.

Lenox (1999) revaluates the performance of logit and probit models and DA approaches on a sample of 949 UK listed companies. His two most important findings are that (i) cash-flow and leverage have non-linear effects on bankruptcy probability, and that (ii) well identified logit and probit models are able to identify bankrupting companies more accurately than DA models. This difference is even larger if realistic threshold probabilities4 are used and for Type I errors5.

B. *Default Prediction with Hazard Models*

Logit models have been criticized for their static character. Hence, they are unable to capture any dynamics of a company's financial structure. In order to use a broad range of information and to allow for time varying covariates Shumway (2001) suggests a hazard model for the estimation of default probabilities that uses both accounting based information and market information.

Shumway (2001) uses a dataset of 300 bankruptcies (1962-1992), for which hazard models yield solid forecasting results. He concludes that many accounting based ratios included in previous static models become insignificant when employed in hazard models. In contrast, market variables such as a firm's market size, past stock returns, and the idiosyncratic standard deviation of its stock returns are found strongly related to default. Additionally, Shumway (2001) reports that his estimates are consistent unlike static DA and logit models.

---

[3] Defined as equity to total assets

[4] If the threshold probability is exceeded by a company, it is predicted to default. Further discussion is provided below.

[5] This error occurs when a company is wrongly classified as non-defaulted, when it in fact fails. Further explanation and discussion is provided in Section III.

Nam et al. (2008) compare a static logit, a hazard model using panel data and a hazard model using macroeconomic variables. On a sample of 367 Korean companies between 1999 and 2000, they show that dynamic models with time-varying covariates yield superior performance relative to static logit. Comparing the two dynamic models, the hazard model with macro-dependencies is more accurate.

Likewise Bellotti and Crook (2009) show that accounting for macro-economic indicators such as interest rate and unemployment index significantly affects default probability and improves model prediction accuracy.

C. *Controversy in Out-of-Sample Forecasting*

A common finding of aforementioned studies is that DA analysis has the weakest predictive power compared to logistic regression. The latter, however, seems to be outperformed by hazard models allowing for the panel character of the bankruptcy data.

Nevertheless, there is a scientific evidence provided by Fuertes, A.-M., and Kalotychou (2006), Rodriguez and Rodriguez (2006), that complex models are only more likely to provide accurate forecasts when tested in-sample, while parsimonious models produce better forecasts in ex-post periods. The authors find this to hold for default probabilities of sovereign debt. Superior out-of-sample predictions of simple logit models over more advanced models is also found by Fantazzini and Silvia (2009) in the field of credit risk default for Small Medium Enterprises.

This controversy can be explained in two different ways. Fuertes, A.-M., and Kalotychou (2006) explain the outperformance by referring to Clements' and Hendry's (1998) theoretical finding that "profligate parameterizations do not help forecasting, particularly the further ahead one forecasts" (Fuertes, A.-M., and Kalotychou (2006))6 . Fantazzini and Silvia (2009) also provide a potential explanation of the outperformance. They prove that even if the static logit may not be suitable for function estimation due to its bias, it can still yield good results in terms of classification rule.[7] (Fantazzini and Silvia (2009)) This means that even though a static logit model might not be the valid model specification, it

[6] For a sound theoretical foundation the interested reader is referred to Clements and Hendry (1998) chapter 12
[7] For the proof see Fantazzini and Silvia (2009)

might still show solid forecasting results. More complex models might therefore provide better guidance for macroeconomic policy or analysis purposes because they might reflect the true data generating process better (Fuertes, A.-M., and Kalotychou (2006)), while logit model is better suited for forecasting.

Although prevailing evidence exists in favor of hazard models, static logit can indeed outperform complex model specifications. Further examination of the predictive power of default models is therefore necessary in order to shed more light on this controversy.

This study uses the newest data, which will allow us to examine predictive power of parsimonious vs. complex models as well as the impact of macro-economic indicators during the abruptly changing period of 2007-2009. We will broadly follow the methodology by Nam et al. (2008).

III. Methodology

This section introduces the models used in this thesis. The first sub-section discusses logit models. Then hazard models, which are widely used in practice, are discussed. The following sub-section describes the selection of variables used in our models. Sub-Section F deals with macroeconomic dependencies. Sub-Section G introduces test frameworks for omitted variables and hetersokedasticity. Finally, several validation tests are presented.

A. *Logit Models*

Because of the various shortcomings of DA (s. section II), researchers focused on logit models in recent studies. The class of logit models is advantageous, as it does not impose the assumption of normality on predictors. Secondly, logit models return a probabilistic output and thus no score has to be converted into a probabilistic measure, which might be an additional source of errors (Ohlson (1980)).

Logit models assume that for any corporation, given a set of attributes, there is a definable probability that it will default. Hence, the probability of default depends conditionally on these attributes. This can be expressed by a function of the following form:

$$Y_i^* = \beta' x_i + \varepsilon_i,$$

where

$$\begin{aligned} Y_i &= 1 \qquad \text{if } Y_i^* \geq 0 \\ Y &= 0 \qquad \text{otherwise,} \end{aligned} \tag{1}$$

where β is a set of regressors, x_i is a set of attributes determining a potential default ($Y_i = 1$). ε_i represents the error term. (Jones and Hensher (2008)) In order to estimate the parameters in the β-vector a log-likelihood function of the following form is maximized:

$$\ln(L) = \sum_i Y_i F(-\beta' x_i) + \sum_i (1 - Y_i)\left(1 - F(-\beta' x_i)\right), \tag{2}$$

where $F(\cdot)$ is the logistic distribution, which is chosen conditional on the assumptions made about ε_i (Jones and Hensher (2008)).

Given the parameter estimates of β the probability of default for firm i can be calculated as follows:

$$P_i(Y_i = 1|x_i) = F(\beta' x_i) = \frac{1}{1 + e^{-\beta' x_i}} \tag{3}$$

There are also multi-period versions of the logit model, which will be described in the next sub-section.

B. *Hazard Models*

In this sub-section, we discuss discrete hazard models. Hazard models are a type of survival models, in which the covariates are related to the time that passes before some event[8] occurs. This model class is closely related to the logit model discussed in the previous sub-section.

Time to failure of a firm is known as the survival time and is denoted as t. The continuous random variable t follows some density function, $f(t_i, x_i, \beta)$, and has some cumulative density function, $F(t_i, x_i, \beta)$. The probability that a company survives the time span t is expressed by the survival function, $S(t_i, x_i, \beta)$:

$$S(t_i, x_i, \beta) = 1 - \sum_{j<t} f(j, x_i, \beta) = 1 - F(t_i, x_i, \beta) \tag{4}$$

Additionally, the model incorporates a hazard function $h(t_i, x_i, \beta)$, which expresses the probability of default at t, given survival until t. In other words, the hazard function (a.k.a. hazard rate) is the ratio of the probability density function $f(t_i, x_i, \beta)$ to the survival function $S(t_i, x_i, \beta)$:

$$h(t_i, x_i, \beta) = \frac{f(t_i, x_i, \beta)}{S(t_i, x_i, \beta)} \tag{5}$$

Hence, the hazard rate might be seen as the instantaneous risk of default. (Nam et al. (2008))

The parameter estimates are obtained by maximizing a likelihood function, which takes the following form for hazard models:

$$L = \prod_i h(t_i, x_i, \beta)^{y_i} S(t_i, x_i, \beta) \tag{6}$$

Shumway (2001) shows that the likelihood function of the multi-period logit model is equivalent to a discrete-time hazard model with a hazard rate $h(\cdot) = F(t_i, x_i, \beta)$, which takes

[8] Here: bankruptcy

the same form as the cumulative probability function of a logit model (s. (7)). Therefore, hazard models can be easily estimated by using the logistic regression technique. At the same time, the model also allows for the incorporation of macroeconomic dependencies. Partitioning β into β_1 and β_2 leads to the following form of the hazard function and helps to understand the technique of incorporating macroeconomic dependencies:

$$h(\cdot) = F(\cdot) = \frac{1}{1 + e^{-(\beta_1^{'}\kappa_t + \beta_2^{'}x_i)}} \tag{7}$$

As it can be seen from (7) the hazard rate consists of some time-dependent κ_t, which is also called as baseline hazard function and represents some macroeconomic variable. It expresses the hazard rate in the absence of covariates (x_i). For example, if the firm i's value of $x_i\beta$ was 1,0986 (=ln(3)), the hazard of that firm would be three times as high as the baseline. Therefore, the probability of default for that firm would therefore be three times higher than for the firm with all x_i equal to zero.

There are different specifications of the baseline hazard function. One example of baseline hazard functions is a function of firm's age[9], ln(age), as suggested in Shumway (2001). Shumway (2001) assumes a certain level of homogeneity across the characteristics of the firms listed in the same period. This seems reasonable, since listing is conditional on fulfillment of certain requirements set by a regulator. Further examples are dummy variables indicating the number of zeros that precede the current observation as used in Beck, Katz, and Tucker (1998), or even macroeconomic variables, such as the rate of recent defaults, volatility of foreign exchange, or changes in interest rates (Hillegeist et al. (2001)).

The second part of the hazard function, $\beta_2^{'}x_i$, is a function of firm specific characteristics represented by financial ratios.

The model can even incorporate time varying covariates by making x dependent on time. This is needed in order to reflect the change in a firm's financial conditions. Ignoring these dynamics will lead to biased and inconsistent estimates. (Shumway (2001))

[9] The starting point of a company's lifespan is defined as the date of listing at the stock exchange.

C. *Hazard versus Logit Models*

The best comparison of static models with hazard models is provided in Shumway (2001). He argues that the static models are inappropriate for default prediction due to the character of bankruptcy data. The frequency of default events is very low, which forces researchers to include several years to obtain a suitable sample size. Underlying characteristics of the majority of firms evolve over time but static models allow only for one set of explanatory variables for each firm. Researchers usually observe bankruptcy data in the year preceding the bankruptcy and disregard the data of sound companies eventually going bankruptcy. This gives ground to a selection bias.

To this end, the biggest advantage of hazard models over static models is that they explicitly account for time (Shumway (2001)). In other words, the explained variable in the hazard model is the length of time a firm spends in the healthy group before leaving it.

Shumway (2001) summarizes the reasons why the hazard model should be preferred over the static model: The first is the failure of the static logit to account for "each firm's period at risk" (Shumway (2001)). The second reason is the incorporation of time-varying explanatory variables, i.e. annual observations of chosen ratios from financial statements. The third reason deals with forecasting: Shumway (2001) argues that since hazard models utilize more data, their out-of-sample predictive power must be necessarily more accurate. This becomes evident if one thinks of a hazard model as logit model incorporating each year of the firm as a separate observation. Our sample consists of observations for the period 2000 to 2009, which means approximately ten times[10] more data than for the static binary model.

Nam et al. (2008) also argue that unlike static models, hazard models can incorporate macroeconomic variables and therefore are much more flexible. The authors argue that controlling for the macro environment is achieved by modifying the shape of the hazard baseline function as explained in the previous sub-section.

A further problem with the basic logit model is the restrictiveness of its IID assumption as criticized by Jones and Hensher (2008). They also point out a logit model's inability to capture firm's heterogeneity not represented by the firm specific characteristics and the errors.

[10] This multiple will be slightly lower, since the bankrupt companies leave the sample

Yet, there is an empirical argument in favor of static logit models. As outlined in section II, the classical logit might produce better out-of-sample prediction of failure compared to advanced models. Evidence and theoretical proof are documented in Fuertes, A.-M., and Kalotychou (2006), Rodriguez and Rodriguez (2006), as well as in Fantazzini and Silvia (2009).

