SVERIGES RIKSBANK
WORKING PAPER SERIES 256

SVERIGES RIKSBANK

Taking the Twists into Account: Predicting Firm Bankruptcy Risk with Splines of Financial Ratios

Paolo Giordani, Tor Jacobson, Erik von Schedvin and Mattias Villani

NOVEMBER 2011

WORKING PAPERS ARE OBTAINABLE FROM

Sveriges Riksbank • Information Riksbank • SE-103 37 Stockholm
Fax international: +46 8 787 05 26
Telephone international: +46 8 787 01 00
E-mail: info@riksbank.se

Taking the Twists into Account: Predicting Firm Bankruptcy Risk with Splines of Financial Ratios*

Paolo Giordani† Tor Jacobson† Erik von Schedvin† Mattias Villani‡

Sveriges Riksbank Working Paper Series, No. 256

Revised January 30, 2013

Abstract

We demonstrate improvements in predictive power when introducing spline functions to take account of highly non-linear relationships between firm failure and leverage, earnings, and liquidity in a logistic bankruptcy model. Our results show that modeling excessive non-linearities yields substantially improved bankruptcy predictions, on the order of 70 to 90 percent, compared with a standard logistic model. The spline model provides several important and surprising insights into non-monotonic bankruptcy relationships. We find that low-leveraged as well as highly profitable firms are riskier than given by a standard model, possibly a manifestation of credit rationing and excess cash-flow volatility.

Keywords: bankruptcy risk model, micro-data, logistic spline regression, financial ratios

JEL: C41, G21, G33

*We are grateful for comments and suggestions from an anonymous referee, Mikael Carlsson, Hans Degryse, Tore Ellingsen, Vasso Ioannidou, Boyan Jovanovic, Jesper Lindé, Judit Montoriol-Garriga, Steven Ongena, Fabiana Penas, Karl Walentin, seminar and conference participants at the Stockholm School of Economics in Riga and the 2011 EC-squared meeting (Florence). We are also grateful for the generous data support provided by Upplysningscentralen AB. Parts of this research were carried out while Tor Jacobson was visiting the Einaudi Institute for Economics and Finance and Erik von Schedvin was visiting the Federal Reserve Bank of Philadelphia, and we gratefully acknowledge their hospitality. We thank Tommy von Brömsen, Mats Levander, and Matias Quiroz for excellent research assistance. We assume full responsibility for any and all errors in the paper. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the views of the Executive Board of Sveriges Riksbank.

†Research Division, Sveriges Riksbank. Email: firstname.lastname@riksbank.se. Phone: + 46 8 787 00 00.

‡Division of Statistics, Department of Computer and Information Science, Linköping University. Email: mattias.villani@liu.se. Phone: + 46 13 281 446

1 Introduction

Bankruptcy is an event of fundamental economic importance. The recent recession has shown that its rate of occurrence in the aggregate have profound influence on the outcomes of economic growth and unemployment, as well as financial stability through the effects on banks and financial markets in general. At the micro level, bankruptcy can be seen as the main driver of credit risk and is hence a primary concern for banks and investors that screen firms and monitor firms' efforts. In spite of its importance, our empirical understanding of the determinants of bankruptcy still has remarkable shortcomings despite the enormous volume of this literature. One such shortcoming, and the focus of this paper, is an empirical exploration of non-linear relationships between firm-level bankruptcy and key financial ratios such as firms' leverage, earnings, and liquidity. For this purpose we employ a recently compiled and extensive panel data set with detailed firm-level information on all incorporated Swedish businesses, both private and public, over the period $1991-2008$. The panel comprises around 4 million firm-year data points, with an average of over 200,000 firms per point in time. Our aim is to demonstrate the substantial gains in explanatory and predictive power that can be achieved by introducing straightforward spline functions into an otherwise standard multi-period logistic modeling framework, as used by Shumway (2001), Chava and Jarrow (2004), Campbell, Hilscher, and Szilagyi (2008), and others. Introducing splines into a logistic regression is a flexible and computationally efficient method for exploring non-linear relationships. It can be described as a simple transformation of the set (or subset) of explanatory variables into an extended covariate set, similar to a plain polynomial extension. In comparison with other flexible approaches, the spline method has the advantage of preserving linearity in the parameters, and the extended model can therefore readily be estimated using maximum likelihood and standard software.[1]

There are good reasons to presume that relationships between firms' financial ratios and bankruptcy risk are non-linear, or even non-monotonic. The functional form of the probability of default implied by Merton's (1974) distance-to-default model suggests that a firm's debt level only yields a modest impact on bankruptcy risk for low debt levels, whereas a substantially enhanced impact occurs when the value of debt approaches the market value of the firm.[2] The economic intuition is straightforward: there is no reason why a debt reduction should have an impact on a firm's bankruptcy risk when solvency is high, while the marginal benefit of a same-sized reduction should be much larger when the firm is indebted close to insolvency. Bharath and Shumway (2008) study the empirical relevance of Merton's bankruptcy probability within a multi-period logistic framework. They do so by including a proxy of the probability measure together with the variables used to construct the proxy jointly in the empirical model and conclude that the statistical significance of the proxy suggests that the non-linear transformation implied by Merton's model leads to enhanced bankruptcy predictions. In other words, their result lends support to the view that the relationship between firms' capital structure and bankruptcy risk indeed is highly non-linear.

In addition to firms' capital structure, other key variables such as firms' earnings ratio

[1] As a point of clarification: the standard logistic model is non-linear in the probability, but linear in the log odds. The spline model, that we propose to use, is non-linear also in the log odds, and will thus allow for more general non-linearities and non-monotonicities. In fact, the logistic model offers a rather restricted functional form that may not necessarily yield a good approximation of the empirical relationship.

[2] See Bharath and Shumway (2008) for a detailed outline of the probability of default implied by Merton's model: $PD_{Merton} = N(-((\ln(V/F) + (\mu - 0.5\sigma_V^2)\,T)\,/(\sigma_V\sqrt{T})))$, where $N(\cdot)$ is the standard normal cdf, V is the market value of the firm, D is the face value of the debt, μ is the return on V, σ_V^2 is the volatility of V, and T is the forecast horizon. The term $\ln(V/F)$ jointly with the standard normal cdf implies that the impact of a firm's leverage ratio, F/V, on its bankruptcy risk is enhanced as the debt level approached the market value of the firm.

and liquidity holdings are also likely to display non-linear relationships with bankruptcy risk. That is, increased earnings are likely to reduce firms' risk of failing on debt payments and ongoing expenditures. However, excessive positive earnings may be the outcome of risky strategies resulting in high cash-flow volatility, which could induce an increased risk of encountering financial distress and bankruptcy (see e.g., Nance, Smith, and Smithson 1993). Furthermore, it is intuitive to assume that increased cash holdings yield a significant impact on firms' bankruptcy risk if their initial cash position is weak, but will be of minor importance for already cash-rich firms. Taken together, these observations motivate a systematic exploration of non-linear relationships between financial ratios and firms' bankruptcy risk. Thus, the focus of this paper is to explore non-linear relationships beyond those imposed by the logistic link function.

Earlier contributions have demonstrated that flexible bankruptcy risk modeling with generalized additive models (GAM) yields significantly improved risk-ranking properties as compared with the standard logistic bankruptcy risk model (see Berg 2007; and Dakovic, Czado, and Berg 2010). This paper contributes to the existing literature by documenting three important features of non-linearities in bankruptcy risk modeling. Firstly, we show that allowing for non-linearities substantially improves both the model's risk-ranking ability and the accuracy of the absolute bankruptcy risk estimates. In terms of a statistical fit-measure, a pseudo-R^2, the improvement in-sample is on the order of 70 to 90 percent which is remarkable given that our approach leaves the information set unchanged. In addition to the improved fit-measure we show that the obtained bankruptcy predictions are unbiased across risk levels. Secondly, thanks to the size of our panel data set, we are able to estimate corresponding spline models for each of the 18 years in our sample period $1991 - 2008$. The resulting relationships are found to be remarkably stable over this time period, suggesting that a non-linear model provides a superior tool for forecasting bankruptcy risk. This is also verified in out-of-sample evaluations of the logistic and logistic spline models. Finally, the estimated non-linear relationships provide important economic insights on the relationships between key financial ratios and bankruptcy risk. More specifically, we document both threshold effects and sign inversions in the relationships between financial ratios and bankruptcy risk. For instance, in line with Merton's model, we observe threshold effects for the relationship between the leverage ratio and bankruptcy risk. The impact of firms' debt levels on their bankruptcy risks is moderate and close to constant for leverage ratios (total liabilities to total assets) in the $30 - 60$ percent region. However, the risk more than quadruples in the $60 - 100$ percent region, and, less intuitively, risks also increase as the leverage ratio decreases towards 0 in the $0 - 30$ percent region. Moreover, we observe sign inversion for the relationship between the earnings ratio (earnings to total assets) and bankruptcy risk. The bankruptcy risk is decreasing in the earnings ratio until the ratio reaches 15 percent and increasing thereafter.

Accounting, or financial ratio, analysis for predicting business failures and bankruptcy risk has a century long tradition and its modern era began in the 1960s with work by Beaver (1966) and by Altman (1968).[3] In an influential paper, Shumway (2001) outlines what has become the dominating method for estimating firm bankruptcy risk. Shumway points to the bias and inconsistencies that arise in static models of bankruptcy due to

[3] Altman's multivariate approach, the seminal Z-score model, which continues to be a benchmark-model widely applied in academia and by the rating industry. Altman (1971, 1984, 2000) has further examined accounting-based modeling, and numerous follow-up papers have been written, noteable ones are Ohlson (1980) and Zmijewski (1984). Altman and Narayanan (1997) and Altman and Saunders (1997) provide surveys of the bankruptcy litterature.

such models' neglectance of ultimately failing firms non-failing behavior in periods prior to bankruptcy.[4] Shumway goes on to show that a multi-period logistic model avoids the bias and inconsistencies in static models, and the approach has since then been a benchmark. He also argues that the significance of many of the financial ratios found by earlier papers does not survive in a multi-period model, in particular if up to date market-driven variables are included. Chava and Jarrow (2004) confirm the superiority of Shumway's approach, and suggest further improvements by controlling for industry effects and by considering a monthly frequency instead of the pre-dominating annual data frequency. Furthermore, Campbell et al. (2008) also applies the Shumway model specification and contributes by considering a wide range of financial ratios as well as market-driven variables in search of optimal models for given forecast horizon. Summing up, it is clear that multi-period logistic models dominate the static approaches. Market-driven variables clearly contribute over-and-above the financial ratios based on firms' financial statements. However, the latter have not entirely played out their roles in bankruptcy modeling, not even for samples of public firms as in the papers mentioned above. For bankruptcy risk in private firms, which is the concern in this paper, financial ratios remain the important information source, since market-driven variables are typically not available.[5]

A large set of various financial ratios have been proposed for modeling bankruptcy risk. We have selected three ratios based on what we judge are frequently used variables

[4] The traditional, static models typically only make use of the last financial statement before a firm goes bankrupt. Or, alternatively, e.g., the next to last financial statement before failure, but in that case discarding the information contained in the last statement. By means of a simple 2-period example Shumway demonstrates how the bias of the maximum likelihood estimator of the probability of default arises in a static model.

[5] Bharath and Shumway (2008) evaluate the out-of-sample accuracy of the Merton (1974) model and find that the distance-to-default measure is not a sufficient statistic for the probability of default in the sense that its accuracy can be surpassed by means of a reduced form model. Hence, suggesting that financial statements contain default-relevant information over-and-above that provided by market-driven variables.

in the literature, reflecting firm characteristics in key areas such as capital structure, performance, and liquidity. The ratios are: total liabilities over total assets (leverage ratio); earnings before interest and taxes over total assets (earnings ratio); cash and liquid assets in relationship to total liabilities (cash ratio). As we will demonstrate, these variables are close to being monotonically related to firm failure for large segments, which explains their long-standing popularity in the bankruptcy risk literature. However, substantially more information about firms' riskiness can be gained by accounting also for non-linear aspects of these variables' relationships with firm failure. Furthermore, in a recent paper, Jacobson, Lindé and Roszbach (2011) examine the empirical role of macroeconomic factors for bankruptcy risk modeling using the same longitudinal data set as in this paper. Their results suggest that macroeconomic factors shift the mean of the bankruptcy risk distribution over time and thereby are the most important determinants of the average level of firm failure. Therefore, in addition to the set of financial ratios, we include two macroeconomic variables in order to capture the important time-varying mean of the failure risk distribution. We also include variables that control for the size and age of the firm.

The remainder of this paper is structured as follows. In the next section, we present the Swedish firm data set. In Section 3 we provide a brief introduction to the statistical framework, and in particular how to introduce spline functions. The empirical results are reported in Section 4 for two versions of the models, one where only the lagged levels of the financial ratios and control variables are included and then another where the model is augmented with spline functions. We undertake in-sample, as well as out-of-sample comparisons of the estimated models along three dimensions: (i) the fit of the models in terms of an adjusted R^2, (ii) the accuracy of the bankruptcy risk ranking, and (iii) the accuracy of the absolute bankruptcy risk estimates. Finally, Section 5 concludes.

