DANMARKS NATIONALBANK
WORKING PAPERS
2005 • 28

Anne Dyrberg Rommer

Financial Markets
Danmarks Nationalbank, Copenhagen
and
Centre for Applied Microeconometrics (CAM),
Institute of Economics, University of Copenhagen

Testing the Assumptions of Credit-scoring Models

21 June 2005

The Working Papers of Danmarks Nationalbank describe research and development, often still ongoing, as a contribution to the professional debate.

The viewpoints and conclusions stated are the responsibility of the individual contributors, and do not necessarily reflect the views of Danmarks Nationalbank.

As a general rule, Working Papers are not translated, but are available in the original language used by the contributor.

Danmarks Nationalbank's Working Papers are published in PDF format at www.nationalbanken.dk. A free electronic subscription is also available at this Web site.

The subscriber receives an e-mail notification whenever a new Working Paper is published.

Please direct any enquiries to
Danmarks Nationalbank, Information Desk, Havnegade 5, DK-1093 Copenhagen K
Denmark
Tel.: +45 33 63 70 00 (direct) or +45 33 63 63 63
Fax : +45 33 63 71 03
E-mail:info@nationalbanken.dk

Nationalbankens Working Papers beskriver forsknings- og udviklingsarbejde, ofte af foreløbig karakter, med henblik på at bidrage til en faglig debat.

Synspunkter og konklusioner står for forfatternes regning og er derfor ikke nødvendigvis udtryk for Nationalbankens holdninger.

Working Papers vil som regel ikke blive oversat, men vil kun foreligge på det sprog, forfatterne har brugt.

Danmarks Nationalbanks Working Papers er tilgængelige på Internettet www.nationalbanken.dk i pdf-format. På webstedet er det muligt at oprette et gratis elektronisk abonnement, der leverer en e-mail notifikation ved enhver udgivelse af et Working Paper.

Henvendelser kan rettes til :
Danmarks Nationalbank, Informationssektionen, Havnegade 5, 1093 København K.
Telefon: 33 63 70 00 (direkte) eller 33 63 63 63
E-mail: info@nationalbanken.dk

ISSN (trykt/print) 1602-1185

ISSN (online) 1602-1193

Resumé

Dette papir diskuterer et antal af de emner, der er relevante, når en kreditrisikomodel sættes op og det tester de antagelser, der gøres i de fleste regnskabsbaserede kreditrisikomodeller. Der foretages en ikke-standard sammenligning af to hazard-modeller med forskelligt specificerede hazard-funktioner: den ene specificeres som en logit model, den anden som en probit model. Specifikationen af kreditrisikomodellen som en hazard-model tillader os at inkludere information om virksomhederne i perioden op til "økonomiske vanskeligheder". Den logistiske fordeling ligner normalfordelingen, undtagen i halerne, hvilket medfører, at sandsynlighederne udregnet vha. af de to modeller ofte er ret ens, undtagen i halerne. Halerne i den logistiske fordeling er "tungere" end halerne i normalfordelingen, dvs. at i halerne af den logistiske fordeling er sandsynlighederne større sammenlignet med normalfordelingen. Sammenligningen af de to fordelinger er relevant, da egenskaberne i halerne af fordelingerne er i fokus i papiret. Logit- og probit-specifikationerne testes formelt mod hinanden vha. to tests, der så vidt vi ved benyttes for første gang i kreditrisiko-litteraturen. I estimationerne antages, at hvis to virksomheder har identiske værdier af de forklarende variabler, så har de også identiske hazard-funktioner, dvs. uobserverbar heterogenitet antages ikke at være til stede. Tilstedeværelsen af uobserverbar heterogenitet kan medføre flere problemer, derfor, som et specifikations-check, udvides estimationerne til også at omfatte uobserverbar heterogenitet.

Herudover diskuteres de forskellige måder, som forskellige typer af exits er behandlet på i litteraturen. Der er nylige eksempler på studier inden for kreditrisiko-litteraturen og inden for den industriøkonomiske litteratur, der stadig ikke skelner mellem forskellige typer af exit. Da det omfattende data set, som benyttes her, tillader sammenligning af forskellige specifikationer, undersøges det, hvad konsekvenserne er af at opstille 1) en hazard-model, der modellerer virksomheder i økonomiske vanskeligheder, og hvor virksomheder, der forlader samplet af andre årsager end økonomiske vanskeligheder, behandles som ikke længere observerede, når de forlader samplet, samt at opstille 2) en hazard-model, hvor den generelle exit hændelse modelleres (dvs. ikke opdelt på typen af exit). Så vidt vi ved er der ingen andre papirer, der sammenligner resultaterne for sådanne estimationer.

Konklusionerne i artiklen er følgende: 1) Det lader ikke til, at der er nogen forskel mellem logit- og probit-specifikationen for hazard-funktionen. 2) Uobserverbar heterogenitet ser ud til ikke at spille en rolle, sandsynligvis fordi et antal af proxy variable er benyttet for uobserverbare faktorer. 3) Resultaterne i modellen afhænger af hvilken type exit, der modelleres (økonomiske vanskeligheder versus pooled exit). Dette gælder både for de estimerede parameter-estimater samt modellernes forudsigelsesevne, uanset om specifikationen for hazard-funktionen er logit- eller probit-specifikationen. Den praktiske implikation af papiret er, at det er vigtigt at tænke grundigt over specifikationen af kreditrisikomodeller. Det er vigtigt at forstå, at resultaterne afhænger af den portefølje, der undersøges, og dermed, at hver modelbygger er nødt til at overveje de forskellige emner.

Abstract

This paper discusses a number of issues that are relevant when setting up a credit-scoring model and tests the assumptions used in accounting-based credit-scoring models. A non-standard comparison of two hazard models with differently specified hazard functions is made: one with a logit specification and the other with a probit specification. The specification of the credit-scoring model as a hazard model allows us to include information leading up to "financial distress". The logistic distribution is similar to the normal, except in the tails, and so the logit and the probit model tend to give similar probabilities, except in the tails. The tails of the logistic distribution are considerably heavier than the tails of the normal distribution, i.e. in the tails of the logistic distribution, the probabilities are larger compared to the normal distribution. The comparison of the two distributions is relevant, as the properties at the tails of the distributions are at focus here. The logit and the probit specification are formally tested against each other using two tests, which are probably used for the first time within the credit-scoring literature. The estimations assume that if two firms have identical values of the covariates, they also have identical hazard functions, that is, unobserved heterogeneity is assumed away. The presence of unobserved heterogeneity can cause several problems, therefore, as a specification check, the hazard functions are extended to also include unobserved heterogeneity.

In addition to investigating the various specifications of the hazard function, the paper discusses the treatment in the literature of different types of exits. There are recent examples of studies within the credit-scoring and the industrial organization literature, which still do not distinguish between exit types. As the extensive data set allows comparisons of different specifications, the paper explores the consequences of setting up 1) a hazard model where the event "financial distress" is modelled and where firms that exit for other reasons than financial distress are treated as censored or no longer observed and 2) a hazard model where the general exit event is modelled (i.e. not split up on exit type). To the best of our knowledge no other paper has provided results from such estimations.

The conclusions in the article are the following: Firstly, there does not seem to be any major difference between the logit and the probit specification. Secondly, unobserved heterogeneity seems to be unimportant, probably because a number of proxies are used for inherently unobservable variables. Thirdly, the results differ depending on the event, which is modelled (financial distress versus pooled exits). This is the case for the estimated parameters as well as the predictive abilities of the models, no matter whether the specification for the hazard functions is the logit or the probit specification. The practical implication of the paper is that it is important to think careful about the specification of credit-scoring models. It is crucial to understand that the results depend on the portfolio under consideration, and hence, that every model builder has to think about the issues.

Anne Dyrberg Rommer*

Testing the Assumptions of Credit-scoring Models

* The author would like to thank Hans Christian Kongsted, Anders Møller Christensen, Jesper Berg, Karsten Biltoft, Bo Honoré, David Lando, Jesper Rangvid and colleagues at the Danish Central Bank for commenting on this or earlier versions of the paper and seminar participants at the Danish Graduate Program in Economics workshop held 13. - 14. November 2003, Centre for Applied Microeconometrics (University of Copenhagen) seminar held 18. November 2003, Danish Central Bank seminar held 27. November 2003. Corresponding address is Anne Dyrberg Rommer, Financial Markets, Danmarks Nationalbank, Havnegade 5, DK-1093 Copenhagen, Denmark. Phone: + 45 33 63 63 63. Email: ady@nationalbanken.dk

1. Introduction

The purpose of this paper is to discuss a number of issues that are relevant when setting up a credit-scoring model and to test the assumptions used in accounting-based credit-scoring models. Specification issues are important to consider, as more powerful models are more profitable than weaker ones, c.f. Stein (2005). The topic is important not only for individual credit institutions, who use credit-scoring models to find out which clients they want to offer loans and to detect clients that are likely to default at an early stage, but also for banking supervisors, who are regulating banks, for central banks, who are analyzing the developments in the financial sector and accordingly assessing financial stability developments, and for other agents, e.g. management, financial analysts, investors and auditors, who also need timely warnings.

In addition to the strong interest in the topic of credit-scoring from the policy side (from central banks and banking supervisors) and from a more practical front (from e.g. credit institutions, managers, investors and auditors), there is also a strong academic interest in the topic. This is seen by the vast amount of literature on credit-scoring models. Some of the recent studies are Hillegeist, Keating, Cram and Lundstedt (2004), Jones and Hensher (2004), Dyrberg (2004) and Grunert, Norden and Weber (2005). Examples of surveys of the developments in the area include Zavgren (1982), Jones (1987), Dimitras, Zanakis and Zopounidis (1996), Altman and Saunders (1998) and Balcaen and Ooghe (2004). Some of the often-quoted parametric credit-rating studies are Altman (1968), Ohlson (1980) and Shumway (2001). Examples of non-parametric credit-rating studies are Frydman, Altman and Kao (1985), Tam and Kiang (1992) and Dimitras, Slowinski, Susmaga and Zopounidis (1999).

Altman (1968), Ohlson (1980) and Shumway (2001), which are some of the often-quoted parametric credit-rating studies, c.f. above, suggest the use of multivariate discriminant analysis, the logit model and the hazard model, respectively. These methods are the standard methods within the parametric credit-scoring literature. Accordingly, they are used in a number of papers. For example, discriminant analysis is used in Bardos (2001), Cifarelli and Corielli (1988), Betts and Belhoul (1987), Dambolena and Khoury (1980) and in Altman, Haldeman and Narayanan (1977), and logit models are used in Moody's Investors Service (2001a), Moody's Investors Service (2001b), Moody's Investors Service (2002), Jiménez and Saurina (2004) and Corcóstegui, González-Mosquera, Marcelo and Trucharte (2003). Hazard models are used in Hillegeist, Keating, Cram and Lundstedt (2004), Campbell, Hilscher and Szilagyi (2005) and Chava and Jarrow (2004). In addition to the standard methods (discriminant analysis, logit models and hazard models), other parametric methods have been used occasionally, e.g. the probit model. Papers, which consider probit models, are e.g. Bunn and Redwood (2003), Skogsvik (1990) and Zmijewski (1985).