D. *Specification of Models*

The first aim of this thesis is to empirically investigate potential improvements attained by using discrete hazard models over of the static model. The second is to investigate potential performance improvements reached by incorporating a macro dependent baseline into a hazard model. The third is to examine significance of market driven variables in estimating default probabilities and to find the optimal combination of accounting and market driven variables able to predict bankruptcy. To address the first two aims, we formulate three competing models as in Nam et al. (2008).

Model I

Static (single-period) logit, in which the covariates do not change over time. The logit model is formulated as:

$$P(y_i = 1) = \frac{1}{1 + e^{-x_i \beta}} \tag{8}$$

Model II

Duration model with time-varying covariates, which can also be expressed in the form of a logit model:

$$P(y_i = 1) = \frac{1}{1 + e^{-x_{i,t} \beta}} \tag{9}$$

This specification does not include the time-variant baseline function reflecting the macro-economic environment. The probability of failure of the firm in this exponential hazard model does not depend on its age either.

Model III

Duration model with time-varying covariates and macro-economic dependencies:

$$P(y_{i,t} = 1) = \frac{1}{1 + e^{-(\beta_1' \kappa_t + \beta_2' x_{i,t})}} \tag{10}$$

This specification includes a baseline hazard function κ_t reflecting macroeconomic environment. Possible variables might include the volatility of interest rate or foreign exchange rate, or any other system-level variable capturing the rate of defaults independent of individual characteristics of the firm. We even test whether a function of firm's age, $ln(age)$, might be used to capture the potential homogeneity of bankrupt firms founded in the same years. We also examine the use of time dummies as a baseline function.

E. *Selection of Variables*

This sub-section explains the choice of variables incorporated in the models. Accounting based ratios and market based ratios are discussed separately.

Accounting Based Ratios

Various financial variables can be used to determine the factors driving the default behavior of manufacturing firms. Prior research suggests many different approaches. The most popular became the scores created by Beaver (1966), Altman (1968), Ohlson (1980), and Zmijewski (1984), who use accounting ratios to find the probability of default using a static model.

Altman (1968) finds that the company is likely to go bankrupt if it is highly leveraged, unprofitable and experiences cash-flow problems. In his O-score Ohlson (1980) identifies four basic factors as being statistically significant in determining the default probability: size of the company (log of total assets), measure of financial structure (total liabilities/ total assets), measures of performance (net income/total assets) and a measure of current liquidity (working capital/ total assets). Zmijewski (1984) confirms this by finding that bankruptcy is a decreasing function of return on assets (ROA), liquidity (current asset/current liabilities) and an increasing function of leverage (debt/total assets).

The aforementioned results seem to be very intuitive. Bankruptcy is often triggered by the inability to serve debt. This is more likely to occur when the company does not have access to external financing or has cash-flow problems. Therefore, we can expect the firm to go bankrupt if its current year cash-flow is insufficient to cover debt obligation. This is a direct connection to free cash-flow theory. Its main objective is to find a trade-off between lowering costs of asymmetric information by distributing cash-flow through dividends and thereby maximizing firm value and the increased probability of default (Jensen (1986)). To this end, the size of companies' indebtedness is an expected determinant of default probability.

Next, large companies are expected to be less likely to have problems accessing external finance markets. The logarithm of total assets as a proxy for firm's size has traditionally been used to avoid the non-stationarity problem.

Further, it is not unrealistic to hypothesize that a firm's profitability is a determinant of bankruptcy. Firms are more likely to default if they go through a low profitability period.

Market Based Ratios

Campbell, Hilscher, and Szilagyi (2008) add to the research on corporate default mainly by proposing the use of market-valued total assets (market value of equity + book value of liabilities) instead of measuring them at book-value. They find that such a measure has slightly better explanatory power than book total assets because market prices better reflect the prospect of the firm. Campbell, Hilscher, and Szilagyi (2008) further suggest the use of the logarithm of market capitalization to that of S&P 500 index instead of log(total assets) as size approximation. The last market driven variable employed in their study is the excess stock return of the firm over S&P 500.

Market variables are also used in Shumway (2001). He uses market equity based measures such as market capitalization to reflect the fact that as the firms approach bankruptcy equity is usually discounted by the market. One further market driven variable included by Shumway (2001) is the volatility of stock return (sigma – standard deviation of each firm's stock return). There is a clear connection between sigma and bankruptcy. Higher volatility of stock is caused by higher volatility of cash-flows, which in turn puts a firm at higher risk of not being able to meet its interest payments.

F. *Macroeconomic Dependencies*

There are several ways to proxy a baseline function accounting for macro-economic development. Incorporating a baseline function allows examining the individual hazard rate for the hypothetical firm with all covariates equal to zero. The idea behind the use of the baseline rate in discrete hazard models is that we want to capture the increase/decrease in hazard rates of bankruptcies during the recession/economic expansion.

The first option is to use time dummies for each of the observation in the sample. Another approach is to employ the Annual Rate[11] developed by Hillegeist et al. (2001). Next, as suggested by Nam (et all 2008) volatility of foreign exchange rate or the change in interest rates can be used as a baseline function. Shumway (2001) suggests the use of $ln(age)$, as discussed above (s. section III.B). We discuss baseline specifications examined in this study in turn.

Time Dummies

To control for the time, in which bankruptcies occur it is possible to employ time dummies as a proxy for the baseline function (Beck, Katz, and Tucker (1998)). The important distinction from the original baseline is the fact that the dummy marks the time that passes between the predetermined period until the current observation. Thus, for example all observations start in the year 2000 regardless of the time they had been at risk before the year 2000. This causes that all the observations of companies' data are left censored[12]. Therefore, it is necessary to assume a constant hazard rate, which means that the individual hazard rate for the firm *i* is independent of survival period.

Further, this "baseline" specification provides a good approximation of economic conditions in the observed periods. However, it will have low or no forecasting power especially in the crisis periods caused by macro-economic factors.

---

[11] Explained below

[12] The starting data point for the firm's age is before 2000 but it is unknown when exactly.

Annual Rates

It is a well known fact that recessions and expansions significantly impact the probability of bankruptcy. Hillegeist et al. (2001) argue that this temporal dependence may lead to understated standard errors (overstated t-statistics) in a logit model. By incorporating a time-varying baseline hazard rate the problem is effectively mitigated. The Annual Rate by Hillegeist et al. (2001) is the ratio of corporate bankruptcies to total number of firms in the sample over the previous year. This autoregressive specification is therefore the actual realization of unconditional baseline hazard rate in the previous period. Such specification should yield unbiased results (Hillegeist et al. (2001)). Further, unlike dummy variables, it is possible to forecast with the Annual Rate as a proxy of baseline function.

However, the forecasting power of this specification is questionable in case of unexpected shocks. Nam et al. (2008) also criticize the Annual Rate and dummy variables approaches stating that such indirect measures are less effective in "*capturing economy wide common effects since the firm's historical survival period cannot properly reflect the overall macro-dependencies and their correlations*."

Macro Variables

The ideal way to incorporate macro dependencies into the baseline function is to directly include them in the form of system-level variables causing the temporal dependence of hazard rate. Nam et al. (2008) suggest the use of volatility of foreign exchange rate or the change in interest rates for this purpose. The appropriate macroeconomic variables must be highly correlated with the default pattern similar to unconditional hazard function (Nam et al. (2008)). The included variables must however not be correlated to avoid the multicollinearity problem.

G. *Tests for Misspecification*

This sub-section describes tests for omitted variables and heteroskedasticity.

Test for Omitted Variables

The most convenient framework for testing for omitted variables is a Lagrange multiplier (LM) framework. Following Verbeek (2009), testing for omitted variables, z_i, is easily done by evaluating whether expression (11) is significantly different from zero.

$$\sum_{i=1}^{N}\left[\frac{y_i - F(x_i'\hat{\beta})}{F(x_i'\hat{\beta})\left(1 - F(x_i'\hat{\beta})\right)} f(x_i'\hat{\beta})\right] z_i \tag{11}$$

If z_i enters the model with zero coefficients, the term in the squared brackets, ε_i^G, and z_i should be uncorrelated.

The LM test statistic is obtained from an auxiliary regression, where a vector of ones is regressed on the $K + J$ variables $\varepsilon_i^G x_i'$ and $\varepsilon_i^G z_i'$. N times the uncentered R^2 does follow a χ_J^2 distribution, where J indicates the degrees of freedom, under the null hypothesis that z_i enters the model with no coefficients. (Verbeek (2009))

Test for Heteroskedasticity

Heteroskedasticity is a serious problem for logit models as it causes parameter estimates to be inconsistent. (Verbeek (2009)) Davidson and MacKinnon (1984) argue further that, because these models are usually estimated using cross-sectional data, it is a problem, which is likely to be encountered quite often.

The variance of the logit model is defined as $\pi^2/3$, in the case of heteroskedasticity the variance of ε_i depends on exogenous variables z_i, which should not contain a constant:

$$V\{\varepsilon_i\} = h\frac{\pi^2}{3}(z_i'\alpha), \tag{12}$$

where h is some function $h > 0$ with $h(0) = 0$, and $h'(0) \neq 0$.

We test the null hypothesis of homoskedasticity against the alternative of hetersokedasticity of the following form (Quantitative Micro Software (2007)):

$$var(\varepsilon_i) = \exp(2z_i'\gamma), \tag{13}$$

where γ is an unknown parameter, z_i is vector of explanatory variables, which are suspected to cause heteroskedasticity. The test static is obtained by evaluating the explained sum of squares of the regression below:

$$\frac{(y_i - \hat{p}_i)}{\sqrt{\hat{p}_i(1-\hat{p}_i)}} = \frac{f\left(-x_i^{'}\hat{\beta}\right)}{\sqrt{\hat{p}_i(1-\hat{p}_i)}} x_i^{'} b_1 + \frac{f\left(-x_i^{'}\hat{\beta}\right)\left(-x_i^{'}\hat{\beta}\right)}{\sqrt{\hat{p}_i(1-\hat{p}_i)}} z_i^{'} b_2 + v_i, \tag{14}$$

where $\hat{p}_i$ indicates the estimated probabilities of default, $\hat{\beta}$ indicates the obtained parameter estimates, and v_i is the error term.

Under the null hypothesis of homoskedasticity the explained sum of squares follows a χ_J^2 distribution where J indicates the number of degrees of freedom, which is equal to the number of variables in z_i.

H. *Model Evaluation Approaches*

The usual procedure in credit risk modeling starts by identifying statistically significant indicators of bankruptcy and specify a correct model. This is described in the previous sub-sections. Second step of credit analysis is to evaluate obtained scores in terms of out-of-sample predictive power. This sub-section presents several tests in order to evaluate our models and to compare models to one another.

Model Accuracy

Model accuracy is the most discussed dimension of model quality; however it is only one of several. Models can fail in two ways. Either the model predicts a company to survive when it actually fails (Type I error). In this case, an investor might lose promised interest payments, principal or both. She might also suffer from a decline in the obligation's market price.

The model might also predict failure when a company in fact survives (Type II error). In this case, the investor might lose interest payments and fees when the loans are turned down or lost through non-competitive bidding. (Sobehart, Keenan, and Stein (2001)) In case of tradable obligations, she might sell them at disadvantageous market price even though the obligation could have been held to maturity without facing risk of default.

Concluding from the above, a model should accurately classify defaulters and non-defaulters. The classification rule used in evaluating the forecast accuracy is as follows:

$$\text{Predict } Y_i = 1 \text{ if } F(\beta' x_i) > P^* \tag{15}$$

where P^* is a threshold probability, which is chosen by the researcher. One prediction rule might be to set P^* equal to 0,5, as one should predict default if it is more likely than not. However, Jones and Hensher (2008) argue that in unbalanced datasets, where there is only a small proportion of ones or zeros for the dependent variable, the aforementioned prediction rule might fail. Another popular rule is setting P^* to the sample proportion of defaults (=defaults/total number of firms in the sample).