2 Data, Institutional Setting, and Descriptive Statistics

The firm data set underlying this paper is an panel consisting of $4,039,183$ yearly observations on the stock of (on average) roughly $200,000$ Swedish *aktiebolag*, or corporate firms, as recorded between January 1, 1991, and December 31, 2008, hence covering a period of 18 years. Some firms enter or exit the data set within the sample period, which makes the panel unbalanced. *Aktiebolag* are by approximation the Swedish equivalent of corporations in the US, or limited liability businesses in the UK. Swedish law requires every *aktiebolag* to hold in equity a minimum of SEK $100,000$ (approximately USD $15,000$) to be eligible for registration at Bolagsverket, the Swedish Companies Registration Office (SCRO). Swedish corporates are also required to submit an annual financial statement to the SCRO, covering balance-sheet and income-statement data in accordance with the European Union standards. The financial statements, provided to us by Upplysningscentralen AB, the leading credit bureau in Sweden, constitute the backbone of the panel data set analysed below.

In Sweden, as in many other countries, firms have considerable discretion in choosing a fiscal year period for their financial statement. Thus, the fiscal years for Swedish corporates are allowed to vary between a minimum of 6 and a maximum of 18 months. Only for about half of the firm-year observations in our sample does the fiscal year coincide with a calendar year. Intuitively, in a multi-period framework, where dynamic behavior is modeled, it is crucial that the financial statements actually cover the time period for which they are supposed to pertain. In the papers above, or elsewhere to the best of our knowledge, this problem has not been acknowledged, presumably because the bankruptcy literature has mostly dealt with samples of large, listed firms for which fiscal and calendar years tend to coincide. We have standardized the financial statements by first transforming them to quarterly observations and then by aggregating over the four quarters of a given year.

The design of the input data set for a study on bankruptcy determination requires deliberation on a number of issues: (*i*) A definition of the population of corporate firms that in a given year are at risk of failure, or alternatively, actually fail that year; (*ii*) A definition of the dependent variable, i.e., the status of bankruptcy; (*iii*) As noted above, Swedish corporates have substantial discretion in their choices of calendar periods for their fiscal years, hence the financial statements need to be suitably standardized prior to estimation; (*iv*) A choice of the set of financial ratios, as well as the treatment of outliers. We also need to determine other control variables to include. As most studies on financial ratios and firm failure deal with samples of publically traded firms, attention has primarily been given to the last item on this list. Since we want to model the universe of Swedish corporates, all four issues become important for our data set.

In defining the population of existing firms in a given year t we include the firms that have issued a financial statement covering that year and are classified as "active", i.e., firms with reported total sales in excess of $10,000$ SEK (roughly USD $1,500$). Unfortunately the resulting sample does not amount to all active firms because some firms fail to submit their compulsary financial statements. This is particularly common for firms in distress, which is quite intuitive. A typical outcome in our data is a firm that ceases submissions of their financial statements and ultimately—could be many years later—enters into bankruptcy. Hence, there are two additional groups of firms that rightfully belong to the population that we wish to make inference about, but due to their lack of financial statements prove difficult to include in the sample. The first group is made up by firms that do not submit statements, nor do they fail. These firms exist, but we have no record of them and must therefore abstract from them. The second group consist of firms that have neglected to submit their statements, but by *de facto* failing in t leave evidence of their existence. In this paper, where the focus is on non-linearities in the relationships between financial ratios and bankruptcy, we have chosen to discard all incomplete data and hence estimate the models on a sample of submitting firms only.

In order to construct a reasonable dependent variable for firm failure events we have obtained from the credit bureau records of corporate firms' payment remarks. These are systematically collected data on events related to firms' payment behaviour from various relevant sources, e.g., the Swedish retail banks, the Swedish tax authorities, and, in particular, the juridical institutions that deal with the legal formalities in firms' bankruptcy processes. We have adopted the following definition of a firm failure from the credit bureau. A firm has failed if any of the following events has occurred: the firm is declared bankrupt in the legal sense, has officially suspended payments, has negotiated a debt composition settlement, is undergoing a re-construction, or, is distrain without assets. In total, $96,091$ firms in the panel enter into bankruptcy, an average failure rate of 2.38 percent. An overwhelming majority of these are due to bankruptcy in court, around 89 percent. For the remainder this event will almost always ultimately occur, but with a lag, hence the practise of using an extended definition of failure, beyond that of legal bankruptcy. The idea is to include events—and their timing—that capture the point-of-no-return for failing firms. In a loose sense, one can think of this definition as corresponding to the union of US Chapter 7 and Chapter 11 (liquidation) filings for bankruptcy. The included events are all reported on a daily frequency, but for this study we will simply set the bankruptcy indicator $y_{i,t}$ to unity if firm i fails on any day in year t, and to zero otherwise.

Since we are interested in modeling effects from financial ratios on bankruptcy in a multi-period model, it is imperative to make the financial statements temporally aligned with the dependent variable. The first step in this process involves safeguarding against partly, or wholly, overlapping fiscal years for a given firm over time. What we want to observe is a string of non-overlapping financial statements. In the second step we construct artificial, standardized financial statements, all for fiscal years beginning on January 1 and ending on December 31. This is achieved by first working out monthly statements and then aggregating these to yearly statements.[6] In the case where two consecutive statements share a month we interpolate linearly. The flow and stock variables of the financial statements have been separately and accordingly adjusted. This problem of divergence between fiscal and calendar year is a non-trivial problem for Swedish corporates, since on average over time close to 10 percent will submit a statement for a period other than a

[6] Suppose a firm has two financial statements pertaining to one given calendar year t. Let N_{t_1} and N_{t_2} be the lengths of the accounting periods (in months) of each statement, and let n_{t_1} and n_{t_2} be the number of coverage months for the two statements in year t (where $n_{t_1}+n_{t_2} = 12$), and let Var_{t_1} and Var_{t_2} be the values of the financial ratios obtained from each statement. We then calculate the artificial statement for year t according to: $(n_{t_1}/N_{t_1}) \times Var_{t_1} + (n_{t_2}/N_{t_2}) \times Var_{t_2}$ for flow variables and $(n_{t_1}/12) \times Var_{t_1} + (n_{t_2}/12) \times Var_{t_2}$ for stock variables. Thus, the artificial statements are weighted averages of the original ones. The same principle is easily applied to the fewer cases where three statements pertain to one given calendar year.

calendar year.[7]

The three financial ratios included in our empirical analysis are frequently used ratios in papers on bankruptcy risk and are chosen to reflect firms' capital structure, profitability, and liquidity. They strongly correlate with our definition and measure of failure and are: total liabilities over total assets (leverage ratio); EBIT over total assets (earnings ratio); cash and liquid assets in relationship to total liabilities (cash ratio).[8] By including the leverage ratio we control for events where firms fail due to balance-sheet-based insolvency (economic distress), i.e., the value of the liabilities exceeds that of the assets. Furthermore, the earnings- and liquidity ratios provide important information related to whether a firm is at the risk of cash-flow-based insolvency (financial distress), i.e., a shortage of liquid assets to cover debt payments and ongoing expenditures. Hence, these financial ratios are important determinants of firm failure risk.

In addition to the financial ratios we also include a set of control variables. These are: firm size as measured by real total sales (deflating by means of consumer prices, with year 2000 prices as base-line); firm age in years since first registered as a corporate; the yearly GDP-growth rate; and the repo-rate, a short-term interest rate set by Sveriges Riksbank (the Central bank of Sweden). The two firm-specific control variables are included to

[7] The annual number of financial statements increases from about 200,000 in the beginning of the sample period, to well over 300,000 in the final two years. The shares of shorter (less than 12 months) and longer (longer than 12 months) statements are both around 5 percent. Whereas shorter than the stipulated 6 months happen, statements covering a longer period than the allowed 18 months are very rare. Over time, the annual shares of shorter/longer statement periods have come down from about 8 percent to currently around 4 percent. Thus, an overwhelming majority of statements concern a period of 12 months. However, out of the 90 percent of the total number of statements, only 48 percentage points coincide with a calendar year, and hence 42 percentage points refer to other 12 month periods. In these calculations we have allowed for a given calendar year to begin in mid-December the previous year, and end in mid-January the following year. Hence, if anything the 48 percent is an exaggeration. Over time, this share of calendar year statements has increased from 45 to 50 percent.

[8] We also considered an alternative definition of the cash ratio by taking cash and liquid assets over total assets. However, we found that the current ratio, where cash and liquid assets are scaled by total liabilities, adds more explanatory power to the model by reflecting additional information related to the firms' net-debt position.

take account of the well-documented results that smaller and younger firms are, ceteris paribus, riskier than older and larger ones. The two aggregate variables were found to be important determinants of average bankruptcy rates in Jacobson et al. (2011). Hence, by including them here we are able control for business cycle fluctuations and avoid their confounding effects on the estimated relationships between financial ratios and bankruptcy. Following Chava and Jarrow (2004) we could also consider controlling for industry effects, but according to the conclusions in Jacobson et al. (2011) industry effects do not appear to be of first order importance for the Swedish corporate sector, hence we leave those out in the interest of simplicity.

When working with data sets of this size, it is quite evident that a portion of the observations is made up of severe outliers. Such observations would distort the estimation results if they were to be included in a standard logistic model, thus winsorization is common in the literature to avoid outliers that are created by near-zero denominators. However, our spline approach is by construction robust towards inclusion of outliers, and therefore makes winsorization less necessary. Nevertheless, in order to make comparisons with the standard logistic model meaningful we treat the data in accordance with Shumway (2001) and Chava and Jarrow (2004) by winsorizing the top and bottom one percent of the financial ratios. Hence, for each financial ratio we set all observations taking values smaller than the 1^{st} percentile equal to the value of the 1^{st} percentile and equivalently for values larger than the 99^{th} percentile. Table 1 reports the empirical distributions for the winsorized data set. The table distinguishes between bankrupt and non-bankrupt firm-years, for the period $1991-2008$.

[Insert Table 1 about here.]

The picture emerging from Table 1 is that there is a clear difference between failing and non-failing firm-years for these variables. On average, non-failing firms are substantially less leveraged, 66 percent as compared to 101 percent. Non-failing firm-years have substantially larger earnings and exhibit higher cash holdings as compared with failing firm-years. The average earnings ratio is 5.2 percent for non-failing firms, compared to a mere -10.4 percent for failing firms. The table also shows that uncondionally, smaller firms are riskier than larger ones. The average sized failed firm has total sales of SEK 6,175,000, whereas the average sized healthy firm has total sales of SEK 18,562,000. The same conclusion applies to firm age where the average age for failed firms is 9 years, while the average non-failed one is 11 years.

3 Empirical Methodology

As we will illustrate in Section 4, our data set shows definite signs of strong non-linearities in the relationships between bankruptcy risk and several of the most commonly used explanatory financial ratios. There is a large number of non-linear models for binary responses in the statistical literature, including neural nets, kernel regression, smooth threshold models, generalized additive models (GAM), and regression trees. See Hastie, Tibshirani, and Friedman (2009) for a general and accessible introduction. Berg (2007) shows that GAMs outperform logistic regression and neural networks in predicting firm bankrupcty in a single-period or static context. Most non-linear models are however computationally impractical on a data set of our size, especially in a multi-period setting. We therefore focus on the model which we believe has the largest potential for empirical bankruptcy risk modeling: the additive spline regression model. This model is essentially a GAM model with a smaller number of knots and without regularization (see below). Our aim here is to show that this powerful model: i) has manageable computational complexity even for very large data sets, ii) can be estimated with standard software packages, iii) is an easily interpretable model for non-linear data and iv) gives accurate out-of-sample predictions. The model should therefore appeal to a broad group of applied economists and is a strong candidate for inclusion in the standard econometric toolkit.

Let $p_{i,t}$ be the probability of bankruptcy for firm i in year t given that the firm was alive in year $t-1$. The basic bankruptcy risk model for the binary responses is of the form:

$$\begin{aligned} y_{i,t}|p_{i,t} \sim Bern(p_{i,t}), i = 1, ..., N \text{ and } t = 2, ..., T, \\ \ln\left(\frac{p_{i,t}}{1-p_{i,t}}\right) = \alpha + \beta' x_{i,t-1} + \gamma' z_{t-1}, \end{aligned} \tag{1}$$

where $y_{i,t} \in \{0, 1\}$ is the binary response variable recording the failure ($y_{i,t} = 1$) of firm i at time t, $x_{i,t-1}$ is a vector of firm-specific variables and z_{t-1} is a set of macroeconomic variables at time $t-1$.