A number of studies compare various estimation strategies, c.f. Altman, Marco and Varetto (1994), Back, Laitinen, Sere and Wezel (1996), Begley, Ming and Watts (1996), Lo (1986), Frydman, Altman and Kao (1985) and Lennox (1999). This paper is a methodological paper along the lines of these papers. Most of the papers have compared logit analysis to other estimation methods such as discriminant analysis and various non-parametric techniques. Here a non-standard comparison of two hazard models with differently specified hazard functions is made: one with a logit specification and the other with a probit specification. The specification of the credit-scoring model as a hazard model allows us to include information leading up to "financial distress". The logistic distribution is similar to the normal, except in the tails, and so the logit and the probit model tend to give similar probabilities, except in the tails. The tails of the logistic distribution are considerably heavier than the tails of the normal distribution, i.e. in the tails of the logistic distribution, the probabilities are larger compared to the normal distribution. The comparison of the two distributions is relevant, as the properties at the tails of the distributions are at focus here. The only other paper we have found, which compares the logit and the probit model (though not in the framework of a hazard model), is Lennox (1999).

We use the Davidson and MacKinnon (1993:492) test-procedure and the Silva (2001) test-procedure to test the two specifications for the hazard function. Lennox (1999) compares the results from the estimation of a logit and a probit model. He does not set up a hazard model and he does not provide any tests for the specification of the credit-scoring model as a logit or probit model. As we test the logit and the probit specification for the hazard function, our approach extends his study. To the best of our knowledge this is the first time, these tests have been used within the credit-scoring literature.

The estimations assume that if two firms have identical values of the covariates, they also have identical hazard functions, that is, unobserved heterogeneity is assumed away. The presence of unobserved heterogeneity can cause several problems, therefore, as a specification check, the hazard functions are extended to also include unobserved heterogeneity. Unobserved heterogeneity is not modelled in Lennox (1999). The only other studies we know of, which models unobserved heterogeneity in a credit-scoring setting, are Kaiser (2001), Bernhardsen (2001) and Jones and Hensher (2004).

In addition to the investigations of the various specifications of the hazard function (logit and probit specification with and without unobserved heterogeneity) the treatment of different types of exits in the literature is discussed. There are recent examples of studies within the credit-scoring and the industrial organization literature, which still do not distinguish between exit types (e.g. Bunn and Redwood (2003), Pérez, Llopis and Llopis (2004), Mata and Portugal (2002) and Kimura and Fujii (2003)). Therefore, it will be shown what the consequences are of setting up

1) a hazard model where the event financial distress is modelled and where firms that exit for other reasons than financial distress are treated as censored or no longer observed and 2) a hazard model where the general exit event is modelled (i.e. not split up on exit type). To the best of our knowledge no other paper provides the estimations of a hazard model, where firms in financial distress are modelled and where the other forms of exits are treated as censored versus a model, which pool the three modes of exit (financial distress, voluntary liquidation and mergers and acquisitions etc.). Dyrberg (2004), Harhoff, Stahl and Woyde (1998) and Schary (1991) are the papers, which are closest to our paper in this respect, c.f. section 2.

The data set used in the estimations is unique compared to most other credit-scoring studies and industrial organisation studies. It includes the whole population of Danish public and private limited liability companies, which existed between 1995 and 2001. Most are small and medium-sized enterprises. The panel data set covers all non-financial sectors of the Danish economy. Included in the estimations are around 30,000 firms and more than 150,000 firm-year observations. There are 2,617 firms in financial distress, 907 voluntarily liquidated firms, and 1,233 firms that are acquired/have merged with other firms, etc. In comparison, Lennox (1999) is not able to follow each firm throughout his observation window. His sample is also smaller than ours. Compared to our more than 30,000 firms and 2,617 defaults, he uses a sample, which includes 949 firms and 90 defaults.

The overall conclusions from the analysis are that there does not seem to be any major difference between the hazard model with the logit and the probit specification, that unobserved heterogeneity seems to be unimportant, but that the results differ depending on the event, which is modelled (financial distress versus pooled exits). The observations leading to these conclusions are discussed more extensively in section 6.

The practical implication of the paper is that it is important to think careful about the specification of credit-scoring models. Here the specification issues are highlighted and investigated using an extensive data set on Danish non-financial sector firms. It is crucial to understand that the results depend on the portfolio under consideration, and hence, that every model builder has to think careful about the issues. This paper provides a framework for such investigations.

The paper is divided into 6 sections. Section 2 presents the data, which is used in the estimations, and it discusses the way different exits are modelled in the literature. In section 3 the estimation problem is set up and tests for the specification of the hazard function are presented. Section 4 reports the results, including the outcome of the tests for the specification of the hazard function and the parameter estimates, which are obtained, when 1) the financial distress event and 2) the exit event is modelled. Section 5 discusses prediction in credit-scoring models, and it evaluates the predictive ability of the models. Finally, section 6 concludes.

2. Data and the Treatment of Exits

The data set used in the estimations covers all Danish public limited liability companies and private limited liability companies that existed in the period from 1995 to 2001. As it covers the whole population of public and private limited liability companies, it differs from the data set available to individual credit institutions, which wants to set up credit-scoring models. Some of the issues, which are necessary for individual credit institutions to think about, but which are not a problem in the set up here in this paper, are sketched in box 2.a (drop-outs) and box 2.b (reject inference).

The main part of the data set used in the estimations is received from the Danish credit-rating agency KOB A/S. It comprises information on financial issues as well as non-financial issues. On top of the information received from KOB A/S, the data base is augmented to also include whether or not the company is 1) an ultimate parent company, 2) a wholly owned subsidiary, 3) quoted on the stock exchange, 4) owned by the public, 5) owned by a fund and 6) a concentration index (measuring the concentration of the various sectors).

Box 2.a: Drop-outs in credit institutions portfolios

Although not discussed in the literature, individual banks have customers that may drop-out of the sample, while they are still active. There can be different reasons for the drop-out. One hypothesis could be that some firms choose another bank, because it offers better service or a better price (perhaps because they are bundling their activities). These companies may be well-performing companies, which show no indication of financial distress. Another hypothesis could be that some companies change to another bank, because they are asked to by their current bank, i.e. that their current bank could suspect that they would soon enter financial distress. If it is not known by banks, why firms drop out of their portfolio, it is crucial for banks to find out what type of drop-outs they observe, and accordingly, what assumptions they can make about the drop-outs in order to obtain consistent estimates, when they set up their credit-scoring models.

Rommer (2005b:17ff) provides a framework for analyzing drop-outs and interested readers are referred to the paper for details on how she analyses drop-outs versus firms that do not drop out by 1) testing for equal means of a number of characteristics, 2) by estimation of drop-out probits and 3) by estimation of credit-scoring models with and without the drop-outs.

Box 2.b: Reject inference

Credit-scoring models are used in credit institutions to evaluate loan applicants in order to assess whether or not a loan should be granted. The output of a credit-scoring model is an estimate of the probability that a specific loan applicant will default within a certain horizon, usually a one-year horizon.

Credit scoring-models have received new interest, as their output, the PD (probability of default) can now serve as an input in the calculation of a credit institutions minimal capital requirement, if the credit institutions choose to calculate their minimal capital requirement using one of two different internal ratings-based approaches (IRB), c.f. BCBS (2004) and Borup, Kurek and Rommer (2005). In the Basel II proposal, BCBS (2004:91), it is specifically stated that "Internal ratings and default and loss estimates must play an essential role in the credit approval, risk management, internal capital allocations, and corporate governance functions of banks using the IRB approach. Ratings systems and estimates designed and implemented exclusively for the purpose of qualifying for the IRB approach and used only to provide IRB inputs are not acceptable. It is recognized that banks will not necessarily be using exactly the same estimates for both IRB and all internal purposes. For example, pricing models are likely to use PDs and LGDs [loss given default] relevant for the life of the asset. Where there are such differences, a bank must document them and demonstrate their reasonableness to the supervisor."

The idea behind Basel II is that the credit institutions should not only develop credit-scoring models and use these to calculate their minimal capital requirements, but that these models should also play an essential role in the credit approval process. One issue, which is not highlighted in the Basel II proposal, is that it may not be ideal to use the same credit-scoring model to calculate the probability of default and to approve new loans. The reason for this is the typical sample selection argument. If only obligors that already have been approved for a loan are taken into account in the estimations, then it is not appropriate to use the same model to consider new applications. If the models estimated using data on already approved applicants are applied to all applicants, then a sample selection bias is introduced.

Feelders (2003), among others, has a theoretical discussion of the sample selection issue, also called the reject inference problem (without making reference to Basel II). The literature is divided into two camps. Some of the estimations show that it does not make a big difference to adjust for the reject inference problem. An example is Crook and Banasik (2004) (consumer credit-scoring problem). Others find that it does make a difference to adjust for the reject inference problem. These include Chen and Åstebro (2003) and Roszbach (2003) (firm credit-scoring problems) and Chen and Åstebro (2001) and Greene (1998) (consumer credit-scoring problems).

In the raw data set there are 603,956 firm-year observations. After the exclusion of holding companies and financial firms and after making some corrections to the database, there are 430,422 firm-year observations left. The panel used in the estimations consists of companies that were incorporated in the period 1995 – 2001 with at least 5 employees the year they are included in the sample and with a balance sheet of at least kr. 500,000 (flow sampled companies), and companies that were active in 1995 but were incorporated before 1995 with at a balance sheet of at least kr. 500.000 and 5 employees in 1995 (the stock sampled companies). The panel consists of 168,778 firm-year observations, covering 32,453 firms. Due to missing variables, the final number of firm-year observations in the estimations is 168,350, covering 32,365 firms.

Note the difference between flow and stock sampled companies. The stock sampled companies are the ones that are active, when they are included in the sample. The flow sampled companies are the ones that are incorporated at some point in the observation window, which spans from 1995-2001. Hence, the difference between flow and stock sampled companies is that the whole history of the stock sampled companies is not observed, whereas the whole history of the flow sampled companies is observed. Both the flow and the stock sampled firms can be right censored. Only stock sampled firms are left truncated (i.e. not observed from the beginning of the spell, but with a known incorporation data). For further details see e.g. D'Addio and Rosholm (2002) and Dyrberg (2004).

Table 2.a: Number of firms

	Financial distress (E1)	Voluntary liquidations (E2)	Mergers and acquisitions etc. (E3)	Active	Total
1995	0	0	0	18853	18853
1996	372	87	177	20684	21320
1997	348	110	156	22008	22622
1998	347	129	195	23422	24093
1999	453	124	211	25000	25788
2000	618	148	226	26415	27407
2001	479	309	268	27639	28695
Total	**2617**	**907**	**1233**	**164021**	**168778**

Source: Dyrberg (2004)

An overview of the data set is seen from table 2.a, which shows the number of active firms every year and the number of firms that exit because of financial distress (E1), voluntary liquidation (E2), and mergers and acquisitions etc. (E3). Most firms exit because they are financially distressed. Voluntary liquidations account for the smallest number of exits. Rommer (2005a) and Phillips and Kirchhoff (1989) discuss the distinction between voluntary and involuntary exits.

Table 2.b summarizes the predictors used in the estimations. These are the same predictors that were used in Dyrberg (2004). They are divided into core variables, proxy variables, and controls. Core variables are variables that are usually used in credit-scoring studies. The proxy variables serve as proxies for inherently unobserved variables. Controls are included to take into account the macroeconomic effects and sector affiliation etc. As the focus in this paper is on

the methodology behind credit-scoring models and not on the selection of predictors and further details about the data set, for more on that, the reader is referred to Dyrberg (2004), which has extensive discussions of both issues.