The following table illustrates the two types of errors.

		Model	
		Default	Non-default
Actual	Default	Correct prediction	Type I error
	Non-default	Type II error	Correct prediction

Table 1: Types of errors

As one could see above, either of the errors is associated with certain costs. Therefore, one strives for keeping both error rates as low as possible. Nevertheless, it should be noted that minimizing one type of error often comes at the expense that the rate of the other type increases.

Cumulative Accuracy Profiles

Cumulative Accuracy Profiles (CAPs) are a convenient way to visualize model performance as regards discrimination. By discrimination we mean the model's ability to rank companies according to the true default/non-default observations. The CAP curve represents the cumulative probability over the entire population of defaulting and non-defaulting companies. (Sobehart, Keenan, and Stein (2001))

In order to generate the CAP curve, the model scores for the companies are ordered in descending order (riskiest to safest score). "The CAP is [then] constructed by plotting the fraction of all defaults that occurred among borrowers rated x or worse against the fraction of all borrowers that are rated x or worse." (Loeffler and Posch (2008)) CAPs disclose information about the predictive accuracy of the model over the entire range of risk scores for a certain point in time. (Sobehart, Keenan, and Stein (2001)) Examples of CAPs are shown in Figure 1.

A good model concentrates defaulters at the riskiest model scores and non defaulters at the lowest risk level. An ideal CAP would thus allocate all the defaulters in the fraction of the population, which is equal to the rate of default in the population. An uninformative model would yield a CAP that coincides with the 45° line. The CAP of an accurate model lies somewhere in between the two aforementioned CAPs.

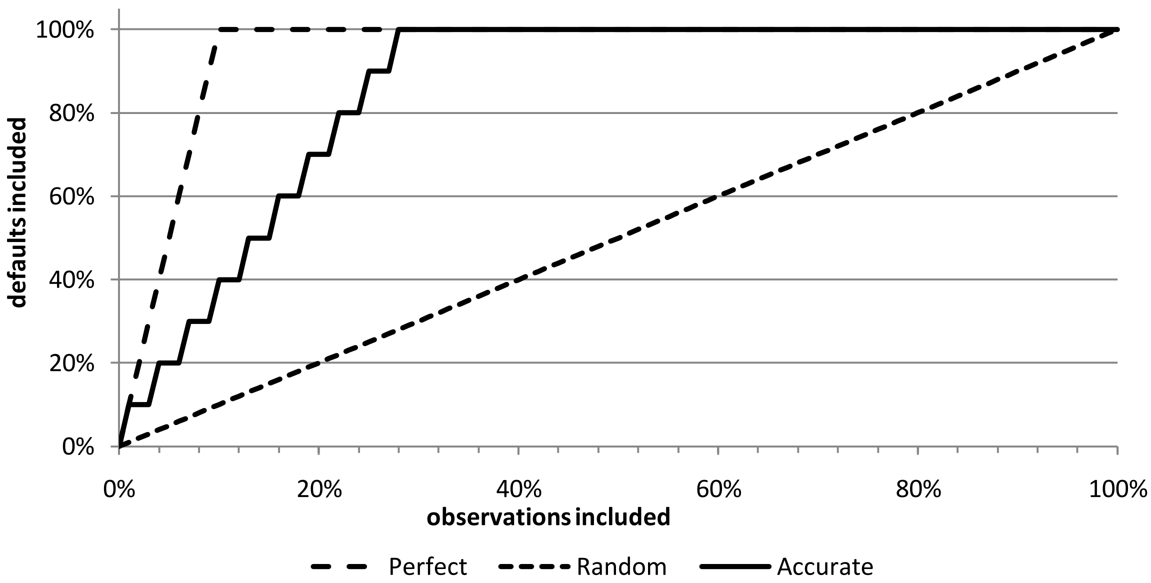

Figure 1: CAP curves for a perfect, an accurate and a random model

The CAP approach has also been criticized for not distinguishing between Type I and Type II errors. CAPs assume implicitly that the costs for both errors are the same even though they might vary in practice. Secondly, it has been argued that CAPs are highly dependent on the data the model has been trained with.

Accuracy Ratios

While CAPs visualize the predictive performance of the models under comparison, accuracy ratios condense the information contained in the CAP curve in one single statistic. (Loeffler and Posch (2008))

The accuracy ratio (AR) can be obtained by relating the area under the CAP but above the diagonal to the maximum area a CAP can enclose above the diagonal.

The test statistic is obtained by calculating the ratio of the area between the model's CAP and the random CAP to the area between the ideal CAP and the random CAP. AR can be algebraically expressed as follows:

$$AR = \frac{2\int_0^1 \varphi(q)dq - 1}{1-\psi} = \frac{1 - 2\int_0^1 \nu(q)dq}{\psi}, \tag{16}$$

where $\varphi(q)$ is the percentage of defaulting companies, which model score is worse than or equal to the one for fraction q. The expression $\nu(q)$ is defined as the percentage of all non defaulters, for which the model score is worse than or equal to the one for fraction q. Finally, ψ is defined as the ratio of the number of defaults in the sample under consideration to the total number of companies. In case the model is ideal, the ratio equals one in case the model is relatively uninformative, the ratio will be close to zero.

The accuracy ratios are only estimates of the discriminatory power. Based on the data used in this study, we calculate confidence intervals for the accuracy ratios using a bootstrapping approach. In order to derive a distribution of the accuracy ratios, the original data used for the initial accuracy ratio estimation is re-sampled. Hence, from the N initial observations of y_i and corresponding $\hat{p}_i$ one draws N observation pairs (y_i and the corresponding $\hat{p}_i$) with replacements. Then a new accuracy ratio for the re-sampled data is calculated. This re-sampling is done many times and thereby a distribution of ARs can be calculated. (Loeffler and Posch (2008))

ROC Curves

Receiver Operating Characteristics (ROC) is a very similar tool for analyzing the model accuracy to Cumulative Accuracy Profiles (CAP) and thus Accuracy ratios. The reason for this is the fact that these measures use and convey the same information. ROC and CAP graphically represent the discriminatory power of the model. The discussion about the statistical evaluation of ROC curves can be found in Engelmann, Hayden, and Tasche (2003).

ROC curve is a plot of *sensitivity* defined as $P(+|\text{default})$[13] versus, *1- specificity* defined as $P(+|\text{non}-\text{default})$[14]. The common part of ROC analysis is computing the area

[13] Probability that a company is classified correctly as defaulted
[14] Probability that a company is wrongly classified as defaulted when it in fact survived

under the ROC curve (AUC), which allows ranking tested models based on their accuracy. An AUC of 1 represents the perfect model while an AUC of 0,5 is a random model not able to distinguish between failing and surviving company at all. There is also a relationship between AR and ROC expressed linearly by:

$$AR = 2 * AUC - 1 \tag{17}$$

Another application of ROC analysis is the Sensitivity/Specificity plots. This projection displays both sensitivity and specificity for all possible cut-off rates, hence unlike simple ROC curve, it allows for distinguishing between Type I and Type II errors. The user can thus decide on the cut-off rate she considers most favorable, since the two errors have different practical impacts (Type I: loss of entire amount lent vs Type II: loss of lending opportunity). The intersection of the two curves is the point of intuitive ideal cut-off point.

Brier Score

The Brier score measures a model's calibration and discrimination. By calibration, we mean how well the estimated probabilities of default are matching the actual default/non-default observations.

The Brier score (BS), when testing default models, is defined as follows:

$$BS = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{p}_i)^2, \tag{18}$$

where i indexes N observations, and $\hat{p}_i$, the estimated probability of default.

The Brier score lies between 0 and 1; better probability forecasts are associated with lower BS values.

IV. Data

This section focuses on the data employed for the estimation of our models. We begin by describing data collection and variables selection process. Next, we present the financial and economic ratios followed by descriptive statistics and correlation analysis. Finally, we explain the choice of the baseline function.

A. *Source of Data*

The bankruptcy data for this study consists of 102 listed US companies, belong to the SIC division *manufacturing* that

(a) filed for Chapter 11 bankruptcy between 2000 and 2009,

(b) have assets worth $100 million or more at the time of filing, measured in 1980 dollars, and

(c) are required to file 10-ks[15] with the SEC.

Furthermore, we only considered those companies, for which financial data was available on Datastream. The cases of defaulted companies were obtained from Lynn M. LoPucki's Bankruptcy Research Database.

For each defaulted company we take randomly one surviving company, which belongs to the same SIC major group[16] as the defaulted company. For the static logit model, companies were also matched by year. This matching principle allows us to keep the data selection process as simple as possible. Simplicity stems from the fact that the proportion of defaulted companies in the population is very small. One further advantage of this matching procedure is that it avoids over-fitting with failing companies, which may lead to biased results. However, a selection bias can still occur, since our entire sample (102 defaulted + 100 surviving) is not drawn randomly from the population but rather created by using described matching procedure. A practical reason for this is that random selection would require substantially more observations than available to get desired variability in explained variable.

[15] A Form 10-K is an annual report required by the U.S. Securities and Exchange Commission (SEC), that gives a comprehensive summary of a public company's performance

[16] The description that applies to the first two digits of the SIC code that appears in SIC Primary of the defaulted company.

We are also aware of the fact that using only manufacturing industry makes it impossible to analyze the effect of industry sector on the probability of default. Financial ratios vary between industries, so to avoid the use of dummy variables most researchers employ the data from only one industry.

B. *Bankruptcy Data*

Our bankruptcy sample consists of 102 listed manufacturing companies that bankrupted between 2000 and 2009. The frequency of defaults is displayed in Table 2 with substantially higher failure rate in 2009.

year	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	Total
# of failing companies	10	16	10	17	9	6	3	1	6	24	102
%-age of total # failing companies	9,80%	15,69%	9,80%	16,67%	8,82%	5,88%	2,94%	0,98%	5,88%	23,53%	100,00%

Table 2: Bankruptcy frequency

The whole sample period is divided into two windows. The data from the period from 2000 until 2006 (7 years) is used as the training dataset for model fitting. The period from 2007 until 2009 (3 years) is used for out-of-sample forecasting and testing of predictive effectiveness. The dataset for the Model I (cross-sectional data) differs from the dataset used to estimate Models II and III (panel data). The sample for the Model I, denoted Sample I, consists of total 202 observations divided into 142 (70%) used for model fitting and remaining 60 observations used for evaluation. The sample for our panel-logit Model II and Model III, denoted Sample II consists of 1544 observations (1160 (75%) in training sample and 384 used as a validation dataset). Sample division and notation is summarized in Table 3 below.

Sample I : Data Sample for Model I

	Training dataset	Evaluation dataset	Total
Defaulted Companies	71	31	102
Non-Defaulted Companies	71	29	100
Total	142	60	202

Sample II : Data Sample for Models II, III

	Training dataset	Evaluation dataset	Total
Defaulted Firm-years	71	31	102
Non-Defaulted Firm-years	1089	353	1442
Total	1160	384	1544

Table 3: Sample division and notation

The following sub-sections describe the financial ratios and macroeconomic proxies considered by our study.

C. *Covariates*

Consulting an extensive review of existing literature on corporate default models in section II, we identify the most common financial ratios that are examined in logit and hazard models including those by Altman (1968), Ohlson (1980), Zmijewsky (1984), Shumway (2001), Campbell, Hilscher, and Szilagyi (2008). We also propose our own leverage and activity ratios to examine if any of these might be significant predictors besides those well established in the literature.