The logistic model in (1) belongs to the class of Generalized Linear Models (GLM), and everything in this section is directly applicable to other members of the GLM class, e.g., response counts following the Poisson distribution, or non-negative, continuous Gamma-distributed response variables. It also applies to models with a more general link function:

$$g(p_{i,t}) = \alpha + \beta' x_{i,t-1} + \gamma' z_{t-1}, \tag{2}$$

where $g(\cdot)$ is any smooth invertible link function. In GLM terminology, the model in (1) is a Bernoulli model with a logit link.

A prominent feature of GLMs is that a transformation of the mean ($p_{i,t}$ in the Bernoulli case) is linear in the explanatory variables. As we show in Section 4, our data set on firm failures are highly non-linear in the log odds, showing a pressing need to go beyond the plain logistic regression model. The most obvious way to introduce non-linearites is by adding polynomial terms. Polynomials are well known to produce an unreasonably global fit and to have poor behavior near the boundaries (see e.g. Hastie, Tibshirani, and Friedman (2009)). After a period of intense research on kernel regression methods, c.f., Li and Racine (2007), the attention has shifted towards the use of splines for non-linear/non-parametric regression. Splines can be viewed as piecewise local polynomials with enforced continuity and higher order smoothness (e.g., continuous first derivatives) at the dividing points. Splines is a local model without the problems associated with global polynomials. A spline model is implemented in the same way as the polynomial regression by extending the covariate set with additional basis function covariates to model the non-linear effects. A particularly popular set of basis functions is the truncated power basis of order S:

$$\{1, x_{i,t-1}, x_{i,t-1}^2, ..., x_{i,t-1}^S, (x_{i,t-1} - k_1)_+^S, ..., (x_{i,t-1} - k_M)_+^S\}, \tag{3}$$

where $k_1, ..., k_M$ are the M dividing points of the local polynomials, usually referred to as *knots*, and

$$(x_{i,t-1} - k)_+^S = \begin{cases} 0 & \text{if } x \leq k \\ (x_{i,t-1} - k)^S & \text{if } x > k \end{cases} . \tag{4}$$

The attraction of this type of basis is that it directly incorporates the continuity constraints so that the function and its $S-1$ first derivatives are all continuous at the knots. The logistic spline regression model is therefore of the form:

$$\begin{gathered} y_{i,t}|p_{i,t} \sim Bern(p_{i,t}), i = 1, ..., N \text{ and } t = 2, ..., T, \\ \ln\left(\frac{p_{i,t}}{1-p_{i,t}}\right) = \alpha + \sum_{s=1}^S \beta_s x_{i,t-1}^s + \sum_{m=1}^M \eta_m (x_{i,t-1} - k_m)_+^S + \gamma' z_{t-1}. \end{gathered} \tag{5}$$

Note that the log odds remains linear in the parameters, and the logistic spline model can therefore be fitted with standard methods, which is absolutely crucial for very large data sets. We use the standard errors proposed by Shumway (2001).

A drawback of the local basis is that the fit can be erratic at the boundaries of the covariates space with ensuing poor extrapolation properties. A *natural spline* mitigates this problem by imposing a linearity constraint beyond the boundary knots. This restriction reduces the variance of the approximating function in these regions, and can be directly imposed on the basis itself. For example, in the case of a truncated power basis the corresponding natural spline basis functions are given by the recursion:

$$h_1(x_{i,t-1}) = x_{i,t-1},\ \ h_{1+m}(x_{i,t-1}) = d_m(x_{i,t-1}) - d_{M-1}(x_{i,t-1}), \tag{6}$$

where

$$d_m(x_{i,t-1}) = \frac{(x_{i,t-1} - k_m)_+^S - (x_{i,t-1} - k_M)_+^S}{k_M - k_m}. \tag{7}$$

To implement the spline regression model one needs to decide on the number of knots (M) and their locations, $k_1, ..., k_M$. With a small, or moderately large, data set these two choices are crucial and much effort in the literature is spent on developing statistical methods to deal with them. With a large number of knots, sometimes even a knot at every observation, it becomes crucial to impose some sort of penalty on model complexity in order to not overfit the data. The most common approaches use either the $L1$ ($\|\eta\|$) or $L2$ ($\eta'\eta$) penalty on the spline coefficients and use cross-validation or Bayesian methods to determine the optimal penalty, see e.g. Hastie, Tibshirani, and Friedman (2009) and Ruppert, Wand and Carroll (2003). An alternative approach is to use Bayesian variable selection to choose among the pre-determined knots, see e.g. Smith and Kohn (1996) and Denison, Mallick, and Smith (2002).

In large data sets, such as ours, allowing for sharply estimated parameters, the exact choice of the number of knot, their locations and optimal model complexity penalty is of lesser importance, and the methods refereed to above are unnecessarily sophisticated.

Moreover, these algorithms are very time-consuming when applied to a very large data sets. We have therefore opted for the following simpler, but effective, strategy. The number of knots is determined by the Bayesian Information Criterion (BIC) as proposed by Schwarz (1978). The BIC chooses the number of knots that maximizes the likelihood function subject to a penalty for model complexity. After choosing the number of knots, the location of the knots is usually fixed at predetermined quantiles of $x_{i,t-1}$, which gives a more dense allocation of knots in regions with many observations. In our case some variables have many observations taking the same value, so that this strategy gives occasional knot duplication or near-duplication, resulting in perfect or high multicollinearity. Rather than eliminating the duplicates we chose to employ the well-known k-means algorithm to determine the location of the knots. Our results are robust to changes in the number of knots and to alternative knot location schemes.

While estimating a flexible mean curve using spline regression with a single covariate is relatively straightforward, it is a substantially harder problem to estimate a flexible mean surface with multiple covariates. The main complication is the curse of dimensionality: any reasonably large number of knots will always be sparse in a high-dimensional covariate space, see Hastie, Tibshirani, and Friedman (2009, p. 23) for an illustration. By far, the most common way of dealing with the curse of dimensionality is to assume away all interactions between covariates. The additive logistic spline model with natural quadratic spline basis functions is then of the form (Hastie and Tibshirani 1990):

$$\begin{gathered} y_{i,t}|p_{i,t} \sim Bern(p_{i,t}),\ i = 1, ..., N \text{ and } t = 2, ..., T, \\ \ln\left(\frac{p_{i,t}}{1-p_{i,t}}\right) = \alpha + \sum_{j=1}^{p} \sum_{m=0}^{M_j} \eta_{j,1+m} h_{j,1+m}(x_{j,i,t-1}) + \gamma' z_{t-1}, \end{gathered} \tag{8}$$

where $h_{j,1+m}(x_{j,i,t-1})$ is the natural spline function for the j^{th} covariate and m^{th} knot. Experiments showed that for our data set, the restriction $M_1 = \ldots = M_p = M$ did not degrade the in-sample fit or the out-of-sample predictive performance. The assumption of additivity also simplifies the interpretation of the model since the marginal effect (the partial derivative) of a covariate is not a function of the other covariates.

In summary, our model is a logistic additive spline model with natural quadratic spline basis functions ($S = 2$) and an equal number of knots for each covariate (henceforth referred to as the logistic spline model). The number of knots is determined by the BIC, and the knots are deployed by separate k-means clustering of each covariate. Given the knot locations, the spline basis functions are computed and used as covariates in a standard logistic regression model. The basis expanded logistic regression is fitted with standard methods, available in almost any statistical software package.

4 Empirical Results

In this section we report empirical results for the logistic and logistic spline models. In particular, we first show the univariate relationships between the firm-specific variables and bankruptcy risk. Secondly, we briefly report results for the standard logistic model, corresponding to the ones reported in Shumway (2001), Chava and Jarrow (2004), Campbell et al. (2008), and others. We then move on to the in-sample results for the spline models. Finally, we document the year-by-year stability in the estimates obtained from the logistic spline model, and evaluate the out-of-sample properties of the logistic and logistic spline models.

4.1 Univariate Relationships

The logistic spline models reported below can be justified by the highly non-linear relationships between the accounting variables and bankruptcy risk that are documented in Figure 1. The figure shows the observed bankruptcy frequency and the estimated relationships obtained from a univariate logistic and a univariate logistic spline model, respectively, for each of the financial ratios and control variables.

[Insert Figure 1 about here.]

From the figure it is clear why the leverage ratio, TL/TA, and earnings ratio, $EBIT/TA$, are extensively used in the bankruptcy literature. The heterogeneity in bankruptcy frequency for the two variables spans between 0.5 and 12 percent and the relationships are almost monotonic in the high density regions. Furthermore, the cash ratio, CH/TL, exhibits a distinct monotonic relationship with bankruptcy risk, but the impact is less pronounced as compared with the leverage- and earnings ratios. For the two control variables we observe a decline in the risk with respect to total sales, $Size$, and a more challenging non-monotonic relationship for firm age, Age, where risk is increasing for young firms and starts to decrease for firms that have been active for more than four years.

The estimated univariate relationships suggest that the logistic spline model substantially improves the fit with respect to all variables, especially in the regions characterized by non-monotonic relationships with bankruptcy risk. Taken together, these results indicate that a multivariate model that controls for these non-linear features is likely to yield enhanced bankruptcy predictions.

4.2 Logistic Models

The first model, which we call the Private Firm Model, incorporates the leverage- and the earnings ratios only. This parsimonious model corresponds to the private firm model reported in Chava and Jarrow (2004). The second model that we consider is an extended version of the Private Firm Model and includes additional variables related to firms' cash holdings, total sales, age, and two macroeconomic variables. The explanatory variables $x_{i,t}$ and z_t enter the model lagged on year, hence they are observed at the time of prediction. That is, the model is designed to capture relationships between a set of explanatory variables describing the characteristics of a firm and its environment at time $t-1$ and the event that the firm fails in period t. In order to get consistent and unbiased estimates of these relationships, it is essential to have explanatory variables that pertain to year $t-1$, and nothing else.[9] By means of our periodization procedure of the financial statements, described in Section 2, we have made sure that this is the case here.

[Insert Table 2 about here.]

Models I and II in Table 2 concern coefficient estimates for the Private Firm Model and the Extended Private Firm Model. For both model specifications, the coefficients for the financial ratios have signs according to intuition. That is, the leverage ratio, TL/TA, has a positive coefficient, whereas the earnings ratio, $EBIT/TA$, and the cash ratio, CH/TL, have negative coefficients. For the Extended Private Firm Model, the reported

[9] Shumway (2001) provides a thorough discussion on requirements for consistency for the maximum likelihood estimator of the intercept and slope parameters. In particular he demonstrates why the static model of bankruptcy, i.e., a model that considers the last financial statement only, discarding information on a firm for other years than the one preceding bankruptcy, will result in biased and inconsistent estimates. It is straightforward to apply Shumway's arguments to the case when fiscal years do not coincide with calendar years so that a given firm has, e.g., submitted a financial statement covering a fiscal year such that x_{t-1} is in effect a linear combination of the standardized statements for x_t, x_{t-1} and x_{t-2}.

marginal effects give that a one-standard-deviation increase in the leverage ratio leads, on average, to a 34 percent increase in the bankruptcy risk, and a one-standard-deviation increase in the earnings ratio and cash ratio leads, on average, to an 18 and 87 percent decline, respectively. Thus, all three ratios exercise an economically significant impact on bankruptcy risk.

For the control variables included in the Extended Private Firm Model we see that total sales, $Size$, has a negative impact and that firm age, Age, has a positive impact on bankruptcy risk. The positive coefficient for firm age is counterintuitive, but serves as an excellent illustration of the spurious results that a linear model can give rise to. Furthermore, the impacts of the two macroeconomic variables are in line with Jacobson et al. (2011). The signs of the macro coefficients are as expected: a negative one for ΔGDP implying a reduced bankruptcy probability when growth is high in the previous year, and a positive coefficient for the $REPO\text{-}RATE$ implying an increased bankruptcy risk when interest rates are high.

The performances of the models are evaluated by means of McFadden's pseudo-R^2 and an additional measure that we denote by ROC (Receiver Operating Characteristics, c.f., Hosmer and Lemshow (2000)). The ROC-measure is equivalent to the one reported in Chava and Jarrow (2004) and quantifies the models ability to rank firms according to their relative riskiness. It spans between 0.5 and 1, where 0.5 corresponds to a model that randomly assigns bankruptcy probabilities to firms, and where 1 corresponds to the perfect model where the highest bankruptcy probabilities are assigned to the firms that *de facto* fail.

Because of differences in the selection of variables and sample composition, it is not straightforward to directly compare our results with earlier estimates in the literature. In particular, Shumway (2001), Chava and Jarrow (2004), and Campbell et al. (2008) are all concerned with listed firms and therefore also incorporate market-driven variables. However, although our models do not incorporate any market-driven variables the pseudo-R^2 coefficients of 6.3 and 8.4 percent are not far from the 11.4 percent that is reported for a model with a one-year forecasting horizon in Campbell et al. (2008). Moreover, the ROC values of 0.76 and 0.77 are in the same vicinity as the out-of-sample values reported in Chava and Jarrow (2004). Thus, the performance of the logistic models are re-assuringly close to the ones of models previously reported in the literature.