Table 2.b: Predictors

<table>
<tr><th></th><th>Variables</th></tr>
<tr><td rowspan="5">CORE VARIABLES</td><td>Firm Age (dummies) (reference category: firms that are one year old)</td></tr>
<tr><td>Short-term debt to total assets</td></tr>
<tr><td>Return on net assets</td></tr>
<tr><td>Solvency ratio</td></tr>
<tr><td>Firm size</td></tr>
<tr><td rowspan="10">PROXIES</td><td>Firms that are not diversified (reference category)
Diversification 2 sectors (related business)
Diversification 3–9 sectors (related business)
Diversification 2 sectors (unrelated business)
Diversification 3–9 sectors (unrelated business)</td></tr>
<tr><td>Local authority group 1 (reference category)
Local authority group 2
Local authority group 3
Local authority group 4
Local authority group 5</td></tr>
<tr><td>Concentration</td></tr>
<tr><td>Firms that are not owned by the public (reference category)
Owned by the public (dummy)</td></tr>
<tr><td>Firms that are not owned by the a fund (reference category)
Owned by a fund (dummy)</td></tr>
<tr><td>Firms that are not ultimate parent companies (reference category)
Ultimate parent companies (dummy)</td></tr>
<tr><td>Firms that are not wholly owned subsidiaries (reference category)
Wholly owned subsidiaries (dummy)</td></tr>
<tr><td>Public limited liability company (reference category)
Private limited liability company (dummy)</td></tr>
<tr><td>Firms that are not publicly traded (reference category)
Publicly traded companies (dummy)</td></tr>
<tr><td>Firms without critical comments from the auditors (reference category)
Critical comments from the auditors (dummy)</td></tr>
<tr><td rowspan="3">CONTROLS</td><td>Year dummies: Year 1996 (reference category), Year 1997, Year 1998, Year 1999, Year 2000, Year 2001</td></tr>
<tr><td>Sector affiliation dummies: Farming, Forestry, Fishing, Mining, Manufacturing (reference category), Energy, Construction, Trade and hotel, Transport, Business service, Public service activities, Organisations, Not stated, Unknown. IT dummy (those firms that are considered as IT or tele-companies have, on top of the sector dummies, an IT dummy equal to 1, all other firms have an IT dummy equal to 0)</td></tr>
<tr><td>Primary bank dummies: Some firms register a Primary Bank connection in one of the following four categories (see Dyrberg (2004:section 7) for further details): Category 1, Category 2, Category 3, Category 4, Some firms do not register a primary bank connection (reference category)</td></tr>
</table>

In the credit-scoring literature, active firms and financially distressed firms only are most often modelled (see e.g. Beaver (1966), Altman (1968) and Ohlson (1980)). Exceptions are Schary (1991) and Dyrberg (2004), who both advocate for a richer discussion of the determinants of exits. They distinguish between bankruptcy, voluntary liquidation, and mergers and acquisitions and estimate the credit-scoring models as a competing-risks model. Shumway (2001) sets up a hazard model and treats firms that exit for other reasons than financial distress as censored or no longer observed, when they leave the sample, i.e. he groups active firms and firms that are voluntarily liquidated, and firms that have merged or have been acquired by other firms etc. Bunn and Redwood (2003) consider a firm as failed in a particular year if its company status is "in receivership, liquidation or dissolved, and its last reported accounts were in the previous year. This definition includes voluntary liquidation and dissolution where there may be no risk of default, but we are unable to distinguish between voluntary and compulsory failures in our data. … We do not consider being taken over to be a failure", i.e. out of necessity Bunn and Redwood (2003) group firms in financial distress and voluntarily liquidated firms.

In the industrial organization literature there has been some attempts to study the factors, which lead firms to exit, split up on type of exit, c.f. Harhoff, Stahl and Woyde (1998), Köke (2001), Prantl (2003) and Bhattacharjee, Higson, Holly and Kattuman (2004). These four mentioned studies distinguish between two of the three mentioned exit types. Despite the fact that these papers show that it is important to distinguish between exit types, a number of recent industrial organization studies still consider exits in general, i.e. not split up on exit type. Some of these recent studies are Pérez, Llopis and Llopis (2004), Mata and Portugal (2002) and Kimura and Fujii (2003). Other studies are Audretsch and Mahmood (1994), Audretsch and Mahmood (1995) and Mata et al. (1995).

As there are recent examples of studies within the credit-scoring and the industrial organization literature, which still do not distinguish between exit types, it will be shown what the consequences are of setting up 1) a hazard model where the event financial distress is modelled and where firms that exit for other reasons are treated as censored or no longer observed and 2) a hazard model where the general exit event, i.e. not split up on exit type, is modelled. To the best of our knowledge no other paper provides such estimations. Dyrberg (2004), Harhoff et al. (1998) and Schary (1991) are the papers, which are closest to our paper. Harhoff et al. (1998) and Schary (1991), who distinguishes between two modes of exit (bankruptcies and voluntary liquidations) and three modes of exit (bankruptcies, voluntary liquidations, and mergers and acquisitions), respectively, estimate a competing-risks model and a model with pooled exits. None of the two studies provide estimates for a hazard model, which models financial distress and treats all other exits as censored or no longer observed, when they leave the sample. Dyrberg (2004), who distinguishes between three modes of exit (bankruptcies, voluntary liquidations, and mergers and acquisitions, etc.), estimates a competing-risks model and a hazard model, where firms in financial distress are the event, and where all other exits are treated as censored or no longer observed, when they leave the data set. She does not provide estimates for a hazard model, which estimates the probability of the pooled exit event.

Another paper, which is along the lines of our paper, is Honjo (2000). He argues that by concentrating on firms in financial distress, he hopes to identify more significant factors, such as post-entry performance, compared to the studies, which model exits, i.e. treat all types of exits the same way.[1] Honjo (2000) does not show what his results would have been, had he included firms that exit for other reasons than financial distress. By reporting the results (parameter estimates and predictive ability) from estimations where we model 1) firms in financial distress and 2) exits, we show what the consequences are of modelling either event.

3. The Estimation Problem

Various statistical estimation methods have been suggested in the credit-scoring literature, c.f. the introduction. In this paper the focus is on the parametric estimation method, which is suggested in Shumway (2001), namely the hazard model. Shumway (2001) suggests the use of the hazard model, as it solves some of the econometric problems of the single-period logit approach that is suggested in Ohlson (1980). A particular important property of the hazard model is that it includes information on how long a firm survives with a set of characteristics.

We will investigate various specifications for the hazard function, namely the logit and the probit specification with and without unobserved heterogeneity.

In the following sections we set up the estimation problem, discuss the various specifications for the hazard functions and present the tests, which are used to test the different specifications against each other.

3.1. Derivation of the Sample Likelihood

The sample likelihood for the hazard model is derived in this section, which builds on Jenkins (1995). We derive the sample likelihood for the case, when one has discrete data. In this respect, note, that the data, which is used in the estimations

---

[1] Honjo (2000:560) is interested in investigating the post-entry performance of new firms, in particular, the business failure of new firms over time: "Following the analytical framework used by Audretsch and Mahmood (1994), (1995) and Mata et al. (1995), we estimate determinants of business failure among new manufacturing firms in Tokyo during 1986-1994. However, whereas these previous studies dealt with exits, we analyze business failures, which may be regarded as special cases of exits, that is, exits without solvency. Firms exit for different reasons: some may exit voluntarily while they are still gaining profits, but others are forced to exit due to business failure. By concentrating on the latter ones, this paper, it is hoped, will be able to identify more significant factors, such as post-entry performance." Honjo (2000) defines firms without solvency as those firms, which have ceased operations with total debt more than 10 million yen. The data on failed firms include the firms that voluntarily compromised with creditors and ceased operations. Honjo (2000) does not consider firms that exit for other reasons than financial distress.

(and presented in section 2), is, what in the literature is called grouped duration data. Information on the firms is obtained once a year, when the financial statement is handed in. With grouped duration data one can assume that durations are intrinsically discrete and treat them accordingly, or alternatively, one may attempt to relate the model to an underlying process in continuous time. Here data is treated as if it was intrinsically discrete.

The probability that a firm ends up as an event (e.g. a financially distressed firm) at time T_1 is

$$prob(T_i = t) = prob(T_1 = t),$$

where T_i is a discrete random variable representing the time at which the end of the spell occurs, also called the event time. This is an unconditional probability, which can be rewritten as a product of conditional probabilities. Define the hazard rate, h_{it}, as

$$h_{it} = prob(T_i = t | T_i \geq t; X_{it}),$$

where X_{it} is a vector of regressor variables which may vary with time. For both the stock and flow sampled firms, the probability of experiencing an event in year t, the unconditional probability, is

$$prob(T_i = t) = h_{it} \prod_{\tau=1}^{t-1} (1 - h_{i\tau}) = [h_{it} / (1 - h_{it}) \prod_{\tau=1}^{t} (1 - h_{i\tau})]$$

where τ denotes the incorporation date of the firm.

The probability of surviving beyond period t, the survivor function, is

$$prob(T_i > t) = \prod_{\tau=1}^{t} (1 - h_{i\tau})$$

To motivate the derivation of the sample likelihood for the stock sampled (left truncated) firms consider a firm, i, which was incorporated in 1992, and which is still active in 1998. The *unconditional probability* of the firm still being active in 1998 is then

$$prob(T_i > t) = \prod_{\tau=1}^{t} (1 - h_{i\tau}) = (1 - h_{i,1998})(1 - h_{i,1997})...(1 - h_{i,1992}),$$

whereas the probability of still being active, *conditional* on not having left the sample before 1995, when the firm is observed for the first time (called time *b*) is

$$prob(T_i > t + s_i | T_i > b - 1) = \prod_{t=b}^{b+s_i} (1 - h_{it}) = \frac{(1 - h_{i,1998})(1 - h_{i,1997})...(1 - h_{i,1992})}{(1 - h_{i,1994})...(1 - h_{i,1992})}$$

= $(1-h_{i,1998})(1-h_{i,1997})(1-h_{i,1996})(1-h_{i,1995})$

s_i indicates how many years each firms is observed.

The probability of experiencing the event in 1998, conditional on not having left the sample before 1995, is

$$prob(T_i = t + s_i | T_i > b-1) = \frac{h_{ib+s_i}}{1-h_{ib+s_i}} \prod_{t=b}^{b+s_i}(1-h_{it}) = \frac{(h_{i,1998})(1-h_{i,1997})...(1-h_{i,1992})}{(1-h_{i,1994})...(1-h_{i,1992})}$$

= $(h_{i,1998})(1-h_{i,1997})(1-h_{i,1996})(1-h_{i,1995})$

It is seen that the conditioning of the survivor probabilities for the stock sampled firms is handled via a "cancelling" of terms. It is important to note, that a firm only contributes with as many observations as there is in the data set. The flow sampled firms are observed from the incorporation date (=when the firm is first observed at time b) and contributes from that time on, altogether s_i years. This is also seen from the equations.

The likelihood of observing the event history data for the whole sample can now be constructed. The likelihood is

$$L = \prod_{i=1}^{n}\left[\left[\frac{h_{ib+s_i}}{1-h_{ib+s_i}}\prod_{t=b}^{b+s_i}(1-h_{it})\right]^{\delta_i}\left[\prod_{t=b}^{b+s_i}(1-h_{it})\right]^{(1-\delta_i)}\right]$$

$\delta_i = 1$ is defined as firms with completed spells and $\delta_i = 0$ is defined as those with uncompleted spells. When the event financial distress is modelled, the firms with uncompleted spells are the ones, which are still active at the end of the observation window, or those which gets censored, e.g. because they leave the sample for other reasons than financial distress. When the exit event is modelled, the firms with uncompleted spells are the ones, which are active at the end of the observation window.

s_i indicates how many years each firms is observed and $b + s_i$ refers to the point in time, when a firm leaves the database.

The log likelihood function is then

$$\log L = \sum_{i=1}^{n}\delta_i \log\left[\frac{h_{ib+s_i}}{1-h_{ib+s_i}}\right] + \sum_{i=1}^{n}\sum_{t=b}^{b+s_i}\log(1-h_{it})$$

This specification of the likelihood function gives the probability that a particular firm experiences an event (either financial distress or exit, depending on which of the two we are modelling).