Table 4 holds the summary of all the considered independent variables including the notation and the origin. We divide the ratios into the seven categories used by Beaver (1966). We add one more category, which comprises of market variables used in previous studies.

We winsorized each of the explanatory variables using 5/95 percentile interval in order to remove outliers and smoothen our data. In other words, we converted the values of outliers into the values given by percentile thresholds. As in (Campbell, Hilscher, and Szilagyi (2008)) we did not remove outliers from Sigma (*r31*) and Price (*r32*).

CASH FLOW RATIOS	Notation	Origin	LIQUID ASSET RATIOS	Notation	Origin
Cash Flow to Total Liabilities	r1	Lenox (2002) (adj)	Cash and Bank to Total Assets	r19	Nam (2008)
Cash Flow to Financial Expenditures	r2	Zeitun (2007)	Working Capital to Total Assets	r20	Altman (1968)
PROFITABILITY RATIOS			**SHORT TERM SOLVENCY RATIOS**		
Net profit to Net Sales	r3	Park (2002)	Current Asset less Inventory to Current Liabilities	r21	Zmijewsky (1984)
Operating Income to Net Sales	r4	Own	Current Asset to Current Liabilities	r22	Zmijewsky (1984)
Net Income to Total Asset	r5	Zmijewsky (1984)	**ACTIVITY RATIOS**		
EBIT to Total Asset	r6	Altman (1968)	Cash to Sales	r23	Own
Net Income to Total Equity	r7	Park (2002)	Current Asset to Sales	r24	Own
Net Income to Total Liabilities	r8	Park (2002)	Working Capital to Sales	r25	Own
Retained earnings to Total Assets	r9	Altman (1968)	Total Asset to Sales	r26	Altman (1984)
LEVERAGE RATIOS			Stockholder’s Equity Turnover	r27	Park (2002)
Total Current Liabilities to Total Assets	r10	Zmijewsky (1984)	**MARKET VARIABLES**		
Total Debt to Total Asset	r11	Zmijewsky (1984)	Market to Book ratio	r28	Campbell (2008)
Debt to Equity	r12	Zeitun (2007)	Log(Market Capitalization to Market		
EBIT to Interest Expenditures	r13	Own	Capitalization S&P 500)	r29	Campbell (2008)
Equity to Asset	r14	Own	Market Equity to Total Liabilities	r30	Altman (1984)
Sales to Debt	r15	Own	Sigma-yearly volatility of stock prices	r31	Shumway (2001)
SIZE RATIOS			Log(Price per Share)	r32	Campbell (2008)
Total Assets	r16	Park(2002)	Excess return over S&P 500	r33	Campbell (2008)
Log of Total Assets	r17	Ohlson (1980)			
Number of Employees	r18	Lenox(2002)			

Table 4: Ratios used default analysis

D. *Variable Selection*

The variable selection process should be largely based on the existing theory. The field of bankruptcy prediction, however, suffers from a lack of agreement as for which variables should be used. The first step in our empirical search for the best model is therefore the correlation analysis. If high correlation[17] is detected, we prioritize the most commonly used and best performing ratios in the literature and market driven ratios over our own or less established ones. Some ratios are deleted, as they are an obvious transformation of other ratios.

Next, the choice of variables entering our models is made by looking at the significance of ratios. The score used for prediction consists only of significant or marginally significant variables selected by stepwise estimation under the significance level of 10%[18]. The final choice of variables is further dependent upon the heteroskedasticity and misspecification tests.

The ratios included in our static and the dynamic models are shown in Table 5.

Static logit model (Model I)			Hazard models (Model II and III)		
Variable		Exp. Sign	Variable		Exp. Sign
Cash Flow to Financial Expenditures Net Income to	r2	-	Net Profit to Net Sales	r3	-
Total Assets	r5	-	EBIT to Total Assets	r6	-
EBIT to Total Assets	r6	-	Total Current Liabilities to Total Assets	r10	+
Cash and Bank to Total Assets	r19	-	Cash and Bank to Total Assets	r19	-
Current Assets to Current Liabilities	r22	-	Current Assets to Current Liabilities	r22	-
Market to Book ratio	r28	-	Log(Market Capitalization to Market	r29	-
Log(Price per Share)	r32	-	Capitalization S&P 500)		
Excess return over S&P 500	r33	-	Market Equity to Total Liabilities	r30	-
			Log(Price per Share)	r32	-
			Excess return over S&P 500	r33	-

Table 5: Ratios included in the static logit model and the hazard models

The next two tables present summary statistics for all the variables used to forecast bankruptcy. Table 6 holds descriptive statistics of covariates used in Model I and Table 7 describes variables used in the Models II and III, in which each firm-year represents one observation. Summaries in Table 6 and Table 7 are the distributions of the Sample I and Sample II respectively. The training datasets are used to produce these tables. Whole sample properties can be found in the Appendix, sub-section B. Correlation tables of the selected ratios can be found in Appendix B.

[17] We consider two variables highly correlated if their correlation coefficient is larger than 0,6.
[18] One exception was made in the case of *r29* in the Model III where the p-value equals to 0,109.

	Defaulted Companies Number of observations = 71				Non-Defaulted Companies Number of observations = 71			
Variable	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
r2	4,89	35,03	-7,56	295,16	43,57	289,48	-116,25	2428,00
r5	-0,19	0,28	-1,30	0,41	0,03	0,09	-0,29	0,22
r6	-0,02	0,34	-0,40	2,29	3,15	13,27	-0,86	107,70
r19	-0,19	0,32	-1,47	0,21	0,04	0,24	-0,84	0,86
r22	1,01	0,57	0,60	3,49	2,61	2,63	0,02	15,31
r28	0,76	1,84	-1,12	14,61	2,37	4,99	-16,32	36,48
r32	0,46	0,43	0,14	1,81	1,07	0,61	-0,72	3,45
r33	-0,27	0,31	-0,64	0,35	0,04	0,18	-0,49	0,45

Table 6: Sample I: Descriptive statistics, training sample Model I

	Defaulted Companies Number of observations = 71				Non-Defaulted Companies Number of observations = 1089			
Variable	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
r3	-0,15	0,16	-0,51	0,18	-0,02	0,15	-0,51	0,18
r6	-0,02	0,34	-0,40	2,29	0,87	2,44	-0,40	9,68
r10	0,39	0,23	0,08	0,69	0,24	0,15	0,08	0,69
r22	1,01	0,57	0,60	3,49	2,04	1,36	0,60	6,04
r29	-4,38	1,15	-6,11	-2,92	-4,31	0,98	-6,11	-2,92
r30	0,19	0,41	0,00	2,00	2,14	3,37	0,00	12,98
r32	0,46	0,43	0,14	1,81	0,89	0,53	0,14	1,81
r33	-0,27	0,31	-0,64	0,35	-0,01	0,19	-0,64	0,35

Table 7: Sample II: Descriptive statistics, training sample Model II and Model III

The analysis of the basic statistical properties of the explanatory variables gives support to hypotheses drawn by previous research as regards size and signs. It is, however, important to note that each firm has an equal weight when interpreting these distributions. This means that smaller firms may dominate the distribution and that in case of panel datasets the distributions will reflect both time-series and cross-sectional fluctuations. The averages of the panel sample of non-defaulted firms (Table 7) will also be driven down by the firm-years close to default, but still classified as non-failed such as two or more years prior to default. Table 6 and Table 7 show the means of the variables for the companies prior to default (less than one year) as well as the means for the non-defaulting companies (non-defaulted or more than a year before filing) for the training sample[19]. We observe much more pessimistic results for defaulted companies, which is hardly surprising. The comparison of the left and right sides of Table 6 and Table 7 reveals the differences between failed companies and the rest of the

[19] Note that the properties are equal for both Samples, since the same observations of defaulted companies are used.

sample. A huge gap between defaulted and non-defaulted firms in cash-flow to financial (*r2*) expenditures indicates the importance of having a healthy cash position relative to short-term financial expenditures. The average value of EBIT to total asset (*r6*) also differs substantially across the two groups suggesting the prevalence of unprofitable companies in the default group. Another interesting point is that the defaulted companies have on average current liabilities to total assets (*r10*) of 39% compared to 24% by non-defaulted companies. We do not expect size to be highly important determinant of bankruptcies, since there is hardly any difference in our size measure (*r29*) between the two groups. Another striking difference is the measure of market performance represented by excess return over S&P500 (*r33*). The observations of the defaulted companies have underperformed the index on average by 24%, while the defaulted companies underperformed S&P500 on average by only 1% in Sample I and overperformed it by approximately 4% in Sample II[20].

E. *Macroeconomic Dependencies*

As mentioned in section III (Methodology), there are several ways to proxy a baseline function to account for macro-economic development. It has been argued that the ideal choice is to incorporate system-level variables capturing the development of macro dependencies. However, we find such specification of baseline consistently insignificant in predicting bankruptcy. In contrast to Nam et al. (2008), who found change in interest rate correlated with actual hazard rate, we have not observed such dependence. Neither did the foreign exchange rate USD/EUR show the desired pattern. As regards the use of year dummies, we exclude them from the model because they cannot be used for forecasting the crisis period.

We decided to follow the works of Hillegeist et al. (2001) and Shumway (2001) suggesting that Annual Rates capturing the actual realization of unconditional baseline hazard rate in the previous period should yield unbiased results. The baseline hazard rate is graphed in Figure 2. It has been argued that such a specification is a good proxy for macroeconomic environment.

[20] EXP(-0,2746084)= 0,75986963 => 24% underperformance, EXP(-0,0075434)= 0,99248498 =>1% underperformance, EXP(0,0389008)= 1,039667344 => 4% outperformance

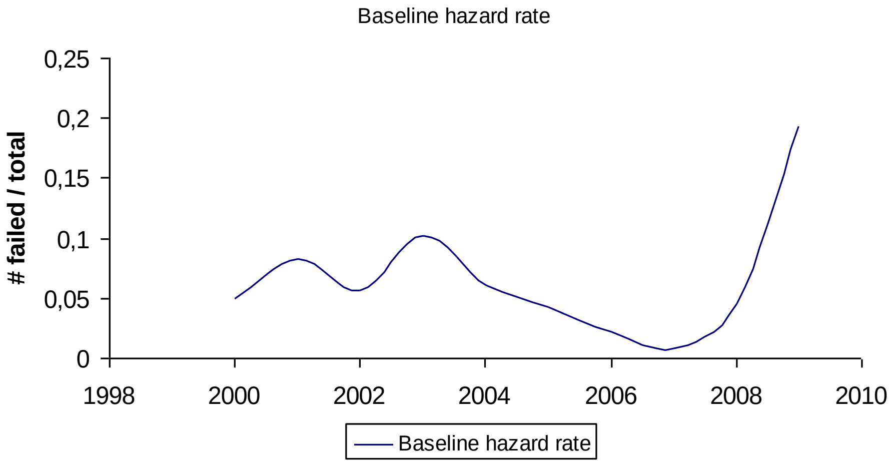

Figure 2: Baseline hazard rate

Hazard models may potentially account for some form of duration dependence. Therefore, we match the Annual Rate with the function of firm's age, *ln(firm age)*, as suggested by Shumway (2001), thereby obtaining a combined baseline function. Shumway (2001) also postulates that if the duration dependence is present, firm age might be a very important explanatory variable.

Figure 3 plots the hazard function as a function of firm's age. An increasing hazard function indicates the existence of duration dependence. The hazard is larger for large values of *t/*age and smaller for smaller values of *age.* The shape of the function means that with increasing age the firms in our sample are more likely to go bankrupt. Function is, however, not monotonious because if the firm survives until the age of approximately 23 year, its hazard is decreasing. This finding is in line with Shumway (2001).

