

Wahlstrøm, Ranik Raaen; Becker, Linn-Kristin; Fornes, Trude Nonstad

Article

Enhancing credit risk assessments of SMEs with non-financial information

Cogent Economics & Finance

Provided in Cooperation with:

Taylor & Francis Group

Suggested Citation: Wahlstrøm, Ranik Raaen; Becker, Linn-Kristin; Fornes, Trude Nonstad (2024) : Enhancing credit risk assessments of SMEs with non-financial information, Cogent Economics & Finance, ISSN 2332-2039, Taylor & Francis, Abingdon, Vol. 12, Iss. 1, pp. 1-33, <https://doi.org/10.1080/23322039.2024.2418910>

This Version is available at:

<https://hdl.handle.net/10419/321640>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>

Enhancing credit risk assessments of SMEs with non-financial information

Ranik Raaen Wahlstrøm, Linn-Kristin Becker & Trude Nonstad Fornes

To cite this article: Ranik Raaen Wahlstrøm, Linn-Kristin Becker & Trude Nonstad Fornes (2024) Enhancing credit risk assessments of SMEs with non-financial information, Cogent Economics & Finance, 12:1, 2418910, DOI: [10.1080/23322039.2024.2418910](https://doi.org/10.1080/23322039.2024.2418910)

To link to this article: <https://doi.org/10.1080/23322039.2024.2418910>



© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 05 Nov 2024.



[Submit your article to this journal](#)



Article views: 1486



[View related articles](#)



[View Crossmark data](#)



Citing articles: 2 [View citing articles](#)

Enhancing credit risk assessments of SMEs with non-financial information

Ranik Raaen Wahlstrøm , Linn-Kristin Becker and Trude Nonstad Fornes

NTNU Business School, Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT

We investigate non-financial variables for predicting bankruptcy in small and medium-sized enterprises (SMEs). The variables encompass management, board and ownership structures and are sourced from universally accessible information, rendering them available to all stakeholders and allowing for the analysis of all SMEs within a market. Using a large and recent sample of SMEs, we empirically examine the variables that predict bankruptcy over time horizons of one, two and three years. Our analysis incorporates state-of-the-art discrete hazard models, the least absolute shrinkage and selection operator (LASSO), extreme gradient boosting (XGBoost), adaptive boosting (AdaBoost), bagging and random forest. We also test robustness using balanced datasets generated using the synthetic minority oversampling technique (SMOTE). We find that including non-financial variables enhances bankruptcy predictions compared to using financial variables alone. Moreover, our results show that among our variables, the most significant non-financial predictors of bankruptcy are the age of chief executive officers (CEOs), chairpersons and board members, as well as ownership share and place of the board members' residences.

IMPACT STATEMENT

This research highlights the critical role of integrating non-financial information with traditional financial variables to enhance the prediction of SME bankruptcy. While financial variables remain the most significant predictors, the inclusion of nonfinancial factors significantly improves the accuracy of the assessment of SMEs' financial health, benefiting investors, policymakers, and financial institutions by enabling better risk management and more effective support schemes. The findings underscore the importance of diverse board composition and local engagement in reducing bankruptcy risk, offering valuable insights for improving SME governance and stability.

ARTICLE HISTORY

Received 17 July 2024
Revised 2 October 2024
Accepted 15 October 2024

KEYWORDS

Small and medium-sized enterprises (SMEs);
bankruptcy prediction;
corporate governance;
non-financial predictors;
LASSO

JEL CLASSIFICATION CODES

C25; C53; G17; G20;
G33; M41


SUBJECTS

Economics; Finance;
Business, Management and
Accounting

1. Introduction

Improved assessments of the financial standing of small and medium-sized enterprises (SMEs) are important for many stakeholders. For example, better assessment helps SMEs access formal capital. This is crucial because SMEs exhibit a higher level of information asymmetry than larger corporations, making it more difficult for them to access formal capital, as they face challenges in conveying their standing to lenders and investors (Beck & Demirguc-Kunt, 2006; Beck et al., 2008; Berger & Udell, 1998, 2002; Block et al., 2018; Masiak et al., 2019). This is particularly pressing now as they face challenges due to the war in Ukraine and being at the center of the economic shocks of the COVID-19 pandemic (G20 & OECD, 2015; Organisation for Economic Co-operation and Development, 2022). SMEs also benefit from an improved assessment of their financial standing when conveying credibility to customers and suppliers. Furthermore, banks and investors benefit from better assessments of SMEs' financial standing, as they improve risk management and reduce lending errors through more accurate evaluations of potential

CONTACT Ranik Raaen Wahlstrøm  ranik.raaen.wahlstrom@ntnu.no  NTNU Business School, Norwegian University of Science and Technology, Trondheim, Norway

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/23322039.2024.2418910>.

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

© 2024 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

new borrowers and enhanced monitoring of existing borrowers. Additionally, better assessment enables a more precise determination of the risk-weighted value of loan portfolios. In summary, this may increase lending to SMEs, which significantly positively affects the profitability and diversification of loan portfolios (Altman & Sabato, 2007; Dietsch & Petey, 2004). Moreover, assessing SMEs' financial standing is important for regulators, who perform on-site supervision of banks by analyzing their loan portfolios. Similarly, enhanced assessments of SMEs' financial standing facilitate a better analysis of the financial standing of large populations of firms encompassing entire markets. This is crucial for investors in these markets and policymakers when designing support schemes for businesses or predicting losses in tax revenues resulting from bankruptcy. In summary, improved assessments of SMEs' financial standing benefit the economy as a whole, especially because SMEs make up a large proportion of the economy.¹

Firm's standings are typically evaluated through credit risk modeling, which commonly relies on either historical accounting-based financial variables or securities market information (Altman et al., 2010). Early studies on credit risk modeling used only accounting-based financial variables (eg Altman, 1968; Beaver, 1966). However, using only accounting-based financial variables for credit risk modeling can be problematic for several reasons (Agarwal & Taffler, 2008; Hillegeist et al., 2004). First, because accounts represent the past, they may not be useful for predicting future credit risk. Second, the true value of firms' assets may differ from the book value because of accounting principles and conservatism. This may lead to an incorrect assessment of a firm's standing. Third, accounts are subject to management manipulation. Finally, accounts are prepared on a going-concern basis and have limited utility in predicting bankruptcy. SMEs may be particularly exposed to these challenges because they usually have less detailed and transparent accounting-based financial data than larger corporations (Berger & Frame, 2007). Moreover, as SMEs have fewer obligations regarding accounting data disclosure, inherently smaller accounting figures and are more vulnerable to external events than larger companies, financial ratios may be weak predictors of SME bankruptcy (Ciampi, 2015; Ciampi et al., 2021).

The limitations of using accounting-based financial variables can be overcome by using market-based variables. The literature has proven that such variables effectively increase the power of bankruptcy prediction models (Beaver et al., 2005; Chava & Jarrow, 2004; Hillegeist et al., 2004; Merton, 1974; Shumway, 2001; Tian et al., 2015). However, the market data for SMEs are unavailable. Thus, it is crucial to investigate non-financial variables to predict the bankruptcy of such companies (Altman et al., 2010; Ciampi et al., 2020, 2021).

This study investigates the non-financial predictors of bankruptcy in privately held SMEs. Soft non-financial information, context-dependent qualitative data that are not easily transferable, has been found to improve credit assessments (Cornée, 2019; Grunert et al., 2005). However, such information is typically exclusively accessible to banks that acquire it through close relationships with borrowers (Cornée, 2019; Liberti & Petersen, 2019). Moreover, soft non-financial information may be unsuitable for major decisions involving multiple decision-makers because of the challenges in transferring it from the collector to others and for minor decisions due to the high labor costs associated with its collection (Cornée, 2019). However, hard non-financial information, which refers to quantitative, explicit and context-independent data, is usually documented as numbers or can be easily converted into a numerical form, ensuring that it can be communicated to others without any loss of detail. Related studies that use hard non-financial information to assess SME credit risk include Wilson and Altanlar (2014), who examined newly incorporated companies, including SMEs and other firms, and found that the board of directors' characteristics contribute to bankruptcy prediction. However, they do not consider financial variables because they focus solely on newly established firms with no prior accounts available. In contrast, we focus on improving credit assessments for all SMEs in an economy using all available bankruptcy predictors derived from both financial and non-financial information. Another study related to ours is that of Altman et al. (2023a, 2023b), who found that models based on financial variables show improved predictive power when payment behavior, management-related and employee-related variables are incorporated. Moreover, Ciampi (2015) demonstrated that including management-related variables in existing models based on financial variables enhances the prediction of default among small enterprises. However, both of these studies use non-financial variables that are not readily available to all stakeholders and consequently analyze subsamples of the economy. For instance, employee-related information about SMEs is typically only available internally within SMEs, whereas payment behavior variables are usually only accessible to banks that have long-standing customer relationships with the SMEs. Moreover, even for banks, payment behavior variables are not always available—for example, when assessing

loan applications from new borrowers. These limitations make such variables unsuitable for assessing large populations of SMEs, such as when investors want to assess the financial standing of SMEs in a larger population or when policymakers are designing business support schemes or forecasting losses in tax revenues resulting from bankruptcies. By contrast, we examine non-financial predictors derived from universally available information on all SMEs. These predictors enable stakeholders to assess the financial standing of SMEs across an entire economy. Thus, we contribute to the literature by investigating non-financial predictors of SME bankruptcy that not only assist banks in reducing lending errors and determining risk-weighted assets but also support other users of bankruptcy prediction models, including regulators involved in determining banks' capital requirements, policymakers, private and public investors and credit rating agencies (Altman et al., 2017; Rajan et al., 2015). Based on this, we derive the following research question:

Can non-financial information available for all SMEs in an economy improve assessments of their collective financial standing?

In general, by focusing on privately held SMEs, we address an important gap in the literature, as most existing research on credit risk assessment investigates large corporations (da Silva Mattos & Shasha, 2024; Kuizinienė et al., 2022; Matenda et al., 2022; Zhao et al., 2024). This focus is crucial given the economic significance of SMEs and their distinct financial frameworks and challenges compared with larger corporations. Overall, consistent with the future research avenue proposed by the existing literature on credit risk modeling (Ciampi et al., 2021; Habib et al., 2020), we make a valuable contribution by demonstrating how the integration of non-financial variables into bankruptcy prediction models mitigates the information asymmetry surrounding SMEs and enhances the ability of all stakeholders to evaluate their standing.

We employ a new and unique sample of 818,927 SME financial statements from 2014 to 2019. For each financial statement, we derive the common financial variables found in Altman (1968), Altman and Sabato (2007) and Paraschiv et al. (2023) as benchmarks, as well as 20 non-financial variables representing management, board and ownership structures. We use these variables to predict SME bankruptcy over one, two and three years. We employ the least absolute shrinkage and selection operator (LASSO) method to explore the importance of the variables and select the most appropriate variable sets. This method is commonly used in finance literature (eg Calomiris & Mamaysky, 2019; Chae, 2024; Chincio et al., 2019; Coad & Srhoj, 2020; Hautsch et al., 2015; Ogneva et al., 2020; Tian et al., 2015); and is found by Paraschiv et al. (2023) to be superior to other methods for selecting corporate bankruptcy predictors. This method has also been used in recent literature on SME bankruptcy and default prediction (Altman et al., 2023; Paraschiv et al., 2023). We evaluate the selected variable sets in- and out-of-sample when used in discrete hazard models with logistic regression (LR) (Shumway, 2001).

Our empirical results show that financial variables are important predictors of SME bankruptcy. Furthermore, we find that both the in-sample fit and out-of-sample prediction performance improve when non-financial variables are included. Among these, the age of chief executive officers (CEOs), chairpersons and board members, along with the ownership share of board members and whether they reside in the same county as the SMEs' headquarters (HQs), are the most important non-financial bankruptcy predictors. However, we also find that bankruptcy prediction models consisting solely of non-financial variables do not achieve acceptable performance. This underscores the importance of financial variables and aligns with previous studies on predicting bankruptcy in larger companies (eg Liang et al., 2020).

Our findings are robust across 18 permutations of three financial variable sets as benchmarks, two years as test populations (2018 and 2019) and three different horizons for predicting bankruptcy (one, two and three years). Furthermore, our results are robust to using balanced datasets generated with the synthetic minority oversampling technique (SMOTE). Finally, our findings are robust to employing the machine learning methods extreme gradient boosting (XGBoost), adaptive boosting (AdaBoost), bagging and random forest. These methods model nonlinear relationships that have been shown to substantially improve bankruptcy predictions (Lohmann et al., 2023; Lohmann & Ohliger, 2018).

2. Background

The initial discussions on bankruptcy prediction in the literature focused on analyzing companies' accounting figures (Smith & Winakor, 1930), while Beaver (1966) demonstrated how individual financial

ratios can predict company bankruptcy in univariate models. Altman (1968) introduced the first multivariate bankruptcy prediction model, the Z-score model, which continues to be widely used by practitioners and academics (e.g., Blöchlinger & Leippold, 2018; Campello et al., 2018; Chang et al., 2019; Chava & Jarrow, 2004; Tian et al., 2015). The model comprises five financial variables that are primary aspects of a company's financial profile: liquidity, profitability, leverage, solvency (coverage) and activity. Altman's (1968) model was initially designed to predict the bankruptcy of large listed companies and includes a variable derived in part from the market value of equity. However, it was later refined to target private firms by incorporating only financial variables derived exclusively from accounting values (Altman et al., 1977, 2019).

Edmister (1972) was the first to highlight that early bankruptcy prediction models largely ignored SMEs. Motivated by this, he developed a model to predict small business defaults. Altman and Sabato (2007) expand on Edmister's (1972) work by developing a bankruptcy prediction model specifically for SMEs. Similar to the variable set in Altman (1968), the set of variables in Altman and Sabato (2007) consisted of five financial variables categorized as the main aspects of a company's financial profile. Paraschiv et al. (2023) also considered SME bankruptcy prediction and employ the LASSO method to empirically identify the set of ten variables out of 155 financial variables that yield the best predictions. The authors further demonstrated that these variables improve credit risk assessments, resulting in significantly higher bank profits. The ten selected variables capture SMEs' leverage, liquidity, solvency (coverage), age and profitability. Moreover, financial variables have been applied to derive proxies for earnings management, which have been found to improve SME bankruptcy predictions (Séverin & Veganzones, 2021).

In the context of non-financial bankruptcy predictors, we differentiate between those derived from soft and hard non-financial information (Cornée, 2019; Liberti & Petersen, 2019). Soft non-financial information refers to context-dependent qualitative information that is not easily transferable and is typically collected by banks about borrowers through close relationships. The literature suggests that such information is valuable for assessing the credit risk of small and opaque borrowers. For example, Berger and Udell (2002) argued that relationship lending, conditional on soft non-financial information, reduces information problems in small firms. Furthermore, Berger et al. (2005) found that smaller banks lend more to smaller firms because they have a comparative advantage in collecting and acting on soft non-financial information. Moreover, Stein (2002) suggested that consolidation in the banking industry leads to a decline in small business lending because soft information cannot be credibly transmitted within larger hierarchies. This is consistent with the findings of Rajan et al. (2015) that the models used for predicting the defaults of securitized subprime mortgages before the U.S. subprime mortgage crisis in 2007 severely underestimated the likelihood of default for opaque borrowers. Furthermore, using a sample of German SMEs, Grunert et al. (2005) found evidence suggesting that combining financial and soft non-financial variables leads to more accurate default predictions than using financial or soft non-financial variables. Cornée (2019) replicated Grunert et al. (2005) using credit files from a social relational bank specializing in providing external debt funding to genuinely small and opaque firms that prioritize social over financial goals. Cornée (2019) found that including soft non-financial variables yields better predictions than using only financial variables. However, compared to Grunert et al. (2005), he revealed that soft non-financial information tends to be more valuable than financial variables. He argued that this may be because he uses a smaller and more opaque sample of borrowers than that of Grunert et al. (2005). This suggests that the larger and more transparent the borrower, the lower the predictive value of soft non-financial information compared to financial variables.

In contrast, hard non-financial information is quantitative, explicit and context-independent. It is typically recorded as or can be easily reduced to numbers, making it easy to convey to others without losing information. Variables derived from such information include payment behavior variables, that is, indicators of late payments to creditors, which previous studies found to increase bankruptcy prediction accuracy compared to using only financial variables (eg Altman et al., 2023; Back, 2005; Laitinen, 1999; Wilson et al., 2000). Furthermore, the variables derived from hard non-financial information include management-related variables, which are indicators based on the characteristics of the management and board of directors. For example, Wilson and Altanlar (2014) reported that the board of directors' characteristics contribute to predicting the bankruptcy of newly incorporated companies with limited publicly available

data, including SMEs and other firms. Moreover, Ciampi (2015) found that default predictions among small enterprises improve when management-related variables are included in existing models based on financial variables. Altman et al. (2023a, 2023b) developed an SME default predictor by considering financial variables in combination with payment behavior, management-related and employee-related variables and found that the predictive power of models based on financial variables improves when introducing these additional variables. Other hard non-financial variables used to assess SME credit risk are derived from the local banking market (Arcuri & Levratto, 2020), published legal judgments (Yin et al., 2020) and bag-of-words models applied to content scraped from corporate websites (Crosato et al., 2023). Additional hard non-financial factors that may help convey the credit risk include receiving government grants. Srhoj et al. (2021a, 2021b) showed that receiving such grants can have a certification effect, increasing the likelihood of obtaining a long-term bank loan by being 'certified' by the government. Moreover, Lohmann and Ohliger (2020) argued that including qualitative information from firms' annual reports, such as structural and linguistic characteristics, enhances the discriminatory power of bankruptcy prediction models based on financial variables.

Integrating non-financial predictors can also introduce ethical challenges that must be carefully considered as they can perpetuate or exacerbate discrimination. For instance, Fuster et al. (2022) highlighted that machine learning techniques exhibit improved accuracy in predicting mortgage defaults when considering borrowers' ethnicity, leading to adverse consequences for specific borrower groups. The legal implications are also significant. Reliance on non-financial predictors could expose stakeholders to legal challenges if they lead to discriminatory practices in lending or investment decisions. Therefore, we argue that the non-financial variables used in our study should be employed to assess the collective financial standing of all firms in a larger population and should be used with caution when applied to decision support at the individual level. We further mitigate the bias in the non-financial variables by not considering them in isolation but with other financial and non-financial variables.

3. Variables

We investigate the 20 non-financial bankruptcy predictors presented in Table 1, both individually and in conjunction with the benchmark sets of accounting-based financial predictors in Altman (1968), Altman and Sabato (2007) and Paraschiv et al. (2023), as presented in Table 2.