3.2. Derivation of an easy Estimation Method

Using the above log likelihood function, Jenkins (1995:133) shows how to derive an easy estimation method. He defines a variable $y_{it}=1$ if $t=b+s_i$ and $\delta_i=1$, and $y_{it}=0$ otherwise. Using the y_{it} as indicator variable, the log-likelihood function can now be rewritten as

$$\log L=\sum_{i=1}^{n}\sum_{t=b}^{b+s_i} y_{it}\log[h_{it}/(1-h_{it})]+\sum_{i=1}^{n}\sum_{t=b}^{b+s_i}\log[(1-h_{it})].$$

There are two important things to notice. One is (as mentioned above) that each firm contributes with as many observations as it has years at risk of exiting. The other is that the creation of multiple observations from each firm follows directly from factoring the likelihood function for the data. Each of the terms may be treated as though it came from a distinct, independent observation.

The log likelihood function has the same form as the standard log-likelihood function for regression analysis of a binary variable, in this case y_{it}. Straightforward calculations give that the log likelihood function can be written as

$$\log L=\sum_{i=1}^{n}\sum_{t=b}^{b+s_i} y_{it}\log(h_{it})+\sum_{i=1}^{n}\sum_{t=b}^{b+s_i}(1-y_{it})\log(1-h_{it})$$

3.3. Expression for the Hazard Rate

Before the model can be estimated, the specification has to be completed by assuming an expression for the hazard rate. Below two different specifications will be presented. For now it is assumed that the specifications for the hazard functions capture all differences between firms using observed explanatory variables.

Let X_{it} characterize the covariates, including the baseline hazard function. The baseline hazard function can be specified parametrically and non-parametrically. A non-parametric specification is preferred as any inconsistency caused by misspecification is then avoided. In the estimations, dummies for each age (up to a dummy with companies that are 30 years old or more) are included (the reference category is firms that are 1 year old). Note the difference between the age of the specific companies and the time dummies that control for the macroeconomic environment. Let β denote the coefficients of the covariates, which may be time-varying.

Having introduced X_{it} and β, the specifications for the hazard function can now be presented. The suggested specifications are the logit and the probit. The logistic distribution is similar to the normal (except in the tails, which are considerably heavier, i.e. in the tails of the logistic distribution, the probabilities are larger compared to the normal distribution), and so the logit and the probit model tend to give similar probabilities, except in the tails, c.f. Greene (2003:667). Greene

(2003:667) also notes, that one would expect different predictions from the logit and the probit model, if the sample contains very few responses or non-responses. As we are in the tails of the distribution and as our sample contains very few responses compared to non-responses, it is worth investigating, whether the usual conclusion holds, namely that it is difficult to provide practical generalities on which model to choose (Greene (2003:667)).

The logit specification for the hazard function is:

$$h_{it} = 1/[1+\exp[-\beta' X_{it}]] \Leftrightarrow \log[h_{it}/(1-h_{it})] = \beta' X_{it}$$

The probit specification for the hazard function is:

$$\Phi^{-1}(h_{it}) = \beta' X_{it}$$

where $\Phi^{-1}(h_{it})$ is the inverse cumulative distribution function of a standard normal variable.

As the true hazard function is not known, there is no guarantee that any of the specifications represent the true specification of the hazard.

3.4. Unobserved Heterogeneity and its Consequences

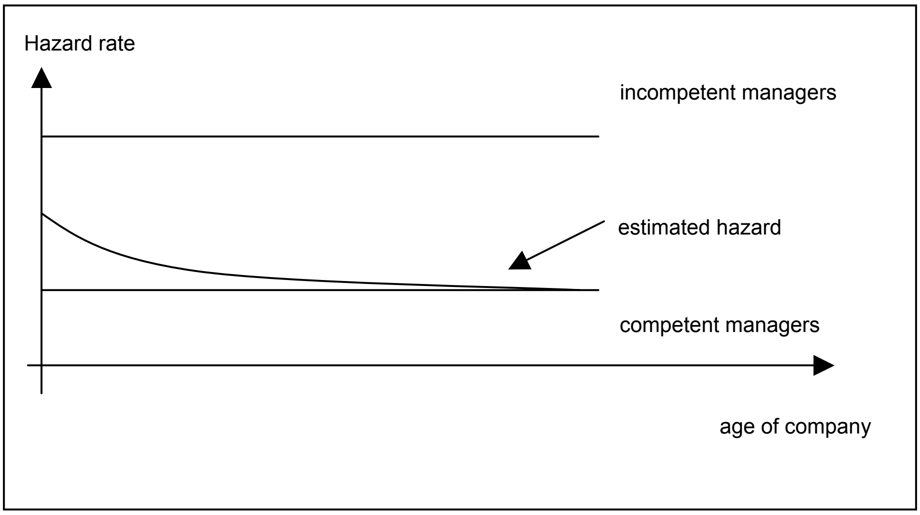
The hazard models presented above assumes that if two firms have identical values on the covariates, they also have identical hazard functions, that is, all differences between firms are assumed to be captured using observed explanatory variables, or, in other words, unobserved heterogeneity is assumed away. In Dyrberg (2004) it is argued that the assumption perhaps is more reasonable in estimations, which uses the data set presented and used in Dyrberg (2004) (as well as here) than in most cases, due to the fact several proxies are used for the variables that are inherently unobservable.

The probit and the logit specification for the hazard function are extended to also include unobserved heterogeneity. The estimation with unobserved heterogeneity can be seen as a specification check. The only other studies we know of, which models unobserved heterogeneity in a credit-scoring setting, are Kaiser (2001), Bernhardsen (2001) and Jones and Hensher (2004).

The presence of unobserved heterogeneity can cause several problems. The most serious problem is that the unobserved heterogeneity tends to produce estimated hazard functions that decline with age even when the true hazard is not declining for any individual in the sample (Kiefer (1988:671f)). Figure 3.4 sketches the effect of unobserved heterogeneity. The figure illustrates a case where all firms can be divided into two groups. One group consists of firms with a high and constant hazard (and with incompetent managers), and the other group consists of firms with a low and constant hazard (and with competent managers). When estimating the hazard function for the whole sample, the empirical hazard function starts out midway between the two hazards and then declines until it approaches the lowest hazard as an asymptote. The estimated hazard captures a composition effect (the remaining sample is increasingly made up by firms with low hazards). The estimated hazard rate declines, even though the true hazards are constant.

A natural question is whether or not one can rely on the estimates of β in the presence of unobserved heterogeneity (see Allison (2001)). If the *unobserved components are correlated* with the measured covariates, then the coefficients may be severely biased. If the *unobserved components are independent* of the measured covariates (random effects) then the unobserved heterogeneity tends to attenuate the estimated coefficients towards 0. Both situations are undesireable. Nonetheless, the last case is “better” than the former, as we at least know, how the unobserved heterogeneity affects the parameter estimates.

Figure 3.4: Unobserved heterogeneity (composition effect)



3.5. Expression for the Hazard Rate with Unobserved Heterogeneity

The specifications of the hazard functions are extended to incorporate unobserved heterogeneity (denoted u_i). The probit specification for the hazard function is now:

$$\Phi^{-1}(h_{it}) = \beta' X_{it} + u_i ,$$

where $\Phi^{-1}(h_{it})$ is the inverse cumulative distribution function of a standard normal variable.

The logit specification for the hazard function is now:

$$\log[h_{it}/(1-h_{it})] = \beta' X_{it} + u_i \,.$$

In the specifications the unobserved components are assumed to be uncorrelated with the measured covariates (random effects model) and the censoring times.

The estimation strategy is more complicated compared to the situation without unobserved heterogeneity, as there are as many firm specific effects as there are firms in the data set, and so there are not enough degrees of freedom left to fit these parameters. The way to get around the problem is to assume that the distribution of *u* has a shape whose functional form is summarized in terms of only a few key parameters. The way to proceed is then to write the likelihood function so that it refers to the distributional parameter (rather than each *u*). This method, which is known as ”integrating out the individual random effect”, is feasible because *u* is independent of the covariates and the censoring times. Details and derivations are found in Wooldridge (2002) and Greene (2003).

The importance of unobserved heterogeneity can be tested using a likelihood ratio test. *u* is assumed to follow a normal distribution, and the null hypothesis is, that the ratio of the heterogeneity variance to one plus the heterogeneity variance is equal to zero, i.e. that there is no cross-period correlation. If the null hypothesis that the ratio is zero cannot be rejected then unobserved heterogeneity is unimportant.

The way unobserved heterogeneity is tested for can be criticized. The specifications hinge on distributional assumptions. If they are not correct, one might incorrectly find that unobserved heterogeneity is important (or that it is not important). The approach is a parametric approach. Heckman and Singer (1984) have suggested a non-parametric approach to characterize the distribution of the unobserved heterogeneity. The idea is that one fits an arbitrary distribution using a set of parameters, and that these parameters comprise a set of “mass points” and the probabilities of a specific firm being located at each mass point. This should make the choice of distributional shape for *u* less arbitrary. Nonetheless, it is not clear how many mass points to fit. Usually one considers two mass points, but there are often no a priori reasons to believe that there will not be more mass points than two.

3.6. Specification Tests

The specifications for the hazard function can be tested using the Davidson and MacKinnon (1993:492) test-procedure and the Silva (2001) test-procedure. Using these tests it is investigated whether the logit or the probit specification has the best fit. This is done in section 4. This section explains how the tests work. To the best of our knowledge this is the first time, these tests have been used within the credit-scoring literature. As is mentioned above, a number of studies use either the logit or the probit model, however, the only other study, which to the best of our knowledge estimates credit-scoring models using both the logit and the probit model in order to compare the results using the two specifications, but not to perform any tests, is Lennox (1999). In section 6, where all the results from the comparison of the hazard models are discussed, we will return to the outcome of his investigations and compare them with the outcome of our tests.

Davidson and MacKinnon (1993:492) explain in their textbook how one can test non-nested models such as the logit and the probit. The test procedure is sketched here using an example: Let the model with the logit specification for the hazard function have a maximized value of the log likelihood function of -16,300 ("the worst fitting model"), and the model with the probit specification for the hazard function have a maximized value of the likelihood of -16,100 ("the best fitting model"). The null hypothesis is that the hazard follows the worst fitting model. The alternative hypothesis is that the hazard follows a "larger model", which is a weighted average of the two models. This means that the larger model must fit at least as well as whichever of the two models that fits best, since the larger model can choose to weight the worst fitting model with zero. Compute the likelihood ratio statistic, which in this case is

2(-16,100-(-16,300)) = 400.

The likelihood-ratio test is distributed $\chi^2(1)$. If the null is rejected, the test rejects that the hazard of financial distress follows the worst fitting model. Since 400 clearly exceeds the 5% critical value for a one-degree-of-freedom test, the worst fitting model is rejected at some level smaller than 5% if it is tested against the larger model, and so it is concluded that the model with the probit specification has a better fit in this example. Note that the worst fitting model is tested against the larger model, even though it is not estimated. The test does not say anything about the best fitting model, which might be rejected too, if one were to test it against the larger model.

Silva (2001) suggests another test for non-nested models, which is also discussed in Greene (2003:682ff). Let P_1 and P_2 denote the probability of an event (given the explanatory variables) under the models defined by L_1 and L_2, e.g. the logit and the probit model, respectively. Let Model 1 denote the null specification and Model 2 denote the alternative. The likelihood of a “super model”, which combines the two alternative binary choice models, can be written using two mixing parameters α and ρ, c.f. Silva (2001) and Greene (2003:682) (the likelihood, which depends on both α and ρ, is not written here). The score test, which Silva (2001) derives, is the test for the hypothesis that $\alpha = 0$ for any particular value of ρ.