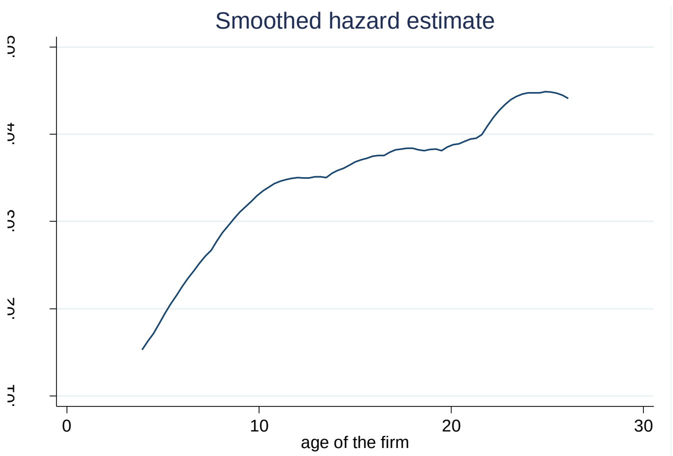

Figure 3: Plot of estimated hazard function by hazard contributions

V. Empirical Results

In this section, the empirical results are presented and discussed. Our primary analysis compares the prediction power of three competing models summarized in Table 8.

	Description	Hazard function	Expression as logit model
Model I	Simple logit model	$h_i(t) = h_0 \cdot e^{x_i \beta}$	$P(y_i = 1) = \frac{1}{1 + e^{-x_i \beta}}$
Model II	Hazard model without macro-dependent baseline function	$h_i(t) = h_0 \cdot e^{x_{i,t} \beta}$	$P(y_i = 1) = \frac{1}{1 + e^{-x_{i,t} \beta}}$
Model III	Hazard model with macro-dependent baseline function	$h_i(t) = h_0(t) \cdot e^{x_{i,t} \beta}$	$P(y_{i,t} = 1) = \frac{1}{1 + e^{-(\beta_1' \kappa_t + \beta_2' x_i)}}$

Table 8: Summary of competing models (Nam et al. (2008))

In Sample I (cross-sectional data), each bankrupt firm contributes one failure observation (y=1). The data for each bankrupt firm comes from the year prior to its bankruptcy, which is matched by one observation of surviving firm (y=0) from the same year. All bankruptcies happened between 2000 and 2009.

In Sample II (panel data), the observation window starts in the year 2000 for all firms. For example, if the firm bankrupted in year 2002 it would contribute three firm-year observations to the multi-period logit models (Model II, III). The dependent variable associated with the first two years would be equal to zero ($y_i = 0$). Each failed firm thus contributes only one failure observation ($y_i = 1$) to the model indicating the bankruptcy occurred.

A. *Results*

Table 9 gives the maximum likelihood estimates for all three models along with the tests of misspecification. First, we test the models for omitted variables by following the procedure described in sub-section G. For this purpose, we consider squared terms of the explanatory variables of the models, as Saunders and Allen (2002) and Lenox (1999) find that the path to bankruptcy might be non-linear. Second, we test for heteroskedasticity, as this might be a serious problem for logit models. In the presence of heteroskedasticity, parameter estimates become inconsistent (Verbeek (2009)).

A first visual inspection of plots of the residuals against the fitted values (s. Appendix C) suggests that our models might be exposed to heteroskedasticity. Therefore, further statistical tests are employed in order to examine the assumption of homoskedasticity in the errors. Using the testing procedure described in the Methodology section, we obtain the results presented in Table 9.

Explanatory Variables	Expected sign	Model Ia	Model Ib	Model II	Model III
r2	-	0,011***			
r3	-			-2,26**	-1,94*
r5	-	-10,06***	-22,78***		
r6	-	-3,70***	-4,42***	-1,42**	-1,38**
r10	+			2,44***	2,43***
r19	-	-5,21*	-10,46**		
r22	-	-1,50***	-1,97***	-0,92***	-1,00***
r28	-/+	-0,57***	-0,41**		
r29	-			-0,27*	-0,26
r30	-			-0,99***	-0,98***
r32	-	-2,65***		-2,00***	-2,42***
r33	-	-7,06***	-7,55**	-5,45***	-5,51***
r2sq	-		-0,00***		
r19sq	-		-14,63**		
r33sq	+		17,99***		
ln(age)	+				1,71***
AnnualRate	+				4,73***
_cons		5.36***	2,58	-2,28***	-2,88***
Omitted variable test					
LM test		18,13***	1,08	11,04	11,86
Null hypothesis		rejected	not rejected	not rejected	not rejected
Heteroskedasticity test					
LM test		105,36***	10,270	8,298	10,318
Null hypothesis		rejected	not rejected	not rejected	not rejected

legend: * p<.1; ** p<.05; *** p<.01

Table 9: Maximum likelihood estimates and test results for all three models.

Model I

Model I is estimated using the ratios extracted from the financial statements issued the year preceding the bankruptcy filing. Totally, eight variables are indentified as significant. The initially estimated model is denoted as Model Ia in Table 9. However, the LM test statistics in the Table 9 indicates the presence of heteroskedasticity in Model Ia. Further examination by the means of LM-test of this model revealed that there are also potentially omitted variables in the model causing bias.

The LM test for heteroskedasticity points out that the variance of residuals is of the non-constant functional form. We re-estimated the Model Ia again, this time including quadratic terms of all explanatory variables showing that heteroskedasticity was caused by incorrect assumption of a linear functional form. In particular, Cash Flow to Financial Expenditures (*r2*), Cash and Bank to Total Assets (*r19*) and Excess return over S&P500 (*r33*) were found to be non-linear determinants of bankruptcy for the static logit model. These terms have very high and significant effect on bankruptcy probability. After taking into account the non-linear functional form of *r2*, *r19* and *r33*, neither the null hypothesis of homoskedasticity nor the null of no-omitted variables can be rejected. The model with included quadratic terms is denoted as Model Ib and its results are presented in the second column of Table 9.

The results for the re-estimated Model Ia (denoted Model Ib) show that the bankruptcy is more likely if the firm experiences a decrease in profitability as indicated by the coefficient of *r5* and *r6*. The significant polynomial term of Cash and Bank to Total Assets (*r19)* indicates the company's cash position is another important determinant of bankruptcy. A significant sign for squared Cash Flow to Financial Expenditures (*r2sq*) supports the cash flow theory, yet the coefficient is close to zero. The short-term solvency ratio Current Asset to Current Liabilities (*r22*) is also important, indicating the company is less likely to go bankrupt if it holds enough short-term assets to cover short-term liabilities such as interest payments. As regards Market-to-book ratio (*r28*), both signs are possible depending on market forces. Since the bankrupt firms often experienced losses, the book value worsens, which pushes market-to-book ratio up. In contrast to this, the bankruptcy is quite effectively anticipated by traders discounting the market value, which drives down the market-to-book ratio. The negative sign for the Market-to-Book ratio (*r28*) underlines the effectiveness of stock market valuation of the firm. Consequently, the lower the market value of the firm relative to its book value the more likely the bankruptcy. Finally, although the sign of the coefficients for the quadratic log excess return relative to the S&P 500 index might seems counterintuitive, closer examination reveals that we can indeed expect the positive sign. The mean of the log excess return for the entire training sample is negative as shown in the Table 10 below, so it is reasonable to expect its quadratic term, which is always positive, to have positive effect on probability of bankruptcy.

Variable	Obs	Mean	Std. Dev.	Min	Max
r33	142	-0,23	0,57	-2,53	0,98

Table 10: Descriptive statistics log excess returns

Models II and III

Models II and III are hazard models taking multiple periods of each firm into account. The results of both models are very similar to the classical static model. Table 9 shows that most of the estimated coefficients of the bankruptcy function under Model II and Model III are significant at 1%. The only marginally significant or insignificant coefficient is our relative size measure (*r29*). Judging from its signs, we can conclude that the bankrupt companies tend to be relatively small. The reason for the relative insignificance of size might partly be the nature of our sample. We employed a matched pair technique[21] so the size is not expected to drive the default probability in our sample. Nevertheless, in reality it most probably does. The coefficients for Net profit to Net Sales (*r3*) and Ebit to Total Assets (*r6*) confirm our previous findings as regards the positive effect of profitability decrease on bankruptcy. Unlike the static model, our hazard models find the leverage measure, Current liabilities to Total Assets (*r10*), significant in predicting defaults. The coefficient for *r10* is positive as expected. Short-term solvency measured by Current Assets to Current Liabilities (*r22*) is also found significant on 1%. Thus, a firm with a high *r22* holds enough liquid and short-term assets to cover interest payments, which might help it to postpone or avoid bankruptcy. Significant effect of Market equity to Total Liabilities (*r30*) on the sample bankruptcy probability has the same interpretation as the Market-to-Book ratio. Traders are able to anticipate the bankruptcy and discount the market value of equity. The remaining two market variables, *r32* and *r33*, take the expected sign. Both variables capture the effect of past stock performance and their coefficients indicate that they are important determinants of bankruptcy.

[21] The procedure involved approximate matching the companies based on their size of total assets

To get a better picture of marginal effects of changes in explanatory variables on the probability of failure, we provide a table of proportional impacts of one quarter of standard deviation increase in each of the explanatory variables. A similar approach is used by Campbell, Hilscher, and Szilagyi (2008). This analysis is presented in Table 11.We study the ceteris paribus effect on the firm with initial average sample values of explanatory variables.

Variable	Model Ib	Model II	Model III
r2	-8%		
r3		-8%	-7%
r5	-72%		
r6	-100%	-57%	-56%
r10		10%	10%
r19	-53%		
r22	-62%	-27%	-28%
r28	-29%		
r29		-6%	-6%
r30		-56%	-55%
r32		-24%	-28%
r33	-93%	-26%	-26%

Table 11: Impact of 1/4 increase of standard deviation of each of covariates on default probability

Referring to Table 11 Model Ib, a one-quarter of standard deviation increase in our cash-flow ratio (*r2*) profitability leads to an 8% decrease of it initial values. This is also the lowest marginal effect. The same increase in profitability reduces the probability by 72% for *r5* and by 100% for *r6*. Likewise, the remaining effects of the abovementioned increase are a 53% decrease for cash holdings (*r19*), a 62% decrease for short-term solvency (*r22*), a 29% decrease for Market-to-Book ratio and a 93% decrease for past excess returns (*r33*). To summarize, static logit results, the largest negative relationship between the predictor and the probability of default is found for profitability, solvency, excess market returns and cash holdings.

The predicted probability has similar sensitivity to the mentioned increase for both of the hazard models (Models II and III). Increases in profitability (*r6*) and market equity to total liabilities (*r30*) are found to be most important yielding over 50% reduction in failure probability. Of medium importance are the increases in short-term solvency (*r22*), price per share (*r32*) and excess return (*r33*) causing reduction of bankruptcy probability from its initial level by 24 to 28%.

Surprisingly, a corresponding increase in leverage accounted only for 10% increase. As expected, the variation in size has only a small effect on the risk of bankruptcy (6% decrease), due to our sampling procedure. Thus for the hazard models, the variations in profitability, market value relative to book total liabilities, stock-market performance and short-tem solvency are more important than the variations in size and leverage.

VI. Model Evaluation

This section addresses the forecasting power of the three different models during the recent financial crisis using an evaluation period of three years between 2007 and 2009. The most important question is whether the models are able to discriminate between defaulting and non-defaulting companies.

A. *Model Accuracy*

In the following sub-section, we present the sensitivity and specificity as well as Type I and Type II errors for the out-of-sample period for the models estimated in this thesis.