The 20 non-financial variables listed in Table 1 are divided into three categories. First, they include management structure variables indicating the characteristics of SMEs' CEOs. Second, they constitute board structure variables indicating board characteristics that can significantly impact firm performance by supporting, controlling and evaluating management (Krause et al., 2016; Withers & Fitza, 2017). Finally, Table 1 includes ownership structure variables.

Our non-financial variables indicate the ages of the CEOs, chairpersons and board members. These are included because previous literature argues that persons of higher and lower ages in these roles can result in lower bankruptcy probabilities (Platt & Platt, 2012). On the one hand, older age may mean greater experience, which can help avoid bankruptcy. However, as older age may bring about conservatism, younger CEOs and board members may help avoid bankruptcy by being more willing to try new ideas and adapt to changing business environments. Platt and Platt (2012) found that firms that went bankrupt had, on average, younger CEOs and board members than those that did not go bankrupt. Furthermore, Antulov-Fantulin et al. (2021) investigated the predictability of municipal bankruptcy and found that the age of council members is one of the most important predictors.

Furthermore, we hypothesize that SME governance improves if its management and board are rooted in the same local community. Thus, we include in Table 1 variables that indicate whether CEOs, chairpersons and board members reside in the same county as the SME HQs. This is consistent with Wilson and Altanlar (2014), which found that bankruptcy was associated with fewer board members living in the same county as the company's registered address.

Table 1 also includes variables that indicate the genders of CEOs, chairpersons and board members. We include this, as previous literature suggests that women tend to be more risk-averse than men (eg Borghans et al., 2009; Charness & Gneezy, 2012; Dwyer et al., 2002; Jianakoplos & Bernasek, 1998). The literature on bankruptcy prediction confirms this assumption by suggesting that more women among

Table 1. Non-financial variables.

	Variable name	Description
Management structure	CEO age	The natural logarithm of the age of the CEO
	CEO woman	Dummy; one if the CEO is a woman
	CEO duality	Dummy for CEO duality; one if a single individual serves as both the CEO and chairperson of the board
	CEO on board	Dummy; one if the CEO sits on the board
	CEO county	Dummy; one if the CEO resides in the same county as the SME's headquarters (HQs)
Board structure	Two CEOs	Dummy; one if the company has two CEOs
	Chairperson age	The natural logarithm of the age of the chairperson of the board
	Chairperson woman	Dummy; one if the chairperson of the board is a woman
	Chairperson county	Dummy; one if the chairperson of the board resides in the same county as the SME's headquarters (HQs)
	Board size	The natural logarithm of the number of board members
	Board age avg	The natural logarithm of the average age among all board members
	Board age std	Standard deviation of age among board members
	Board women	Proportion of board members who are women
	Board county	Proportion of board members who resides in the same county as the SME's headquarters (HQs)
Ownership structure	Board non-owners	Proportion of board members who are not shareholders
	Ownership concentration 1	Average holdings of shareholders
	Ownership concentration 2	Standard deviation of shareholders' holdings
	Ownership CEO	Ownership share with the CEO
	Ownership chairperson	Ownership share with the chairperson of the board
	Ownership board	Ownership share with the board members

Note: The 20 non-financial variables of interest divided into categories representing the management, board and ownership structures.

Table 2. Benchmark financial variable sets.

	Variable name	Category
Altman (1968)	EBIT/total assets	Coverage
	Retained earnings/total assets	Profitability
	Sales/total assets	Activity
	Total equity/total liabilities	Leverage
	Working capital/total assets	Liquidity
Altman and Sabato (2007)	Current liabilities/total equity	Leverage
	EBITDA/interest expense	Activity
	EBITDA/total assets	Profitability
	Retained earnings/total assets	Coverage
	Short-term liquidity/total assets	Liquidity
Paraschiv et al. (2023)	(Current liabilities - short-term liquidity)/total assets	Leverage
	Accounts payable/total assets	Liquidity
	Dummy; one if paid-in equity is less than total equity	Solvency
	Dummy; one if total liability exceeds total assets	Leverage
	Interest expenses/total assets	Solvency
	Inventory/current assets	Liquidity
	Log(age in years)	Age
	Net income/total assets	Profitability
	Public taxes payable/total assets	Liquidity
	Short-term liquidity/current assets	Liquidity

Notes: The benchmark variable sets of accounting-based financial variables are those of Altman (1968), Altman and Sabato (2007) and Paraschiv et al. (2023). The variables are sorted alphabetically.

the management and board members yield lower probabilities of bankruptcy. For example, Cho et al. (2021) found that among Chinese firms during 2005–2016, the likelihood of bankruptcy was negatively associated with the proportion of female executives. C. J. García and Herrero (2021) found that the likelihood of bankruptcy among EU firms during 2002–2019 is negatively associated with greater board gender diversity. Wilson and Altanlar (2014) found that bankrupt companies have fewer female board members. Moreover, Antulov-Fantulin et al. (2021) show that the gender of council members is among the most important variables for predicting default in municipalities.

Moreover, we include in Table 1 variables indicating whether the CEO sits on the board, whether the SME has two CEOs and whether a single individual serves as the SME's CEO and chairperson, referred to as CEO duality (Krause et al., 2014). The existing literature highlights the potential advantages and disadvantages of CEO duality. On the one hand, it may be unfortunate because it reduces the independence between the board and management, reducing the board's ability to control the management (Jensen,

1993). However, CEO duality may also be beneficial if it leads to a more flexible leadership, which improves company efficiency (Combs et al., 2007; Dowell et al., 2011). The empirical evidence on how CEO duality affects the probability of bankruptcy is inconclusive. For instance, Ciampi (2015) found that CEO duality reduces the risk of default. Duru et al. (2016) found that CEO duality negatively affects firm performance when independent board members account for a small proportion of the board. However, this negative effect is mitigated as the proportion of independent board members increases. Eventually, the authors found that the impact becomes positive when the proportion of independent board members increases even more. By contrast, Platt and Platt (2012) and Manzaneque et al. (2016) found no significant effect of CEO duality on the likelihood of company bankruptcy.

We also include non-financial variables that measure board size and the proportion of board members who are not shareholders. Having more board members results in greater diversity and access to information, thereby increasing the board's efficiency and independence by boosting its ability to control management and direct the company in the right direction (Dalton et al., 1999; Manzaneque et al., 2016; Pearce & Zahra, 1992). However, board efficiency can decrease with more members if this results in a poorer flow of information (Guest, 2009). Platt and Platt (2012) and Manzaneque et al. (2016) found a negative relationship between the size of a company's board and the likelihood of bankruptcy. By contrast, Ciampi (2015) reported that board size did not significantly help predict defaults.

Furthermore, we include two non-financial variables that measure ownership concentration.² Theoretically, high ownership concentration has pros and cons (Ciampi, 2015). On the one hand, larger shareholders can benefit firm performance because they typically have more expertise relevant to the firm than smaller shareholders. They also have greater incentives to monitor management and boards effectively. However, they may also promote firm inefficiency if they exercise their control rights for private benefits. Ciampi (2015) found that a higher ownership concentration, where a single shareholder has the majority of shares, reduces the probability of an SME default. Tang et al. (2020) and Liang et al. (2020) revealed that the ownership stake of the majority of stakeholders helps predict corporate bankruptcy. In contrast, Manzaneque et al. (2016) suggested that ownership concentration does not significantly impact financial distress.

Finally, Table 1 presents three non-financial variables indicating the proportion of ownership by CEOs, chairpersons and board members. We include these variables because we hypothesize that the probability of bankruptcy is negatively associated with the ownership share of SMEs' key personnel; higher ownership should provide greater incentives to avoid bankruptcy. This corresponds with Lilienfeld-Toal and Ruenzi (2014), who found that CEO ownership increases the performance of listed firms.

4. Data

Our data consist of all unconsolidated annual financial statements of privately held Norwegian limited-liability SMEs for the accounting years 2014–2019. Although related studies also focus on data from only one economy and time period,³ we recognize the limitations of the generalizability of our analyses, specifically on Norwegian SMEs. However, we consider Norway over 2014–2019 to be a feasible test environment for several reasons. During our sample years, Norway experienced stable economic conditions and a lack of significant business cycle fluctuations, falling between the aftermath of the European debt crisis and the onset of the financial shock caused by the coronavirus crisis. Moreover, Norway is integrated with the European internal market as part of the European Economic Area (EEA) and is considered a high-income European country.⁴ Additionally, 99% of all Norwegian firms are SMEs, comparable to other European countries, employing 56% of the workforce.⁵

The annual financial statements for our data were provided by the Norwegian government agency Brønnøysund Register Centre (BRC).⁶ It is mandatory for all Norwegian limited liability companies to report their annual financial statements to the authorities, and these are subsequently stored with the BRC. Additionally, the BRC has provided us with the industry classification of the firms when reporting their annual financial statements by the Norwegian Standard Industrial Classification (SIC2007).⁷ Moreover, the BRC has supplied us with the dates of bankruptcy filings for all firms in our data that have filed for bankruptcy. Furthermore, information on firms' CEOs, chairpersons, board members, and owners used to derive our non-financial variables was provided by Enin AS.⁸

Following the EU recommendation 2003/361, we define a company as an SME if its total assets are less than EUR 43 million or its turnover is less than EUR 50 million.⁹ Furthermore, following Paraschiv et al. (2023), we restrict our sample to firms with total assets above NOK 500,000. This is because very small firms have little or no debt. In addition, very small firms are prone to data error problems and outliers because the accounting numbers used to derive variables are often zero. Moreover, we exclude financial statements of companies in the industries ‘Real estate activities’, ‘Financial and insurance activities’, ‘Public administration and defense’, ‘Electricity and gas supply’ and ‘Water supply, sewerage, waste’ as this is a common practice in the literature (eg Paraschiv et al., 2023). We also exclude financial statements of companies for investment and holding purposes only and those of companies that have not reported any industry. After these exclusions, our final sample consists of 818,927 financial statements from 192,118 SMEs. Breakdown by accounting year is shown in Table 3. Furthermore, Table A1 in the Appendix presents the breakdown per industry at the one- and two-digit levels. The industries with the highest representation at the one-digit level in our sample are ‘Construction’ and ‘Domestic trade, car repair shop’, each accounting for 22% of the unique firms in our sample. These industries also have the highest bankruptcy rates in our sample, except of ‘Accommodation, food service’, which has an even higher rate.

For our financial variables (see Table 2), we have complete information and do not have to deal with missing values. Before deriving the financial variables, we convert the accounts of all financial statements to NOK using the most recent daily exchange rate on or before the balance sheet date, sourced from Norges Bank.¹⁰ To mitigate the effect of outliers on financial ratios, we follow prior bankruptcy prediction studies by winsorizing them at the 1st and 99th percentiles within each accounting year (Chava & Jarrow, 2004; Paraschiv et al., 2023; Shumway, 2001; Tian et al., 2015). Furthermore, if the denominator for a financial ratio is zero, we set the variable value to zero if the numerator is also equal to zero and the 1st or 99th percentiles if the numerator is negative or positive, respectively. The variable ‘*log(age in years)*’ is log-transformed.

However, our non-financial variables (see Table 1) have missing values. For variables derived from several persons, such as ‘board age avg’ and ‘board county’, we calculate their values based on available data. Thus, these are missing only if the values for all the persons they are based are missing. Subsequently, we impute the missing values of all variables per accounting year using a common approach based on the *k*-nearest neighbor (*k*-NN) algorithm (Troyanskaya et al., 2001). When imputing missing values for continuous non-financial variables, we set them to the mean value of the three nearest observations in terms of the inverse of the Euclidean distance in the vector space of the financial variables in Paraschiv et al. (2023), as detailed in Table 2. We use the same procedure for missing values for the dummy variables, which need to be set to one or zero, but we use only the nearest observation for imputing.

Table 4 presents the mean, median and standard deviation for each non-financial variable across all 818,927 financial statements in our sample, as well as the subsets of observations categorized as non-bankrupted and observations categorized as bankrupted over one-year, two-year and three-year horizons. We observe that the average age among the CEOs and chairpersons are $e^{3.90} \approx 49$ and $e^{3.94} \approx 51$, respectively.¹¹ Furthermore, we observe that 15% and 12% of CEOs and chairpersons are women, whereas 16% of SMEs’ board members, on average, are women. Moreover, 78% of the CEOs sit on their company’s board, while the average board size is $e^{0.52} \approx 1.68$ board members. On average, 48% of

Table 3. The sample.

Accounting year	Financial statements	Number of companies	Bankruptcy frequency		
			One-year horizon	Two-year horizon	Three-year horizon
2014	124,944	124,944	0.42%	1.07%	1.05%
2015	129,323	129,323	0.38%	1.08%	1.11%
2016	133,858	133,858	0.43%	1.13%	1.17%
2017	139,068	139,068	0.42%	1.16%	1.10%
2018	143,802	143,802	0.44%	1.12%	0.77%
2019	147,932	147,932	0.29%	0.69%	0.18%
Total	818,927	192,118	0.39%	1.04%	0.88%

Note: Our sample of SMEs’ financial statements, number of unique companies and bankruptcy frequency over different horizons per accounting year.

Table 4. Descriptive statistics.

	Imputed	All			Non-bankrupted			Bankrupted over a one-year horizon			Bankrupted over a two-year horizon			Bankrupted over a three-year horizon		
		Mean	Median	Std	Mean	Median	Std	Mean	Median	Std	Mean	Median	Std	Mean	Median	Std
Management structure																
CEO age	10.01%	3.90	3.91	0.22	3.90	3.91	0.22	3.79	3.83	0.23	3.79	3.81	0.24	3.79	3.82	0.24
CEO woman	7.06%	0.15	0.00	0.36	0.15	0.00	0.36	0.19	0.00	0.40	0.18	0.00	0.39	0.18	0.00	0.38
CEO duality	13.26%	0.60	1.00	0.49	0.60	1.00	0.49	0.56	1.00	0.50	0.56	1.00	0.50	0.57	1.00	0.50
CEO on board	9.45%	0.78	1.00	0.41	0.78	1.00	0.41	0.75	1.00	0.43	0.76	1.00	0.43	0.75	1.00	0.43
CEO county	16.11%	0.89	1.00	0.31	0.89	1.00	0.31	0.82	1.00	0.38	0.83	1.00	0.38	0.83	1.00	0.38
Two CEOs	7.06%	0.01	0.00	0.09	0.01	0.00	0.09	0.00	0.00	0.02	0.00	0.00	0.05	0.00	0.00	0.04
Chairperson age	11.25%	3.94	3.95	0.22	3.94	3.95	0.22	3.83	3.85	0.25	3.82	3.85	0.25	3.83	3.85	0.25
Chairperson woman	7.42%	0.12	0.00	0.33	0.12	0.00	0.33	0.16	0.00	0.37	0.15	0.00	0.35	0.14	0.00	0.35
Chairperson county	19.24%	0.87	1.00	0.34	0.87	1.00	0.34	0.80	1.00	0.40	0.80	1.00	0.40	0.80	1.00	0.40
Board size	3.29%	0.52	0.69	0.55	0.52	0.69	0.55	0.46	0.00	0.53	0.46	0.00	0.52	0.45	0.00	0.52
Board age avg	5.18%	3.92	3.93	0.20	3.92	3.93	0.20	3.82	3.85	0.23	3.81	3.83	0.23	3.82	3.84	0.23
Board age std	5.18%	3.11	0.00	4.57	3.12	0.00	4.58	2.60	0.00	4.11	2.65	0.00	4.22	2.64	0.00	4.25
Board women	3.29%	0.16	0.00	0.30	0.16	0.00	0.30	0.19	0.00	0.33	0.18	0.00	0.32	0.17	0.00	0.32
Board county	6.90%	0.77	1.00	0.36	0.77	1.00	0.36	0.61	0.67	0.38	0.63	0.71	0.39	0.62	0.71	0.40
Board non-owners	4.40%	0.48	0.50	0.46	0.48	0.50	0.46	0.44	0.33	0.44	0.44	0.33	0.45	0.44	0.33	0.45
Ownership concentration 1	1.16%	0.74	1.00	0.32	0.75	1.00	0.32	0.74	1.00	0.32	0.73	1.00	0.31	0.73	1.00	0.31
Ownership concentration 2	1.16%	0.05	0.00	0.10	0.05	0.00	0.10	0.05	0.00	0.11	0.05	0.00	0.10	0.05	0.00	0.10
Ownership CEO	8.08%	0.43	0.34	0.43	0.43	0.34	0.43	0.44	0.40	0.41	0.44	0.40	0.41	0.45	0.43	0.41
Ownership chairperson	8.39%	0.42	0.33	0.43	0.42	0.33	0.43	0.44	0.37	0.43	0.43	0.35	0.43	0.43	0.38	0.43
Ownership board	4.40%	0.51	0.50	0.46	0.50	0.50	0.46	0.54	0.64	0.45	0.54	0.58	0.45	0.54	0.57	0.45

Notes: The mean, median and standard deviation (Std) for the non-financial variables across all observations in our sample of 818,927 financial statements. Our sample is detailed in Table 3. The non-financial variables are described in Table 1.

board members are non-owners, meaning that about half of SME board members in our sample are shareholders. The average shareholder holdings are, on average, 74% across all financial statements. On average, the ownership shares of CEOs, chairpersons and board members are 43%, 42% and 51%, respectively. The latter is naturally the highest as the chairpersons, and, in many cases, the CEOs also sit on the board. On average, 89%, 87% and 77% of CEOs, chairpersons and board members reside in the same county as the SMEs' HQs. When comparing the observations categorized as non-bankrupt to those categorized as bankrupt, we observe that the latter have, on average, CEOs, chairpersons and board members who are younger, less likely to reside in the same county as SMEs' HQs and have higher ownership shares. Table 4 also details the fraction of missing values imputed with k -NN for each non-financial variable. We observe that the highest fraction of missing values is for the three variables that indicate whether the CEOs, chairpersons and board members reside in the same county as the SMEs' HQs.