Silva (2001) suggests to set up the test in the following way for binary choice models. First parameter estimates of the competing models are computed by maximum likelihood and predicted probabilities for the events are computed. Then the variable $z(\rho)$ is calculated for the null model. $z(\rho)$ is defined as:

$$z(\rho)=\left[\frac{\left(\frac{P_2}{P_1}\right)^{\rho}-1}{\rho}-\frac{\left(\frac{1-P_2}{1-P_1}\right)^{\rho}-1}{\rho}\right]\frac{P_1[1-P_1]}{P_1^*}$$

where

P_1 and P_2 denote the probability of an event (given the explanatory variables) under the models defined by L_1 and L_2, and where P_1^* denotes the derivative of $P(y=1|X)$ with respect to the index.

In the limit, when $\rho \to 0$ the formulae for $z(\rho)$ reduces to

$$z(0)=\ln\left[\frac{P_1(1-P_2)}{P_2(1-P_1)}\right]\frac{P_1[1-P_1]}{P_1^*}$$

In the limit, when $\rho \to 1$ the formulae reduces to

$$z(1)=\frac{(P_2-P_1)}{P_1^*}$$

A priori it is not possible to know which value of ρ will lead to the test with the best performance. Silva (2001:580) notes, that (in many cases) setting $\rho=0$ or $\rho=1$ can be based on computational convenience. Using either $z(0)$ or $z(1)$ the two models can be tested against each other, when Model 1 is re-estimated with $z(\rho)$ as an additional variable. If the coefficient to $z(\rho)$ is not significantly different from zero, when the model is re-estimated, then the null hypothesis, that $\alpha=1$, is accepted, and Model 1 is favored. A rejection of the null hypothesis favors Model 2.

To conclude, this section presented two tests for non-nested models, such as the logit and the probit model. These tests are performed and the outcome of the tests is reported in section 4.

4. Results

This section reports the results from the estimations of the various specifications of the hazard models. First it is tested whether the logit or the probit specification for the hazard function has the best fit. Then it is investigated whether or not unobserved heterogeneity plays a role. Finally, the parameter estimates of the preferred specification are presented and discussed. All results are obtained from the estimations of 1) a model, where the event is financial distress and where firms that exit for other reasons than financial distress are treated as censored or no longer observed, and from the estimations of 2) a model, where the event is exit (i.e. not split up on exit type).

4.1. The Specification of the Hazard Function

In this section the specification of the hazard function is tested. First, it is tested whether the preferred specification of the hazard function is the logit or the probit specification, when the event is the financial distress event (i.e. the E1 event). Then it is tested whether the preferred specification of the hazard functions is the logit or probit specification, when the event is the pooled exit event.

The relevant calculations for the Davidson and MacKinnon test are:

E1 event (logit or probit):

Likelihood ratio test = 2("the worst fitting model" – "the best fitting model")

= 2 (value of the maximized log likelihood function of the logit model – value of the maximized log likelihood function of the probit model)

= 2(-10,461-(-10,585)) = 248.

The null hypothesis is that the hazard follows the worst fitting model, which in this case is the logit model. The alternative hypothesis is that the hazard follows a "larger model", which is a weighted average of the two models. The likelihood-ratio test statistic is 248. It is distributed $\chi^2(1)$. Since 248 clearly exceeds the 5% critical value for a one-degree-of-freedom test, the worst fitting model is rejected at some level smaller than 5% if it is tested against the larger model, and so it is concluded that the model with the probit specification has a better fit.

For the pooled exit event, in the same way, it is found that the model with the probit specification has a better fit.

To sum up, according to the Davidson and MacKinnon (1993:492) test-procedure the best specification for the pooled exit event as well as the E1 event is the probit specification, c.f. table 4.1.a.

The relevant calculations for the Silva test are:

E1 event (logit or probit):

As a start the logit model is taken as Model 1 and the probit model as Model 2. The result of the test is that $z(\rho)$ is significantly different from zero, and so Model 2 is favored (the probit model). Then the probit model is taken as Model 1 and the logit model as Model 2. The result of this test is that $z(\rho)$ is significantly different from zero, and so Model 2 is favored (the logit model).

For the model estimated with the pooled exit event, in the same way, it is not possible to conclude whether the logit or the probit specification has the best fit.

To sum up, based on the Silva (2001) test, it is not possible to conclude, which specification of the hazard function is the preferred specification, c.f. table 4.1.b.

All in all, the Davidson and MacKinnon test showed that the probit model is preferred over the logit model, however, the Silva (2001) test did not give any guidance on what model to choose. The outcome of these tests (along with the other results obtained in this paper) is discussed in further details in section 6.

Table 4.1.a: The preferred specification for the hazard function: Davidson and MacKinnon test

Data treatment	The specification of the hazard function and the resulting value of the maximized log likelihood function	The preferred specification for the hazard function
E1 exit is modelled	Logistic: -10,585	The best fit: Probit
	Probit: -10,461	
E1, E2 and E3 treated as one exit type	Logistic: -18,682	The best fit: Probit
	Probit: -18,638	

Table 4.1.b: The preferred specification for the hazard function: Silva test

Data treatment	The null model is the logit model	The null model is the probit model	The preferred specification for the hazard function
E1 exit is modelled	The coefficient to $z(1)$ is 2.4761. The coefficient is significantly different from zero at the 1 pct. significance level, and so the alternative model is favoured (the probit)	The coefficient to $z(1)$ is -2.1879. The coefficient is significantly different from zero at the 1 pct. significance level, and so the alternative model is favoured (the logit)	The best fit: Logit or Probit?
E1, E2 and E3 treated as one exit type	The coefficient to $z(1)$ is 2.0710 The coefficient is significantly different from zero at the 1 pct. significance level, and so the alternative model is favoured (the probit)	The coefficient to $z(1)$ is -0.7839. The coefficient is significantly different from zero at the 1 pct. significance level, and so the alternative model is favoured (the logit)	The best fit: Logit or Probit?

4.2. Testing for Unobserved Heterogeneity

Unobserved heterogeneity can be important in specifications where proxies are not used for inherently unobserved variables. An example where things could go wrong, if proxies are not included, is sketched in Rommer (2005a), which uses the same data set as us in her estimations: "Say, for example, that some of the companies in the sample are willing to take a lot of chances and engage in risky investment projects. If no proxies are included for these firms, then these firms cannot be distinguished from other firms, which have the same levels of their explanatory variables as the risky firms have. When a negative shock is hitting, the problem is then, that a larger number of the risky firms are likely to enter financial distress compared to other firms. Since riskiness could be correlated with explanatory variables that are included, the parameter estimates on the latter are likely to be inconsistent, as they will then be correlated with the error term, which includes information on whether the firm is risky or not. The above situation is usually a problem in credit-scoring studies, which usually do not include proxies for inherently unobservable variables. Here the situation, which is sketched with the above example, is less of a problem, as we have included a number of proxies in the estimations. In connection to the above situation, two of the very important proxies are a dummy, which measures whether or not "illegal loans have been adopted", there are "inconsistencies in the profit and loss account", "the financial statement is incomplete" etc. and a dummy, which indicates whether or not the company is a private or public limited liability company. These proxies indicate the willingness to take on risk, the ability of the entrepreneur etc. and do as such indicate something about whether or not the firm is likely to engage in risky activities." Note also, as Rommer (2005a) points out, "that Jenkins (2003:102) notes, that the effects of unobserved heterogeneity are mitigated, and thence estimates are more robust, if a flexible baseline-hazard specification is used (as it is in this case), and that the topic of unobserved heterogeneity underscores the importance of getting good data, including a wide range of explanatory variables that summarize well the differences between, in this case, the firms."

Even though we do not suspect to find that unobserved heterogeneity is important (for the above mentioned reasons), the Davidson and MacKinnon (1993:492) test-procedure discussed above is used to test whether or not unobserved heterogeneity is important, c.f. table 4.2. The conclusion is that the probit model without unobserved heterogeneity does the best job, and so unobserved heterogeneity seems not to be important.

The following observations point to the conclusion:

E1 event (with and without unobserved heterogeneity):

There is no difference between the estimation of the probit model with and without unobserved heterogeneity, as the maximized log likelihoods in both cases yield the same result. This means that unobserved heterogeneity is not important.

The tests of the probit model without unobserved heterogeneity against the logit model with unobserved heterogeneity and the logit model without unobserved heterogeneity, respectively, are based on the following calculations:

Likelihood ratio test of probit model without unobserved heterogeneity against logit model with unobserved heterogeneity = 2("the worst fitting model" – "the best fitting model")

= 2(value of the maximized log likelihood function of the logit model with unobserved heterogeneity – value of the maximized log likelihood function of the probit model)

= 2(-10,579-(-10,461)) = 236.

The null hypothesis is that the hazard follows the worst fitting model. The alternative hypothesis is that the hazard follows a "larger model". The worst fitting model is rejected at some level smaller than 5% if it is tested against the larger model, and so it is concluded that the model with the probit specification has a better fit.

Likelihood ratio test of probit model without unobserved heterogeneity against logit model without unobserved heterogeneity is done in section 4.1.

The conclusion from the test is that the model with the probit specification has a better fit.

Pooled exit event (with and without unobserved heterogeneity):

As for the E1 event, there is no difference between the estimation of the probit model with and without unobserved heterogeneity, as the maximized log likelihoods in both cases yield the same result. This means that unobserved heterogeneity is not important. The same result holds for the logit model, when the pooled exit event is estimated, and so it is only necessary to compare the probit and logit model without unobserved heterogeneity. This was done in section 4.1. The result is that the probit specification has a better fit.

To conclude, the overall result is that unobserved heterogeneity seems not to be important. The hazard specification, which has the best fit, no matter whether we model firms in financial distress or exits in general, is the probit specification. Further comments to the result are provided in section 6.

Table 4.2: The preferred specification for the hazard function

Data treatment	The specification of the hazard function and the resulting value of the maximized log likelihood (without and with unobserved heterogeneity)	The preferred specification for the hazard function
E1 exit is modelled	Logistic (without): -10,585 Logistic (with): -10,579	The best fit: Probit (without)
	Probit (without): -10,461 Probit (with): - 10,461*	
E1, E2 and E3 treated as one exit type	Logistic (without): -18,682 Logistic (with): -18,682*	The best fit: Probit (without)
	Probit (without): -18,638 Probit (with): -18,638*	

Note: * indicates that unobserved heterogeneity is unimportant in these specifications. In these cases, the maximized value of the log likelihood function is equal to the models without unobserved heterogeneity.

4.3. Parameter Estimates

The parameter estimates that are obtained when the pooled exit event and the E1 event are estimated using the probit specification without unobserved heterogeneity for the hazard function, respectively, are reported in figure 4.3 and tables 4.3.a ("core variables") and 4.3.b ("proxies").[2] It is chosen to report the significance, sign and magnitude of the coefficients. One could also have calculated the change in the probability of financial distress or the probability of exit with respect to one of the right-hand-side variables, but this is not done here (the way to do this is discussed in Johnston and DiNardo (1997:422)).

Models with the pooled exit event and the E1 event are also estimated using the logit specification for the hazard function. The significant parameter estimates, that are obtained when these events are estimated, have the same sign, as when the probit specification for the hazard function is used (i.e. as reported in the tables).