4.3 Logistic Spline Models

We now shift our attention to the logistic spline model in Eq. (8). Before reporting the empirical results we will briefly outline how the number of knots and their placements are determined. As outlined in Section 3, we have adopted a basic approach relying on a standard information criterion and a clustering procedure to decide on these issues. The transformation $x_{j,i,t-1} \rightarrow h_{j,1+m}(x_{j,i,t-1})$ is carried out for each number of knots, M. The augmented model (where the extended set of covariates is based on an equal number of knots for each variable) is then estimated and the BIC information criterion is calculated. This is repeated for $M = 0, 1..., M_{\max}$, where $M_{\max}$ is set to 25, to make reasonably sure that a global minimum is identified. In our case the combination of a very large sample and what turns out to be highly non-linear functions, results in the BIC choosing 10 and 11 knots as the optimal number for the spline versions of the Private Firm Model and the Extended Private Firm Model, respectively. However, we noticed that a large improvement in fit is obtained with a handful of knots, say 4 or 5, suggesting that this approach need not be restricted to large sample sizes only.[10]

Models III and IV in Table 2 concern coefficient estimates for the spline models. Since the models involve a large set of estimated parameters (1+10 and 1+11 for each of the 5 firm-specific variables) that in themselves offer very limited intuition, we summarize the relationships by reporting parameter estimates that are averaged over the sample.[11] Nevertheless, the average coefficient estimates are, albeit being a crude approximation and certainly not sufficient to account for the full relationships, in line with the linear model and offer some intuition. The three financial ratios have coefficient estimates with the same signs as previously, but the magnitude of the estimates have increased considerably. Furthermore, the coefficients for total sales, *Size*, is of the same magnitude as before and the impact of firm age, *Age*, stays positive and increases in magnitude. The large positive coefficient for firm age is counterintuitive, but this picture changes as we consider the total effect reported below. Finally, the impact of the two macroeconomic variables is of the same magnitude as for the plain logistic model.

Turning to the goodness of fit, it is striking that the pseudo-R^2 improves from 6.3 to 10.6 for the Private Firm Model, and from 8.4 to 16.2 percent for the Extended Private Firm Model. These results indicate that a model that allows for flexible non-linear relationships exhibit a substantially improved performance by assigning bankruptcy probabilities to firms that more accurately correspond to the actual outcomes. The increase

[10] The pseudo-R^2 coefficients for the Private firm model and the Extended private firm model with 5 knots are 10.5 and 16.0 percent, respectively.

[11] Following (8), we report the average coefficient: $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} d\theta_{i,t}/dx_j$ for the logistic spline models.

in the pseudo-R^2 is remarkable given that the underlying information set is unchanged. Also, the ROC-measures indicate that the spline models ranking properties are enhanced. However, this is most evident for the Extended Private Firm Model, which may be explained by the highly non-monotonic relationship between firm age and bankruptcy risk, c.f., Figure 1.

[Insert Figure 2 about here.]

Next we make an assessment of the in-sample properties of the models in an absolute sense. For a given model we sort all firms with respect to their estimated bankruptcy probabilities into one hundred percentiles and then calculate the average bankruptcy probability in each percentile. We then compare the average estimated bankruptcy probability with the actual failure frequency for the firms in each percentile. If the estimated models were to perfectly predict the absolute riskiness of the firms within each percentile, a scatter plot of the two variables would line up along the 45-degree line, corresponding to a slope coefficient of unity and an intercept equal to zero. The two panels in Figure 2 show the accuracy of the bankruptcy predictions obtained from the logistic and the logistic spline versions of the Private Firm Model and the Extended Private Firm Model. Graph (I) in Panel A shows the relationship for the logistic version of the Private Firm Model. The relationship indicates that there is a considerable divergence between the estimated bankruptcy probabilities and the actual outcomes. This is also demonstrated in Graph (II) illustrating the same relationship but on a logarithmic scale, which offers a greater resolution of the left-hand tail of the distribution. On average, the graphs show that the predictions obtained from the logistic version of the Private Firm Model tend to overestimate the bankruptcy risk in the $1-2.5$ and > 11 percent intervals, and underestimate the risk in the < 1 and $2.5-11$ percent intervals. The deviation is substantial where, for example, the group of firms that is assigned a bankruptcy risk of 5 percent exhibit an observed bankruptcy frequency of around 10 percent. In contrast, Graph (III) and (IV) in Panel A show that the estimated risk obtained from the logistic spline version of the Private Firm Model almost overlap the 45-degree line which indicates that the assigned bankruptcy risks correspond remarkably well with the actual outcomes. Furthermore, a similar message is given by the graphs in Panel B where the estimates obtained from the logistic and logistic spline versions of the Extended Private Firm Models are evaluated.[12] However, the estimated bankruptcy probabilities from the Extended private models exhibit a larger heterogeneity, which indicates that the additional variables included in the extended model are important determinants of firms' bankruptcy risk. Taken together, we thus conclude that the accuracy of the bankruptcy predictions in an absolute sense substantially improves once we allow for flexible non-linear relationships. They now appear unbiased irrespective of risk level.

4.3.1 Non-Linear Relationships

In Figure 3 we illustrate the estimated non-linear relationships between bankruptcy risk and the firm-specific variables obtained for the spline version of the Extended Private Firm Model, Model IV in Table 2. Each variable is displayed in two ways. Firstly, in the left-hand panels, we illustrate the bankruptcy probability as a function of each of the five firm-specific variables in a multivariate conditional setting where the other variables are set to their sample means. Secondly, in the right-hand panels, we document the

[12] To the extent that banks rely on a logistic bankruptcy model for screening and monitoring of borrowers the biases in the estimated failure probabilities may lead to systematic errors. Moreover, if banks' buffer capital is calculated according to the IRB-approach, i.e., risks in loans are evaluated using an internal risk model, then using a logistic model could yield inadequate buffer capital.

derivatives of the logit function ($d\theta/dx_j$, c.f., Eq. (8) above) across all segments of the variables. The conditional mean function shows the relationship between each variable and the level of the bankruptcy risk and the logit derivative complements by showing for which segments the marginal impact of each variable is statistically different from zero. The graphs also include confidence bands reflecting estimation uncertainty.[13]

[Insert Figure 3 about here.]

Judging by Figure 3, the univariate relationships displayed in Figure 1 largely hold for all the financial ratios. For the relationship between the leverage ratio, TL/TA, and bankruptcy risk, illustrated in Panels (I.A) and (I.B) in Figure 3, we observe a distinct non-monotonic relationship. The graphs show that both low and high leverage levels are associated with a higher bankruptcy risk. That is, firms with a leverage ratio below 30 percent exhibit increasing risks as the ratio decreases. In the other direction we find that more leveraged firms display a straightforward and intuitive failure-relationship. Hence, the $60 - 100$ percent interval of the leverage ratio is characterized by a sharp increase in bankruptcy risk; it more than quadruples from 0.7 to around 3 percent. The maximum impact occurs at a leverage level around 95 percent, c.f., Panel (I.B). The observed threshold effect is in line with the functional form of the probability of default implied by Merton's distance-to-default model. Thus, quite intuitively, the marginal reduction in failure risk from reducing firms' debts is larger for highly leveraged firms that are close to balance-sheet-based insolvency.

[Insert Table 3 about here.]

[13] The increased uncertainty around each knot in the confidence intervals of the derivatives of the logit function is a consequence of the truncation of the quadratic splines, which produces discontinuous second derivatives.

The increased bankruptcy risk for low-leveraged firms may be driven by cases where firms fail because lenders are unable to resolve asymmetric information problems and therefore restrict their credit supply (Stiglitz and Weiss 1981). Limited access to external financing makes firms more vulnerable to liquidity shocks which induce an increased distress risk (see, e.g., Opler, Pinkowitz, Stulz, and Williamson 1999). In Table 3, Panel A, we show that a low leverage level is likely to be supply driven. That is, the table shows that firms with low leverage ratios on average are substantially smaller as compared with firms having a medium, or a high, leverage ratio. Small firms are more likely to suffer from information asymmetries, which make them more exposed to financial frictions (see, e.g., Almeida, Campello, and Weisbach 2004). Furthermore, in the table we also show that low-leveraged firms on average pay substantially higher interest rates. Higher interest rates are likely to be driven by lenders limiting the credit accessibility by both contracting the supply of credit, and by increasing its price.[14] Thus, these results suggest that a low leverage ratio is likely to be the outcome of limited credit supply which explains the increased bankruptcy risk.[15]

Panels (II.A) and (II.B) in Figure 3 show an apparent non-monotonic relationship between the earnings ratio, $EBIT/TA$, and bankruptcy risk, where both low and high earnings ratios are associated with increased risks. More specifically, bankruptcy risks are high for earnings ratios below -40 percent, and then sharply decline for ratios in the $-40-10$ percent interval. The impact is reversed for high earnings ratios, so that above 15 percent they exhibit a statistically significant positive impact on bankruptcy risk. A

[14] The credit rationed firm has to rely on short-term and expensive trade credit as an alternative to regular financing.

[15] An alternative—or complementary—explanation for the increased risk for low-leveraged firms is that leverage improves managerial incentives and reduces free cash flow that could be invested in low net present value projects (see, e.g., Jensen 1986).

negative relationship between the earnings ratio and bankruptcy risk is intuitive since higher earnings decrease the risk of failing on debt payments and ongoing expenditures. However, one explanation for the observed non-monotonic relationship could be that high earnings are associated with high cash-flow volatility. Firms that exhibit high cash-flow volatility are more likely to experience a cash-flow shortfall, which in turn may trigger financial distress, see e.g., Nance, Smith, and Smithson (1993). In Table 3, Panel B, we report the three and five year firm-specific cash-flow volatility for firm-year observations associated with low, medium, and high earnings ratios. The table shows that low and high earnings ratios (as opposed to medium) are associated with a substantially higher volatility. Moreover, comparing failing with non-failing firm-years, we see that failing firms overall have more volatile earnings ratios, and this feature is emphasized for firms with high earnings ratios, i.e., above 15 percent. Thus, high earnings ratios are associated with a higher volatility, which may be a factor that lies behind the observed non-monotonic relationship between the earnings ratio and bankruptcy risk.

Financial frictions are an additional underlying factor that may play a role for the non-monotonic relationship in the earnings ratio. In Table 3, Panel B, we show that bankrupt firms with high earnings ratios on average are smaller and have higher interest expenditures. That is, failing firms with a high earnings ratio pay an interest rate spread twice that of similar non-failing firms, and four times larger than firms in the medium earnings ratio segment.[16] These results are in line with earlier findings in the literature, showing that high cash-flow volatility is associated with lower investments, a greater need for external financing, and higher costs for external financing (Minton and Schrande 1999). Moreover, Table 3, Panel B, also shows that bankrupt firms with high earnings ratios tend

[16] The spread is calculated as intertest expenditures over total liabilities minus the *REPO-RATE*.

to experience a reduction in their fixed assets (property, plants, and machinery). This suggests that high earnings ratios may be a manifestation of asset redeployment, where constrained firms sell fixed assets in a secondary market in order to generate funding when such is unavailable, or expensive in capital markets (see Lang, Poulsen, and Stulz 1995). Thus, bankrupt firms with high earnings ratios face limited or costly financing, which may trigger a higher failure risk. Taken together, high cash-flow volatility in combination with limited and costly financing are factors that potentially induce the documented positive relationship between excessive earnings and firm failure.

Returning to Figure 3, the panels (III.A) and (III.B) show that the cash ratio, CH/TL, features a clear threshold effect. We see that bankruptcy risks decrease sharply as the cash ratio increases from 0 to 50 percent. For a cash ratio exceeding 50 percent we find that bankruptcy risks are stable around 0.5 percent. The documented relationship is intuitive and illustrates that the marginal benefit of increased cash holdings is large for firms with low cash holdings and of less importance for cash-rich firms.

Finally, the two control variables *Age* and *Size* display relationships with bankruptcy risk that closely correspond to the ones outlined in the univariate cases above, c.f., Figure 1. In Figure 3, panel (V.A), bankruptcy risk and *Age* display a distinct hump-shaped relationship such that risk is increasing in *Age* until the firm reaches the age of 4 years and then risks fall steadily beyond that age until around the age of 16 years where it becomes constant. In the right-hand panel the derivative of the logit with respect to *Age* is thus positive for ages up to 4 years, negative between 4 and 16 years, and then impact is insignificantly different from zero for firm ages beyond 16 years. This result is in agreement with the predictions in the classical work of Jovanovic (1982), where firms upon entry gradually learn about their efficiencies in operation and may, when sufficient information has been accumulated, come to the conclusion that exit is the rational choice.[17] Turning to $Size$ as a determinant of firm failure, panel (IV.A) suggests an almost monotonically negative relationship, confirming intuition and common wizdom. However, in panel (IV.B) the derivative of the logit with respect to Size reveals a very modest impact overall, and it is only for the very largest firms that we find a negative derivate that is significantly different from zero.