Table 5 presents the Pearson correlations for the full sample for 2014–2019. Panel A presents the correlations between the management structure variables and non-financial and financial variables. Panels B and C show the same results for the board and ownership structure variables, respectively. We observe a positive correlation between 'CEO on board' and 'CEO duality'. This is as expected; if a CEO is also the chairperson, they also sit on the board, although this is not necessarily true the other way around. Furthermore, we observe positive correlations between (i) the age of CEOs, chairpersons and board members, (ii) the gender of CEOs, chairpersons, and board members and (iii) whether the CEOs, chairpersons and board members reside in the same county as the SMEs' HQs. This testifies to an age, gender and residential homogeneity within SMEs. Moreover, we observe a positive relationship between the number of board members ('board size') and the diversity in board members' ages ('board age std'). We also observe a positive relationship between 'CEO duality' and 'ownership CEO', as well as between 'CEO duality' and 'ownership chairperson'. This indicates a higher likelihood of a single individual serving as CEO and chairperson in SMEs, where this person has a larger ownership share. Furthermore, we observe a negative relationship between the proportion of board members who are not shareholders ('board non-owners') and the ownership shares of CEOs, chairpersons and board members. We also observe positive correlations among the ownership shares of CEOs, chairpersons and board members. In unreported correlation analyses, we find that these correlations are small and insignificant when the dummies 'CEO on board' and 'CEO duality' are zero, indicating that the positive correlations are due to the same individuals holding multiple roles.

5. Methodology

5.1. Test setting

We predict the probability of bankruptcy over different time horizons. First, we predict a one-year horizon. This aligns with the focus of Basel III on forecasting horizons for corporate exposures one year ahead and is also the most commonly used horizon among practitioners and academics (Hillegeist et al., 2004; Paraschiv et al., 2023; Tian et al., 2015). We consider financial statements bankrupt if the underlying company files for bankruptcy within one year of the balance sheet date. Additionally, recognizing that owners and managers often have incentives to delay bankruptcy proceedings (Srhoj et al., 2023), we predict the probability of bankruptcy over time horizons of two and three years following the approach of Ciampi et al. (2020). Table 3 presents the bankruptcy frequency across the different prediction horizons for each accounting year.

Furthermore, we employ an out-of-sample testing approach and assess the robustness of our findings using two distinct test populations: all financial statements from 2018 and 2019. To prevent any data leakage, we employ a four-year rolling window approach to select variables and train models using all financial statements from the four accounting years preceding the test populations. Specifically, using the financial statements of 2018 and 2019 as test populations, we select variables and train the models using all financial statements from the accounting years 2014–2017 and 2015–2018, respectively.

We use discrete hazard models for bankruptcy prediction, as suggested by Shumway (2001). Hazard models are superior to other approaches (Bauer & Agarwal, 2014) and their use is a convention in recent

Table 5. Pairwise correlations.
Panel A. Management structure non-financial variables.

Variables	(1)	(2)	(3)	(4)	(5)	(6)
(1) CEO age						
(2) CEO woman	-0.08					
(3) CEO duality	0.12	-0.11				
(4) CEO on board	0.06	-0.09	0.55***			
(5) CEO county	0.03	-0.01	0.11**	0.11**		
(6) Two CEOs	-0.03	0.03	-0.00	0.01	-0.00	
Chairperson age	0.67***	-0.03	-0.07	-0.06	0.01	-0.02
Chairperson woman	-0.04	0.52***	-0.01	0.00	0.01	0.02
Chairperson county	0.03	-0.01	0.18***	0.14***	0.51***	0.00
Board size	-0.02	0.02	-0.49***	-0.14***	-0.07*	0.03
Board age avg	0.71***	-0.04	-0.02	-0.05	0.01	-0.02
Board age std	0.00	0.01	-0.31***	-0.05***	-0.03	0.01
Board women	-0.01	0.57***	-0.02	-0.02	0.01	0.02
Board county	0.02	-0.01	0.25***	0.32***	0.50***	-0.02
Board non-owners	-0.01	-0.04	-0.35***	-0.33***	-0.13**	0.02
Ownership concentration 1	-0.00	-0.02	0.28***	0.10**	0.02	-0.01
Ownership concentration 2	0.04	-0.02	-0.10**	-0.05*	-0.02	-0.00
Ownership CEO	0.02	0.00	0.48***	0.38***	0.14**	0.01
Ownership chairperson	0.02	0.01	0.51***	0.31***	0.12**	-0.01
Ownership board	-0.00	0.05	0.36***	0.40***	0.13**	-0.00
(Current liabilities - short-term liquidity) / total assets	-0.13*	-0.02	-0.12**	-0.09*	-0.07	0.00
EBIT / total assets	-0.01	-0.02	0.10	0.10	0.05	0.01
EBITDA / interest expense	0.04	-0.01	0.04	0.03	0.00	-0.00
EBITDA / total assets	-0.02	-0.02	0.09	0.09	0.05	0.01
Accounts payable / total assets	-0.14**	0.01	-0.10**	-0.07*	-0.05	-0.00
Current liabilities / total equity	-0.04	0.00	-0.03	-0.03	-0.03	0.00
Dummy; one if paid-in equity is less than total equity	0.06	-0.01	0.09*	0.08	0.08	0.01
Dummy; one if total liability exceeds total assets	-0.07	0.02	-0.01	-0.00	-0.04	-0.00
Interest expenses / total assets	0.01	-0.03	-0.01	-0.01	-0.02	-0.00
Inventory / current assets	-0.04	0.08	-0.07*	-0.02	0.01	0.00
Log(age in years)	0.35***	-0.02	-0.06	-0.08	0.01	0.01
Net income / total assets	0.00	-0.01	0.10	0.09	0.05	0.01
Public taxes payable / total assets	-0.16***	0.08	-0.03	0.00	-0.00	0.01
Retained earnings / total assets	0.10	-0.02	0.14**	0.11**	0.09	0.00
Sales / total assets	-0.18***	0.08	-0.08	-0.04	-0.02	0.01
Short-term liquidity / current assets	0.03	0.06	0.11**	0.09*	0.05	0.00
Short-term liquidity / total assets	0.02	0.09	0.07	0.06	0.03	0.01
Total equity / total liabilities	0.01	-0.01	0.01	0.01	0.00	0.00
Working capital / total assets	0.11	0.01	0.04	0.04	0.05	-0.00

Panel B. Board structure non-financial variables.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) Chairperson age									
(2) Chairperson woman	-0.09								
(3) Chairperson county	-0.00	0.01							
(4) Board size	0.09	-0.03	-0.12**						
(5) Board age avg	0.84***	-0.06	0.00	0.01					
(6) Board age std	0.16	-0.02	-0.07**	0.67***	0.04				
(7) Board women	-0.03	0.75***	0.01	0.09	-0.05	0.06			
(8) Board county	-0.02	0.01	0.60***	-0.19***	-0.00	-0.09***	-0.01		
(9) Board non-owners	0.08	-0.09	-0.17***	0.37***	0.05	0.23***	-0.06	-0.31***	
CEO age	0.67***	-0.04	0.03	-0.02	0.71***	0.00	-0.01	0.02	-0.01
CEO woman	-0.03	0.52***	-0.01	0.02	-0.04	0.01	0.57***	-0.01	-0.04
CEO duality	-0.07	-0.01	0.18***	-0.49***	-0.02	-0.31***	-0.02	0.25***	-0.35***
CEO on board	-0.06	0.00	0.14***	-0.14***	-0.05	-0.05***	-0.02	0.32***	-0.33***
CEO county	0.01	0.01	0.51***	-0.07*	0.01	-0.03	0.01	0.50***	-0.13**
Two CEOs	-0.02	0.02	0.00	0.03	-0.02	0.01	0.02	-0.02	0.02
Ownership concentration 1	-0.06	0.01	0.04	-0.42***	-0.03	-0.30***	-0.03	0.04	0.07*
Ownership concentration 2	0.09	-0.03	-0.03	0.18***	0.06	0.17***	-0.01	-0.04	0.01
Ownership CEO	-0.09	0.07	0.16***	-0.43***	-0.05	-0.27***	0.05	0.26***	-0.73***
Ownership chairperson	-0.10	0.07	0.18***	-0.49***	-0.06	-0.32***	0.05	0.27***	-0.79***
Ownership board	-0.09	0.10	0.17***	-0.27***	-0.07	-0.14***	0.10	0.30***	-0.89***
(Current liabilities - short-term liquidity) / total assets	-0.10	-0.03	-0.08	0.07	-0.11	0.03	-0.04	-0.11**	0.18**
EBIT / total assets	-0.03	-0.00	0.06	-0.09	-0.03	-0.06	-0.00	0.09	-0.12*
EBITDA / interest expense	0.02	-0.00	0.01	-0.03	0.03	-0.02	-0.00	0.03	-0.03
EBITDA / total assets	-0.05	-0.00	0.06	-0.09	-0.05	-0.06	-0.00	0.09	-0.13*
Accounts payable / total assets	-0.11	-0.00	-0.06	0.06	-0.11*	0.04	-0.01	-0.09**	0.11**
Current liabilities / total equity	-0.04	-0.00	-0.03	0.01	-0.04	-0.00	-0.00	-0.04	0.08
	0.05	0.00	0.08	-0.07*	0.05	-0.04	0.01	0.11*	-0.17**

(continued)

Table 5 Continued.

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dummy; one if paid-in equity is less than total equity									
Dummy; one if total liability exceeds total assets	-0.07	0.02	-0.03	-0.01	-0.06	-0.01	0.01	-0.05	0.05
Interest expenses / total assets	0.01	-0.04	-0.02	0.01	0.01	0.00	-0.04	-0.04	0.03
Inventory / current assets	-0.01	0.06	-0.01	0.06	-0.01	0.06	0.07	-0.01	0.08**
Log(age in years)	0.37***	-0.03	-0.00	0.07	0.37***	0.08	0.00	-0.03	0.03
Net income / total assets	-0.02	0.00	0.06	-0.08	-0.02	-0.06	0.00	0.10	-0.12*
Public taxes payable / total assets	-0.15***	0.07	0.01	-0.01	-0.15***	-0.02	0.07	0.00	-0.09
Retained earnings / total assets	0.06	0.00	0.10	-0.11**	0.07	-0.06	0.01	0.14**	-0.23***
Sales / total assets	-0.15***	0.06	-0.02	0.03	-0.16***	0.01	0.06	-0.04	0.00
Short-term liquidity / current assets	0.00	0.06	0.06	-0.08*	0.01	-0.05	0.07	0.08*	-0.17***
Short-term liquidity / total assets	0.00	0.09	0.05	-0.07	0.01	-0.05	0.10	0.05	-0.17**
Total equity / total liabilities	0.01	-0.00	0.00	-0.01	0.01	-0.01	-0.00	0.00	-0.01
Working capital / total assets	0.10	0.02	0.05	-0.02	0.11	-0.01	0.02	0.06	-0.09*

Panel C. Ownership structure non-financial variables.

Variables	(1)	(2)	(3)	(4)	(5)
(1) Ownership concentration 1					
(2) Ownership concentration 2	-0.55***				
(3) Ownership CEO	0.28***	-0.12**			
(4) Ownership chairperson	0.29***	-0.14**	0.84***		
(5) Ownership board	0.10**	-0.09*	0.82***	0.86***	
CEO age	-0.00	0.04	0.02	0.02	-0.00
CEO woman	-0.02	-0.02	0.00	0.01	0.05
CEO duality	0.28***	-0.10**	0.48***	0.51***	0.36***
CEO on board	0.10**	-0.05*	0.38***	0.31***	0.40***
CEO county	0.02	-0.02	0.14**	0.12**	0.13**
Two CEOs	-0.01	-0.00	0.01	-0.01	-0.00
Chairperson age	-0.06	0.09	-0.09	-0.10	-0.09
Chairperson woman	0.01	-0.03	0.07	0.07	0.10
Chairperson county	0.04	-0.03	0.16***	0.18***	0.17***
Board size	-0.42***	0.18***	-0.43***	-0.49***	-0.27***
Board age avg	-0.03	0.06	-0.05	-0.06	-0.07
Board age std	-0.30***	0.17***	-0.27***	-0.32***	-0.14***
Board women	-0.03	-0.01	0.05	0.05	0.10
Board county	0.04	-0.04	0.26***	0.27***	0.30***
Board non-owners	0.07*	0.01	-0.73***	-0.79***	-0.89***
(Current liabilities - short-term liquidity) / total assets	0.00	0.02	-0.19***	-0.19**	-0.19**
EBIT / total assets	0.05	-0.04	0.14*	0.14*	0.13*
EBITDA / interest expense	0.02	-0.02	0.04	0.04	0.03
EBITDA / total assets	0.05	-0.04	0.14*	0.14*	0.13*
Accounts payable / total assets	-0.03	0.03	-0.12**	-0.12**	-0.11**
Current liabilities / total equity	0.03	-0.02	-0.07	-0.07	-0.08
Dummy; one if paid-in equity is less than total equity	0.03	-0.02	0.17**	0.17**	0.18**
Dummy; one if total liability exceeds total assets	-0.01	0.01	-0.05	-0.05	-0.05
Interest expenses / total assets	-0.01	0.02	-0.03	-0.03	-0.04
Inventory / current assets	-0.03	0.03	-0.09**	-0.10**	-0.07**
Log(age in years)	-0.04	0.09	-0.05	-0.05	-0.04
Net income / total assets	0.05	-0.04	0.13*	0.13*	0.13*
Public taxes payable / total assets	-0.04	0.02	0.05	0.05	0.09
Retained earnings / total assets	0.06	-0.04	0.24***	0.24***	0.24***
Sales / total assets	-0.02	0.02	-0.03	-0.04	-0.00
Short-term liquidity / current assets	-0.02	-0.03	0.17***	0.17***	0.18***
Short-term liquidity / total assets	-0.01	-0.02	0.16**	0.16**	0.17**
Total equity / total liabilities	0.01	-0.01	0.01	0.01	0.01
Working capital / total assets	0.01	-0.01	0.09*	0.09	0.09

Notes: Panels A, B and C display the Pearson correlations for the full sample during 2014–2019 between the management, board and ownership structure variables, respectively, and all other non-financial and financial variables. The variables are defined in Tables 1 and 2. *, ** and *** denote the significance levels at 10%, 5% and 1%, respectively.

literature (eg Campbell et al., 2008; Chava & Jarrow, 2004; Paraschiv et al., 2023; Tian et al., 2015). Earlier studies (eg Altman, 1968) employed only one observation for each firm. This makes the models prone to producing biased results, as they ignore changes in firms over time. By contrast, discrete hazard models consider bankruptcy hazards over multiple years. This makes them preferable, as they incorporate time-varying variable estimates, adjust for the period at risk and employ more data.

We share the code for running all our analyses and generating the results at: <https://enhancing-credit-risk-2024.ranik.no>.

5.2. LASSO and logistic regression

Following the existing literature (eg Altman et al., 2010; Campbell et al., 2008; Paraschiv et al., 2023; Shumway, 2001), we model bankruptcy prediction using LR models given by

$$\hat{\mathbf{y}} = \mathbf{1} \odot (\mathbf{1} + e^{-\mathbf{X}\boldsymbol{\beta} - \mathbf{1}\beta_0}), \quad (1)$$

where $\hat{\mathbf{y}} = \{\hat{y}_n\}_{n=1, \dots, N} \in [0, 1]^N$ is the vector of predicted probabilities of bankruptcy for each financial statement n , $\mathbf{X} = \{x_{(n,i)}\}_{n=1, \dots, N, i=1, \dots, I}$ is the matrix of values for the input variables i , $\mathbf{1}$ is an $N \times 1$ vector of ones, e denotes the exponential function and \odot denotes Hadamard (element-wise) division. We estimate the coefficients $\boldsymbol{\beta} = \{\beta_i\}_{i=1, \dots, I}$ and β_0 by minimizing the negative of the log-likelihood function $\ell(\boldsymbol{\beta}, \beta_0)$ given by

$$\ell(\boldsymbol{\beta}, \beta_0) = \sum_{n=1}^N [\mathbf{y} \odot (\mathbf{X}\boldsymbol{\beta} + \mathbf{1}\beta_0) - \ln(\mathbf{1} + e^{\mathbf{X}\boldsymbol{\beta} + \mathbf{1}\beta_0})], \quad (2)$$

where $\mathbf{y} = \{y_n\}_{n=1, \dots, N} \in \{0, 1\}^N$ is the vector of actual classifications of bankrupt (1) and non-bankrupt (0), \ln denotes the natural logarithm and \odot denotes Hadamard (element-wise) multiplication. We select variables for the LR models using the LASSO feature selection method popularized by Tibshirani (1996). This method minimizes

$$-\ell(\boldsymbol{\beta}, \beta_0) + \lambda \|\boldsymbol{\beta}\|_1, \quad (3)$$

where $\|\boldsymbol{\beta}\|_1 = \sum_{i=1}^I |\beta_i|$ is the l_1 -norm of $\boldsymbol{\beta}$, λ is a positive tuning parameter, and all variables are standardized to zero mean and unit variance for comparison on an equal scale. When using the LASSO method to select variable sets for LR models, we empirically determine λ values for Equation (3) by performing a 10-fold cross-validation, following Friedman et al. (2010). In this process, the dataset is divided into 10 subsets, and the model is trained 10 times, each time using a different subset as the validation set and the remaining data as the training set. For each fold, the performance of the trained LASSO model (Equation (3)) is evaluated by the AUC value (see Section 5.4) on the validation set across a range of 100 different λ values in a geometric sequence from the smallest possible value for λ where all the coefficients are still zero to this λ value multiplied by 10^{-6} . For each λ value, the cross-validated AUC is calculated as the mean of the AUC values across the 10 folds and the λ value that maximizes this cross-validated AUC is identified. The standard error of the cross-validated AUC is also computed at each λ . To make a bias-variance tradeoff and reduce model complexity, we follow the practice of choosing the largest λ value that keeps the cross-validated AUC within one standard deviation of the maximum cross-validated AUC. Finally, we use the chosen λ value in Equation (3), minimize it with all training data, and let all variables with non-zero coefficients be the ones selected by our LASSO method.

We also use the LASSO method to show the variable importance. This is done with LASSO path plots derived by minimizing Equation (3) several times with different values of λ : First, λ is set sufficiently large such that the term $\lambda \|\boldsymbol{\beta}\|_1$ becomes so dominant that all the estimated coefficients become zero. Furthermore, the λ value is gradually decreased such that more coefficients gradually become non-zero and thus enter the model. Variables becoming non-zero at higher λ values have stronger predictive power and, thus, higher importance.

The statistical significance of the estimated coefficients of the LR models is determined by their z-scores derived from the Wald statistic (Hosmer et al., 2013). We acknowledge that these have post-selection inference problems because statistical inference is derived from the data after the variables have been selected based on the same data. This can lead to biased and invalid confidence intervals because the model selection process introduces additional uncertainty. Several approaches have been proposed to address this issue, including stability selection and debiased LASSO (eg Lee et al., 2016; Meinshausen & Bühlmann, 2010; Taylor & Tibshirani, 2018; van de Geer et al., 2014; Zhang & Zhang, 2014). However,

given the large data sample size and the memory-intensive and computationally complex nature of these methods, they are impractical without significant computational resources that are not available to us. Moreover, we do not derive coefficient estimates and their statistical significance from the LASSO method but merely use it to select variables that are later used in LR models for which we derive coefficient estimates and their statistical significance. However, we correct the standard errors used to derive z-scores by clustering them at the firm level.

