Received: 17 November 2021 | Revised: 29 March 2023 | Accepted: 7 May 2023
DOI: 10.1002/ijfe.2840

RESEARCH ARTICLE

WILEY

Can market information outperform hard and soft information in predicting corporate defaults?

Stefano Filomeni[1] | Udichibarna Bose[1] | Anastasios Megaritis[2] | Athanasios Triantafyllou[3]

[1]University of Essex, Essex Business School, Finance Group, Colchester, UK
[2]Keele University, Keele Business School, Keele, UK
[3]IESEG School of Management, Univ. Lille, CNRS, UMR 9221 - LEM - Lille Economie Management, F-59000, Lille, France

Correspondence
Stefano Filomeni, University of Essex, Essex Business School, Finance Group, Colchester, UK.
Email: stefano.filomeni@essex.ac.uk

Abstract
Recent evidence has shown that hybrid models for credit ratings are important when assessing the risk of firms. Within this stream of literature, we aim to provide novel evidence on how hard (quantitative), soft (qualitative), and market information predict corporate defaults for unlisted firms by implementing the Cox proportional hazard model. We address this research question by exploiting a unique proprietary dataset comprising of detailed information on internal credit ratings of European unlisted mid-sized firms and compute their Merton's distance-to-default (DD) measure of credit risk with market data collected on comparable publicly listed companies. Our results show that the bank's use of hard, soft, and market information when assessing the credit ratings of borrowers has a significant influence on the prediction of their defaults. Further, we investigate the significant influence of soft information in predicting corporate defaults by drawing on two separate processes through which loan officers can inject soft information in credit scoring, that is, 'codified' and 'uncodified' discretion. Finally, when we distinguish between the loan officer's discretion to upgrade or downgrade an applicant's credit score, we find that it is the upgrade that is likely to predict a lower probability of a firm defaulting. This study contributes to the policy debate on safeguarding the banking sector's continuity by positing that integrating market information into banks' hybrid methods of credit rating helps to improve the accuracy in predicting unlisted firms' credit risk that is useful to policy makers for the design of future forward-looking financial risk management frameworks.

KEYWORDS
credit rating, corporate default, distance-to-default, hard information, Merton model, soft information

1 | INTRODUCTION

The type of information used by banks in predicting credit risks represents a source of concern for policymakers as inaccurate information can greatly threaten the stability and intermediation role of the banking sector. In this regard, Basel II has proposed a step forward that explicitly recognises the importance of going beyond traditional borrowers' quantitative hard information by allowing banks to incorporate qualitative soft information in their internal credit ratings (BCBS, 2001). The aim is to produce internal ratings that more accurately predict borrower-specific credit risks. In this regard, the literature on default prediction so far has identified three main approaches to predict corporate credit risks: accounting-based approaches that make use of purely quantitative accounting information, hybrid methods of credit rating which combine quantitative hard and qualitative soft information, and market information-based approaches that are based on default predictive models, such as the Merton's (1974) distance-to-default (DD).

Accounting-based approaches attempt to forecast defaults by using accounting information based on the firm's balance-sheet data (Altman, 1968; Altman & Sabato, 2008; Beaver, 1966; Louzada et al., 2016; Ohlson, 1980). Hybrid models attempt to predict defaults by combining hard and 'hardened' soft information about borrowers in their credit ratings (Brown et al., 2012, 2015; Casu et al., 2022; Filomeni et al., 2021; Liberti & Mian, 2009; Liberti & Petersen, 2019; Roy & Shaw, 2021; Xu et al., 2018). The DD, calculated on the basis of observable equity market data, is widely used for predicting the insolvency risk of listed companies, especially in the US equity market (Bharath & Shumway, 2008; Byström, 2006; Vassalou & Xing, 2004). Studies have also extensively shown the superiority of the Merton model for predicting defaults of US listed firms (Agarwal & Taffler, 2008; Altman, 1968; Bauer & Agarwal, 2014; Das et al., 2009; Doumpos et al., 2015; Gadenne & Iselin, 2000; Hillegeist et al., 2004). Moreover, some recent empirical findings in the literature show that the hybrid models integrating market and accounting information can significantly improve the predictability of default risks of SMEs and unlisted firms (Andrikopolous & Khorasgani, 2018; Roy & Shaw, 2021). However, none of these studies effectively combine firm-specific accounting (quantitative) and soft (qualitative) information with public-peer market information.[1] Therefore, in our best knowledge, this study is the first to integrate market information simultaneously with hard and soft information in predicting default probabilities of European unlisted firms.