A model's sensitivity describes the probability that a model classifies a company as defaulted (+), given a specified threshold probability (P^*), when it actually defaulted (D). A model's specificity is defined as the probability that a model classifies a company as non-defaulted (-), given a specified threshold probability, when it actually does not default ($\sim D$). Type I and Type II errors are used as defined in section III.H: A Type I error occurs in case the model predicts a company to survive when it actually fails. A Type II error would occur if the model predicts a failure when a company in fact survives (Type II error)

From Table 12 below it can be seen that Model Ib scores the highest sensitivity for all three threshold probabilities, while it ranks last for the specificity measure. This indicates that Model Ib is particularly good in identifying defaulting companies even at high threshold probabilities, such as $P^* = 0{,}5$. On the other hand, Model Ib has a lower power when it comes to classifying a company as non-defaulted when it actually did not default. In other words, Model Ib suffers from a higher false alarm rate.

Model II and Model III are worse than Model Ib as regards sensitivity measures. However, Model III, incorporating macroeconomic dependencies, performs slightly better on threshold levels of 0,5 and 0,1 and significantly better when $P^* = 0{,}3$. As for specificity, Model II is slightly better than Model III. This suggests that Model II and Model III are worse in predicting actual defaults but better in predicting actual non-defaults than Model I. The latter finding is hardly surprising given the unbalanced structure of the Sample II consisting mostly of non-bankrupt firms.

Model Ib

		True		
		default (D)	non-default (~D)	Total
Classified	default (+)	30	8	38
	non-default (-)	1	21	22
	Total	31	29	60

Classified "default" if predicted Pr(D)>=0,5

Sensitivity Pr(+|D) 96,77%
Specificity Pr(-|~D) 72,41%

Type I (False ~D rate for true D) Pr(-|~D) 3,23%
Type II (False D rate for true ~D) Pr(+|~D) 27,59%

		True		
		default (D)	non-default (~D)	Total
Classified	default (+)	30	8	38
	non-default (-)	1	21	22
	Total	31	29	60

Classified "default" if predicted Pr(D)>=0,3

Sensitivity Pr(+|D) 96,77%
Specificity Pr(-|~D) 72,41%

Type I (False ~D rate for true D) Pr(-|~D) 3,23%
Type II (False D rate for true ~D) Pr(+|~D) 27,59%

		True		
		default (D)	non-default (~D)	Total
Classified	default (+)	30	11	41
	non-default (-)	1	18	19
	Total	31	29	60

Classified "default" if predicted Pr(D)>=0,3

Sensitivity Pr(+|D) 96,77%
Specificity Pr(-|~D) 62,07%

Type I (False ~D rate for true D) Pr(-|~D) 3,23%
Type II (False D rate for true ~D) Pr(+|~D) 37,93%

Model II

		True		
		default (D)	non-default (~D)	Total
Classified	default (+)	17	6	23
	non-default (-)	14	347	361
	Total	31	353	384

Classified "default" if predicted Pr(D)>=0,5

Sensitivity Pr(+|D) 54,84%
Specificity Pr(-|~D) 98,30%

Type I (False ~D rate for true D) Pr(-|~D) 45,16%
Type II (False D rate for true ~D) Pr(+|~D) 1,70%

		True		
		default (D)	non-default (~D)	Total
Classified	default (+)	18	10	28
	non-default (-)	13	343	356
	Total	31	353	384

Classified "default" if predicted Pr(D)>=0,3

Sensitivity Pr(+|D) 58,06%
Specificity Pr(-|~D) 97,17%

Type I (False ~D rate for true D) Pr(-|~D) 41,94%
Type II (False D rate for true ~D) Pr(+|~D) 2,83%

		True		
		default (D)	non-default (~D)	Total
Classified	default (+)	24	27	51
	non-default (-)	7	326	333
	Total	31	353	384

Classified "default" if predicted Pr(D)>=0,1

Sensitivity Pr(+|D) 77,42%
Specificity Pr(-|~D) 92,35%

Type I (False ~D rate for true D) Pr(-|~D) 22,58%
Type II (False D rate for true ~D) Pr(+|~D) 7,65%

Model III

		True		
		default (D)	non-default (~D)	Total
Classified	default (+)	18	9	27
	non-default (-)	13	344	357
	Total	31	353	384

Classified "default" if predicted Pr(D)>=0,5

Sensitivity Pr(+|D) 58,06%
Specificity Pr(-|~D) 97,45%

Type I (False ~D rate for true D) Pr(-|~D) 41,94%
Type II (False D rate for true ~D) Pr(+|~D) 2,55%

		True		
		default (D)	non-default (~D)	Total
Classified	default (+)	22	16	38
	non-default (-)	9	337	346
	Total	31	353	384

Classified "default" if predicted Pr(D)>=0,3

Sensitivity Pr(+|D) 70,97%
Specificity Pr(-|~D) 95,47%

Type I (False ~D rate for true D) Pr(-|~D) 29,03%
Type II (False D rate for true ~D) Pr(+|~D) 4,53%

		True		
		default (D)	non-default (~D)	Total
Classified	default (+)	26	29	55
	non-default (-)	5	324	329
	Total	31	353	384

Classified "default" if predicted Pr(D)>=0,1

Sensitivity Pr(+|D) 83,87%
Specificity Pr(-|~D) 91,78%

Type I (False ~D rate for true D) Pr(-|~D) 16,13%
Type II (False D rate for true ~D) Pr(+|~D) 8,22%

Table 12: Type I and Type II errors for Model Ib, II and III

In Appendix D sensitivity/specificity plots can be found. The graphs plot sensitivities/specificities against all possible cut-off probabilities. The intersection of the two curves can be interpreted as the optimal cut-off probability. For Model Ib we find a probability close to 100% as optimal. This seemingly strange result is caused by the fact that Model Ib assigns probabilities to the bankrupt firms very close to 1 and to non-bankrupt firms very close to 0. On the contrary, the optimal cut-off probabilities for Model II and Model III are much lower meaning that to capture most of the bankruptcies a very low cut-off point (app. 5%) must be used. This is, however, not detrimental to models specificity as indicated by the curves.

The results discussed above are also reflected by the scores for Type I and Type II errors. Model Ib scores the lowest Type I error and is thus the best in indentifying true defaulters. Type I errors are significantly larger for Model II and Model III. Both model's performance regarding Type I errors can be improved by decreasing the threshold probability from 0,5 to 0,1. For Model II this yields a reduction from 45,16% to 22,58% and for Model III decreasing the threshold from 0,5 to 0,1 yields a decrease in Type I errors from 41,94% to 16,13%. This improvement is reached solely by incorporating macro dependent baseline function and comes only at the marginal expense of increasing Type II errors, in which Model II and Model III perform better than Model Ib.

It should be noted that larger Type II errors might be accepted by the agent, as the costs of Type I errors can be expected to be higher than those of Type II errors. Costs of Type I errors involve losing parts of or the whole principal as well as promised interest payments on the loan.

However, even the lowest Type I error rates for Model II and Model III, we report here, are still quite high compared to error rates reported in existing literature. Nam et al. (2008) for example report Type I error rates of 13,89% for a duration-independent hazard model (Model II) and 8,33% for a duration-dependent hazard model (Model III). Conversely, we find a lower Type I error rate for the simple logit model (Model I: 3,23%) than Nam et al. (2008) who report 8,33%.

B. *Cumulative Accuracy Profiles*

Figure 4 below presents the cumulative accuracy profiles for all three models in the out-of-sample period. It can be clearly seen that all of the models perform significantly better than a random assignment of default probabilities. However, Model Ib performs the best across all the models. Model III, allowing for macro-economic dependencies, performs slightly better than Model II.

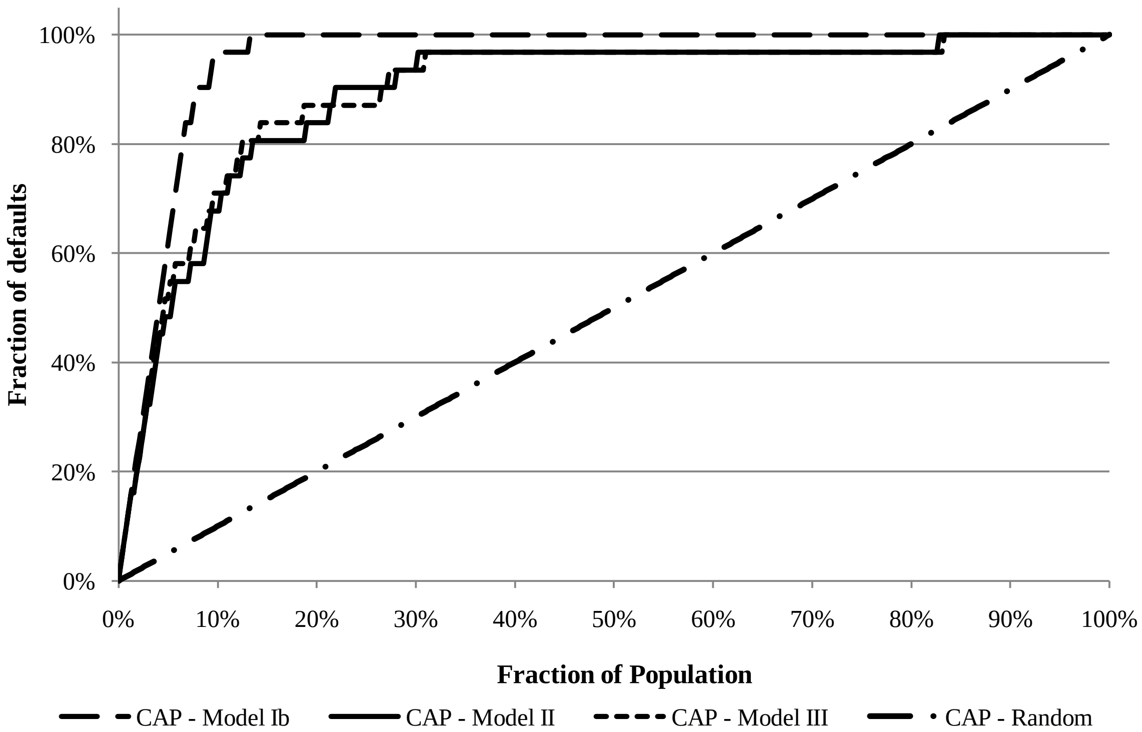

Figure 4: Cumulative Accuracy Profiles for bankruptcy models

All three models are able to detect defaults in the bottom 5% of the population equally well, since the shapes of the CAPs coincide for all three models. Beyond the bottom 5% of the population the static model (Model Ib) is noticeably better in discriminating defaults in the middle range of firms.

The CAPs of Model II and Model III intersect in the range of 20% to 30% on the horizontal axis. Therefore, based on the visual inspection, one cannot unambiguously claim that Model III outperforms Model II. Whilst Model III is better at distinguishing between non-defaulters and defaulters in lower score ranges (s. Figure 4), this conclusion does not hold for firms of higher quality, which will usually be found on the left hand side of Figure 4.

C. *Accuracy Ratios*

"Accuracy ratios condense the information contained in CAP curves into a single number." (Loeffler and Posch (2008)) The accuracy ratios for the models used in our thesis and the corresponding confidence intervals (after 1000 re-samplings in the bootstrapping procedure) are presented in Table 13.

		Confidence interval (95%)	
Model	Accuracy ratio	Lower bound	Upper bound
Model Ib	0,9155	0,7980	0,9911
Model II	0,8576	0,7281	0,9436
Model III	0,8687	0,7464	0,9526

Table 13: Accuracy Ratios and confidence intervals

Loeffler and Posch (2008) find that ARs for rating models used in practice usually lie between 50% and 90%. All three models achieve $AR > 0{,}5$ Model Ib even exceeds aforementioned range marginally by yielding an AR of over 0,91. Model III performs next best and Model II ranks last.