5.3. Machine learning methods

We also test the robustness of our results by predicting bankruptcy using machine learning methods instead of LASSO and LR, otherwise applying the same test setting (see [Section 5.1](#)). Specifically, following da Silva Mattos and Shasha (2024), we employ four machine learning methods within three classes of ensemble methods: boosting (XGBoost and AdaBoost), bagging and random forest. Indeed, these three classes of ensemble methods have been found to outperform other methods for bankruptcy prediction (Barboza et al., 2017). XGBoost (Chen & Guestrin, 2016), which was also used to test the robustness in Altman et al. (2023a) and AdaBoost (Freund & Schapire, 1997) build ensembles of decision trees sequentially, where each tree corrects the errors of the previous ones. Furthermore, bagging, or bootstrap aggregating (Breiman, 1996), trains multiple decision tree classifiers on subsets of the training data selected randomly with replacement and then aggregates their predictions to improve stability and accuracy. Finally, random forest (Breiman, 2001) trains multiple independent decision trees with randomly selected subsets of variables and outputs the mode of the classes for classification. For each machine learning method, we tune the hyperparameters using a grid search with a cross-validation scheme.¹² To avoid data leakage, tuning is performed on the training data. For further description of the machine learning methods, we refer to their descriptions in the bankruptcy prediction studies by Barboza et al. (2017) and Radovanovic and Haas (2023).

5.4. Evaluation metrics

We evaluate our bankruptcy prediction models using several evaluation metrics. First, we evaluate them using average precision (AP) and the area under the receiver operating characteristic curve (AUC), which are widely used for binary classification problems, including bankruptcy prediction (eg Altman et al., 2023a; Paraschiv et al., 2023; Tian et al., 2015). The AUC is the area under the plot of the false positive rate against the true positive rate across all observations when varying the discrimination threshold across the two limit values of 0 and 1 (Hosmer et al., 2013). AP is calculated similarly, but based on the plot of precision against recall and is recommended over AUC for evaluating model performance with imbalanced datasets (Saito & Rehmsmeier, 2015). Higher AUC and AP values indicate a better-performing model. As suggested by Paraschiv et al. (2023), we follow Hosmer et al. (2013) by considering $AUC \in [0.7, 0.8)$ acceptable, $AUC \in [0.8, 0.9)$ excellent and $AUC \geq 0.9$ outstanding.

Furthermore, they were evaluated based on the accuracy ratio (AR), a performance metric derived from the cumulative accuracy profile (CAP) curve (Engelmann et al., 2003; Mai et al., 2019). It is calculated as the ratio of the area between the CAP curve of the model and the random model to the area between the CAP curve of the perfect model and the random model. Moreover, we use the Kolmogorov-Smirnov (KS) statistic (Hodges, 1958), a non-parametric test to compare the distributions of two samples by the maximum difference between their cumulative distribution functions. In the context of bankruptcy prediction model evaluation, the KS statistic can be used to compare the predicted probabilities for bankrupt and non-bankrupt classes. Higher AR and KS statistics indicate better model performance, reflecting a greater ability to distinguish between bankrupt and non-bankrupt cases.

Additionally, we evaluate them based on Hinge loss (Crammer & Singer, 2001), measuring the distance between the predicted and actual classifications encoded as -1 and 1 , as well as the logistic loss (Log loss), which is the value of the log-likelihood function shown in [Equation \(2\)](#), divided by the number of observations. Moreover, following Tian et al. (2015) and Paraschiv et al. (2023), we use the Akaike (1974) Information Criterion (AIC), evaluating the goodness of fit of the model similarly to the log loss but penalized for the number of parameters to prevent overfitting, balancing model complexity and fit.

We also use the Bayesian Information Criterion (BIC), which is similar to the AIC but introduces a stronger penalty for the number of parameters. Additionally, we assess the models using Brier (1950) score, which quantifies the mean-squared difference between the actual and predicted bankruptcy probabilities. Lower Hinge loss, Log loss, AIC, BIC and Brier scores indicate better models.

We also report the Brier Skill Score (BSS), which compares the Brier score of a predictive model with that of a reference model (Roulston, 2007). As a reference model, we use a model that always predicts the average of $\mathbf{y} \in \{0, 1\}^N$, which is the bankruptcy frequency, in the training data. Furthermore, we follow bankruptcy prediction literature (eg Campbell et al., 2008; Paraschiv et al., 2023; Tian et al., 2015) by evaluating the in-sample fit of LR models using McFadden's (1974) pseudo-R squared $R^2 = 1 - \frac{\ell(\beta, \beta_0)}{\ell(\beta_0)} \in [0, 1]$ where the denominator is the log-likelihood of a model containing only the intercept coefficient β_0 . Higher BSS and R^2 values indicate better models.

Finally, following the bankruptcy prediction literature (eg Chava & Jarrow, 2004; Paraschiv et al., 2023; Shumway, 2001; Tian et al., 2015), we evaluate them based on decile rankings. When applying this method, the observations are divided into deciles based on their predicted probability of bankruptcy provided by the model, and the proportion of actual bankruptcies within each decile is then reported. This method allows a clear assessment of the model's discriminatory power by showing how well bankrupt and non-bankrupt firms can be distinguished across different risk levels.

6. Results

6.1. Main results

Tables 6, 7 and 8 present the estimation results of the LR models when predicting bankruptcy over horizons of one, two and three years, respectively, using variable sets selected by the LASSO method from the population of non-financial variables in Table 1 and the set of financial variables in Altman (1968), Altman and Sabato (2007) and Paraschiv et al. (2023) (see Table 2). Additionally, we report the results using all the financial statements from 2018 and 2019, respectively, as out-of-sample test samples. First, we use the accounting year 2018 as our test sample and let the LASSO method use all the financial statements from the four previous years to select the best variable set among the population of 25 variables, consisting of the five financial variables in Altman (1968) (see Table 2) and the 20 non-financial variables in Table 1. The selected variables and the estimation results when used in LR models are presented in the first columns of Tables 6, 7 and 8. Next, we repeat the procedure using the 2019 accounting year as our test sample and report the results in the second column. We then repeat the procedure with the variables in Altman and Sabato (2007) instead of those in Altman (1968) and show the results in the third and fourth columns. Finally, the procedure is repeated with a population of 30 variables, including those in Paraschiv et al. (2023) and the 20 non-financial variables in Table 1. The results are presented in the last two columns of Tables 6, 7 and 8. In each table, Panel A presents the variables selected by the LASSO method and their LR coefficient estimates and z-scores in parentheses. Panels B and C report the in-sample fit and out-of-sample prediction performance, respectively, when using the variable sets selected by the LASSO method, as shown in Panel A, in the top rows, and when using exclusively financial variables in the bottom rows.

6.1.1. The importance of financial variables

The tables show that financial variables are undoubtedly important because the LASSO method selects most financial variables across all variable sets, prediction horizons and accounting years as our test sample. Furthermore, the selected financial variables are consistent across all settings. Specifically, the same financial variables are selected in all settings, except for 'interest expenses / total assets' which is selected when using the financial variable set of Paraschiv et al. (2023) only when 2019 is the test period and when predicting bankruptcy over a horizon of three years (see Panel A of Table 8).

The importance of the financial variables is further confirmed by the associated LASSO path plots. Figures IA.1, IA.2 and IA.3 in the Internet Appendix present the LASSO path plots of the variables selected among the non-financial variables and those in Altman (1968), Altman and Sabato (2007) and Paraschiv et al. (2023), respectively, when predicting bankruptcy over a horizon of one year. Panels A

Table 6. Estimation results when predicting bankruptcy over a horizon of one year.**Panel A.** Variables selected using the LASSO method and their LR coefficient estimates when predicting bankruptcy over a horizon of one year.

	Altman (1968)		Altman and Sabato (2007)		Paraschiv et al. (2023)	
	2018	2019	2018	2019	2018	2019
CEO age	−0.65(−5.18)	−0.57(−4.31)	−0.84(−6.26)	−0.78(−6.03)	−0.35(−2.73)	−0.35(−2.65)
Chairperson age	−0.78(−6.30)	−0.48(−2.74)	−0.84(−4.84)	−0.54(−3.19)	−0.40(−3.13)	−0.23(−1.32)
Board age avg		−0.56(−2.81)	−0.21(−1.07)	−0.74(−3.85)		−0.40(−1.97)
Board county	−0.82(−14.97)	−0.92(−17.46)	−0.87(−15.83)	−0.97(−18.33)	−0.78(−14.16)	−0.88(−16.77)
Board non-owners		−0.29(−2.88)		−0.40(−3.92)		−0.30(−2.93)
Ownership board	0.72(13.50)	0.50(4.96)	0.95(18.04)	0.66(6.56)	0.60(11.31)	0.41(4.02)
EBIT / total assets	−1.72(−21.24)	−1.78(−22.87)				
Retained earnings / total assets	−0.38(−9.75)	−0.33(−8.97)	−0.70(−22.88)	−0.66(−22.71)		
Sales / total assets	0.14(15.07)	0.15(16.59)				
Working capital / total assets	−0.77(−13.80)	−0.78(−14.57)				
EBITDA / total assets			−1.99(−23.86)	−2.07(−25.82)		
Short-term liquidity / total assets			−1.77(−15.06)	−1.75(−15.53)		
(Current liabilities - short-term liquidity) / total assets					0.12(1.76)	0.20(3.06)
Accounts payable / total assets					1.27(12.77)	1.18(12.31)
Dummy; one if paid-in equity is less than total equity					−0.69(−9.44)	−0.78(−10.84)
Dummy; one if total liability exceeds total assets					0.65(9.03)	0.59(8.42)
Inventory / current assets					0.43(5.66)	0.44(6.02)
Log(age in years)					−0.20(−8.93)	−0.19(−8.57)
Net income / total assets					−1.08(−13.71)	−1.10(−14.71)
Public taxes payable / total assets					4.12(22.33)	4.03(22.84)
Short-term liquidity / current assets					−1.49(−13.69)	−1.37(−13.25)
Intercept	−0.04(−0.10)	0.87(2.20)	2.23(5.62)	3.34(8.75)	−2.28(−5.30)	−1.06(−2.48)

Panel B: In-sample fit when predicting bankruptcy over a horizon of one year.

		Altman (1968)		Altman and Sabato (2007)		Paraschiv et al. (2023)	
		2018	2019	2018	2019	2018	2019
Financial and non-financial variables	R^2	0.137	0.151	0.129	0.141	0.197	0.209
	AUC	0.834	0.846	0.827	0.838	0.876	0.885
	AP	0.036	0.042	0.031	0.035	0.057	0.062
	AR	0.664	0.690	0.650	0.672	0.749	0.767
	KS statistic	0.536	0.558	0.511	0.527	0.612	0.631
	Hinge loss	1.004	1.004	1.004	1.004	1.004	1.004
	Log loss	0.023	0.023	0.023	0.023	0.021	0.021
	AIC	24,345	25,113	24,573	25,421	22,684	23,415
	BIC	24,446	25,237	24,673	25,533	22,841	23,595
	Brier score	0.004	0.004	0.004	0.004	0.004	0.004
	BSS	0.012	0.014	0.011	0.012	0.028	0.030
	Decile 1	0.578	0.592	0.548	0.571	0.672	0.690
	Decile 2	0.156	0.160	0.154	0.150	0.137	0.136
	Decile 3	0.074	0.081	0.097	0.090	0.065	0.055
	Decile 4	0.048	0.045	0.052	0.052	0.035	0.038
	Decile 5	0.042	0.035	0.046	0.043	0.028	0.025
	Decile 6-10	0.103	0.088	0.104	0.095	0.064	0.055
Exclusively financial variables	R^2	0.118	0.127	0.099	0.104	0.185	0.193
	AUC	0.816	0.826	0.794	0.800	0.868	0.874
	AP	0.031	0.034	0.024	0.025	0.050	0.053
	AR	0.628	0.649	0.585	0.597	0.733	0.745
	KS statistic	0.514	0.530	0.470	0.486	0.603	0.607
	Hinge loss	1.004	1.004	1.004	1.004	1.004	1.004
	Log loss	0.024	0.024	0.024	0.024	0.022	0.022
	AIC	24,892	25,815	25,411	26,503	22,999	23,866

(continued)

Table 6 Continued.

		Altman (1968)		Altman and Sabato (2007)		Paraschiv et al. (2023)	
		2018	2019	2018	2019	2018	2019
	BIC	24,959	25,882	25,478	26,570	23,122	23,990
	Brier score	0.004	0.004	0.004	0.004	0.004	0.004
	BSS	0.009	0.011	0.007	0.008	0.024	0.026
	Decile 1	0.546	0.562	0.549	0.560	0.647	0.660
	Decile 2	0.157	0.160	0.116	0.117	0.150	0.142
	Decile 3	0.095	0.089	0.065	0.064	0.068	0.071
	Decile 4	0.046	0.048	0.068	0.062	0.041	0.038
	Decile 5	0.035	0.033	0.055	0.050	0.024	0.025
	Decile 6-10	0.121	0.109	0.148	0.147	0.070	0.063

Panel C. Out-of-sample prediction performance when predicting bankruptcy over a horizon of one year.

		Altman (1968)		Altman and Sabato (2007)		Paraschiv et al. (2023)	
		2018	2019	2018	2019	2018	2019
Financial and non-financial variables	AUC	0.855	0.830	0.842	0.804	0.889	0.860
	AP	0.043	0.025	0.035	0.022	0.064	0.041
	AR	0.707	0.657	0.680	0.607	0.775	0.718
	KS statistic	0.587	0.519	0.536	0.466	0.632	0.600
	Hinge loss	1.004	1.004	1.004	1.004	1.004	1.003
	Log loss	0.024	0.017	0.024	0.018	0.022	0.017
	AIC	6866	5166	6976	5274	6398	4922
	BIC	6955	5275	7065	5373	6537	5080
	Brier score	0.004	0.003	0.004	0.003	0.004	0.003
	BSS	0.015	−0.005	0.011	−0.001	0.032	0.011
	Decile 1	0.586	0.563	0.575	0.505	0.683	0.617
	Decile 2	0.192	0.150	0.152	0.148	0.143	0.150
	Decile 3	0.070	0.089	0.081	0.103	0.063	0.099
	Decile 4	0.049	0.070	0.062	0.056	0.046	0.023
Exclusively financial variables	Decile 5	0.029	0.021	0.041	0.056	0.017	0.031
	Decile 6-10	0.075	0.106	0.089	0.131	0.048	0.080
	AUC	0.833	0.835	0.786	0.787	0.876	0.860
	AP	0.035	0.022	0.026	0.016	0.055	0.037
	AR	0.664	0.668	0.570	0.572	0.749	0.718
	KS statistic	0.538	0.541	0.486	0.435	0.601	0.583
	Hinge loss	1.004	1.004	1.004	1.004	1.004	1.004
	Log loss	0.025	0.018	0.026	0.018	0.023	0.017
	AIC	7078	5234	7443	5408	6543	4953
	BIC	7137	5293	7502	5468	6651	5062
	Brier score	0.004	0.003	0.004	0.003	0.004	0.003
	BSS	0.009	−0.011	0.005	−0.005	0.027	0.008
	Decile 1	0.560	0.577	0.573	0.500	0.667	0.608
	Decile 2	0.167	0.157	0.102	0.129	0.125	0.169
	Decile 3	0.097	0.082	0.049	0.080	0.084	0.082
	Decile 4	0.051	0.052	0.065	0.092	0.041	0.040
	Decile 5	0.032	0.038	0.060	0.061	0.024	0.023
	Decile 6-10	0.094	0.094	0.151	0.138	0.059	0.077

Notes: Estimation results of LR models that predict bankruptcy over a one-year horizon, using variable sets selected by the LASSO method from a population of non-financial variables in Table 1 and a set of financial variables. The columns display the results of permutations using all financial statements from 2018 and 2019 as out-of-sample test samples and the variables in Altman (1968), Altman and Sabato (2007) and Paraschiv et al. (2023), presented in Table 2, as the set of financial variables. Panel A presents the variables selected using the LASSO method, along with their coefficient estimates and z-scores in parentheses, derived from standard errors clustered at the firm level. Panels B and C report the in-sample fit and out-of-sample prediction performance, respectively, using R^2 , AUC, AP, AR, KS statistic, Hinge loss, Log loss, AIC, BIC, Brier score, BSS and decile rankings. The first rows in Panels B and C report metric values when using the variable sets selected by the LASSO method, shown in Panel A, from the population of both non-financial and financial variables. The bottom rows show the metric values when the financial variables are used exclusively per benchmark set, as shown in 2. To train the models, a four-year rolling window approach was followed in which the models were trained on all financial statements from the four accounting years preceding the test populations. These data are presented in 3.

and B in each figure present the plots using 2018 and 2019 as the out-of-sample test samples. In summary, Figures IA.1, IA.2 and IA.3 have six panels presenting the LASSO path plots for the six modeling permutations shown in the six columns of Table 6. Similarly, the six panels of Figures IA.5, IA.6 and IA.7 in the Internet Appendix present the LASSO path plots for the six modeling permutations shown in the

Table 7. Estimation results when predicting bankruptcy over a horizon of two years.**Panel A.** Variables selected using the LASSO method and their LR coefficient estimates when predicting bankruptcy over a horizon of two years.