The results can be grouped into four groups: 1) variables that are significant in one specification (e.g. the E1 event model) and insignificant in the other (e.g. pooled exit event model) (the dummy for being a private limited liability company, owned by the public and the ultimate parent company dummy), 2) variables that are significant in both cases, but with different signs (dummy for being a wholly owned subsidiary), 3) variables that are significant in both cases and with the same sign (but sometimes of quite different magnitude) (e.g. solvency ratio and short-term debt to total assets) and 4) variables that are insignificant in both specifications (owned by a fund and concentration index).

[2] The results, which are reported, are the results of the specifications, where insignificant age dummies are tested away.

Table 4.3.a: Pooled exit versus the E1 event: Core Variables

Variables	Expected Effect	E1 event	Pooled exit
Firm Age (dummies)	Negative	See figure 4.3	See figure 4.3
Short-term debt to total assets	Positive	0,2150*	0,3508*
Return on net assets	Negative	-0,6910*	-0,7873*
Solvency ratio	Negative	-1,1831*	-0,2010*
Firm size	Negative	-0,0612*	-0,0532*

Note: The parameter estimates, that are obtained when the E1 event and the pooled exit event are estimated, are reported. The dummy for publicly traded companies is not included in the estimations as no publicly traded firm enters financial distress. Concerning the controls: There is controlled for the macroeconomic environment and for the various sectors. In the estimations, farming and forestry is included in the same sector affiliation category, as the data were too sparse otherwise. The same is true for mining, energy and construction. The primary bank categories have been altered: As the data was too sparse otherwise, firms that register a bank in category 3 or 4 are now in the same group. * indicates that the variable is significant at the 1 pct. level. ** indicates that the variable is significant at the 5 pct. level.

Table 4.3.b: Pooled exit versus the E1 event: Proxies

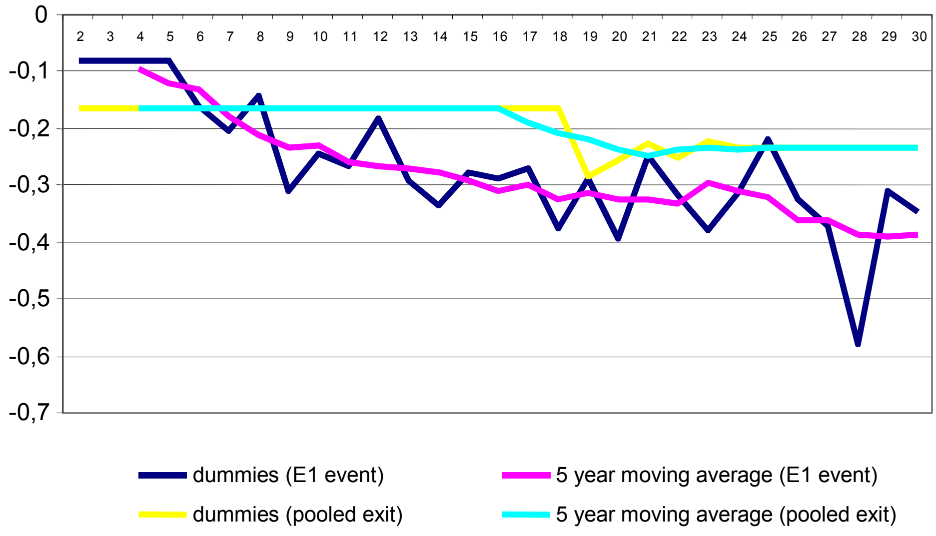
Variables	Expected Effect	E1 event	Pooled exit
Owned by the public	?	0,0737 (not sign.)	0,4161*
Owned by a fund	?	-0,5014 (not sign.)	-0,1344 (not sign.)
Diversification 2 sectors (related business) (dummy)	Negative	-0,1434*	-0,1854*
Diversification 3–9 sectors (related business) (dummy)	Negative	-0,1520*	-0,2790*
Diversification 2 sectors (unrelated business) (dummy)	Negative	-0,0843*	-0,1324*
Diversification 3–9 sectors (unrelated business) (dummy)	Negative	-0,1520*	-0,2022*
Local authority group 1 (reference category)			
Local authority group 2 (dummy)	?	-0,1497*	-0,0720*
Local authority group 3 (dummy)	?	-0,1108*	-0,1209*
Local authority group 4 (dummy)	?	-0,0712**	-0,0702*
Local authority group 5 (dummy)	?	-0,1006*	-0,1178*
Concentration	?	-0,0017 (not sign.)	0,0033 (not sign.)
Critical comments from the auditors (dummy)	Positive	0,4846*	0,3991*
Ultimate parent companies (dummy)	?	0,2008*	0,0547 (not sign.)
Wholly owned subsidiaries (dummy)	Negative	-0,1281*	0,2144*
Private limited liability company (dummy)	Positive	0,1707*	0,0283 (not sign.)
Public limited liability company (reference category)			

Note: For further details, see the note to table 4.3.a.

The effects of the controls differ (results are not reported). When estimating the E1 event the dummy for reporting a group 2 bank was significant and positive, and when estimating the pooled exit event the dummy was not significant. The dummies on the different sector affiliation categories also have different effects. When estimating the E1 event the dummies on trade and hotel, transport, business service, public service activities and organizations were significant and negative, whereas the dummy unknown was significant and positive. When estimating the pooled exit event only two dummies were significantly different from zero (and positive), namely the dummies unknown and not stated.

The overall conclusion from the comparison of parameter estimates is that they differ to quite an extent depending on which event is modelled. Therefore it is important to think careful about the specification of the model in order not to mix "apples and pears". Further comments to the results are provided in section 6.

Figure 4.3: The pooled exit event and the E1 event: Duration dependence



Note: Reference category: 1 year old firms. The last dummy is called 30 years old or more. Most dummies are significant at the 1 pct. level. The exceptions are: 1) When estimating the E1 event the dummy for age 2 – 5 is significant at the 5 pct. level. 2) When estimating the pooled exit event the dummy for age 2 – 18 and the dummy for age 21 are significant at the 5 pct. level.

5. Prediction

The predictive abilities of the model set-ups are evaluated in this section. There exists a whole literature on validation of credit-scoring models, including how to asses the discriminatory power of credit-scoring models, see e.g. BCBS (2005) for a good overview, or sources such as Sobehart, Keenan and Stein (2000), Stein (2002), Engelmann, Hayden and Tasche (2003a) and Engelmann, Hayden and Tasche (2003b). We are presenting the Receiver Operating Characteristics curve (ROC curve) and the Accuracy ratio, as these are two of the most commonly used measures, and as the Validation Group[3], which is the author of BCBS (2005), has

[3] The Validation Group is a subgroup under the Research Task Force formed by the Basel Committee on Banking Supervision. The Validation Group was established in anticipation of the need for more knowledge regarding validation methodologies. The Validation Group consists of representatives from eleven countries.

found, that these two measures appear more meaningful than the other measures, which they present.

Before the measures are presented, first, the notion of type I and type II errors is explained. The important thing to notice is that when a statistical credit-scoring model, such as a hazard model, is used for prediction, it assigns a probability of default (PD) to all the firms in the sample. The PD varies between 0 and 1, and so the model builder has to decide when to declare a firm as an event (= financial distress) and when to declare it as a non-event (e.g. active firm). If the sample, that the model builder uses, contains en equal amount of events and non-events, it would seem natural to choose a cut-off of 0.5, i.e. to predict all firms with a PD above or equal to 0.5 as events and to predict all firms with a PD below 0.5 as non-events. However, if the sample is skewed (in the sense, that it entails a larger number of non-events compared to events), as it most often is, the naïve cut-off level of 0.5 is often modified, e.g. to reflect the proportion of events over non-events. If the model builder did not modify the cut-off level, he or she would only predict very few events as events, and so the number of type I errors (missing prediction) would be very high. With a modified cut-off level, which is lower than 0.5, the number of type I errors would decrease, but this would be at the cost of an increased number of type II errors (false alarms), i.e. a larger number of firms would be predicted to be events, but would be non-events. By adjusting the cut-off level up and down, the model builder can adjust the number of type I and type II errors. The adjustment of the cut-off level will always be at the cost of one of the two types of errors: If type I errors decreases, then type II errors increases and vice versa.

It is very important that the cut-off level that is used by the model builder reflects his or her assessment of the cost of making type I and type II errors, respectively. Because of the trade-off between incorrectly classifying a firm that does not exit because of financial distress as a financially distressed firm (type II error corresponding to the wrong signal) compared to not classifying a financially distressed firm as financially distressed (type I error corresponding to a missing prediction), the user of the model has to assess how bad it is to incorrectly classify a firm that does not exit because of financial distress as a financially distressed firm compared to not classifying a financially distressed firm as financially distressed. A few papers discuss the costs of errors in lending. Altman (1980) investigates how to specify the cost of lending errors for commercial banks and how to more accurately specify the optimal cut-off-score approach to credit-scoring, Weiss and Capkun (2004) compares the prediction of different models based on the net profit each would generate and Stein (2005) shows how the simple cut-off approach can be extended to a more complete pricing approach, and he demonstrates that more powerful models are more profitable than weaker ones.

The beauty of the Receiver Operating Characteristics curve (ROC curve), which we will use as one of our measures for the discriminatory power of the various models, is that it does not depend on the chosen cut-off level (see figure 5.b for the ROC curves generated by prediction with our models). Instead the curve depicts, for all cut-off levels, the type II errors (missing prediction) on the x-axis and the hit rate (correctly called events) on the y-axis. A credit-scoring model performs better the closer the ROC curve is to the upper left hand corner, i.e. to the point (0,1). This point corresponds to a perfect fit, which is indicated by no false alarms and a hit rate equal to 1. Two or more credit-scoring models can be compared and assessed using ROC curves. The credit-scoring model, which produces the curve, which is to the left of the curve of the other model, has the best fit.

One of the summary indices of ROC, the ROC measure (or Area Under the Curve, AUC) is an indicator of the quality of a rating model. It has been shown to be a linear transformation of the Accuracy Ratio (see e.g. Engelmann, Hayden and Tasche (2003b)). We will now introduce the Accuracy Ratio.

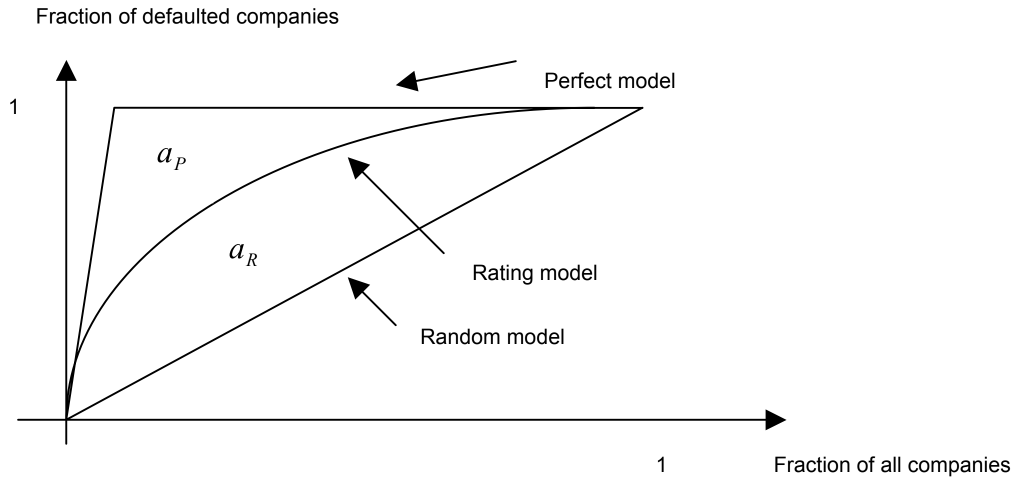
The Accuracy Ratio, which is also called the Gini-coefficient, is a summary index of the Cumulative Accuracy Profile (CAP), which is also known as the Gini curve, the power curve or the Lorenz curve. To obtain the CAP curve, which is illustrated in figure 5.a, all debtors are first ordered by their respective scores from riskiest to safest, that is, from the debtor with the lowest score to the debtor with the highest score. The CAP curve is then determined by plotting the cumulative percentage of all borrowers on the x-axis (and so the x-axis measures the fraction of borrowers with a lower-than-specified score within all defaulters) and the cumulative percentage of all defaulters on the y-axis. The quality of a rating system is measured by the accuracy ratio AR. It is defined as the ratio of the area a_R between the CAP curve of the rating model, which is validated, and the diagonal (the random model), and the area a_p between the CAP curve of the perfect rating model and the CAP curve of the diagonal (the random model), that is:

$$AR = \frac{a_R}{a_p} .$$

A rating system is more accurate the closer the AR is to one.