D. *ROC Curves*

To compare forecasting performance of the models, we also produce ROC curves. Figure 6 in Appendix D shows ROC curves for all three models and confirms the best performance of Model Ib in *ex post* period in terms of its discriminatory power. Another statistics evident from the Figure 6 is the ROC AUC, the area under the ROC curves. This area quantifies the model's ability to discriminate between failed and surviving companies.

Model	AUC
Model Ib	0,9577
Model II	0,9447
Model III	0,9343

Table 14: Area under the ROC curves

Thus, incorporating Annual Rate and duration dependence does not improve the overall discriminatory power of the model according to this measure. However, the macro baseline function improves hazard model's sensitivity, i.e. the ability to correctly identify failing companies

E. *Brier Score*

Table 15 presents the Brier Score for all three models. The Brier Score does not only test for discrimination but also for calibration. Lower scores indicate superior model performance.

Model	Brier Score
Model Ib	0,1393
Model II	0,0421
Model III	0,0403

Table 15: Brier Scores

Table 15 reveals that the lowest score is achieved by Model III, followed by Model II and Model Ib. Whilst Model Ib outperformed the remaining two models when only discrimination was examined, Table 15 indicates that when accounted for both discrimination and calibration Model Ib performs worse than hazard models allowing for time-varying company data and macro-economic dependencies.

Nevertheless this outcome is not very illustrative. The estimated probabilities of default of the hazard models are generally much closer to zero than those of the static logit model because they have lower optimal cut-off points. In case the model classifies a company as defaulted ($\hat{p}_i$) when the company actually does not default ($y_i = 0$), the squared difference $(y_i - \hat{p}_i)^2$ will be smaller for the hazard models than for the static logit model (s. also section III.H). In other words, the impact of Type II error on BS is higher for the static logit model than for hazard models. This drives BS down for the hazard models.

VII. Conclusion

The importance of accurate credit risk measures gained prominence in the recent periods of increased default frequencies in all sectors. We identified a controversy in previous research concerning the out-of-sample accuracy between the two classes of models. In this study, we empirically assess the classical static logit model versus two specifications of discrete hazard models on the newest bankruptcy data. This methodology provides three different maximum likelihood estimations of score vectors used for quantifying bankruptcy probabilities.

We study firm-specific as well as macroeconomic determinants of corporate bankruptcies on a sample of 202 manufacturing companies between 2000 and 2009, with a testing period of 2007-2009. The findings of this paper can mainly be used by lenders in assessing failure rates of their borrowers in loan approval processes.

Our three key research questions are:

1) Is the out-of-sample forecasting more accurate for hazard model as suggested by number of researcher?
2) Are there any improvements attained by incorporating macro-dependent variables in the hazard model?
3) What is the optimal combination of market driven variables and accounting based ratios?

Despite many theoretical arguments in favor of discrete hazard models, we do not find evidence that multi-period logit models outperform standard logit model as for out-of-sample performance. With exception of Brier Score, the static logit model dominates in all specified evaluation methods. This findings confirm the conclusions by Fuertes, A.-M., and Kalotychou (2006), Rodriguez and Rodriguez (2006) and Fantazzini and Silvia (2009). These authors suggest that simple models yield better forecast accuracy because the observable ratios indicating bankruptcy are rather “noisy” prior to bankruptcy, which is also true for our sample. This in other words means that our panel data sample contains many past observation of firm-years marked as non-bankrupt. These firm-years were associated with bankrupted firms prior to filing. The fact that failing companies deteriorate over longer period of time impairs the hazard models’ ability to distinguish between the non-failed and failed companies.

Improved accuracy is reached by extremely low cut-off probability rates set for hazard models.

Within the class of hazard models, we find the macro-dependent baseline function contributing to hazard models accuracy mainly in terms of the ability to correctly identify failing companies. We proxy macro-dependencies by specifying autoregressive failure rate, Annual Rate, which has proved to capture system–level changes in economical environment influencing bankruptcies. Incorporating macro-dependent baseline Annual Rate decreases Type I error, but it is detrimental in terms of overall discriminatory power as indicated by ROC curves. In contrast to ROC, accuracy ratios rank Model III higher than Model II. We also find a firm's age to have a significant impact on default probability. Older companies are generally associated with higher risk of failure in our sample.

As regards the determinants of bankruptcies, we confirm that common observable accounting ratios do not suffice for bankruptcy prediction. On the contrary, five market-driven variables were identified to be significantly related to default probability. Beside the excess market return, which is the most significant market factor, relative market size, past stock prices and market-to-book ratio all predict bankruptcy. The estimated models are combinations of these measures with traditional accounting based profitability, leverage, short-term solvency and liquidity ratios. Setting the cut-off point close to optimal level the estimated static model could accurately predict over 96% of failures, hazard model without macro-dependent baseline over 77% and if macro baseline is incorporated, 84% of bankruptcies.

Our sample does not allow us to study the effects of industry or country on the risk of failure. Larger sample across different sectors and countries would remedy this shortcoming. Further improvement of our analysis would be attained by using more frequent data such as monthly or quarterly instead of yearly. Since our models show increased sensitivity to equity based measures, further incorporation of the "market" would add to models' explanatory power. E.g. it would be possible to proxy market value of Total Assets by measuring its equity component at market value (total market assets = market equity + book liabilities).

In case of simple logit model, there is a risk of a sample selection bias. Since defaulters were not chosen randomly from the whole population, a possible structural difference in included and absenting firms may lead to model properties that misjudge the true probability of default. The violation of the true data generating process is, however, outweighed by higher forecasting accuracy of static model. More extensive dataset would allow us to draw more general and consistent conclusions. We leave this as one of the areas for further research.

References

Altman, E. I., 1968, Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy, *Journal of Finance* 23, 589–609.

Beaver, William H., 1966, Financial ratios as predictors of failure, *Journal of Accounting Research* 4, 71–111.

Beck, N., J. N. Katz, and R. Tucker, 1998, Taking time seriously: time-series-cross-section analysis with a binary dependent variable, *American Journal of Politicial Science* 42, 1260–1288.

Bellotti, Tony, and Jonathan Crook, 2009, Credit Scoring With Macroeconomic Variables Using Survival Analysis, *Journal of Operational Research Society* 60, 1699-1707(9).

Campbell, J. Y., J. Hilscher, and J. Szilagyi, 2008, In search of distress risk, *The Journal of Finance* LXIII, 2899–2939.

Clements, M. P., and D. F. Hendry, 1998, *Forecasting Economic Time Series* (Cambridge University Press, Cambridge).

Davidson, R., and J. G. MacKinnon, 1984, Convenient specification tests for logit and probit models, *Journal of Econometrics* 25, 241–262.

Engelmann, B., E. Hayden, and D. Tasche, 2003, Testing Rating Accuracy, *Risk* 16, 82–86.

Fantazzini, Dean, and Figini Silvia, 2009, Random Survival Forests Models for SME Credit Risk Measurement, *Methodology and Computing in Applied Probability* 11, 24–45.

Fuertes, A.-M., and E. Kalotychou, 2006, Early Warning Systems for Sovereign Debt Crises: The Role of Heterogeneity, *Computational Statistics and Data Analysis* 51, 1420–1441.

Hillegeist, S. A., E. K. Keating, D. P. Cram, and K. G. Lundstedt, 2001, *Corporate bankruptcy: do debt covenant and disclosure quality measures provide information beyond options and other market variables?*, Kellogg Graduate School of Management .

Jensen, M., 1986, Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers, *American Economic Review* 76, 323–329.

Jones, Stewart, and David A. Hensher, 2008, *Advances in Credit Risk Modelling and Corporate Bankruptcy Prediction* (Cambridge University Press, Cambridge).

Lenox, Clive, 1999, Identifying Failing Companies: A Reevaluation of the Logit, Probit and DA Approaches, *Journal of Economics and Business* 51, 347–364.

Loeffler, Gunter, and Peter N. Posch, 2008, *Credit risk modeling using Excel and VBA* (Wiley, Chichester).

Mester, L., 1997, *What's the Point of Credit Scoring? , Business Review.*

Nam, Chae W., Tong S. Kim, Nam J. Park, and Hoe K. Lee, 2008, Bankruptcy Prediction Using a Discrete-Time Duration Model Incorporating Temporal and Macroeconomic Dependencies, *Journal of Forecasting* 27, 493–506.

Ohlson, James A., 1980, Financial Ratios and the Probabilistic Prediction of Bankruptcy, *Journal of Accounting Research* 18, 109–131.

Peresetsky, Anatoly, Alexandr Karminsky, and Sergei Golovan, 2004, *Probability of default models of Russian banks* (BOFIT, Helsinki).

Quantitative Micro Software, LLC, 2007, *EViews 6 User's Guide II* (Irvine CA, USA).

Rodriguez, A., and P. N. Rodriguez, 2006, Understanding and predicting sovereign debt rescheduling: a comparison of the areas under receiver operating characteristic curves, *Journal of Forecasting* 25, 459–479.

Saunders, Anthony, and Linda Allen, 2002, *Credit risk measurement. New approaches to value at risk and other paradigms* (Wiley, New York, NY).

Shumway, Tyler, 2001, Bankruptcy More Accurately: A Simple Hazard Model, *The Journal of Buisness* 74, 101–124.

Sobehart, Jorge, Sean Keenan, and Roger Stein, 2001, Benchmarking Quantitative Default Risk Models: A Validation Methodology, *Algo Research Quarterly* 4, 57–71.

Verbeek, Marno, 2009, *A guide to modern econometrics* (Wiley, Chichester).

Westgaard, Sjur, and Nico van der Wijst, 2001, Default probabilities in a corporate bank portfolio: A logistic model approach, *European Journal of Operational Research* 135, 338–349.

Wiginton, J. C., 1980, A note on the Comparison of Logit and Discriminant Models of Consumer Credit Behavior, *The Journal of Financial and Quantitative Analysis* 15, 757–770.

Zmijewski, Mark E., 1984, Methodological Issues Related to the Estimation of Financial Distress Prediction Models, *Journal of Accounting Research* 22, 59–82.