Taken together, the Figure 3 panels show both threshold effects and sign inversions in the relationships between the three financial ratios and firm failure risk. Relationship characterization by information asymmetries are likely to be more pronounced in our sample—consisting mostly of small private firms, as opposed to larger public ones. Nevertheless, information asymmetry presumably plays a role also for larger public firms.

4.3.2 Stability of the Non-Linear Relationships

The extensive panel data set allows us to examine the stability over time for the estimated non-linear relationships documented above. Since the panel comprises around 4 million firm-year observations, we can estimate the spline model for each of the 18 years in the sample period and make use of more than $200,000$ observations in each year. This is a robustness check that renders credibility to the results outlined above, and potentially demonstrates the time-invariance of the documented non-linear features. The specifications of the 18 yearly models coincide with that of Model (IV) in Table 2, but now

[17] "Efficient firms grow and survive; inefficient firms decline and fail", p. 469, Jovanovic (1982). Agarwal and Gort (1996) estimate firm survivorship using data on manufacturing firms of new products, and take into account the various maturity stages of the products' life cycles. Agarwal and Gort find similar hump-shaped hazard functions for all maturity stages as we do, cf., Figure 2, p 496. They argue that firm survival is a function of both age and endowment, and over time effects of favorable endowment will tend to vanish and result in increasing failure risks. This effect is manifested in the five stages of maturity that Agarwal and Gort consider, but clearly absent in our model.

without the macroeconomic variables included. To take account of time-varying average bankruptcy risks the intercept is set to $[-3.956 - 0.078 \times \Delta GDP_{t-1} + 0.057 \times REPO_{t-1}]$, where the coefficients are estimated in a model where only the two macroeconomic variables are included.

Figure 4 documents the estimated derivative-curves for the five firm-specific variables as given by the 18 yearly spline models. The overall picture is one of remarkable stability in these variables' effects on firm failure risk over the period $1991 - 2008$. In particular, we see that the yearly variation in the logit derivatives for the leverage ratio, *TL/TA*, the earnings ratio, *EBIT/TA*, the cash ratio, *CH/TL*, and firm age, *Age*, are very small. The yearly models' logit derivatives closely coincide with the ones outlined in Figure 3. Given the importance of the leverage- and the earnings ratios, in particular, for bancruptcy predictions, this robustness feature is quite re-assuring. In the case of the *Size* variable, we find that the logit derivative curves display somewhat larger variation over the years.

To further study the yearly logit derivative curves in Figure 4, they are divided into two regimes, $1991 - 1995$ and $1996 - 2008$. On the whole, the effects in both regimes coincide for all variables. However, during the Swedish banking crisis, occurring in the first regime, $1991 - 1995$, the bankruptcy relationship for firm size, and to some extent for leverage, shift as manifested by the variables *Size* and *TL/TA* in Figure 4. The banking crises episode saw exceptionally many firm failures, and unusually large firms going under, so it is not surprising to note the shifts in the derivative effects for these years.

[Insert Figure 4 about here.]

Overall, we conclude that the documented time-invariance in the logit derivatives suggests that the non-linear relationships between the variables and the bankruptcy risk are a persistent feature. Thus, these results indicate that the observed in-sample improvements obtained by allowing for flexible non-linear relationships in the logistic bankruptcy model also are likely to hold for the forecasting properties of the model, which we further document in the proceeding section.

4.4 Out-of-Sample Evaluation

We will next evaluate the forecasting accuracy of the logistic and logistic spline models. The out-of-sample evaluation follows the same approach as the ones reported in Shumway (2001) and Chava and Jarrow (2004). That is, we split the sample period in half and estimate the models on data for the period $1991-1999$ and use the subsequent period $2000-2008$ to gauge the models' forecasting performance.

[Insert Table 4 about here.]

Table 4 documents out-of-sample results for the logistic and logistic spline versions of the Private Firm Model and the Extended Private Firm Model. The reported in-sample pseudo-R^2 coefficients are slightly smaller as compared to the ones reported for the full sample, but the relative improvement obtained by including splines is of the same magnitude, or even slightly enhanced. Furthermore, the reported out-of-sample pseudo-R^2 coefficients are calculated as $1-L_1/L_0$, where L_1 is the log likelihood obtained for the out-of-sample period using the in-sample estimates and L_0 is the log likelihood for an intercept model estimated for the out-of-sample period. As for the improvement in-sample, it is striking that the out-of-sample pseudo-R^2 improves from 2.7 to 7.5 percent for the Private Firm Model, and from 7.2 to 14.8 percent for the Extended Private Firm

Model. These results imply that controlling for non-linear relationships substantially improves the models' forecasting accuracy.

Turning to our measure of relative risk, the reported ROC-measures assess the models' ability to rank firms according to their riskiness in terms of *ex post* bankruptcy frequencies. The documented values indicate that the spline versions of the models exhibit enhanced ranking properties. However, similar to the results reported for the in-sample period, the more striking improvement is observed for the logistic spline version of the Extended Private Firm Model which has a ROC-value of 0.82. Furthermore, a similar message is presented in terms of a decile test, where firms have been sorted into deciles according to their predicted bankruptcy risks. The table shows that the spline version of the Extended Private Firm Model is the best performing model classifying 48 percent of the bankrupt firms in the riskiest decile, as compared to around 41 percent for the other models. These results are close to the ones reported for a comparable private firm model in Chava and Jarrow (2004), where the average ROC-measure spans between 0.72 and 0.77 and the fraction of failing firms in the riskiest decile ranges between 31 and 44 percent.

Finally, we assess the out-of-sample properties of the logistic and logistic spline version of the Extended Private Firm Model in an absolute sense by comparing the predicted failure probabilities with the actual *ex post* bankruptcies (similar to Figure 2). In Figure 5, we present graphs of such predicted and realized failure frequencies on both a probability scale (left-hand side panel) and a logarithmic scale (right-hand side panel). If the estimated models were to perfectly predict the absolute riskiness of the firms within each percentile, all circles would line up along the 45-degree line, corresponding to a slope coefficient of unity and an intercept equal to zero. As can be seen, on average, Graph

(I) and (II) show that the logistic model tends to overestimate the bankruptcy risk in the $0.25 - 2.5$ and $10 <$ percent intervals, and underestimate the risk in the < 0.25 and $2.5 - 10$ percent intervals. In contrast, Graph (III) and (IV) show that the logistic spline version of the model almost overlaps the 45-degree line in the $0 - 4$ and $12.5 <$ percent segments, and shows very moderate deviations from the ideal 45-degree line in the $4 - 12.5$ percent segment. In sum, the out-of-sample exercise shows that allowing for non-linear relationships in the logistic model leads to a substantial improvement in forecasting accuracy.

[Insert Figure 5 about here.]

5 Conclusions

In this paper, we gauge non-linear relationships between financial ratios and firm bankruptcy risk at the microeconomic level using a standard logistic model augmented by natural quadratic splines. Our approach allows for an exploration of threshold and non-monotonic effects beyond those imposed by the logistic link function and the theoretical predictions given by Merton's (1974) distance-to-default model.

Our contribution can be summarized in four main findings. Firstly, the accuracy in the in-sample estimated absolute risk measure is enhanced in the logistic spline model. Increases in model fit (pseudo-R^2) is one manifestation, but perhaps more importantly, our 45-degree plots reveal that the failure probabilities are unbiased over the entire risk distribution, in contrast with the standard logistic case. Secondly, by using a very wide panel data set we are able to estimate separate models for the 18 years in our sample period and find that the estimated non-linear relationships are remarkably stable over time, i.e., they are a persistent feature. This finding is important for two reasons. It suggests that a model that accounts for non-linearities should be a superior forecasting device. It also suggests that these relationships are of a structural nature and hence provide stylized characterizations of financial ratios effects on firms' bankruptcy risk. Thirdly, the out-of-sample analysis confirms the spline model's predictive abilities. Thus, it outperforms the logistic model both in terms of relative risk ranking, and in the accuracy of the predicted absolute risk estimates. Also, the unbiased property across the entire risk distribution is preserved out-of-sample. Finally, our analysis document three interesting features for the leverage ratio and earnings ratio. Consistent with Merton's (1974) distance-to-default model, we find that the marginal reduction in failure risk from reducing firms' debts is at its highest for firms that are on the edge of balance-sheets-based insolvency. We also document that low-leveraged firms exhibit increased failure risk, possibly reflecting credit rationing. Furthermore, firms reporting earnings ratios above 15 percent are associated with higher failure risk and we find evidence suggesting that this is driven by high cash-flow risk in combination with limited and costly external financing.

Our best-fitting model falls short of Shumway's (2001) multi-period logistic model in terms of explanatory power, since our sample, being almost exclusively composed of private firms, cannot consider market determined variables. Nevertheless, we think that the approach suggested here would improve any bankruptcy prediction model. There is no reason, a priori, to not think that also market based information is non-linearly related to firm failure. Hence, a non-linear approach is of general interest, well beyond private firms and financial ratios.

References

Agarwal, Rajshree, and Michael Gort, 1996, The Evolution of Markets and Entry, Exit and Survival of Firms, *The Review of Economics and Statistics*, 78, 489-498.

Almeida, Heitor, Murillo Campello, and Michael S. Weisbach, 2004, The Cash Flow Sensitivity of Cash, *Journal of Finance*, 59, 1777-1804.

Altman, Edward I., 1968, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *Journal of Finance*, 23, 589-611.

Altman, Edward I., 1971, Railroad bankruptcy propensity, *Journal of Finance*, 26, 333-345.

Altman, Edward I., 1984, The success of business failure prediction models, *Journal of Banking and Finance*, 8, 171-198.

Altman, Edward I., 2000, Predicting Financial Distress of Companies: Revisiting the Z-Score and Zeta Models, Working Paper, Stern School of Business, New York University.

Altman, Edward I., and Anthony Saunders, 1997, Credit risk measurement: developments over the last twenty years, *Journal of Banking and Finance*, 21, 1721-42.

Beaver, William, 1966, Financial Ratios as Predictors of Failure, *Journal of Accounting Research,* 4, 71-111.

Berg, Daniel, 2007, Bankruptcy prediction by generalized additive models, *Applied Stochastic Models in Business and Industry*, 23, 129–43.

Bharath, Sreedhar T., and Tyler Shumway, 2008, Forecasting Default with the Merton Distance to Default Model, *Review of Financial Studies*, 21, 1339-1369.

Campbell, John Y., Jens Hilscher, and Jan Szilagyi, 2008, In Search of Distress Risk, *Journal of Finance*, 63, 2899-2939.

Chava, Sudheer, and Robert A. Jarrow, 2004, Bankruptcy Prediction with Industry Effects, *Review of Finance*, 8, 537-569.

Dakovic, Rada, Claudia Czado, and Daniel Berg, 2010, Bankruptcy prediction in Norway: a comparison study, *Applied Economics Letters*, 17, 1739–1746.

Denison, David G. T., Bani K. Mallick, and Adrian F. M. Smith, 2003, *Bayesian methods for nonlinear classification and regression*, Wiley, Chicester.

Duffie, Darell, Leandro Saita and Ke Wang, 2007, Multi-Period Corporate Default Prediction, *Journal of Financial Economics,* 83, 635-665.

Frydman, Halina, Edward I. Altman, and Duen-Li Kao, 1985, Introducing recursive partitioning for financial classification: the case of financial distress, *Journal of Finance*, 40, 269-291.

Hastie, Trevor J., and Robert J. Tibshirani, 1990, Generalized Additive Models, Chapman & Hall, Florida.

Hastie, Trevor, Robert J. Tibshirani, and Jerome Friedman, 2009, The Elements of Statistical Learning, Second edition, Springer Series in Statistics.

Hosmer, David W., and Stanley Lemesbow, 2000, *Applied Logistic Regression*, Second edition, Wiley, New York.

Jacobson, Tor, Jesper Lindé, and Kasper Roszbach, 2011, Firm Default and Aggregate Fluctuations, *Journal of the European Economic Association,* forthcomming.

Jensen, Michael, 1986, Agency costs of free cash flow, corporate finance and takeovers, *American Economic Review*, 76, 323-329.

Jovanovic, Boyan, 1982, Selection and the Evolution of Industry, *Econometrica,* 50, 649-670.

Lang, Larry, Anette Poulsen, and René Stulz, 1995, Asset sales, firm performance, and the agency costs of managerial discretion, *Journal of Financial Economics*, 37, 3 – 37.

Li, Qi and Jeffrey S. Racine, 2007, *Nonparametric econometrics: Theory and practice*, Princeton University Press, New Jersey.

Liang, Kung-Yee. and Scott L. Zeger, 1986, Longitudinal data analysis using generalized linear models, *Biometrika*, 73, 13-22.

McFadden, Daniel, (1974), The Measurement of Urban Travel Demand, *Journal of Public Economics*, 3, 303-328.

Merton, Robert C., 1974, On the Pricing of Corporate Debt: the Risk Structure of Interest Rates, *Journal of Finance,* 29, 449-470.