	Altman (1968)		Altman and Sabato (2007)		Paraschiv et al. (2023)	
	2018	2019	2018	2019	2018	2019
CEO age	−0.58(−6.96)	−0.62(−7.79)	−0.78(−9.42)	−0.88(−10.93)		−0.36(−4.30)
Chairperson age	−0.65(−5.95)	−0.77(−7.39)	−0.73(−6.92)	−0.69(−6.62)	−0.35(−3.17)	−0.34(−3.10)
Board age avg	−0.70(−5.66)	−0.69(−5.84)	−0.83(−6.97)	−0.77(−6.51)	−0.60(−5.10)	−0.38(−3.01)
Board county	−0.92(−27.23)	−0.71(−22.61)	−0.97(−28.66)	−0.92(−28.12)	−0.88(−25.92)	−0.83(−25.41)
Board non-owners				−0.38(−6.04)		−0.51(−15.89)
Ownership board	0.62(19.15)		0.84(25.83)	0.42(6.62)	0.56(17.03)	
EBIT / total assets	−1.24(−24.63)	−1.17(−24.15)				
Retained earnings / total assets	−0.37(−14.53)	−0.28(−11.81)	−0.68(−34.59)	−0.65(−35.15)		
Sales / total assets	0.15(26.36)	0.17(29.56)				
Working capital / total assets	−0.73(−20.57)	−0.76(−22.36)				
EBITDA / total assets			−1.42(−27.13)	−1.36(−27.02)		
Short-term liquidity / total assets			−1.54(−22.57)	−1.46(−22.51)		
(Current liabilities - short-term liquidity) / total assets					0.24(5.39)	0.20(4.57)
Accounts payable / total assets					1.44(22.63)	1.46(23.44)
Dummy; one if paid-in equity is less than total equity					−0.57(−13.77)	−0.60(−15.00)
Dummy; one if total liability exceeds total assets					0.35(8.03)	0.36(8.41)
Inventory / current assets					0.53(11.24)	0.45(9.73)
Log(age in years)					−0.34(−26.40)	−0.32(−24.94)
Net income / total assets					−0.63(−11.94)	−0.61(−12.13)
Public taxes payable / total assets					4.33(36.71)	4.17(36.60)
Short-term liquidity / current assets					−1.37(−20.74)	−1.36(−21.44)
Intercept	3.00(12.10)	3.74(15.76)	5.16(21.56)	5.50(23.36)	−0.19(−0.72)	0.84(3.24)

Panel B: In-sample fit when predicting bankruptcy over a horizon of two years.

		Altman (1968)		Altman and Sabato (2007)		Paraschiv et al. (2023)	
		2018	2019	2018	2019	2018	2019
Financial and non-financial variables	R^2	0.126	0.119	0.117	0.112	0.195	0.189
	AUC	0.814	0.810	0.802	0.797	0.867	0.862
	AP	0.063	0.059	0.055	0.054	0.098	0.095
	AR	0.622	0.612	0.597	0.587	0.725	0.717
	KS statistic	0.479	0.475	0.459	0.451	0.583	0.571
	Hinge loss	1.010	1.010	1.010	1.010	1.010	1.010
	Log loss	0.053	0.054	0.054	0.055	0.049	0.050
	AIC	56,248	59,217	56,875	59,691	51,846	54,557
	BIC	56,359	59,318	56,976	59,803	52,003	54,726
	Brier score	0.011	0.011	0.011	0.011	0.010	0.011
	BSS	0.020	0.018	0.019	0.019	0.045	0.043
	Decile 1	0.485	0.467	0.457	0.455	0.603	0.591
	Decile 2	0.178	0.182	0.182	0.174	0.171	0.174
	Decile 3	0.106	0.118	0.115	0.113	0.083	0.089
Exclusively financial variables	Decile 4	0.077	0.074	0.075	0.081	0.054	0.053
	Decile 5	0.051	0.053	0.060	0.056	0.032	0.034
	Decile 6-10	0.102	0.106	0.112	0.120	0.057	0.061
	R^2	0.097	0.096	0.076	0.073	0.179	0.174
	AUC	0.789	0.790	0.756	0.754	0.858	0.855
	AP	0.051	0.051	0.040	0.039	0.085	0.083
	AR	0.572	0.573	0.507	0.503	0.709	0.702
	KS statistic	0.458	0.461	0.387	0.380	0.564	0.558
	Hinge loss	1.010	1.010	1.011	1.011	1.010	1.010
	Log loss	0.055	0.056	0.056	0.057	0.050	0.051
	AIC	58,131	60,759	59,500	62,369	52,833	55,530
	BIC	58,198	60,827	59,567	62,437	52,956	55,653
	Brier score	0.011	0.011	0.011	0.011	0.011	0.011
	BSS	0.013	0.012	0.010	0.009	0.037	0.035
	Decile 1	0.430	0.432	0.427	0.421	0.576	0.568
	Decile 2	0.197	0.195	0.145	0.145	0.176	0.179
	Decile 3	0.125	0.128	0.109	0.109	0.093	0.096
	Decile 4	0.073	0.071	0.082	0.087	0.058	0.056
	Decile 5	0.046	0.044	0.062	0.063	0.037	0.038
	Decile 6-10	0.129	0.130	0.175	0.175	0.061	0.064

Panel C. Out-of-sample prediction performance when predicting bankruptcy over a horizon of two years.

		Altman (1968)		Altman and Sabato (2007)		Paraschiv et al. (2023)	
		2018	2019	2018	2019	2018	2019
Financial and non-financial variables	AUC	0.797	0.799	0.776	0.782	0.847	0.852
	AP	0.056	0.035	0.045	0.029	0.086	0.056
	AR	0.587	0.593	0.545	0.561	0.686	0.700
	KS statistic	0.455	0.467	0.414	0.430	0.546	0.554
	Hinge loss	1.010	1.010	1.010	1.010	1.009	1.009
	Log loss	0.055	0.038	0.056	0.038	0.051	0.035
	AIC	15,771	11,270	16,075	11,388	14,740	10,484
	BIC	15,870	11,359	16,164	11,487	14,878	10,633
	Brier score	0.011	0.007	0.011	0.007	0.011	0.007
	BSS	0.013	−0.011	0.010	−0.007	0.036	0.009
	Decile 1	0.450	0.448	0.405	0.415	0.548	0.555
	Decile 2	0.187	0.193	0.182	0.192	0.182	0.186
	Decile 3	0.102	0.116	0.117	0.112	0.091	0.094
	Decile 4	0.078	0.078	0.086	0.079	0.063	0.070
	Decile 5	0.060	0.055	0.068	0.065	0.043	0.029
Exclusively financial variables	Decile 6-10	0.122	0.110	0.142	0.137	0.073	0.065
	AUC	0.797	0.792	0.751	0.749	0.846	0.850
	AP	0.049	0.033	0.038	0.024	0.078	0.054
	AR	0.588	0.579	0.497	0.494	0.684	0.696
	KS statistic	0.478	0.474	0.386	0.368	0.544	0.541
	Hinge loss	1.011	1.011	1.010	1.011	1.010	1.010
	Log loss	0.056	0.039	0.060	0.040	0.051	0.036
	AIC	16,021	11,474	17,368	11,797	14,823	10,579
	BIC	16,080	11,533	17,428	11,856	14,931	10,688
	Brier score	0.011	0.007	0.011	0.007	0.011	0.007
	BSS	0.007	−0.015	0.004	−0.008	0.030	0.005
	Decile 1	0.439	0.450	0.417	0.427	0.543	0.553
	Decile 2	0.194	0.188	0.150	0.125	0.189	0.173
	Decile 3	0.138	0.126	0.114	0.105	0.099	0.103
	Decile 4	0.073	0.067	0.088	0.089	0.056	0.067
	Decile 5	0.037	0.041	0.065	0.072	0.041	0.038
	Decile 6-10	0.119	0.127	0.167	0.182	0.073	0.066

Notes: Estimation results of LR models that predict bankruptcy over a two-year horizon, using variable sets selected by the LASSO method from a population of non-financial variables in Table 1 and a set of financial variables. The columns display the results of permutations using all financial statements from 2018 and 2019 as out-of-sample test samples and the variables in Altman (1968), Altman and Sabato (2007) and Paraschiv et al. (2023), presented in Table 2, as the set of financial variables. Panel A presents the variables selected using the LASSO method, along with their coefficient estimates and z-scores in parentheses, derived from standard errors clustered at the firm level. Panels B and C report the in-sample fit and out-of-sample prediction performance, respectively, using R^2 , AUC, AP, AR, KS statistic, Hinge loss, Log loss, AIC, BIC, Brier score, BSS and decile rankings. The first rows in Panels B and C report metric values when using the variable sets selected by the LASSO method, shown in Panel A, from the population of both non-financial and financial variables. The bottom rows show the metric values when the financial variables are used exclusively per benchmark set, as shown in 2. To train the models, a four-year rolling window approach was followed in which the models were trained on all financial statements from the four accounting years preceding the test populations. These data are presented in 3.

six columns of Table 7. Moreover, the six panels of Figures IA.9, IA.10 and IA.11 show the LASSO path plots for the six modeling permutations shown in the six columns of Table 8. As explained in Section 5.2, the LASSO path plots are generated by repeatedly minimizing Equation (3) with varying λ values. Initially, λ is set high enough that the term $\lambda \|\beta\|_1$ dominates, causing all estimated coefficients to be zero. This can be observed on the far-left side of the plots. As λ gradually decreases, moving to the right on the plots, more coefficients become non-zero and enter the model. Variables that become non-zero at higher λ values (further to the left in the plots) have stronger predictive power and thus greater importance compared to variables that become non-zero at lower λ values (further to the right in the plots). We observe that the LASSO method selects financial variables before non-financial ones, as financial variables become non-zero at higher λ values (further to the left in the plots). Moreover, in most cases, the financial variables are selected at much higher λ values than the non-financial variables, further indicating their relative importance.

We also observe that the coefficient signs of all financial variables in Tables 6, 7 and 8 follow the expected directions.¹³ The only exception being the positive sign of 'sales / total assets' for the variable set in Altman (1968). However, this variable is excluded from the revised versions of Altman's (1968)

model, which targets private companies across all industries, because it is highly industry-sensitive (Altman, 2018; Altman et al., 2019).

Moreover, Panels B and C of Tables 6, 7 and 8 show that among the three financial variable sets, the one in Paraschiv et al. (2023) yields the highest in-sample fit and out-of-sample prediction performance across all prediction horizons, years as our test sample and whether using financial variables exclusively or enriching them with non-financial variables. This is expected because Paraschiv et al. (2023) selected their variable set with a sample similar in many respects to that we use in our study.

6.1.2. Enhancing models with non-financial variables

Furthermore, we observe that enriching financial variables with non-financial ones improves bankruptcy prediction models. Specifically, as shown in Panels B and C of Tables 6, 7 and 8, using variable sets containing both financial and non-financial variables (top rows in the panels) results in a better in-sample fit and out-of-sample prediction performance across all evaluation metrics compared with using exclusively financial variables (bottom rows of the panel).

The improved model performance is due to the six non-financial variables selected by the LASSO method. Specifically, the LASSO method consistently selects ‘chairperson age’ across all 18 permutations of the financial variable sets, years used as the test sample and prediction horizons. Additionally, ‘CEO age’ and ‘board age avg’ are selected in all except one and two, respectively, cases. In all cases, we observe a negative coefficient for these variables, indicating that an older age results in a lower probability of bankruptcy. This finding is consistent with Platt and Platt (2012) finding that older CEOs and board members are associated with less bankruptcy. Furthermore, the LASSO method consistently selects the variable ‘board county’ in all cases, which indicates whether board members reside in the same county as the SMEs’ HQs. The coefficients of this variable are negative in all cases, indicating a lower probability of bankruptcy if board members reside in the same county as the SME. Finally, in most cases, the LASSO method selects the variables ‘board non-owners’ and ‘ownership board’. The estimated coefficients of these variables consistently have negative and positive signs, respectively. This finding indicates that a higher proportion of board members who are not shareholders decreases the SME’s likelihood of bankruptcy, and a lower share of ownership by board members also decreases the likelihood of bankruptcy. This may be because boards with a lower ownership share have more outsiders among board members, which can strengthen the board by providing more diversity than boards with high ownership in the SME.

6.1.3. Only non-financial variables

Next, we test the performance of variable sets consisting only of non-financial variables. Such variable sets are useful for assessing, for example, newly established firms without their first financial statement. We create these variable sets by allowing the LASSO method to consider only the 20 non-financial variables in Table 1. Table 9, Panel A shows the selected variables and estimation results when they are used in LR models using 2018 and 2019 as test samples and when predicting bankruptcy over horizons of one, two and three years. Panel B presents the in-sample fit and out-of-sample prediction performance.

While Panel B of Table 9 indicates a better prediction than random when exclusively using non-financial variables; specifically, $AUC > 0.5$, the prediction performance is not considered acceptable as $AUC < 0.7$ (see Section 5.4). Furthermore, we observe that across all evaluation metrics, the in-sample fit and out-of-sample prediction performance are higher when using variable sets consisting exclusively of financial variables or financial and non-financial variables combined (see Panels B and C of Tables 6, 7 and 8) than when using only non-financial variables (see Panel B of Table 9). This further proves the importance of financial variables for bankruptcy prediction, while non-financial variables should be used only in combination with financial variables.

Furthermore, the importance of the non-financial variables in Tables 6, 7 and 8 is supported as all these variables are present in Table 9 with the same sign for the estimated coefficient values. The only exception is ‘board non-owners’, which has another sign for the estimated coefficient in some cases in Table 9 compared with Tables 6, 7 and 8. However, the estimated coefficients are not statistically significant in these cases. Moreover, we observe in Figures IA.4, IA.8 and IA.12 in the Internet Appendix, which

Table 8. Estimation results when predicting bankruptcy over a horizon of three years.**Panel A.** Variables selected using the LASSO method and their LR coefficient estimates when predicting bankruptcy over a horizon of three years.

	Altman (1968)		Altman and Sabato (2007)		Paraschiv et al. (2023)	
	2018	2019	2018	2019	2018	2019
CEO age	−0.70(−8.56)	−0.72(−8.57)	−0.92(−11.19)	−0.91(−10.99)	−0.45(−5.29)	−0.45(−5.30)
Chairperson age	−0.93(−8.82)	−0.69(−6.27)	−0.86(−8.09)	−0.75(−6.92)	−0.51(−4.67)	−0.42(−3.74)
Board age avg	−0.34(−2.89)	−0.40(−3.20)	−0.38(−3.14)	−0.54(−4.34)	−0.06(−0.47)	−0.18(−1.41)
Board county	−0.74(−23.24)	−0.80(−23.77)	−0.97(−28.85)	−0.85(−25.04)	−0.88(−26.25)	−0.77(−22.86)
Board non-owners		−0.47(−14.49)	−0.33(−5.19)	−0.65(−20.20)	−0.28(−4.41)	−0.50(−15.33)
Ownership board			0.42(6.57)		0.27(4.19)	
EBIT / total assets	−0.88(−17.33)	−0.84(−16.61)				
Retained earnings / total assets	−0.31(−12.14)	−0.32(−12.54)	−0.62(−31.11)	−0.59(−30.05)		
Sales / total assets	0.15(25.77)	0.14(23.26)				
Working capital / total assets	−0.60(−16.67)	−0.63(−17.41)				
EBITDA / total assets			−0.95(−17.96)	−0.90(−16.85)		
Short-term liquidity / total assets			−1.34(−20.57)	−1.36(−20.71)		
(Current liabilities - short-term liquidity) / total assets					0.29(6.40)	0.29(6.60)
Accounts payable / total assets					1.20(18.33)	1.21(18.46)
Dummy; one if paid-in equity is less than total equity					−0.53(−13.53)	−0.52(−12.89)
Dummy; one if total liability exceeds total assets					0.17(3.93)	0.21(4.59)
Interest expenses / total assets						1.99(6.44)
Inventory / current assets					0.48(10.16)	0.35(7.37)
Log(age in years)					−0.29(−22.49)	−0.31(−23.55)
Net income / total assets					−0.39(−7.15)	−0.27(−4.89)
Public taxes payable / total assets					3.34(27.74)	3.08(25.52)
Short-term liquidity / current assets					−1.13(−18.11)	−1.10(−17.72)
Intercept	3.43(13.89)	3.04(12.48)	4.81(19.60)	5.20(22.21)	0.48(1.76)	0.76(2.84)

Panel B. In-sample fit when predicting bankruptcy over a horizon of three years.

		Altman (1968)		Altman and Sabato (2007)		Paraschiv et al. (2023)	
		2018	2019	2018	2019	2018	2019
Financial and non-financial variables	R^2	0.088	0.088	0.085	0.081	0.137	0.133
	AUC	0.777	0.777	0.767	0.764	0.826	0.825
	AP	0.041	0.040	0.040	0.035	0.060	0.055
	AR	0.548	0.549	0.528	0.522	0.644	0.643
	KS statistic	0.410	0.416	0.398	0.391	0.505	0.500
	Hinge loss	1.010	1.010	1.010	1.010	1.010	1.010
	Log loss	0.056	0.052	0.056	0.053	0.053	0.050
	AIC	58,635	57,231	58,773	57,678	55,455	54,415
	BIC	58,736	57,343	58,885	57,779	55,634	54,594
	Brier score	0.011	0.010	0.011	0.010	0.011	0.010
	BSS	0.010	0.010	0.012	0.010	0.023	0.021
	Decile 1	0.393	0.405	0.397	0.383	0.499	0.488
	Decile 2	0.189	0.188	0.175	0.182	0.192	0.195
	Decile 3	0.122	0.114	0.121	0.119	0.105	0.109
Exclusively financial variables	Decile 4	0.091	0.092	0.086	0.095	0.071	0.075
	Decile 5	0.073	0.064	0.070	0.068	0.046	0.049
	Decile 6-10	0.132	0.137	0.151	0.153	0.088	0.084
	R^2	0.066	0.064	0.050	0.048	0.122	0.119
	AUC	0.758	0.753	0.723	0.719	0.814	0.813
	AP	0.035	0.033	0.028	0.026	0.052	0.048
	AR	0.509	0.502	0.441	0.434	0.621	0.620
	KS statistic	0.400	0.394	0.325	0.317	0.486	0.484
	Hinge loss	1.011	1.010	1.011	1.010	1.010	1.010
	Log loss	0.057	0.054	0.058	0.055	0.054	0.051
	AIC	60,037	58,707	61,014	59,739	56,434	55,286
	BIC	60,104	58,774	61,081	59,806	56,557	55,409
	Brier score	0.011	0.010	0.011	0.010	0.011	0.010
	BSS	0.006	0.006	0.005	0.004	0.018	0.017
	Decile 1	0.348	0.351	0.331	0.328	0.463	0.452
	Decile 2	0.205	0.195	0.162	0.154	0.197	0.203
	Decile 3	0.139	0.140	0.125	0.122	0.119	0.123
	Decile 4	0.092	0.093	0.100	0.106	0.073	0.076
	Decile 5	0.062	0.061	0.082	0.086	0.052	0.053
	Decile 6-10	0.153	0.160	0.201	0.204	0.096	0.093

Panel C. Out-of-sample prediction performance when predicting bankruptcy over a horizon of three years.