The statistical properties of the ROC measure (and therefore also for the AR measure) can also be used as the point of departure for a formal test (which compares the ROC measure of a rating system with that of a random rating and for comparing two or more rating systems), c.f. Engelmann, Hayden and Tasche (2003a).

Figure 5.a: Cumulative Accuracy Profile



The ROC curves and the Accuracy Ratios for our estimated models are presented in table 5 and figure 5.b. The table and the figure show that the two measures are very close to each other. In fact, there are virtually no differences between estimating the E1 event with the probit and the logit specification for the hazard function, just as well as there are virtually no differences between estimating the pooled exit event with the probit and the logit specification for the hazard function. To sum up, based on the predictive ability of the models, there is hardly any difference between estimating a hazard model with a probit specification for the hazard functions and a hazard model with the logit specification for the hazard function.

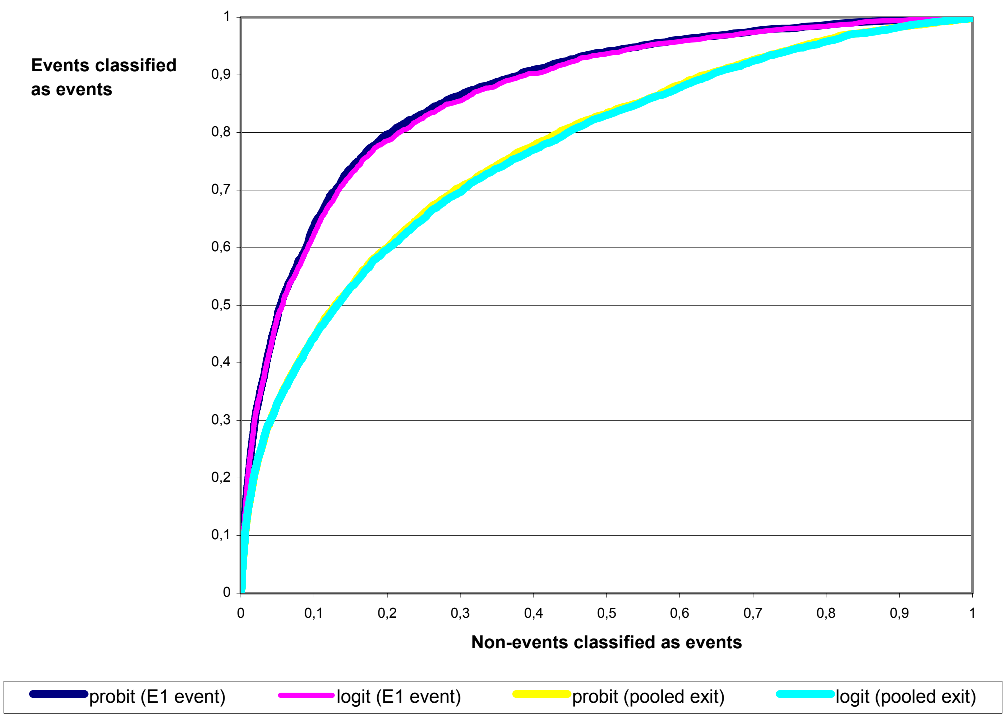
There is, however, a big difference between the predictive ability of the models that estimate the E1 event versus the models that estimate the pooled exit event. Figure 5.b and table 5 show that the models, which model the pooled exit event and the E1 event, respectively, generate quite different predictions. The assumptions behind the E1 event model leads to the highest proportion of correct predictions at all points of the curve (that is, for all cut-off values) for both the probit and the logit specification, and so the overall conclusion is that depending on the event, which is modelled, the predictions vary to quite an extent.

Further comments to the results are provided in section 6.

Table 5: Area under the Receiver Operating Characteristics curve (ROC curve) and the Accuracy Ratio

	E1 event		Pooled exit	
	Probit	Logit	Probit	Logit
Area under ROC curve	0.870	0.865	0.771	0.768
Accuracy Ratio = 2*AUC-1	0.74	0.73	0.542	0.536

Figure 5.b: ROC curves: The logit and the probit model



6. Conclusion

The purpose of this paper is to discuss a number of issues that are relevant when setting up a credit-scoring model and to test the assumptions used in accounting-based credit-scoring models. A non-standard comparison of two hazard models with differently specified hazard functions is made: one with a logit specification and the other with a probit specification. The probit and the logit specification for the hazard function are extended to also include unobserved heterogeneity. The estimation with unobserved heterogeneity can be seen as a specification check. In addition to the investigations of the various specifications of the hazard function the consequences of different treatment of different types of exits is investigated. It is shown what the consequences are of setting up 1) a hazard model where the event financial distress is modelled and where firms that exit for other reasons than financial distress are treated as censored or no longer observed and 2) a hazard model where the general exit event is modelled (i.e. not split up on exit type).

The overall conclusions are that 1) there does not seem to be any major difference between the logit and the probit specification, 2) that unobserved heterogeneity seems to be unimportant, but 3) that the results differ depending on the event, which is modelled (financial distress versus pooled exits). The observations leading to these conclusions are the following:

The first result, namely, that there does not seem to be any major difference between the hazard model with the logit and the probit specification for the hazard function, is based on several observations. As a start, the two specifications were tested against each other using the Davidson and MacKinnon-test (1993) and the Silva (2001) test. The Davidson and MacKinnon-test (1993) showed that the probit model is preferred over the logit model, and the Silva (2001) test did not give any guidance on what model to choose. Despite the result of the Davidson and MacKinnon-test (1993), but along the lines of the result of the Silva (2001) test, the parameter estimates, which are obtained when the hazard model is estimated with the logit and the probit specification for the hazard function, respectively, are compared and show that the two specifications deliver significant parameters, which have the same sign (both when the financial distress event and the pooled exit event is modelled). Furthermore, the predictive abilities of the two model set ups are very alike, when the financial distress event is modelled and the hazard specification is either logit or probit, and when the pooled exit event is modelled and the hazard specification is either the logit or the probit specification. In fact, the predictive ability of the hazard model with the logit and the probit specification for the hazard function are almost identical.

The overall conclusion from the investigations is therefore that even in our case (where we are in the tails of the distribution and where the sample contains few responses to non-responses) it is difficult to provide generalities on which model to choose. Despite the fact that our tests gave conflicting results, the full analysis (which includes the tests, the estimated parameter estimates and the predictive abilities of the models) confirms that it is difficult to distinguish between the logit and the probit specification for the hazard model, and so the main conclusion is, that there does not seem to be a big difference between the logit and the probit specification. This result is in line with Lennox (1999). He does not propose to test the two specifications against each other. Instead he concludes that the results for the two models “are very similar, indicating that there is little to choose between the probit and logit approaches” (Lennox (1999:355)). In addition to Lennox (1999), we are aware of no other credit-scoring study, which compares estimation results from a logit and a probit model.

The second result is that unobserved heterogeneity seems not to be important. This result is obtained using the Davidson and MacKinnon (1993) test, when the probit and the logit specification for the hazard function are extended to also include unobserved heterogeneity. The estimation with unobserved heterogeneity can be seen as a specification check. We did not expect to find that unobserved heterogeneity is important. Unobserved heterogeneity can be important in specifications where proxies are not used for inherently unobserved variables, but as we discussed, a number of proxies are used in this paper. Furthermore, a flexible baseline-hazard specification is used, which should also mitigate the effects of unobserved heterogeneity, c.f. Jenkins (2003:102). The only other studies we know of, which models unobserved heterogeneity in a credit-scoring setting, are Kaiser (2001), Bernhardsen (2001) and Jones and Hensher (2004). The result in this paper differs from the result in Jones and Hensher (2004), who find that unobserved heterogeneity is important in their specification, but is along the lines of Kaiser (2001) and Bernhardsen (2001), who find that unobserved heterogeneity is not important in their specifications.

The third result is that the findings differ depending on the event, which is modelled (financial distress versus pooled exits). This is the case for the estimated parameters as well as for the predictive abilities of the models (no matter whether the specification for the hazard function is the logit or the probit specification). In this way the results highlight that it is important to think careful about the specification of the model in order not to mix "apples and pears". Recent papers within the credit-scoring and industrial organization literature still do not distinguish between exit types. Examples are Bunn and Redwood (2003), Pérez, Llopis and Llopis (2004), Mata and Portugal (2002) and Kimura and Fujii (2003).

The practical implication of the paper is that it is important to think careful about the specification of credit-scoring models. Here the specification issues are highlighted and investigated using an extensive data set on Danish non-financial sector firms. The preferred specification in this set up is a hazard function with either the logit or the probit specification. Unobserved heterogeneity seems not to be present. However, it is important to think careful about the modelled event. It is crucial to understand that the results depend on the portfolio under consideration, and hence, that every model builder has to think careful about the issues. This paper provides a framework for such investigations.

7. LITERATURE

D'Addio, A. and M. Rosholm, 2002. *Left-Censoring in Duration Data: Theory and Applications*. Working Paper No. 2002-5, Department of Economics, School of Economics and Management, University of Aarhus

Allison, P. D., 2001. *Logistic Regression using the SAS System. Theory and Application.* Cary, NC, USA: SAS Institute

Altman, E. I., 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *Journal of Finance*, vol. 23, no. 4, pp. 589-609

Altman, E. I., Haldeman, R. G. and P. Narayanan, 1977. ZETA Analysis. A new model to identify bankruptcy risk of corporations. *Journal of Banking and Finance*, vol. 1, pp. 29-54

Altman, E. I., 1980. Commercial Bank Lending: Process, Credit Scoring, and Costs of Errors in Lending. *The Journal of Financial and Quantitative Analysis*, vol. 15, no. 4, pp. 813-832

Altman, E. I., G. Marco and F. Varetto, 1994. Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking and Finance*, vol. 18., pp. 505-529

Altman, E. I. and A. Saunders, 1998. Credit risk measurement: Developments over the last 20 years. *Journal of Banking and Finance*, vol. 21, pp. 1721-1742

Audretsch, D. and T. Mahmood, 1994. The rate of hazard confronting new firms and plants in US manufacturing. *Review of Industrial Organisation*, vol. 9, pp. 41-56

Audretsch, D. and T. Mahmood, 1995. New firm survival: new results using a hazard function. *Review of Economics and Statistics*, vol. 73, pp. 97-103

BCBS, Basel Committee on Banking Supervision, 2004. *International Convergence of Capital Measurement and Capital Standards. A Revised Framework.* Bank for International Settlements, June 2004

BCBS, Basel Committee on Banking Supervision, 2005. *Studies on the Validation of Internal Rating Systems.* Working paper no. 14, Basel Committee on Banking Supervision, February 2005

Back, B., Laitinen, T., Sere, K. and M. v. Wezel, 1996. *Choosing Bankruptcy Predictors Using Discriminant Analysis, Logit Analysis, and Genetic Algorithms.* Turke Centre for Computer Science, Technical Report No. 40

Balcaen, S. and H. Ooghe, 2004. *Alternative methodologies in studies on business failure: do they produce better results than the classical statistical methods?* Working Paper 2004/249, Faculteit Economie en Bedrijfskunde

Bardos, M., 2001. *Recent developments in the Banque de France's scoring method.* Banque de France Bulletin Digest, no. 93, September 2001

Beaver, W., 1966. Financial Ratios as Predictors of Bankruptcy. *Journal of Accounting Research*, vol. 6, pp. 71-102

Beaver, W., 1968. Market Prices, Financial Ratios, and the Prediction of Failure. *Journal of Accounting Research*, vol. 8, pp. 179-92

Begley, J., Ming, J. and S. Watts, 1996. Bankruptcy Classification Errors in the 1980s: An Empirical Analysis of Altman's and Ohlson's Models. *Review of Accounting Studies*, vol. 1, pp. 267-284

Betts, J. and D. Belhoul, 1987. The Effecitveness of incorporating Stability Measures in Company Failure Models. *Journal of Business Finance and Accounting,* vol. 14, pp. 323-334

Bernhardsen, E., 2001. *A Model of Bankruptcy Prediction.* Norges Bank Working Paper*,* ANO 2001/10

Bhattacharjee, A., C. Higson, S. Holly and P. Kattuman, 2004. *Business Failure in UK and US Quoted Firms: Impact of Macroeconomic Instability and the Role of Legal Institutions.* Unpublished manuscript, January 2004.