Minton, Bernadette A., and Catherine Schrand, 1999, The impact of cash flow volatility on discretionary investment and the costs of debt and equity financing, *Journal of Financial Economics*, 54, 423-460

Nance, Deana R., Clifford W. Smith, Jr., and Charles W. Smithson, 1993, On the Determinants of Corporate Hedging, Journal of Finance, 48, 267-284.

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

Opler, Tim, Lee Pinkowitz, René Stulz, and Rohan Williamson, 1999, The determinants and implications of corporate cash holdings, *Journal of Financial Economics*, 52, 3-46.

Ruppert, David, Wand, M. and Carroll, R. J., 2003, Semiparametric Regression. Cambridge University Press.

Schwarz, Gideon., 1978, Estimating the dimension of a model, *Annals of Statistics*, 6, 461-464.

Shumway, Tyler, 2001, Forecasting bankruptcy more accurately: a simple hazard model, *Journal of Business*, 74, 101-124.

Smith, Michael, and Robert Kohn, 1996, Nonparametric regression using Bayesian variable selection, *Journal of Econometrics*, 75, 317-343.

Stiglitz, Joseph E., and Andrew Weiss, 1981, Credit Rationing in Markets with Imperfect Information, *American Economic Review*, 71, 393-410

Zmijewski, Mark. E., 1984, Methodological issues related to the estimation of financial distress models, *Journal of Accounting Research*, 22, 59-82.

Figure 1: The graphs illustrate the realized bankruptcy frequencies (circles) and estimated bankruptcy probabilities, obtained from univariate logistic (dashed line) and univariate logistic spline models (solid line), for the five firm-specific variables over the full sample period 1991-2008. For each variable the data has been sorted and grouped into 300 equally sized groups. For each group, we calculate the realized bankruptcy frequency as the share of bankrupt firms over all firms, and then an average of the observations of the firm-specific variable at hand. The 300 group-data points are then plotted against each other to yield the circles. For each variable the reported logistic spline fit is calculated based on a univariate spline model incorporating 11 knots, and likewise, the logistic fit is based on a univariate logistic model. The shaded areas in the graphs mark out regions containing 90 percent of the observations. The thicker tick marks on the horizontal axes indicate the location of the spline knots. *Size* is log of total sales. *Age* measures log of firm age (+ 1 year) in number of years sincefirst registrated as a corporate.

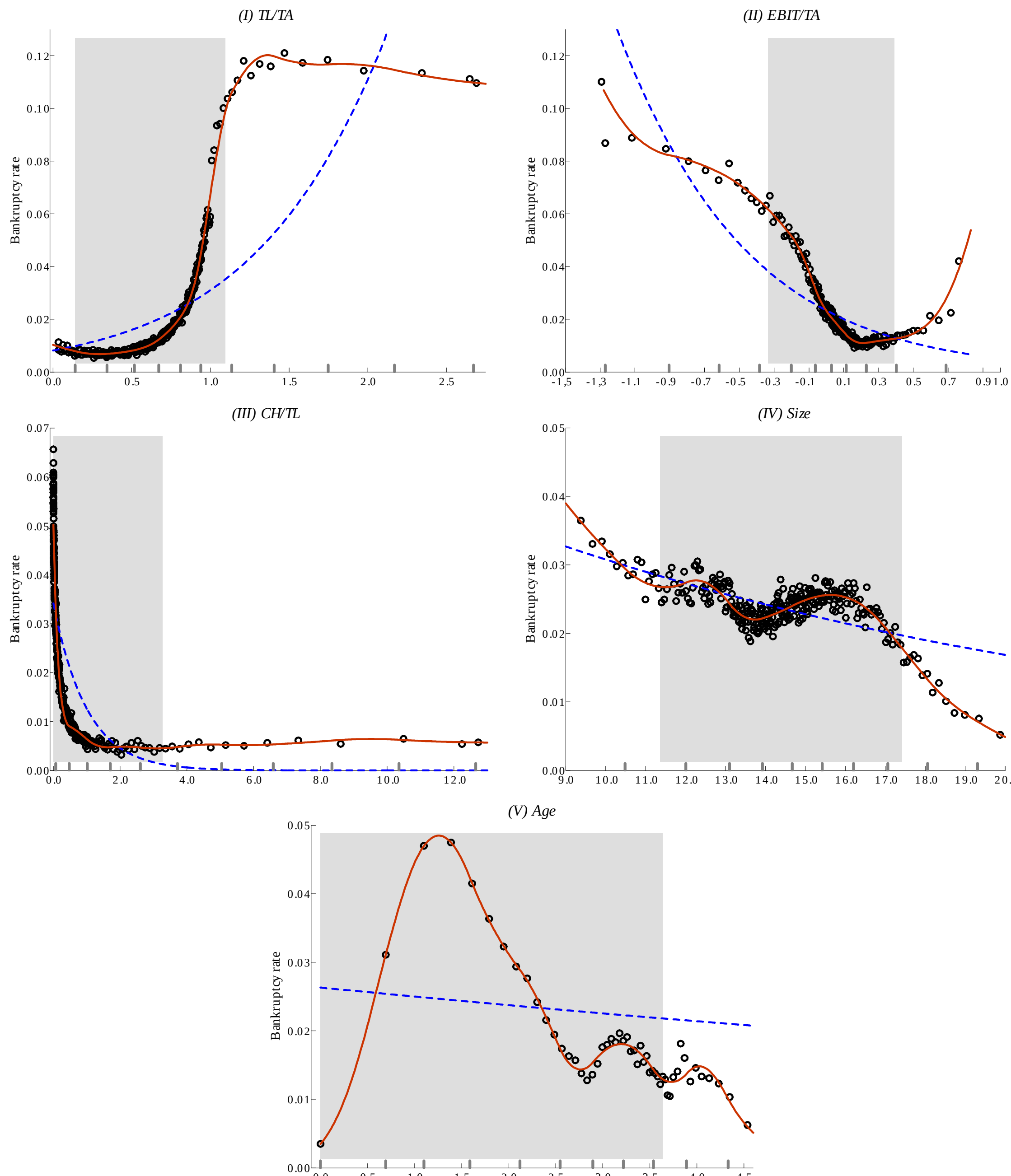

Figure 2: The graphs illustrate in-sample estimated bankruptcy probabilities versus realized bankruptcy frequencies for the period 1990-2008. The graphs correspond to, Panel A: the Private firm model, and Panel B: the Extended private firm model, in Table 2. For each model we sort all firm-year observations with respect to the size of their estimated bankruptcy probability, and divide them into equally sized percentiles. We then calculate the average probability of bankruptcy and the share of realized bankruptcies within each percentile. The circles correspond to the pairs of estimated bankruptcy probabilities versus realized bankruptcy shares, and the 45-degree line illustrates a perfect fit. We have graphed the relationships using a probability scale (left-hand side), and a logarithmic scale (right-hand side).

Panel A

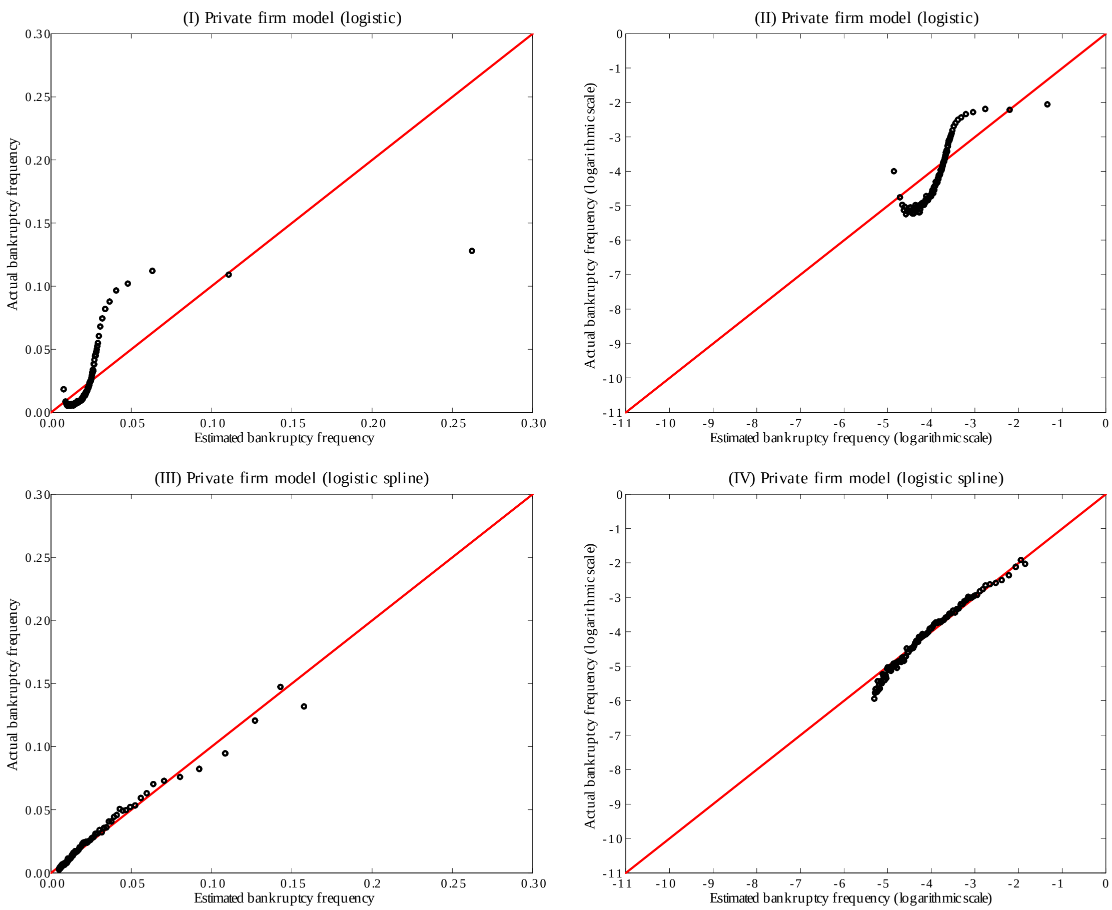

Figure 2 continues on the next page.

Figure 2 continued

Panel B

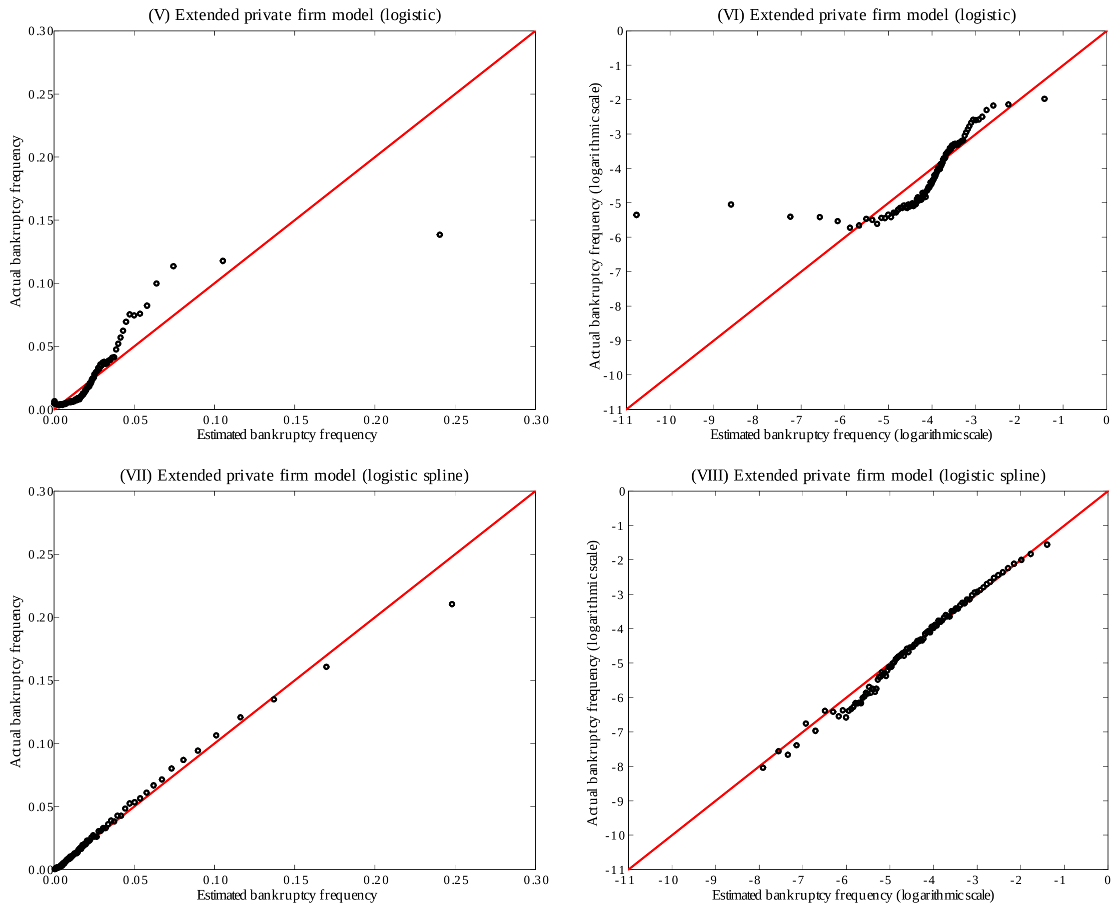

Figure 3: The graphs illustrate the conditional mean function (left-hand side panels) and the derivative of the logit function ($d\theta/dx_j$; right-hand side panels), across all segments, as given by the logistic spline version of the Extended private firm model in Table 2. The conditional mean function for each explanatory variable is calculated by setting the other variables to their sample means. The dashed areas correspond to the 95 and 99 percent confidence intervals. The intervals between the vertical dashed lines in the graphs mark out regions containing 90 percent of the observations. The thicker tick marks on the horizontal axes indicate the location of the spline knots. *Size* is log of total sales. *Age* measures log of firm age (+ 1 year) in number of years since first registrated as a corporate. The confidence bands are calculated using a sample-size adjustment for the covariance matrix where the elements are scaled by the average number of firm-years per firm, so as to account for the dependence over time in firms' observations, c.f., Shumway (2001).