		Altman (1968)		Altman and Sabato (2007)		Paraschiv et al. (2023)	
		2018	2019	2018	2019	2018	2019
Financial and non-financial variables	AUC	0.752	0.785	0.741	0.770	0.810	0.840
	AP	0.025	0.009	0.023	0.008	0.035	0.015
	AR	0.499	0.569	0.479	0.538	0.615	0.679
	KS statistic	0.375	0.436	0.358	0.406	0.487	0.531
	Hinge loss	1.010	1.010	1.010	1.009	1.010	1.009
	Log loss	0.043	0.017	0.043	0.017	0.040	0.016
	AIC	12,270	5,058	12,326	5,064	11,679	4,780
	BIC	12,359	5,157	12,425	5,153	11,837	4,939
	Brier score	0.008	0.002	0.008	0.002	0.008	0.002
	BSS	−0.007	−0.119	−0.006	−0.093	0.000	−0.117
	Decile 1	0.342	0.445	0.348	0.415	0.440	0.570
	Decile 2	0.182	0.166	0.174	0.147	0.208	0.155
	Decile 3	0.138	0.117	0.127	0.125	0.129	0.091
	Decile 4	0.106	0.083	0.102	0.102	0.072	0.072
Exclusively financial variables	Decile 5	0.073	0.057	0.068	0.087	0.058	0.045
	Decile 6-10	0.158	0.132	0.182	0.125	0.092	0.068
	AUC	0.741	0.768	0.700	0.726	0.809	0.840
	AP	0.024	0.007	0.019	0.005	0.036	0.013
	AR	0.479	0.536	0.396	0.451	0.613	0.679
	KS statistic	0.380	0.426	0.305	0.328	0.475	0.539
	Hinge loss	1.011	1.010	1.010	1.010	1.010	1.010
	Log loss	0.043	0.018	0.046	0.018	0.041	0.017
	AIC	12,427	5,290	13,248	5,392	11,718	4,950
	BIC	12,486	5,349	13,307	5,451	11,827	5,059
	Brier score	0.008	0.002	0.008	0.002	0.008	0.002
	BSS	−0.007	−0.124	−0.005	−0.086	0.003	−0.127
	Decile 1	0.348	0.411	0.335	0.351	0.430	0.551
	Decile 2	0.179	0.192	0.133	0.162	0.206	0.185
	Decile 3	0.148	0.109	0.124	0.098	0.131	0.087
	Decile 4	0.083	0.083	0.099	0.109	0.082	0.060
	Decile 5	0.068	0.045	0.100	0.068	0.057	0.026
	Decile 6-10	0.174	0.158	0.209	0.211	0.093	0.091

Notes: Estimation results of LR models that predict bankruptcy over a three-year horizon, using variable sets selected by the LASSO method from a population of non-financial variables in Table 1 and a set of financial variables. The columns display the results of permutations using all financial statements from 2018 and 2019 as out-of-sample test samples, and the variables in Altman (1968), Altman and Sabato (2007) and Paraschiv et al. (2023), presented in Table 2, as the set of financial variables. Panel A presents the variables selected using the LASSO method, along with their coefficient estimates and z-scores in parentheses, derived from standard errors clustered at the firm level. Panels B and C report the in-sample fit and out-of-sample prediction performance, respectively, using R^2 , AUC, AP, AR, KS statistic, Hinge loss, Log loss, AIC, BIC, Brier score, BSS and decile rankings. The first rows in Panels B and C report metric values when using the variable sets selected by the LASSO method, shown in Panel A, from the population of both non-financial and financial variables. The bottom rows show the metric values when the financial variables are used exclusively per benchmark set, as shown in 2. To train the models, a four-year rolling window approach was followed in which the models were trained on all financial statements from the four accounting years preceding the test populations. These data are presented in 3.

present the LASSO path plots for six modeling permutations shown in the six columns of Table 9, that 'board non-owners' is among the last chosen by the LASSO method, that is, at low λ values.

6.2. Robustness tests

6.2.1. Robustness using balanced datasets

As with any bankruptcy prediction dataset, our sample is highly imbalanced; the rate of financial statements categorized as bankrupt is much lower than those categorized as non-bankrupt (Beaver et al., 2010). We acknowledge that sampling to create a balanced training dataset would distort the capabilities of a bankruptcy prediction model because the ratio of non-bankrupt to bankrupt observations in the data used for development deviates from the real-world population (Zmijewski, 1984). However, in this section, we investigate the robustness of our main results when using balanced datasets because this approach has been shown to improve classification accuracy significantly (J. Garcia, 2022). Specifically, we use SMOTE (Fernandez et al., 2018) to generate synthetic samples for the minority class by interpolating between existing minority class samples. SMOTE selects a minority class sample, identifies its k -nearest neighbors and creates new synthetic samples along the line segments joining the

selected sample and its neighbors. This approach helps balance the class distribution without duplicating the existing minority class samples. We use this approach on the data for each accounting year thrice, depending on whether we categorize bankruptcy over one, two, or three-year horizons, resulting in three datasets with 1,631,396, 1,620,862 and 1,623,426 financial statements, respectively, each with 50% of observations categorized as bankrupt per accounting year.

Tables IA.1, IA.2 and IA.3 in Internet Appendix IA.4 present the model performance when using balanced datasets in predicting bankruptcy over horizons of one year, two years and three years, respectively. In each table, Panels A and B exhibit the in-sample fit and out-of-sample prediction performance, respectively. As before, we let the LASSO method select variables from a population of non-financial variables in Table 1 and a set of financial variables, and we measure the performance of using the selected models in LR models. The performance of these models is displayed in the tables' top rows. The bottom rows show the metric values when using exclusively the financial variables per benchmark set. The tables' columns display the results of permutations, using all financial statements from 2018 and 2019 as out-of-sample test samples and the variables in Altman (1968), Altman and Sabato (2007) and Paraschiv et al. (2023) as the set of financial variables. We also follow the same four-year rolling window approach. In summary, Tables IA.1, IA.2 and IA.3 are comparable to Tables 6, 7 and 8, but for when using balanced datasets.

We observe that Tables IA.1, IA.2 and IA.3 confirm our findings. Specifically, across all metrics and test settings, variable sets that include financial and non-financial variables (top rows of panels) yield superior in-sample fit and out-of-sample prediction performance compared to sets that use only the financial variables (bottom rows of panels). We also note that among the three financial variable sets, the one in Paraschiv et al. (2023) consistently yields the highest in-sample fit and out-of-sample prediction performance. Additionally, we observe that while the AUC is at the same level as when not employing a balanced dataset (see Tables 6, 7 and 8), the AP is higher when using balanced datasets. This is expected because in the context of highly imbalanced datasets, it is common to observe that the AP is significantly lower than the AUC. This discrepancy arises from the different sensitivities of these metrics to class imbalances. The AUC evaluates the model's ability to distinguish between classes by considering both true positive and false positive rates, which can result in deceptively high values owing to the large number of true negatives. In contrast, AP focuses on the precision-recall trade-off, directly reflecting the model's performance on the minority class. Consequently, the AP is more sensitive to the model's ability to correctly identify the minority class, often resulting in lower values in imbalanced datasets.

Furthermore, Table IA.4 in the Internet Appendix displays the model performance when considering exclusively non-financial variables and using the balanced datasets. As with not using a balanced dataset (see Table 9), the in-sample fit and out-of-sample prediction performance is overall higher across all evaluation metrics when using variable sets of exclusively financial variables or combined financial and non-financial variables (see Tables IA.1, IA.2 and IA.3), compared with using only non-financial variables (see Table IA.4).

Figures IA.13, IA.14 and IA.15 in Internet Appendix IA.4 present the associated LASSO path plots, when using the balanced datasets, of the variables selected among the non-financial ones and those in Altman (1968), Altman and Sabato (2007) and Paraschiv et al. (2023), respectively, when predicting bankruptcy over a horizon of one year. In each figure, Panels A and B present the plots using 2018 and 2019 as out-of-sample test samples. Furthermore, Figures IA.17, IA.18 and IA.19 present the same but for predicting bankruptcy over a horizon of two years, whereas Figures IA.21, IA.22 and IA.23 present the same but for predicting bankruptcy over a horizon of three years. Moreover, Figures IA.16, IA.20 and IA.24 present the LASSO path plots of the variables selected among only the non-financial variables when using a balanced dataset and predicting bankruptcy over horizons of one year, two years and three years, respectively.

As with not using balanced datasets, we observe from the LASSO path plots that financial variables are selected before non-financial ones, as financial variables become non-zero at higher λ values (further to the left in the plots), and that in most cases, the financial variables are selected at much higher λ values. Thus, the LASSO path plots also confirm the importance of the financial variables when the balanced datasets were used. Moreover, we observe that when using the balanced datasets, more non-financial variables are used than when using non-balanced datasets (see Section 6.1.2). However, the

Table 9. Estimation results of only non-financial variables.**Panel A.** Variables selected using the LASSO method and their LR coefficient estimates when exclusively considering non-financial variables.

	Bankrupted over a one-year horizon		Bankrupted over a two-year horizon		Bankrupted over a three-year horizon	
	2018	2019	2018	2019	2018	2019
CEO age	−1.04(−7.58)	−0.97(−7.31)	−0.86(−10.53)	−0.95(−11.90)	−0.98(−11.88)	−1.02(−12.08)
CEO woman	0.18(3.24)	0.09(1.41)	0.07(1.61)	0.08(2.10)	0.11(3.26)	0.12(3.46)
CEO duality	−0.23(−4.24)	−0.21(−4.09)	−0.17(−5.12)	−0.16(−5.16)	−0.16(−5.00)	
Two CEOs	−2.95(−2.95)	−2.40(−3.40)	−1.50(−4.96)	−1.41(−5.26)	−1.66(−5.22)	−1.46(−5.05)
Chairperson age	−0.94(−5.04)	−0.68(−3.85)	−0.83(−7.71)	−0.77(−7.27)	−0.96(−8.83)	−0.73(−6.49)
Chairperson woman		0.15(2.15)	0.05(1.19)	0.06(1.36)		
Chairperson county		0.30(4.78)			0.21(5.38)	
Board size	−0.22(−3.82)	−0.18(−3.05)	−0.29(−9.50)	−0.27(−7.85)	−0.30(−8.65)	−0.33(−9.69)
Board age avg	−0.10(−0.48)	−0.65(−3.35)	−0.78(−6.48)	−0.72(−6.12)	−0.35(−2.91)	−0.52(−4.17)
Board age std	−0.02(−2.57)	−0.01(−1.34)				
Board county	−1.12(−21.08)	−1.36(−22.05)	−1.14(−34.85)	−1.07(−33.66)	−1.21(−31.10)	−1.00(−30.24)
Board non-owners	0.10(0.85)	−0.09(−0.79)		0.05(0.69)	0.14(1.82)	0.13(1.60)
Ownership concentration 1			−0.08(−1.35)	−0.26(−4.28)	−0.29(−4.65)	−0.35(−5.49)
Ownership concentration 2	0.75(3.53)	1.15(5.74)	0.72(4.84)	0.44(2.97)	0.41(2.67)	0.28(1.81)
Ownership CEO			−0.47(−8.08)			
Ownership board	0.56(5.03)	0.45(4.21)	0.76(15.66)	0.38(5.25)	0.51(6.88)	0.36(4.74)
Intercept	3.18(8.09)	4.01(10.71)	5.86(24.82)	5.87(25.48)	5.08(21.00)	5.05(20.95)

Panel B: In-sample fit and out-of-sample prediction performance when considering non-financial variables exclusively.

		Bankrupted over a one-year horizon		Bankrupted over a two-year horizon		Bankrupted over a three-year horizon	
		2018	2019	2018	2019	2018	2019
In-sample fit	R^2	0.036	0.043	0.049	0.045	0.042	0.039
	AUC	0.686	0.697	0.696	0.688	0.681	0.677
	AP	0.010	0.012	0.030	0.030	0.028	0.025
	AR	0.370	0.393	0.388	0.371	0.359	0.350
	KS statistic	0.275	0.291	0.287	0.271	0.261	0.260
	Hinge loss	1.004	1.004	1.011	1.011	1.011	1.010
	Log loss	0.026	0.026	0.058	0.059	0.058	0.055
	AIC	27,209	28,319	61,242	64,207	61,555	60,321
	BIC	27,354	28,487	61,398	64,363	61,711	60,456
	Brier score	0.004	0.004	0.011	0.011	0.011	0.010
	BSS	0.002	0.003	0.008	0.008	0.007	0.006
	Decile 1	0.269	0.294	0.299	0.290	0.283	0.275
	Decile 2	0.157	0.158	0.157	0.158	0.154	0.154
	Decile 3	0.138	0.123	0.122	0.118	0.119	0.117
	Decile 4	0.098	0.113	0.103	0.100	0.095	0.111
	Decile 5	0.089	0.082	0.077	0.081	0.086	0.076
Out-of-sample prediction performance	Decile 6-10	0.248	0.229	0.241	0.254	0.262	0.267
	AUC	0.698	0.637	0.656	0.653	0.643	0.679
	AP	0.014	0.007	0.027	0.014	0.015	0.004
	AR	0.396	0.274	0.308	0.306	0.285	0.358
	KS statistic	0.295	0.212	0.225	0.247	0.221	0.271
	Hinge loss	1.004	1.004	1.010	1.010	1.010	1.009
	Log loss	0.027	0.019	0.060	0.041	0.044	0.018
	AIC	7,768	5,778	17,174	12,174	12,824	5,204
	BIC	7,896	5,926	17,312	12,313	12,963	5,323
	Brier score	0.004	0.003	0.011	0.007	0.008	0.002
	BSS	0.004	0.001	0.006	−0.002	−0.000	−0.053
	Decile 1	0.324	0.249	0.258	0.232	0.224	0.245
	Decile 2	0.148	0.136	0.141	0.138	0.135	0.151
	Decile 3	0.110	0.103	0.116	0.132	0.129	0.151
	Decile 4	0.095	0.089	0.101	0.137	0.126	0.113
	Decile 5	0.084	0.103	0.084	0.088	0.085	0.102
	Decile 6-10	0.240	0.319	0.300	0.273	0.301	0.238

Notes: The estimation results of the LR models using variable sets selected by the LASSO method from a population consisting exclusively of the non-financial variables in Table 1. The columns display the results of six permutations using all financial statements from 2018 and 2019 as out-of-sample test samples and predicting bankruptcy over horizons of one, two and three years. Panel A presents the variables selected using the LASSO method, along with their coefficient estimates and z-scores in parentheses, derived from standard errors clustered at the firm level. Panel B reports the in-sample fit and out-of-sample prediction performance using R^2 , AUC, AP, AR, KS statistic, Hinge loss, Log loss, AIC, BIC, Brier score, BSS and decile rankings. The first rows in Panel B report the in-sample fit, whereas the bottom rows show the out-of-sample prediction performance. To train the models, a four-year rolling window approach was followed in which the models were trained on all financial statements from the four accounting years preceding the test populations. These data are presented in 3.

importance of the six non-financial variables selected when unbalanced datasets are not used is confirmed when using balanced datasets, as they are still the most important among the non-financial variables, being selected at higher λ values (further to the left in the LASSO path plots). Furthermore, the signs of their estimated coefficients are in the same direction when using balanced datasets, as indicated by their values on the vertical axis of the LASSO path plots.

6.2.2. Robustness using machine learning methods

In this section, we test the robustness of our findings by alternatively using the machine learning methods described in Section 5.3. Tables IA.5, IA.6, IA.7 and IA.8 in the Internet Appendix display the out-of-sample prediction performance when using XGBoost, AdaBoost, bagging and random forest, respectively. In each table, Panels A, B and C show the prediction performance for bankruptcy over horizons of one, two and three years, respectively. Furthermore, all tables have six columns representing the six modeling permutations using one of the three financial variable sets and 2018 and 2019 as out-of-sample test samples. Additionally, each table shows the prediction performance using variable sets containing financial and non-financial variables in the top rows and exclusively financial variables in the bottom rows.

Compared to LR models, in which we first select variables using LASSO, we do not perform feature selection before applying machine learning methods. This is due to their high flexibility and their ability to capture nonlinear relationships, allowing the use of many variables, especially when the dataset is large. Thus, when applying machine learning methods, we use all the variables, either all the financial variables in one of the financial variable sets or all the financial variables in a financial variable set in combination with all 20 non-financial variables.

Furthermore, in Figures IA.25, IA.26 and IA.27 in the Internet Appendix, we show the variable importance in the XGBoost models for predicting bankruptcy over horizons of one year, two years and three years, respectively. Similarly, Figures IA.28, IA.29 and IA.30 show variable importance in the AdaBoost models; Figures IA.31, IA.32 and IA.33 exhibit variable importance when using bagging, whereas Figures IA.34, IA.35 and IA.36 exhibit variable importance when employing random forest. Each figure has six panels for the six modeling permutations using one of the three financial variable sets and 2018 and 2019 as out-of-sample test samples. Moreover, we show only the ten most important variables in each panel, while the financial and non-financial variables are marked with green and purple bars, respectively. For the XGBoost method, feature importance is determined by the average gain across all splits where the variable is used. For AdaBoost and Random Forest, the importance of a variable is the normalized total reduction in Gini impurities. For bagging, the feature importance is assessed using permutation importance (Breiman, 2001), which involves evaluating the reduction in the model's performance when the relationship between the feature and the actual outcome is disrupted.

We observe that the results when using the machine learning methods confirm our finding that including both financial and non-financial variables (top rows in the panels) overall yields better prediction performance than using only financial variables (bottom rows of the panels). Furthermore, the results confirm the importance of financial variables, as they exhibit higher importance in machine learning models than non-financial variables. Finally, the non-financial variables with the highest importance in the machine learning models are overall the same six as those selected by the LASSO method in the earlier results (see Panel A in Tables 6, 7 and 8). In other words, when using machine learning methods instead of LR and LASSO, the most important non-financial variables overall remain those indicating the age of CEOs, board members and chairpersons as well as ownership share with the board members and whether they reside in the same county as the SMEs' HQs.

7. Conclusions and implications

The theoretical implication of our study is that non-financial information available to all SMEs in an economy can improve assessments of their collective financial standing. While the literature has highlighted problems with using accounting-based financial variables for SME bankruptcy prediction due to their less detailed and more opaque nature when produced by SMEs compared to larger firms (Agarwal & Taffler, 2008; Berger & Udell, 2007; Ciampi, 2015; Ciampi et al., 2021; Hillegeist et al., 2004), our study demonstrates that such variables remain the most significant factors for predicting SME bankruptcy

across an entire economy. However, integrating both financial and non-financial variables into SME bankruptcy prediction models enhances their performance compared to solely employing financial variables. Moreover, we find that models based solely on non-financial variables do not yield sufficient performance, further highlighting the importance of using them together with financial variables. Our findings are robust across different financial variable sets as benchmarks, estimation methods and machine learning models, years as out-of-sample test populations, horizons for predicting bankruptcy and for using balanced and non-balanced datasets.