Appendix

A. *Descriptive Statistics for the Whole Sample*

	Defaulted Companies Number of observations = 102				Non-Defaulted Companies Number of observations = 100			
Variable	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
r2	1,76	34,22	-175,32	295,16	45,77	251,80	-116,25	2428,00
r5	-0,16	0,25	-1,30	0,41	0,04	0,13	-0,45	0,80
r6	-0,03	0,48	-2,34	2,33	5,97	20,89	-0,86	114,40
r19	-0,14	0,31	-1,47	0,60	0,06	0,23	-0,84	0,86
r22	1,10	0,78	0,60	5,39	2,56	2,49	0,00	15,31
r28	0,66	1,64	-1,12	14,61	2,11	4,35	-16,32	36,48
r32	0,52	0,41	0,14	1,81	1,09	0,65	-0,72	3,45
r33	-0,34	0,31	-0,64	0,35	-0,02	0,23	-0,88	0,45

Table 16: Descriptive statistics, Whole sample Model I

	Defaulted Companies Number of observations = 102				Non-Defaulted Companies Number of observations = 1442			
Variable	Mean	Std. Dev.	Min	Max	Mean	Std. Dev.	Min	Max
r3	-0,13	0,16	-0,51	0,18	-0,01	0,15	-0,51	0,18
r6	0,01	0,39	-0,40	2,33	0,92	2,45	-0,40	9,68
r10	0,36	0,23	0,08	0,69	0,24	0,15	0,08	0,69
r22	1,10	0,78	0,60	5,39	2,09	1,36	0,60	6,04
r29	-4,49	1,05	-6,11	-2,92	-4,34	0,96	-6,11	-2,92
r30	0,30	0,79	0,00	7,04	2,30	3,35	0,00	12,98
r32	0,52	0,41	0,14	1,81	0,94	0,54	0,14	1,81
r33	-0,34	0,31	-0,64	0,35	-0,02	0,20	-0,64	0,35

Table 17: Descriptive statistics, Whole sample Model II and Model III

B. *Correlation Ananlysis*

e(V)	r5	r6	r19	r22	r28	r33	r2sq	r19sq	r33sq
r5	1.0000								
r6	0.2792	1.0000							
r19	0.2533	0.4480	1.0000						
r22	0.3404	0.5151	0.0140	1.0000					
r28	0.1422	0.6690	0.4563	0.1897	1.0000				
r33	-0.0802	0.5456	0.0469	0.5071	0.3027	1.0000			
r2sq	-0.3873	-0.9358	-0.5540	-0.5663	-0.6327	-0.4980	1.0000		
r19sq	**0.5050**	0.4094	0.3809	0.2271	0.3369	0.0477	-0.5622	1.0000	
r33sq	-0.4828	-0.4429	-0.0027	-0.5617	-0.1830	-0.3381	0.4582	-0.1804	1.0000

Table 18: Correlation matrix of covariates of Model Ib

e(V)	r3	r6	r10	r22	r29	r30	r32	r33
r3	1.0000							
r6	-0.0204	1.0000						
r10	0.2294	-0.1007	1.0000					
r22	0.0565	0.0288	0.3272	1.0000				
r29	-0.0090	0.1227	0.2484	0.1173	1.0000			
r30	0.0482	0.0871	-0.1025	-0.0854	0.2384	1.0000		
r32	-0.2203	-0.0055	-0.0518	-0.1106	0.0686	-0.1793	1.0000	
r33	-0.1478	-0.0441	-0.0361	0.0517	-0.0056	-0.0621	**0.5997**	1.0000

Table 19: Correlation matrix of covariates of Model II and Model III

C. *Heteroskedasticity Plots*

Below the residuals of the models were plotted against fitted values for visual inspection of heteroskedasticity.

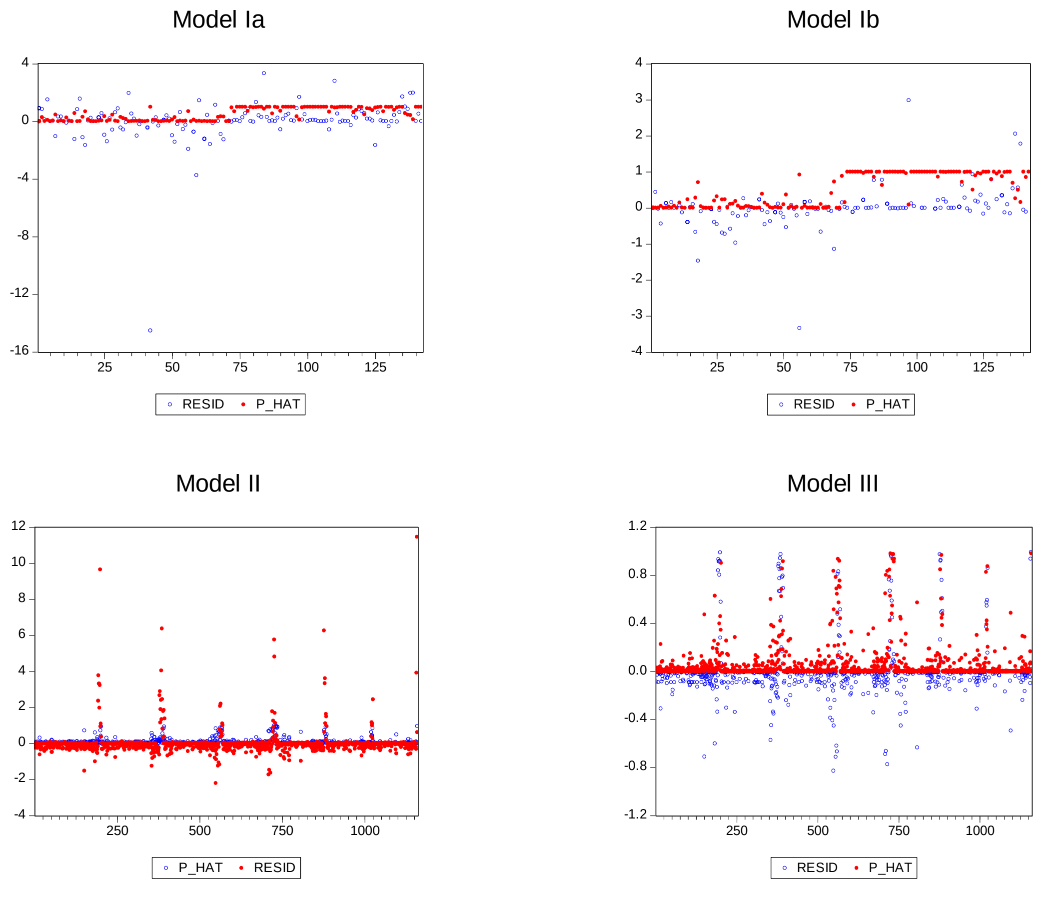

Figure 5: Heteroskedasticity Plots

D. *ROC Curves*

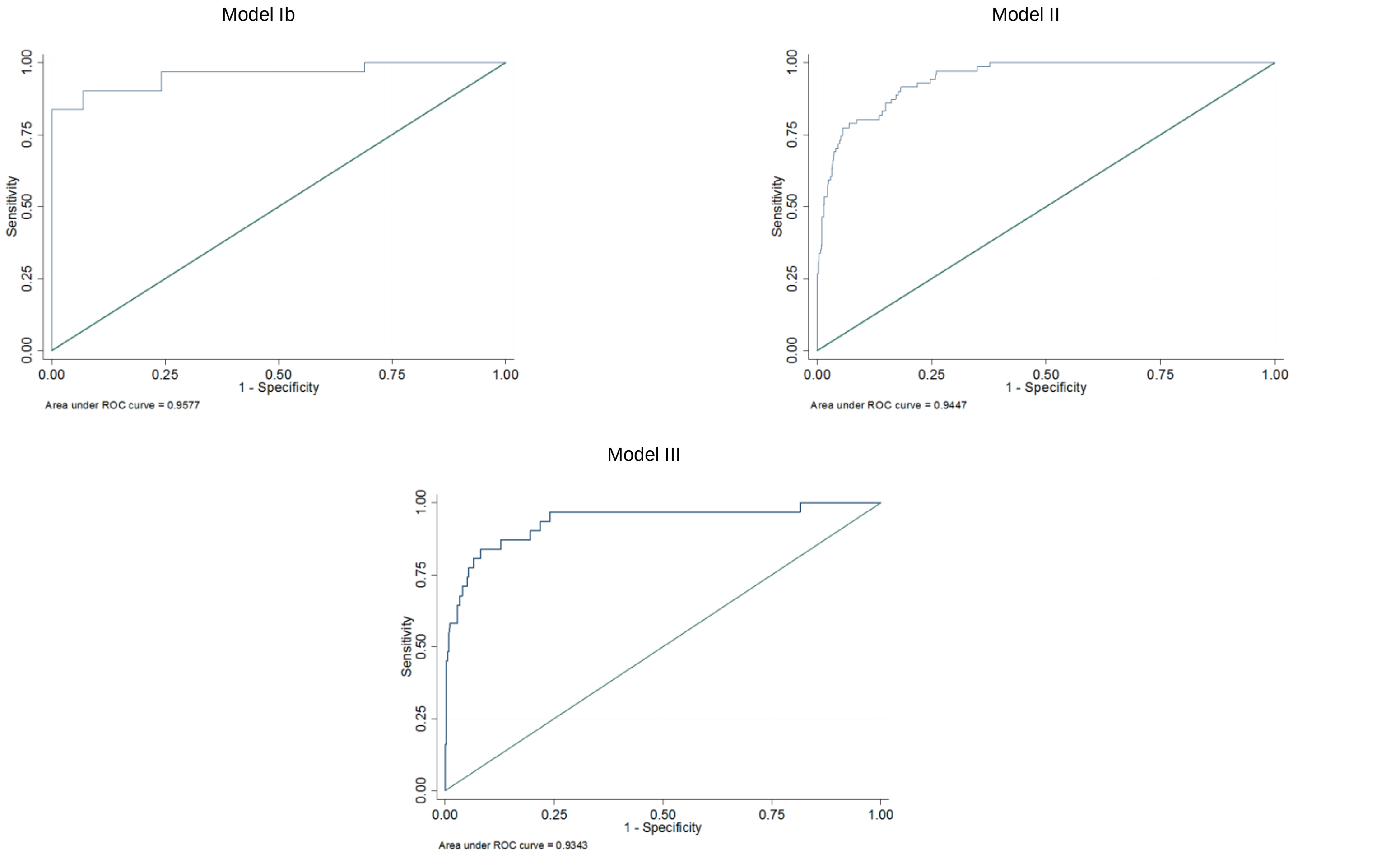

Figure 6: ROC curves for all three models

E. *Sensitivity/Specificity Plots*

In this sub-section Sensitivity/Specificity plots for all three models are displayed. The intersection between the sensitivity and the specificity curve indicates the optimal threshold probability for the respective model.

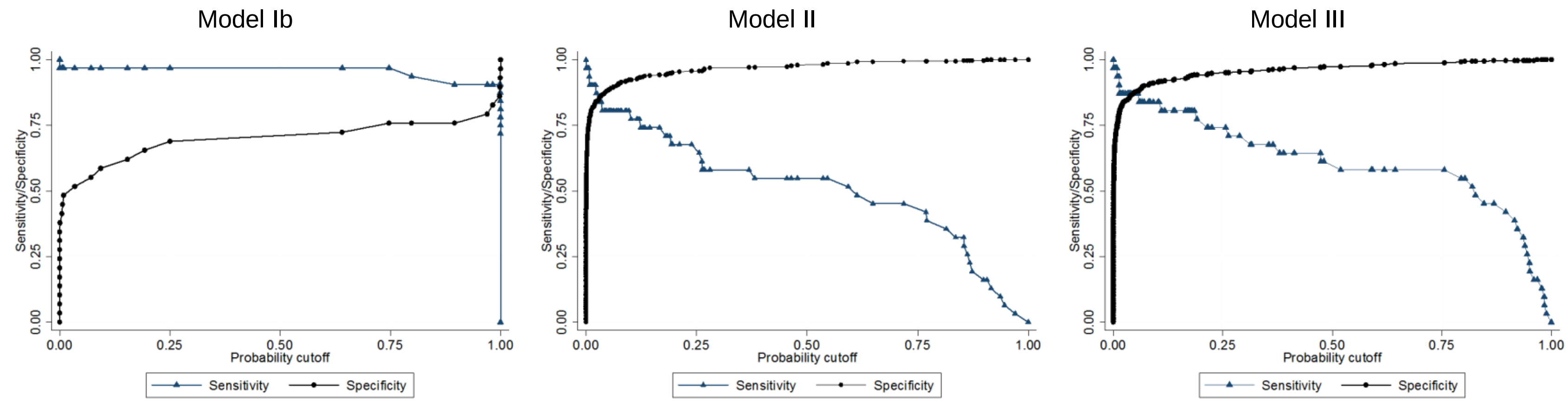

Figure 7: Sensitivity/Specificity plots for all three models