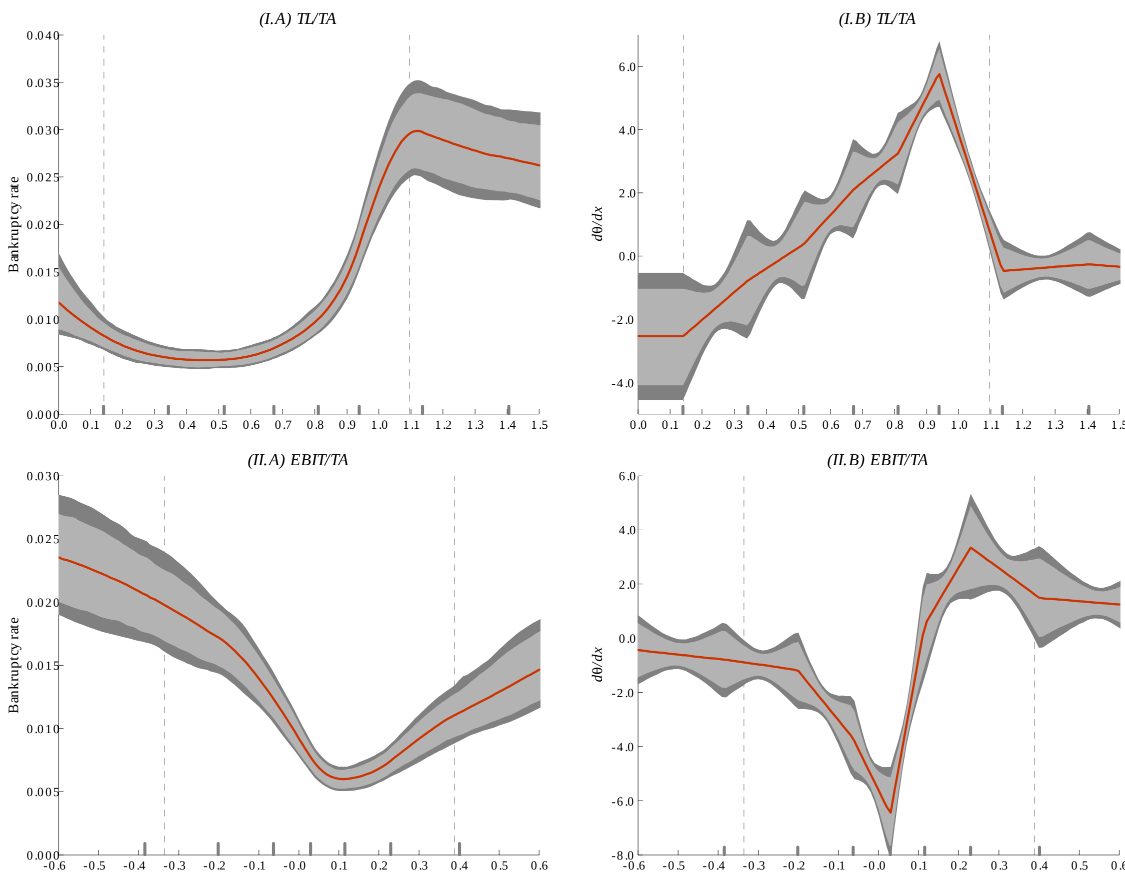

Figure 3 continues on the next page.

Figure 3 continued

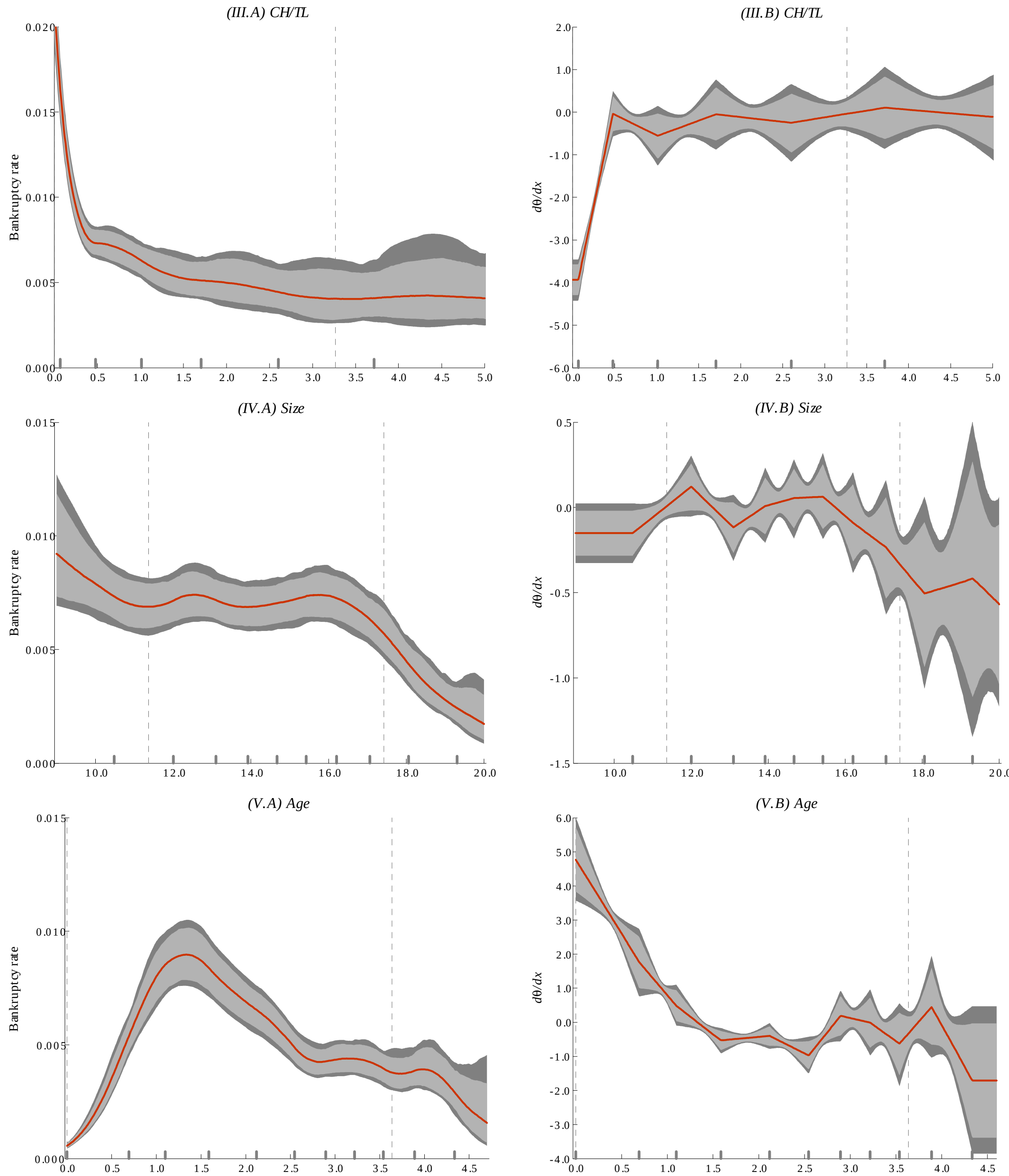

Figure 4: The graphs illustrate the derivative of the logit function ($d\theta/dx_j$) across all segments, for each variable and year, in the period 1991-2008. The 18 years have been divided into two regimes: the banking crises period in 1991-1995 (dashed lines) and then the remaining period 1996-2008 (solid lines). The specifications of the 18 yearly models coincide with that of Model (IV) in Table 2, except now excluding the macroeconomic variables and setting the intercept to [-3.956-0.078$*\triangle GDPG_{t-1}$+0.057$*REPO_{t-1}$], where the coefficients correspond to a model where only the two macroeconomic variables are included, so as to take account of the time-varying mean bankruptcy risk. The number of knots is optimally determined for each year, ranging between 3 and 7. $Size$ is log of total sales. Age measures log of firm age (+ 1 year) in number of years since first registrated as a corporate.

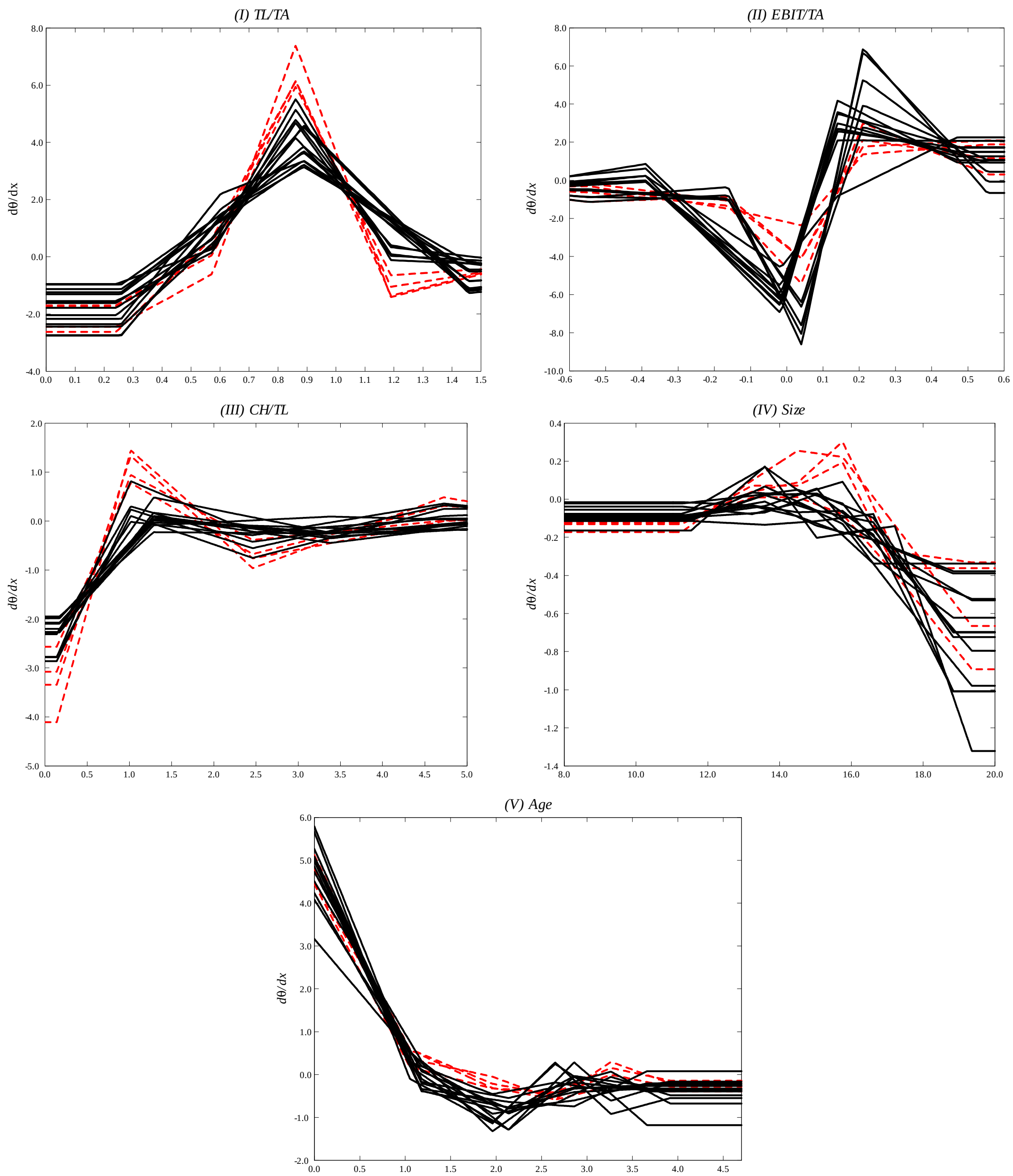

Figure 5: The graphs illustrate out-of-sample predicted bankruptcy probabilities versus realized bankruptcy frequencies for the period 2000-2008. Predicted probabilities are generated by the Extended private firm model, estimated for the in-sample period 1991-1999 using the logistic and logistic spline approaches. For each model we sort all firm-year observations with respect to the size of their estimated bankruptcy probability, and divide them into equally sized percentiles. We then calculate the average probability of bankruptcy and the share of realized bankruptcies within each percentile. The circles correspond to the pairs of estimated bankruptcy probabilities versus realized bankruptcy shares, and the 45-degree line illustrates a perfect fit. We have graphed the relationships using a probability scale (left-hand side), and a logarithmic scale (right-hand side).