Additionally, our results indicate the most important non-financial predictors of SME bankruptcy in a large population. First, variables related to the age of CEOs, chairpersons and board members are important, and our results indicate that an older age results in a lower probability of bankruptcy. This aligns with previous research, which indicates that older CEOs and board members are associated with a lower incidence of bankruptcy (Platt & Platt, 2012). Second, our findings indicate that fewer board members who are shareholders and a lower ownership share among board members are linked to a lower likelihood of bankruptcy. This contradicts the idea that higher ownership provides greater incentives and previous literature that has found that CEO ownership increases the performance of listed firms (Lilienfeld-Toal & Ruenzi, 2014). However, the dynamics of ownership with board members could be different from that of CEOs. Specifically, while the CEO works with the firm on a daily basis and could thus be considered internal regardless of ownership share, lower ownership shares by board members, especially in SMEs, indicate more external members, which can enhance the board's effectiveness by offering greater diversity. Finally, our findings indicate that more board members residing in the same county as the SME is linked with a lower probability of bankruptcy. This aligns with previous research, which shows that bankruptcy is linked to having fewer board members residing in the same county as the firm's registered address (Wilson & Altanlar, 2014).

The practical implications of our findings are manifold: Better assessments of the financial standing of large populations of firms, which our non-financial variables allow for, benefit investors and policy-makers, aiding in the design of support schemes for businesses and predicting losses in tax revenues due to bankruptcies. Furthermore, enhanced SME bankruptcy prediction models benefit banks and investors through improved risk management and reduced lending errors owing to more accurate assessments of potential new borrowers and better monitoring of existing borrowers. Additionally, improved bankruptcy predictions enable more precise determination of the risk-weighted value of loan portfolios. Enhanced bankruptcy prediction models increase access to formal capital for SMEs by reducing information asymmetry. Better credit assessments are also important for firms in general when assessing credit exposure to customers and suppliers and for regulators conducting on-site supervision of banks. In conclusion, better evaluations of SMEs' financial health benefit the entire economy, particularly because SMEs constitute a significant portion of it.

Future research should explore additional non-financial variables that may influence SME bankruptcy, such as industry-specific factors, market conditions and socioeconomic variables. Longitudinal studies can provide deeper insight into how these variables interact over time. Additionally, further refinement of the prediction models using advanced machine learning techniques can enhance their accuracy and applicability across different contexts. To this end, we grant access to all our codes for all our test settings and distinct methodological frameworks. This transparency allows academics and practitioners to replicate our study, improve prediction models and evaluate the significance of bankruptcy predictors based on available internal and external information.

Notes

1. SMEs account for over half of Europe's GDP (European Commission, 2020) and employ approximately 48% of private-sector employees in the United States (OECD, 2022).
2. Another ownership structure variable available to us is the number of shareholders. However, this highly correlates with our variable for average shareholder holdings ('ownership concentration 1'). Thus, we only include the latter variable.
3. For example, Wilson and Altanlar (2014) focus on newly incorporated UK companies during 2000–2008, Ciampi (2015) focuses on Italian small firms from 2008 in the manufacturing, building and service sectors and Altman et al. (2023a) focus on Croatian SMEs from 2015 to 2019.

4. Accessed September 18, 2024: datatopics.worldbank.org/world-development-indicators/the-world-by-income-and-region.html.
5. Accessed September 18, 2024: www.nho.no/tema/sma-og-mellomstore-bedrifter/tall-og-fakta-om-smb.
6. See www.brreg.no/en/.
7. SIC2007 is based on UN's ISIC Rev. 4 and EU's NACE Rev. 2. For an overview, we refer to www.ssb.no/nace.
8. See www.enin.ai/.
9. The EU recommendation 2003/361 defines a company as an SME based on, first, the number of employees and, second, either total assets or turnover. As the number of employees is unavailable in our data, we rely on the latter classification. Basel III relies on a somewhat similar definition: A company is an SME if its sales are not larger than EUR 50 million.
10. The exchange rates are available at www.norges-bank.no/en/topics/Statistics/exchange_rates.
11. As the variables are logged (see their definition in Table 1), we transform here with the exponential function to derive real values.
12. For XGBoost, we tune step size shrinkage, sample fraction per tree, minimum loss reduction, the L2 regularization term, number of trees and maximum tree depth. For bagging, we tune the number of decision trees and the sample and feature fractions per tree. For random forest, we tune the number of trees, minimum samples to split a node and minimum samples per leaf. For AdaBoost, we tune the number of weak classifiers and their weights per iteration.
13. We refer to the respective studies for a discussion on the coefficient signs of the variables in their sets.

Author contributions

Ranik Raaen Wahlstrøm: Conceptualization, Methodology, Formal analysis, Data curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision.

Linn-Kristin Becker: Methodology, Formal analysis, Writing - Original Draft.

Trude Nonstad Fornes: Methodology, Formal analysis, Writing - Original Draft.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Code availability statement

Code for running all our analyses is available at: <https://enhancing-credit-risk-2024.ranik.no>.

About the authors

Ranik Raaen Wahlstrøm, Ph.D., serves as an Associate Professor at the NTNU Business School. His research interests include data science, machine learning, quantitative finance, bankruptcy and default prediction, FinTech, and accounting.

Linn-Kristin Becker is an analyst at Intrum and holds a master's degree from the NTNU Business School.

Trude Nonstad Fornes is a project manager at Bane NOR and holds a master's degree from the NTNU Business School.

ORCID

Ranik Raaen Wahlstrøm  <http://orcid.org/0000-0001-7563-4380>

Data availability statement

The data that support the findings of this study are available from the Norwegian government agency Brønnøysund Register Centre and Enin AS. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the authors with the permission of the Brønnøysund Register Centre and Enin AS.

References

- Agarwal, V., & Taffler, R. (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance*, 32(8), 1541–1551. <https://doi.org/10.1016/j.jbankfin.2007.07.014>
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589–609. <https://doi.org/10.2307/2978933>
- Altman, E. I. (2018). A fifty-year retrospective on credit risk models, the Altman Z -score family of models and their applications to financial markets and managerial strategies. *The Journal of Credit Risk*, 14(4), 1–34. <https://doi.org/10.21314/JCR.2018.243>
- Altman, E. I., Balzano, M., Giannozzi, A., & Srhoj, S. (2023a). Revisiting SME default predictors: The Omega score. *Journal of Small Business Management*, 61(6), 2383–2417. <https://doi.org/10.1080/00472778.2022.2135718>
- Altman, E. I., Balzano, M., Giannozzi, A., & Srhoj, S. (2023b). The Omega score: An improved tool for SME default predictions. *Journal of the International Council for Small Business*, 4(4), 362–373. <https://doi.org/10.1080/26437015.2023.2186284>
- Altman, E. I., Haldeman, R. G., & Narayanan, P. (1977). ZETA analysis A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance*, 1(1), 29–54. [https://doi.org/10.1016/0378-4266\(77\)90017-6](https://doi.org/10.1016/0378-4266(77)90017-6)
- Altman, E. I., Hotchkiss, E., & Wang, W. (2019). *Corporate financial distress, restructuring, and bankruptcy: Analyze leveraged finance, distressed debt, and bankruptcy* (4th ed.). John Wiley & Sons.
- Altman, E. I., Iwanicz-Drozdzowska, M., Laitinen, E. K., & Suvas, A. (2017). Financial distress prediction in an international context: A review and empirical analysis of Altman's Z-score model. *Journal of International Financial Management & Accounting*, 28(2), 131–171. <https://doi.org/10.1111/jifm.12053>
- Altman, E. I., & Sabato, G. (2007). Modelling credit risk for SMEs: Evidence from the U.S. market. *Abacus*, 43(3), 332–357. <https://doi.org/10.1111/j.1467-6281.2007.00234.x>
- Altman, E. I., Sabato, G., & Wilson, N. (2010). The value of non-financial information in small and medium-sized enterprise risk management. *The Journal of Credit Risk*, 6, 1–33. <https://doi.org/10.21314/JCR.2010.110>
- Antulov-Fantulin, N., Lagravinese, R., & Resce, G. (2021). Predicting bankruptcy of local government: A machine learning approach. *Journal of Economic Behavior & Organization*, 183, 681–699. <https://doi.org/10.1016/j.jebo.2021.01.014>
- Arcuri, G., & Levratto, N. (2020). Early stage SME bankruptcy: Does the local banking market matter? *Small Business Economics*, 54(2), 421–436. <https://doi.org/10.1007/s11187-018-0042-4>
- Back, P. (2005). Explaining financial difficulties based on previous payment behavior, management background variables and financial ratios. *European Accounting Review*, 14(4), 839–868. <https://doi.org/10.1080/09638180500141339>
- Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405–417. <https://doi.org/10.1016/j.eswa.2017.04.006>
- Bauer, J., & Agarwal, V. (2014). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking & Finance*, 40, 432–442. <https://doi.org/10.1016/j.jbankfin.2013.12.013>
- Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research*, 4, 71–111. <https://doi.org/10.2307/2490171>
- Beaver, W. H., Correia, M., & McNichols, M. F. (2010). Financial statement analysis and the prediction of financial distress. *Foundations and Trends in Accounting*, 5(2), 99–173. <https://doi.org/10.1561/14000000018>
- Beaver, W. H., McNichols, M. F., & Rhie, J. W. (2005). Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy. *Review of Accounting Studies*, 10(1), 93–122. <https://doi.org/10.1007/s11142-004-6341-9>
- Beck, T., & Demirguc-Kunt, A. (2006). Small and medium-size enterprises: Access to finance as a growth constraint. *Journal of Banking & Finance*, 30(11), 2931–2943. <https://doi.org/10.1016/j.jbankfin.2006.05.009>
- Beck, T., Demirguc-Kunt, A., & Maksimovic, V. (2008). Financing patterns around the world: Are small firms different? *Journal of Financial Economics*, 89(3), 467–487. <https://doi.org/10.1016/j.jfineco.2007.10.005>
- Berger, A. N., & Udell, G. F. (2007). Small business credit scoring and credit availability. *Journal of Small Business Management*, 45(1), 5–22. <https://doi.org/10.1111/j.1540-627X.2007.00195.x>
- Berger, A. N., Udell, G. F., Petersen, M. A., Udell, G. F., & Stein, J. C. (2005). Does function follow organizational form? Evidence from the lending practices of large and small banks. *Journal of Financial Economics*, 76(2), 237–269. <https://doi.org/10.1016/j.jfineco.2004.06.003>
- Berger, A. N., & Udell, G. F. (1998). The economics of small business finance: The roles of private equity and debt markets in the financial growth cycle. *Journal of Banking & Finance*, 22, 613–673. [https://doi.org/10.1016/S0378-4266\(98\)00038-7](https://doi.org/10.1016/S0378-4266(98)00038-7)
- Berger, A. N., & Udell, G. F. (2002). Small business credit availability and relationship lending: The importance of bank organisational structure. *The Economic Journal*, 112(477), F32–F53. <https://doi.org/10.1111/1468-0297.00682>
- Blöchliger, A., & Leppold, M. (2018). Are ratings the worst form of credit assessment except for all the others? *Journal of Financial and Quantitative Analysis*, 53(1), 299–334. <https://doi.org/10.1017/S0022109017000874>
- Block, J. H., Colombo, M. G., Cumming, D. J., & Vismara, S. (2018). New players in entrepreneurial finance and why they are there. *Small Business Economics*, 50(2), 239–250. <https://doi.org/10.1007/s11187-016-9826-6>

- Borghans, L., Heckman, J. J., Golsteyn, B. H. H., & Meijers, H. (2009). Gender differences in risk aversion and ambiguity aversion. *Journal of the European Economic Association*, 7(2-3), 649–658. <https://doi.org/10.1162/JEEA.2009.7.2-3.649>
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123–140. <https://doi.org/10.1007/BF00058655>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1), 1–3. [https://doi.org/10.1175/1520-0493\(1950\)078<0001:VOFEIT;2.0.CO;2](https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT;2.0.CO;2)
- Calomiris, C. W., & Mamaysky, H. (2019). How news and its context drive risk and returns around the world. *Journal of Financial Economics*, 133(2), 299–336. <https://doi.org/10.1016/j.jfineco.2018.11.009>
- Campbell, J. Y., Hilscher, J., & Szilagyi, J. (2008). In search of distress risk. *The Journal of Finance*, 63(6), 2899–2939. <https://doi.org/10.1111/j.1540-6261.2008.01416.x>
- Campello, M., Gao, J., Qiu, J., & Zhang, Y. (2018). Bankruptcy and the cost of organized labor: Evidence from union elections. *The Review of Financial Studies*, 31(3), 980–1013. <https://doi.org/10.1093/rfs/hhx117>
- Chae, H. C. (2024). In search of gazelles: Machine learning prediction for Korean high-growth firms. *Small Business Economics*, 62(1), 243–284. <https://doi.org/10.1007/s11187-023-00760-8>
- Chang, X., Chen, Y., Wang, S. Q., Zhang, K., & Zhang, W. (2019). Credit default swaps and corporate innovation. *Journal of Financial Economics*, 134(2), 474–500. <https://doi.org/10.1016/j.jfineco.2017.12.012>
- Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, 83(1), 50–58. <https://doi.org/10.1016/j.jebo.2011.06.007>
- Chava, S., & Jarrow, R. A. (2004). Bankruptcy prediction with industry effects. *Review of Finance*, 8(4), 537–569. <https://doi.org/10.1093/rof/8.4.537>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)* (pp. 785–794). Association for Computing Machinery. New York, NY, USA. <https://doi.org/10.1145/2939672.2939785>
- Chinco, A., Clark-Joseph, A. D., & Ye, M. (2019). Sparse signals in the cross-section of returns. *The Journal of Finance*, 74(1), 449–492. <https://doi.org/10.1111/jofi.12733>
- Cho, E., Okafor, C., Ujah, N., & Zhang, L. (2021). Executives' gender-diversity, education, and firm's bankruptcy risk: Evidence from China. *Journal of Behavioral and Experimental Finance*, 30, 100500. <https://doi.org/10.1016/j.jbef.2021.100500>
- Ciampi, F. (2015). Corporate governance characteristics and default prediction modeling for small enterprises. An empirical analysis of Italian firms. *Journal of Business Research*, 68(5), 1012–1025. <https://doi.org/10.1016/j.jbusres.2014.10.003>
- Ciampi, F., Cillo, V., & Fiano, F. (2020). Combining Kohonen maps and prior payment behavior for small enterprise default prediction. *Small Business Economics*, 54(4), 1007–1039. <https://doi.org/10.1007/s11187-018-0117-2>
- Ciampi, F., Giannozzi, A., Marzi, G., & Altman, E. I. (2021). Rethinking SME default prediction: A systematic literature review and future perspectives. *Scientometrics*, 126(3), 2141–2188. <https://doi.org/10.1007/s11192-020-03856-0>
- Coad, A., & Srhoj, S. (2020). Catching Gazelles with a Lasso: Big data techniques for the prediction of high-growth firms. *Small Business Economics*, 55(3), 541–565. <https://doi.org/10.1007/s11187-019-00203-3>
- Combs, J. G., Ketchen, D. J., Jr., Perryman, A. A., & Donahue, M. S. (2007). The moderating effect of CEO power on the board composition–firm performance relationship. *Journal of Management Studies*, 44(8), 1299–1323. <https://doi.org/10.1111/j.1467-6486.2007.00708.x>
- Cornée, S. (2019). The relevance of soft information for predicting small business credit default: Evidence from a social bank. *Journal of Small Business Management*, 57(3), 699–719. <https://doi.org/10.1111/jsbm.12318>
- Crammer, K., & Singer, Y. (2001). On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning Research*, 2, 265–292.
- Crosato, L., Domenech, J., & Liberati, C. (2023). Websites' data: A new asset for enhancing credit risk modeling. *Annals of Operations Research*. <https://doi.org/10.1007/s10479-023-05306-5>
- Dalton, D. R., Daily, C. M., Johnson, J. L., & Ellstrand, A. E. (1999). Number of directors and financial performance: A meta-analysis. *Academy of Management Journal*, 42(6), 674–686. <https://doi.org/10.5465/256988>
- da Silva Mattos, E., & Shasha, D. (2024). Bankruptcy prediction with low-quality financial information. *Expert Systems with Applications*, 237, 121418. <https://doi.org/10.1016/j.eswa.2023.121418>
- Dietsch, M., & Petey, J. (2004). Should SME exposures be treated as retail or corporate exposures? A comparative analysis of default probabilities and asset correlations in French and German SMEs. *Journal of Banking & Finance*, 28(4), 773–788. <https://doi.org/10.1016/j.jbankfin.2003.10.006>
- Dowell, G. W. S., Shackell, M. B., & Stuart, N. V. (2011). Boards, CEOs, and surviving a financial crisis: Evidence from the internet shakeout. *Strategic Management Journal*, 32(10), 1025–1045. <https://doi.org/10.1002/smj.923>
- Duru, A., Iyengar, R. J., & Zampelli, E. M. (2016). The dynamic relationship between CEO duality and firm performance: The moderating role of board independence. *Journal of Business Research*, 69(10), 4269–4277. <https://doi.org/10.1016/j.jbusres.2016.04.001>
- Dwyer, P. D., Gilkeson, J. H., & List, J. A. (2002). Gender differences in revealed risk taking: Evidence from mutual fund investors. *Economics Letters*, 76(2), 151–158. [https://doi.org/10.1016/S0165-1765\(02\)00045-9](https://doi.org/10.1016/S0165-1765(02)00045-9)
- Edmister, R. O. (1972). An empirical test of financial ratio analysis for small business failure prediction. *The Journal of Financial and Quantitative Analysis*, 7(2), 1477–1493. <https://doi.org/10.2307/2329929>