Borup, L., Kurek, D. and A. D. Rommer, 2005. *Assessing the consequences of Basel II: Are there incentives for cherry-picking when banks pool data across countries?* Working Paper no. 27, Danmarks Nationalbank

Bunn, P. and V. Redwood, 2003. *Company-accounts-based modelling of business failure and the implications for financial stability.* Working paper no. 210, Bank of England

Campbell, J. Y., Hilscher, J. and J. Szilagyi, 2005. *In Search of Distress Risk.* Unpublished manuscript

Chava, S. and R. A. Jarrow, 2004. Bankruptcy prediction with industry effects. *Review of Finance*, vol. 8, pp. 537-569

Chen, G. C. and T. Åstebro, 2001. *The Economic Value of Reject Inference in Credit Scoring*. Unpublished manuscript

Chen, G. C. and T. Åstebro, 2003. *Bound and Collapse Baysian Reject Inference When Data are Missing not at Random.* Unpublished manuscript

Cifarelle, D. M. and F. Corielli, 1988. Business Failure Analysis. A Baysian Approach With Italian Firm Data. *Studies in Banking and Finance,* no. 7, pp. 73 - 89

Corcóstegui, C., L. González-Mosquera, A. Marcelo and C. Trucharte, 2003. *Analysis of procyclical effects on capital requirements derived from a rating system.* Paper presented at Basel Committee/Banca d'Italia workshop, 20-21 March 2003

Crook, J. and J. Banasik, 2004. Does reject inference really improve the performance of application scoring models? *Journal of Banking & Finance*, vol. 28, pp. 857-874

Dambolena, I. G. and S. J. Khoury, 1980. Ratio Stability and Corporate Failure. *The Journal of Finance*, vol. XXXV, no. 4, pp. 1017-1026

Davidson, R. and J. G. MacKinnon, 1993. *Estimation and Inference in Econometrics.* New York: Oxford University Press

Dimitras, A. I., Zanakis, S. H. and C. Zopounidis, 1996. A survey of business failures with an emphasis on prediction methods and industrial applications. *European Journal of Operational Research*, vol. 90, pp. 487-513

Dimitras, A. I., Slowinski, R., Susmaga, R. and C. Zopounidis, 1999. Business failure prediction using rough sets. *European Journal of Operational Research*, vol. 114, pp. 263-280

Dyrberg, A., 2004. *Firms in Financial Distress: An Exploratory Analysis.* Working Paper no. 17, Danmarks Nationalbank

Engelmann, B., Hayden E. and D. Tasche, 2003a. Testing rating accuracy. *Risk*, January 2003

Engelmann, B., Hayden E. and D. Tasche, 2003b. *Measuring the Discriminative Power of Rating Systems.* Discussion paper series 2, Banking and Financial Supervision, no. 01/2003

Feelders, A. J., 2003. *An Overview of Model Based Reject Inference for Credit Scoring.* Unpublished manuscript

Frydman, H., Altman, E. I. and D. Kao, 1985. Introducing Recursive Partitioning for Financial Classification: The Case of Financial Distress. *The Journal of Finance*, vol. XL, no. 1, pp. 269-291

Greene, W., 1998. Sample selection in credit-scoring models. *Japan and the World Economy*, vol. 10, pp. 299-316

Greene, W., 2003. *Econometric Analysis.* New Jersey, USA: Prentice Hall

Grunert, J., Norden, L. and M. Weber, 2005. The role of non-financial factors in internal credit ratings. *Journal of Banking and Finance*, vol. 29, pp. 509-531

Harhoff, D., K. Stahl and M. Woywode, 1998. Legal Form, Growth and Exit of West German Firms: Empirical Results for Manufacturing, Trade and Service Industries. *Journal of Industrial Economics,* vol. 46, pp. 453-488

Heckman, J. and B. Singer, 1984. A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data, *Econometrica*, vol. 52, pp. 271-320

Hillegeist, S. A., Keating, E. K., Cram, D. P. and K. G. Lundstedt, 2004. Assessing the Probability of Bankruptcy. *Review of Accounting Studies*, no. 9, pp. 5-34

Honjo, Y., 2000. Business failure of new firms: an empirical analysis using a multiplicative hazards model. *International Journal of Industrial Organization*, vol. 18, pp. 557-574

Jenkins, S., 1995. Easy Estimation Methods for Discrete-Time Duration Models. *Oxford Bulletin of Economics and Statistics*, vol. 57, no. 1, pp. 129-138

Jenkins, S., 2003. *Survival Analysis.* Unpublished manuscript

Johnston, J. and J. DiNardo, 1997. *Econometric Methods*, fourth edition. New York: McGraw-Hill International Editions

Jiménez, G. and J. Saurina, 2004. *Collateral, type of lender and relationship banking as determinants of credit risk.* WP no. 0414, Banco de España

Jones, F. L., 1987. Current Techniques in Bankruptcy Prediction. *Journal of Accounting Literature*, vol. 6, pp. 131-164

Jones, S. and D. A. Hensher, 2004. Predicting Firm Financial Distress: A Mixed Logit Model. *The Accounting Review*, vol. 79, no. 4, pp. 1011-1038

Kaiser, U., 2001. *Moving in and out of Financial Distress: Evidence for Newly Founded Service Sector Firms.* Unpublished manuscript

Kiefer, N. M., 1988. Economic Duration Data and Hazard Functions. *Journal of Economic Literature*, vol. 26, pp. 646-79

Kimura, F. and T. Fujii, 2003. Globalizing activities and the rate of survival: Panel data analysis on Japanese firms. *Journal of Japanese International Economics*, vol. 17, pp. 538-560

Köke, J., 2001. *Determinants of acquisition and failure: Stylized facts and lessons for empirical studies.* Discussion Paper No. 01-30, ZEW

Lennox, C., 1999. Identifying Failing Companies: A Re-evaluation of the Logit, Probit and DA Approaches. *Journal of Economics and Business*, vol. 51, pp. 347-364

Lo, A. W., 1986. Logit versus Discriminant Analysis. A Specification Test and Application to Corporate Bankruptcies. *Journal of Econometrics*, vol. 31, pp. 151-178

Mata, J., Portugal, P. and P. Guimarâes, 1995. The survival of new plants: start-up conditions and post-entry evolution. *International Journal of Industrial Organization*, vol. 13, pp. 459-481

Mata, J. and P. Portugal, 2002. The survival of new domestic and foreign-owned firms. *Strategic Management Journal*, vol. 23, pp. 323-343

Moody's Investors Service, 2001a. *Moody's RiskCalc For Private Companies: Spain. Rating Methodology*, July 2001

Moody's Investors Service 2001b. *Moody's RiskCalc For Private Companies: France. Rating Methodology*, December 2001

Moody's Investors Service 2002. *Moody's RiskCalc For Private Companies: Italy. Rating Methodology*, October 2002

Ohlson, J. A., 1980. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, vol. 19, pp. 109-131

Pérez, S. E., Llopis, A. S. and J. A. S. Llopis, 2004. The Determinants of Survival of Spanish Manufacturing Firms. *Review of Industrial Organization*, vol. 25, pp. 251-273

Phillips, B. D. and B. A. Kirchhoff, 1989. Formation, Growth and Survival; Small Firm Dynamics in the U.S. Economy. *Small Business Economics*, vol. 1, pp. 65-74

Prantl, S., 2003. *Bankruptcy and Voluntary Liquidation: Evidence for New Firms in East and West Germany after Unification.* Discussion Paper No. 03-72, ZEW

Rommer, A. D., 2005 (a). *Firms in Financial Distress: An Exploratory Analysis.* Unpublished manuscript

Rommer, A. D., 2005 (b). *A Comparative Analysis of the Determinants of Financial Distress in French, Italian and Spanish firms.* Working Paper no. 26, Danmarks Nationalbank

Roszbach, K., 2003. *Bank Lending Policy, Credit Scoring and the Survival of Loans*. Working Paper no. 154, Sveriges Riksbank

Schary, M. A., 1991. The probability of exit. *RAND Journal of Economics*, vol. 22, no. 3, pp. 339-353

Shumway, T., 2001. Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *Journal of Business*, vol. 74, no. 1, pp. 101-124

Silva, J. M. C. S., 2001. A Score Test for Non-nested Hypotheses with Applications to Discrete Data Models. *Journal of Applied Econometrics*, vol. 16, pp. 577 - 597

Skogvsik, K., 1990. Current Cost Accounting ratios as Predictors of Business Failure: The Swedish Case. *Journal of Business Finance & Accounting*, vol. 17, no. 1, pp. 137-160

Sobehart, J. R., Keenan, S. C. and R. M. Stein, 2000. *Rating Methodology: Benchmarking quantitative Default Risk Models: A Validation Methodology.* Moody's Investors Service, March 2000

Stein, R. M., 2002. *Benchmarking Default Prediction Models. Pitfalls and Remedies in Model Validation.* Moody's KMV, Technical Report #030124

Stein, R. M., 2005. The relationship between default prediction and lending profits: Integrating ROC analysis and loan pricing. *Journal of Banking & Finance*, vol. 29, pp. 1213-1236

Tam, K. Y. and M. Y. Kiang, 1992. Managerial Applications of Neural Networks: The Case of Bank Failure Predictions. *Management Science*, vol. 38, no. 7, pp. 926-947

Weiss, L. A. and V. Capkun, 2004. *The Impact of Incorporating the Cost of Errors into Bankruptcy Prediction Models.* Unpublished manuscript

Wooldridge, J. M., 2002. *Econometric Analysis of Cross section and Panel data.* Cambridge (Mass.), USA: The MIT Press

Wooldridge, J. M., 2003. *Introductory Econometrics. A Modern Approach*. Ohio, USA: Thomson South-Western

Zavgren, C., 1983. The Prediction of Corporate Failure: The State of the Art. *Journal of Accounting Literature*, vol. 2, pp. 1-38

Zmijewski, M. E., 1985. Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, vol. 22, supplement 1984, pp. 59-82