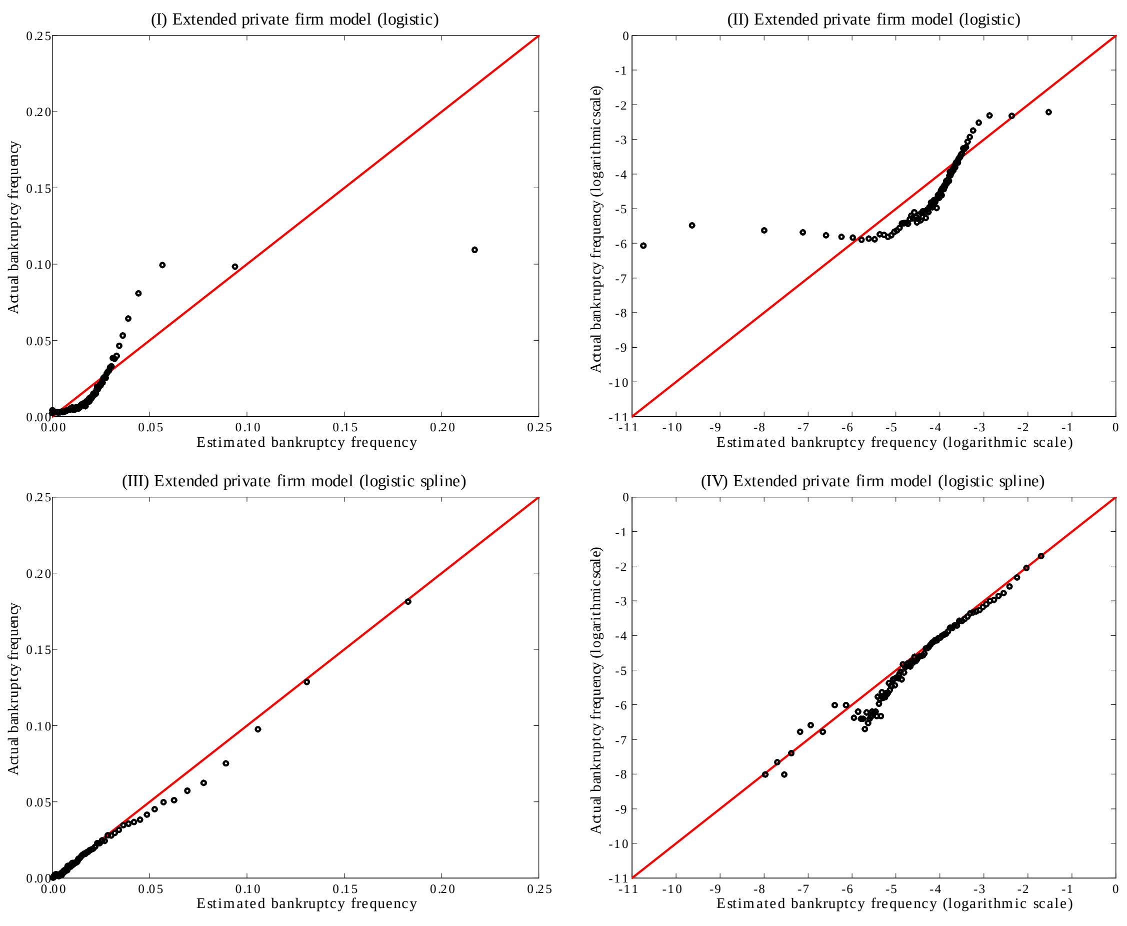

Table 1: This table reports descriptive statistics for the winsorized firm data, 1991-2008. The ratios are total liabilities over total assets (leverage ratio, TL/TA); EBIT over total assets (earnings ratio, $EBIT/TA$); total liabilities over total assets (leverage ratio, TL/TA); cash and liquid assets in relation to total liabilities (cash ratio, CH/TL). *Size* is total sales in SEK 1,000. *Age* measures firm age in number of years since first registrated as a corporate. Differences in means and medians between the samples of non-bankrupt and bankrupt firm-years are assessed using a Student's t-test and Wilcoxon-Mann-Whitney test. The star (*) indicates whether the differences between the corresponding mean- and median values are significantly different at the 1% level.

	Non-bankrupt firm-years					Bankrupt firm-years						
Variables	Mean	Median	Std	Min	Max	Mean		Median		Std	Min	Max
TL/TA	0.663	0.668	0.366	0.034	2.692	1.010	*	0.928	*	0.526	0.034	2.692
EBIT/TA	0.052	0.061	0.250	-1.292	0.761	-0.104	*	0.000	*	0.392	-1.292	0.761
CH/TL	0.791	0.208	1.751	0.000	12.727	0.222	*	0.023	*	0.903	0.000	12.727
Size	18,562	1,683	362,734	10	108,000,000	6,175	*	1,578	*	33,073	10	4,610,817
Age	11	7	13	0	110	9	*	5	*	10	0	110
#Obs			3,943,047						96,091			

Table 2: This table reports coefficient estimates for the logistic and logistic spline models ($S = 2$ and $M = 10$), for the full sample period 1991-2008. The financial ratios are total liabilities over total assets (leverage ratio, TL/TA); EBIT over total assets (earnings ratio, $EBIT/TA$); cash and liquid assets in relation to total liabilities (cash ratio, CH/TL). $Size$ is log of total sales. Age measures log of firm age (+ 1 year) in number of years since first registrated as a corporate. $\triangle GDP$ is growth in real gross domestic output and $REPO\text{-}RATE$ is the short-term interest rate set by the Swedish central bank. $d\theta/dx$ denotes the estimated slope coefficients on the log odds scale ($(1/NT)\sum_{i=1}^{N}\sum_{t=1}^{T} d\theta_{i,t}/dx_j$). dp/dx denotes the average marginal effect. $\#Signf$ denotes the number of significant spline coefficients for each explanatory variable. The pseudo-R^2 is calculated according to McFadden (1974) and ROC denotes the area under the ROC curve. Reported t-values are calculated with standard errors obtained after a sample size adjustment where the covariance matrix is scaled by the average number of firm-years per firm, so as to account for the dependence between firm-year observations, c.f., Shumway (2001). All coefficients for the logistic models are statistically distinct from 0 at the 1% level. The coefficients obtained for the spline basis expansion of each firm-specific explanatory variable ($\eta_{j,1}, \eta_{j,2}, ..., \eta_{j,1+M}$ in Eq. (8)) are jointly distinct from 0 at the 1% level.

	Logistic				Logistic Spline			
	(I) Private Firm Model		(II) Extended Private Firm Model		(III) Private Firm Model		(IV) Extended Private Firm Model	
Variables	$d\theta/dx$	dp/dx	$d\theta/dx$	dp/dx	$d\theta/dx$	*#Sign f*	$d\theta/dx$	*#Sign f*
Intercept	-4.648		-4.013		-4.690		-4.349	
	(-250.6)		(-42.6)		(-6.1)		(-3.6)	
TL/TA	1.170	0.023	0.929	0.014	2.763	7	1.563	6
	(66.6)		(45.7)					
$EBIT/TA$	-0.665	-0.013	-0.746	-0.011	-2.610	8	-1.840	8
	(-24.0)		(-25.5)					
CH/TL			-0.510	-0.007			-2.186	6
			(-21.0)					
$Size$			-0.041	-0.001			-0.039	4
			(-7.1)					
Age			0.064	0.001			0.451	7
			(7.1)					
ΔGDP			-6.982	-0.107			-5.738	
			(-11.1)				(-9.105)	
REPO-RATE			4.840	0.074			3.495	
			(16.8)				(11.723)	
Pseudo-R^2	0.063		0.084		0.106		0.162	
ROC	0.758		0.765		0.771		0.825	
#Knots (M)	-		-		10		11	
#Bankruptcies					96,091			
#Obs					4,039,138			

Table 3: This table reports mean values for the winsorized firm data, 1991-2008. Panel A reports mean values for a set of variables where the firm-year observations are grouped with respect to low, medium, and high leverage ratios, TL/TA. Panel B reports mean values for a set of variables where the firm-year observations are grouped with respect to low, medium, and high earnings ratios, $EBIT/TA$. $\sigma_{EBIT/TA}(t-2,t)$ and $\sigma_{EBIT/TA}(t-4,t)$ are the three and five year firm-specific earnings ratio volatilities, respectively. $Size$ is total sales in SEK 1,000. $Spread$ is firm interest expenditure over total liabilities minus the short-term interest rate set by the central bank. $\triangle FIX(t-2,t)$ and $\triangle FIX(t-4,t)$ are the yearly changes in fixed assets (property, plants, and machinery), averaged over three and five years, respectively.

Panel A.

	Non-bankrupt firm-years				Bankrupt firm-years			
		TL/TA:				*TL/TA:*		
	All	$< 30\%$	$30-60\%$	$60\% <$	All	$< 30\%$	$30-60\%$	$60\% <$
TL/TA	0.663	0.175	0.461	0.882	1.010	0.166	0.471	1.111
$Size$	18,562	10,593	18,671	20,501	6,175	2,957	3,113	6,668
$Spread$	0.019	0.066	0.017	0.012	0.024	0.158	0.041	0.018

Panel B.

	Non-bankrupt firm-years				Bankrupt firm-years			
		EBIT/TA:				*EBIT/TA:*		
	All	$< 0\%$	$0-15\%$	$15\% <$	All	$< 0\%$	$0-15\%$	$15\% <$
$EBIT/TA$	0.053	-0.202	0.067	0.299	-0.102	-0.347	0.060	0.360
$\sigma_{EBIT/TA}(t-2,t)$	0.106	0.154	0.064	0.141	0.179	0.226	0.078	0.261
$\sigma_{EBIT/TA}(t-4,t)$	0.121	0.167	0.080	0.157	0.199	0.239	0.101	0.276
$Size$	18,562	13,137	22,496	16,703	6,175	5,497	7,526	5,079
$Spread$	0.019	0.023	0.014	0.026	0.024	0.022	0.014	0.058
$\Delta FIX(t-2,t)$	-0.016	-0.077	0.006	0.006	-0.036	-0.056	0.016	-0.069
$\Delta FIX(t-4,t)$	-0.006	-0.055	0.013	0.010	-0.020	-0.032	0.020	-0.057

Table 4: This table reports the out-of-sample accuracy of the logistic and logistic spline versions of the Private firm model and the Extended private firm model. The models are estimated for the in-sample period $1991-1999$ and then used to predict failure probabilities that are evaluated for the out-of-sample period $2000-2008$. The in-sample pseudo-R^2 coefficients are calculated according to McFadden (1997). The out-of-sample pseudo-R^2 coefficients are calculated as 1-L_1/L_0, where L_1 is the log likelihood obtained for the out-of-sample period using the in-sample estimates and L_0 is the log likelihood for an intercept model estimated for the out-of-sample period. The out-of-sample *ROC*-measures quantifies the area under the *ROC*-curve. The out-of-sample decile tests are obtained by sorting the estimated failure probabilities, in a descending order, and by computing the fraction of total number of bankruptcies in the different deciles of the sorted data.

	Private Firm Model		Extended Private Firm Model	
	Logistic	Logistic Spline	Logistic	Logistic Spline
In-sample, (1991-1999)				
Pseudo-R^2	0.056	0.099	0.071	0.151
#Knots (M)	-	10	-	10
#Bankruptcies	61,085			
#Obs	1,924,716			
Out-of-sample, (2000-2008)				
Pseudo-R^2	0.027	0.075	0.072	0.148
ROC	0.763	0.772	0.765	0.820
Decile:				
1	0.424	0.411	0.403	0.477
2	0.165	0.170	0.168	0.188
3	0.109	0.120	0.120	0.114
4	0.078	0.086	0.086	0.075
5	0.057	0.062	0.065	0.052
6-10	0.166	0.151	0.158	0.094
#Bankruptcies	35,006			
#Obs	2,114,422			

Earlier Working Papers:

For a complete list of Working Papers published by Sveriges Riksbank, see www.riksbank.se

Sveriges Riksbank
Visiting address: Brunkebergs torg 11
Mail address: se-103 37 Stockholm

Website: www.riksbank.se
Telephone: +46 8 787 00 00, Fax: +46 8 21 05 31
E-mail: registratorn@riksbank.se

ISSN 1402-9103 E-Print, Stockholm 2011