- Engelmann, B., Hayden, E., & Tasche, D. (2003). Measuring the discriminative power of rating systems. *Risk*, 16, 82–86.
- European Commission. (2020). *An SME strategy for a sustainable and digital Europe* (Technical Report).
- Fernandez, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research*, 61, 863–905. <https://doi.org/10.1613/jair.1.11192>
- Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119–139. <https://doi.org/10.1006/jcss.1997.1504>
- Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 22. <https://doi.org/10.18637/jss.v033.i01>
- Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., & Walther, A. (2022). Predictably unequal? The effects of machine learning on credit markets. *The Journal of Finance*, 77(1), 5–47. <https://doi.org/10.1111/jofi.13090>
- G20 & OECD. (2015). *G20/OECD high-level principles on SME financing* (Technical Report).
- García, C. J., & Herrero, B. (2021). Female directors, capital structure, and financial distress. *Journal of Business Research*, 136, 592–601. <https://doi.org/10.1016/j.jbusres.2021.07.061>
- García, J. (2022). Bankruptcy prediction using synthetic sampling. *Machine Learning with Applications*, 9, 100343. <https://doi.org/10.1016/j.mlwa.2022.100343>
- Grunert, J., Norden, L., & Weber, M. (2005). The role of non-financial factors in internal credit ratings. *Journal of Banking & Finance*, 29(2), 509–531. <https://doi.org/10.1016/j.jbankfin.2004.05.017>
- Guest, P. M. (2009). The impact of board size on firm performance: Evidence from the UK. *The European Journal of Finance*, 15(4), 385–404. <https://doi.org/10.1080/13518470802466121>
- Habib, A., Costa, M. D., Huang, H. J., Bhuiyan, M. B. U., & Sun, L. (2020). Determinants and consequences of financial distress: Review of the empirical literature. *Accounting & Finance*, 60(S1), 1023–1075. <https://doi.org/10.1111/acfi.12400>
- Hautsch, N., Schaumburg, J., & Schienle, M. (2015). Financial network systemic risk contributions. *Review of Finance*, 19(2), 685–738. <https://doi.org/10.1093/rof/rfu010>
- Hillegeist, S. A., Keating, E. K., Cram, D. P., & Lundstedt, K. G. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9(1), 5–34. <https://doi.org/10.1023/B:RAST.0000013627.90884.b7>
- Hodges, J., Jr. (1958). The significance probability of the Smirnov two-sample test. *Arkiv För Matematik*, 3(5), 469–486. <https://doi.org/10.1007/BF02589501>
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression*. John Wiley & Sons.
- Jensen, M. C. (1993). The modern industrial revolution, exit, and the failure of internal control systems. *The Journal of Finance*, 48(3), 831–880. <https://doi.org/10.1111/j.1540-6261.1993.tb04022.x>
- Jianakoplos, N. A., & Bernasek, A. (1998). Are women more risk averse? *Economic Inquiry*, 36(4), 620–630. <https://doi.org/10.1111/j.1465-7295.1998.tb01740.x>
- Krause, R., Semadeni, M., & Cannella, A. A. (2014). CEO duality: A review and research agenda. *Journal of Management*, 40(1), 256–286. <https://doi.org/10.1177/0149206313503013>
- Krause, R., Semadeni, M., & Withers, M. C. (2016). That special someone: When the board views its chair as a resource. *Strategic Management Journal*, 37(9), 1990–2002. <https://doi.org/10.1002/smj.2444>
- Kuiziniene, D., Krilavičius, T., Damaševičius, R., & Maskeliūnas, R. (2022). Systematic review of financial distress identification using artificial intelligence methods. *Applied Artificial Intelligence*, 36(1), 2138124. <https://doi.org/10.1080/08839514.2022.2138124>
- Laitinen, E. K. (1999). Predicting a corporate credit analyst's risk estimate by logistic and linear models. *International Review of Financial Analysis*, 8(2), 97–121. [https://doi.org/10.1016/S1057-5219\(99\)00012-5](https://doi.org/10.1016/S1057-5219(99)00012-5)
- Lee, J. D., Sun, D. L., Sun, Y., & Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. *The Annals of Statistics*, 44(3), 907–927. <https://doi.org/10.1214/15-AOS1371>
- Liang, D., Tsai, C. F., Lu, H. Y. R., & Chang, L. S. (2020). Combining corporate governance indicators with stacking ensembles for financial distress prediction. *Journal of Business Research*, 120, 137–146. <https://doi.org/10.1016/j.jbusres.2020.07.052>
- Liberti, J. M., & Petersen, M. A. (2019). Information: Hard and soft. *The Review of Corporate Finance Studies*, 8(1), 1–41. <https://doi.org/10.1093/rcfs/cfy009>
- Lilienfeld-Toal, U. V., & Ruenzi, S. (2014). CEO ownership, stock market performance, and managerial discretion. *The Journal of Finance*, 69(3), 1013–1050. <https://doi.org/10.1111/jofi.12139>
- Lohmann, C., Möllenhoff, S., & Ohliger, T. (2023). Nonlinear relationships in bankruptcy prediction and their effect on the profitability of bankruptcy prediction models. *Journal of Business Economics*, 93(9), 1661–1690. <https://doi.org/10.1007/s11573-022-01130-8>
- Lohmann, C., & Ohliger, T. (2018). Nonlinear relationships in a logistic model of default for a high-default installment portfolio. *Journal of Credit Risk*, 14(1), 45–68. <https://doi.org/10.21314/JCR.2017.232>
- Lohmann, C., & Ohliger, T. (2020). Bankruptcy prediction and the discriminatory power of annual reports: Empirical evidence from financially distressed German companies. *Journal of Business Economics*, 90(1), 137–172. <https://doi.org/10.1007/s11573-019-00938-1>
- Mai, F., Tian, S., Lee, C., & Ma, L. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research*, 274(2), 743–758. <https://doi.org/10.1016/j.ejor.2018.10.024>

- Manzaneeque, M., Priego, A. M., & Merino, E. (2016). Corporate governance effect on financial distress likelihood: Evidence from Spain. *Revista de Contabilidad*, 19(1), 111–121. <https://doi.org/10.1016/j.rcsar.2015.04.001>
- Masiak, C., Block, J. H., Moritz, A., Lang, F., & Kraemer-Eis, H. (2019). How do micro firms differ in their financing patterns from larger SMEs? *Venture Capital*, 21(4), 301–325. <https://doi.org/10.1080/13691066.2019.1569333>
- Matenda, F. R., Sibanda, M., Chikodza, E., & Gumbo, V. (2022). Bankruptcy prediction for private firms in developing economies: A scoping review and guidance for future research. *Management Review Quarterly*, 72(4), 927–966. <https://doi.org/10.1007/s11301-021-00216-x>
- McFadden, D. (1974). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), *Frontiers in econometrics* (Vol. 1, pp. 105–142). Academic Press.
- Meinshausen, N., & Bühlmann, P. (2010). Stability selection. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 72(4), 417–473. <https://doi.org/10.1111/j.1467-9868.2010.00740.x>
- Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance*, 29(2), 449–470. <https://doi.org/10.2307/2978814>
- Ogneva, M., Piotroski, J. D., & Zakolyukina, A. A. (2020). Accounting fundamentals and systematic risk: Corporate failure over the business cycle. *The Accounting Review*, 95(5), 321–350. <https://doi.org/10.2308/accr-52638>
- Organisation for Economic Co-operation and Development. (2022). *Financing SMEs and entrepreneurs 2022: An OECD scoreboard*.
- Paraschiv, F., Schmid, M., & Wahlstrøm, R. R. (2023). *Bankruptcy prediction of privately held SMEs using feature selection methods* (Working Paper). Zeppelin University, Norwegian University of Science and Technology, Swiss Finance Institute (SFI), University of St. Gallen, and University of St. Gallen. <https://doi.org/10.2139/ssrn.3911490>
- Pearce, J. A., & Zahra, S. A. (1992). Board composition from a strategic contingency perspective. *Journal of Management Studies*, 29(4), 411–438. <https://doi.org/10.1111/j.1467-6486.1992.tb00672.x>
- Platt, H., & Platt, M. (2012). Corporate board attributes and bankruptcy. *Journal of Business Research*, 65(8), 1139–1143. <https://doi.org/10.1016/j.jbusres.2011.08.003>
- Radovanovic, J., & Haas, C. (2023). The evaluation of bankruptcy prediction models based on socio-economic costs. *Expert Systems with Applications*, 227, 120275. <https://doi.org/10.1016/j.eswa.2023.120275>
- Rajan, U., Seru, A., & Vig, V. (2015). The failure of models that predict failure: Distance, incentives, and defaults. *Journal of Financial Economics*, 115(2), 237–260. <https://doi.org/10.1016/j.jfineco.2014.09.012>
- Roulston, M. S. (2007). Performance targets and the Brier score. *Meteorological Applications*, 14(2), 185–194. <https://doi.org/10.1002/met.21>
- Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*, 10(3), e0118432. <https://doi.org/10.1371/journal.pone.0118432>
- Séverin, E., & Veganzones, D. (2021). Can earnings management information improve bankruptcy prediction models? *Annals of Operations Research*, 306(1–2), 247–272. <https://doi.org/10.1007/s10479-021-04183-0>
- Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business*, 74(1), 101–124. <https://doi.org/10.1086/209665>
- Smith, R. F., & Winakor, A. H. (1930). *A test analysis of unsuccessful industrial companies* (Bureau of Business Research, Bulletin. 31). University of Illinois Press.
- Srroj, S., Kovač, D., Shapiro, J. N., & Filer, R. K. (2023). The impact of delay: Evidence from formal out-of-court restructuring. *Journal of Corporate Finance*, 78, 102319. <https://doi.org/10.1016/j.jcorpfin.2022.102319>
- Srroj, S., Lapinski, M., & Walde, J. (2021a). Impact evaluation of business development grants on SME performance. *Small Business Economics*, 57(3), 1285–1301. <https://doi.org/10.1007/s11187-020-00348-6>
- Srroj, S., Škrinjarić, B., & Radas, S. (2021b). Bidding against the odds? The impact evaluation of grants for young micro and small firms during the recession. *Small Business Economics*, 56(1), 83–103. <https://doi.org/10.1007/s11187-019-00200-6>
- Stein, J. C. (2002). Information production and capital allocation: Decentralized versus hierarchical firms. *The Journal of Finance*, 57(5), 1891–1921. <https://doi.org/10.1111/0022-1082.00483>
- Tang, X., Li, S., Tan, M., & Shi, W. (2020). Incorporating textual and management factors into financial distress prediction: A comparative study of machine learning methods. *Journal of Forecasting*, 39(5), 769–787. <https://doi.org/10.1002/for.2661>
- Taylor, J., & Tibshirani, R. (2018). Post-selection inference for -penalized likelihood models. *The Canadian Journal of Statistics = Revue Canadienne de Statistique*, 46(1), 41–61. <https://doi.org/10.1002/cjs.11313>
- Tian, S., Yu, Y., & Guo, H. (2015). Variable selection and corporate bankruptcy forecasts. *Journal of Banking & Finance*, 52, 89–100. <https://doi.org/10.1016/j.jbankfin.2014.12.003>
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>
- Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., & Altman, R. B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics*, 17(6), 520–525. <https://doi.org/10.1093/bioinformatics/17.6.520>
- van de Geer, S., Bühlmann, P., Ritov, Y., & Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, 42(3), 1166–1202. <https://doi.org/10.1214/14-AOS1221>
- Wilson, N., & Altanlar, A. (2014). Company failure prediction with limited information: Newly incorporated companies. *Journal of the Operational Research Society*, 65(2), 252–264. <https://doi.org/10.1057/jors.2013.31>

- Wilson, N., Summers, B., & Hope, R. (2000). Using payment behaviour data for credit risk modelling. *International Journal of the Economics of Business*, 7(3), 333–346. <https://doi.org/10.1080/13571510050197230>
- Withers, M. C., & Fitza, M. A. (2017). Do board chairs matter? The influence of board chairs on firm performance. *Strategic Management Journal*, 38(6), 1343–1355. <https://doi.org/10.1002/smj.2587>
- Yin, C., Jiang, C., Jain, H. K., & Wang, Z. (2020). Evaluating the credit risk of SMEs using legal judgments. *Decision Support Systems*, 136, 113364. <https://doi.org/10.1016/j.dss.2020.113364>
- Zhang, C. H., & Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 76(1), 217–242. <https://doi.org/10.1111/rssb.12026>
- Zhao, J., Ouenniche, J., & De Smedt, J. (2024). Survey, classification and critical analysis of the literature on corporate bankruptcy and financial distress prediction. *Machine Learning with Applications*, 15, 100527. <https://doi.org/10.1016/j.mlwa.2024.100527>
- Zmijewski, M. E. (1984). Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research*, 22, 59–82. <https://doi.org/10.2307/2490859>

Appendix A. The sample per industry

Table A1. The sample per industry.

Industry level 1	Industry level 2	Financial statements	Number of companies	Bankruptcy frequency		
				One-year horizon	Two-year horizon	Three-year horizon
Agriculture, forestry and fishing	Crop and animal production, hunting	4765	1084	0.36%	0.78%	0.61%
	Forestry and logging	2461	570	0.33%	0.61%	0.37%
	Fishing and aquaculture	12,989	3155	0.14%	0.32%	0.43%
Mining and quarrying	Mining of coal and lignite	14	3	0.00%	0.00%	0.00%
	Extraction of oil and natural gas	633	147	0.16%	0.16%	0.16%
	Mining of metal ores	75	15	0.00%	0.00%	0.00%
	Other mining and quarrying	2709	538	0.11%	0.33%	0.41%
	Mining support service activities	2769	617	0.14%	0.43%	0.54%
Manufacturing	Food products	9188	2007	0.52%	1.40%	1.09%
	Beverages	1103	285	0.54%	1.45%	1.18%
	Textiles	1353	283	0.00%	0.59%	0.30%
	Wearing apparel	809	197	0.37%	2.35%	1.73%
	Leather and leather products	98	22	0.00%	0.00%	0.00%
	Wood and wood products	5177	1079	0.31%	0.89%	0.83%
	Paper and paper products	340	67	0.29%	0.59%	0.29%
	Printing and reproduction	3156	669	0.29%	0.92%	0.92%
	Refined petroleum products	55	15	0.00%	1.82%	0.00%
	Chemicals, chemical products	925	197	0.32%	0.97%	0.97%
	Pharmaceuticals	201	51	0.50%	0.50%	0.00%
	Rubber and plastic products	1860	374	0.27%	0.75%	0.59%
	Other non-metal mineral products	2818	573	0.00%	0.75%	0.53%
	Basic metals	548	114	0.18%	1.28%	0.55%
	Fabricated metal prod.	8944	1885	0.48%	1.20%	0.91%
	Electronic and optical products	1372	293	0.29%	0.44%	0.36%
	Electrical equipment	1641	353	0.43%	0.98%	1.04%
	Machinery and equipment	4308	892	0.42%	0.79%	0.56%
	Motor vehicles etc.	758	143	0.40%	1.06%	0.92%
	Other transport equipment	2043	456	0.64%	1.52%	1.37%
	Furniture	2143	467	0.84%	1.49%	1.12%
	Other manufacturing	2500	558	0.16%	0.76%	0.68%
	Repair, installation of machinery	6870	1573	0.28%	0.89%	0.74%
Construction	Construction of buildings	96,658	23,772	0.46%	1.26%	1.05%
	Civil engineering	3956	878	0.46%	1.11%	1.21%
	Specialised construction activities	76,835	17,755	0.50%	1.45%	1.20%
Domestic trade, car repair shop	Motor vehicles, trade and repair	30,129	6786	0.42%	1.17%	1.05%
	Wholesale trade	70,096	15,636	0.37%	0.99%	0.89%
Transportation and storage	Retail trade	87,311	20,031	0.73%	1.80%	1.45%
	Land transport, pipeline transport	20,653	4685	0.49%	1.51%	1.35%
	Water transport	10,675	2447	0.10%	0.43%	0.36%
	Air transport	544	113	0.18%	0.37%	0.37%
	Support act. for transportation	10,401	2167	0.12%	0.50%	0.46%
Accommodation, food service, Information and communication	Postal and courier activities	1121	321	1.52%	4.10%	2.59%
	Accommodation	11,848	2534	0.28%	0.86%	0.74%
	Food and beverage service act.	26,109	6949	1.12%	2.95%	2.28%
	Publishing activities	8021	1701	0.20%	0.49%	0.45%
	Motion picture, TV, music prod.	4532	1190	0.22%	0.46%	0.44%
	Programming, broadcasting act.	295	78	0.00%	0.68%	0.34%

(continued)

Table A1. Continued.

Industry level 1	Industry level 2	Financial statements	Number of companies	Bankruptcy frequency		
				One-year horizon	Two-year horizon	Three-year horizon
Profess., scientific, tech. act.	Telecommunications	2405	575	0.17%	0.54%	0.71%
	Computer programming, consultancy	26,886	7062	0.22%	0.53%	0.49%
	Information service activities	2927	738	0.10%	0.48%	0.38%
	Legal and accounting activities	20,103	4350	0.05%	0.17%	0.20%
	Head offices, management consult.	37,651	9148	0.18%	0.29%	0.27%
	Architecture, engineering act.	43,187	10,557	0.20%	0.44%	0.42%
	Scientific research and development	3030	765	0.10%	0.40%	0.40%
	Advertising and market research	6291	1571	0.17%	0.87%	0.92%
	Other prof., scientific, techn. act.	12,268	3187	0.37%	0.82%	0.71%
	Veterinary activities	2216	494	0.09%	0.05%	0.05%
Administrative, support service	Rental and leasing activities	11,403	2667	0.30%	0.73%	0.66%
	Employment activities	6701	1718	0.70%	1.63%	1.49%
	Travel agency, tour operators	5355	1378	0.35%	0.88%	0.90%
	Security, investigation activities	1045	253	0.48%	1.91%	1.72%
	Buildings, landscape service act.	9082	2221	0.53%	1.21%	0.87%
Education	Business support activities	8924	2066	0.18%	0.52%	0.44%
	Education	10,236	2548	0.20%	0.51%	0.52%
	Human health activities	26,538	6115	0.08%	0.14%	0.15%
Human health, social work	Residential care activities	704	152	0.00%	0.43%	0.43%
	Social work without accommodation	9066	1951	0.08%	0.19%	0.17%
Arts, entertainment and recreation	Arts and entertainment activities	5023	1309	0.24%	0.48%	0.40%
	Libraries, museums, other culture	436	85	0.00%	0.00%	0.00%
	Gambling and betting activities	778	165	0.13%	0.64%	0.77%
	Sports, amusement, recreation	9288	2160	0.23%	0.54%	0.55%
Other service activities	Membership organisations	516	120	0.00%	0.00%	0.19%
	Repair, personal, household goods	1037	256	0.87%	1.83%	1.35%
	Other personal service activities	12,022	2782	0.23%	0.62%	0.57%
Total		818,927	192,118	0.39%	1.04%	0.88%

Notes: Our sample of SMEs' financial statements, the number of unique companies and bankruptcy frequency over different time horizons per industry at the one- and two-digit levels, in accordance with the Norwegian Standard Industrial Classification (SIC2007) based on the UN's ISIC Rev. 4 and EU's NACE Rev. 2.