We address this research question by exploiting a unique proprietary dataset comprising of detailed loan-level information on 437 European unlisted firms belonging to the mid corporate segment of the market. This comprehensive dataset allows us to collect information not only on borrowers' internal credit ratings but also on the different components of intermediate ratings. We implement semiparametric Cox regression models[2] which are widely used in finance for survival analysis (Henebry, 1997; Lane et al., 1986; Buehler et al., 2006) to investigate the association between hard, soft, and market information of a given borrowing firm, and its probability of survival after the bank disburses a loan. Next, we further examine the influence of soft information in predicting corporate defaults by distinguishing between two routes loan officers can use to 'harden' soft information in the rating process[3]: questionnaire and override. Questionnaire reflects that type of soft information which is mandatory for the loan officer when assessing borrowers' credit risk and is based on the bank's standardised and codified numerical scale in the qualitative questionnaire which is mandatory as part of the rating process; as such, the answers provided by loan officers in the obligatory qualitative questionnaire are not subject to authentication by the bank's senior managers. We label this type of hardening as 'codified discretion' (Filomeni et al., 2021). An override gives an opportunity to the loan officer to either upgrade or downgrade the final credit score attributed to a given borrowing firm and is closely monitored by the bank's headquarters. We mention this type of hardening as 'uncodified discretion' (Filomeni et al., 2021).[4] Thus, we exploit the information from these two types of discretion to assess their influence on the prediction of firm defaults when they are integrated into the final ratings. Further, we also differentiate between the codified and uncodified discretion to upgrade or downgrade credit ratings to further study the predictive power of 'hardened' soft information on firms' default probabilities. Ratings' deviations can involve one or several rating notches resulting in upgrades or downgrades in either the codified and/or uncodified use of discretion. The reasons for downgrades stem from commercial hazards due to decline in the firm's economic conditions, non-compliance to marketing strategies, or to changes in the regulatory framework that affect firm value. The motivations for upgrades can be related to the borrower-specific credit risk such as entering markets with greater opportunities for socioeconomic development and expanding demand, participating with outstanding commercial partners, or restructuring projects aimed at cost reduction.

This study significantly contributes to the finance theory literature on financial intermediation and traditional corporate finance. The theoretical literature on financial intermediation has been studying the origin and differences of soft versus hard information for a few decades. This literature draws the distinction between these different types of information from the 'modus operandi' of banks that aim to exploit an informational advantage by collecting and processing private soft information in their firm-bank relationships to reach more accurate credit assessments of their borrowing firms (Liberti & Petersen, 2019), differently from the public bond market in which investors rely primarily on publicly disclosed information in their investment decisions. Indeed, a key difference highlighted in the financial intermediation literature is banks' superior capability in gathering and processing information (Allen et al., 2015; Diamond, 1984, 1991; Ramakrishnan & Thakor, 1984), while the public debt markets mostly rely on credit ratings provided by specialised rating agencies to evaluate corporate creditworthiness (Ederington & Goh, 1998).

Building on this theoretical literature, the current empirical studies show that firms' access to external sources of financing depends on the amount of hard and quantitative information financial markets possess about these companies. On the one hand, empirical evidence shows that the amount of hard information available for a publicly traded firm is easy to access, and significantly affects the probability of corporate default which can further exert a significant effect on firms' access to debt capital (Faulkender & Petersen, 2006). On the other hand, it demonstrates that private and unlisted firms are the most informationally opaque firms due to limited availability of hard information as they have no track record and often operate in emergent industries (Liberti & Petersen, 2019). Therefore, we contribute to the current theoretical and empirical literature on financial intermediation by exploiting a unique proprietary dataset comprising of European unlisted mid-sized firms to study whether the inclusion of market information in banks' hybrid credit scoring models that make use of both hard and soft information can significantly improve the predictability of corporate defaults for unlisted firms.

Further, we also contribute to the theoretical literature on corporate finance that highlights the role of asymmetric information in accessing external debt financing (Demetriades & Devereux, 2000; Fazzari et al., 1988; Myers & Majluf, 1984). In this respect, existing studies provide evidence that greater information asymmetry generate barriers, especially for smaller and more informationally opaque unlisted firms, to tap into external sources of financing to fund investments (Filomeni et al., 2023). To mitigate the problem of asymmetric information, banks gather valuable qualitative soft information about the economic prospects of corporations through a monitoring activity that involves 'human touch' between the parties that contributes to reduce borrowers' moral hazard (Diamond, 1984; Qian et al., 2015). However, we argue in the paper that, in addition to hard and soft information, integrating market information in unlisted firms' credit scoring process can help to accurately measure and capture changes in the level of default risk in a timely manner as market data is likely to react more rapidly to newly disclosed information about borrowers' creditworthiness than accounting measures which react slowly, consistent with Tinoco and Wilson (2013) in the context of listed firms. Moreover, existing studies (such as Das et al., 2009; Doumpos et al., 2015) are focused on integrating, rather than comparing, accounting with market information in forecasting corporate defaults. Hence, this study contributes to the literature by arguing that, in addition to hard and soft information, market information collected from publicly traded peers can be useful in predicting the defaults of their comparable unlisted firms. This is also the first study that attempts to use market information to assess the credit risks of unlisted firms through a survival analysis, since the literature so far has used the Merton (1974) model for measuring credit risks of private listed firms (Andrikopolous & Khorasgani, 2018; Rikkers & Thibeault, 2009). We calculate the DD for unlisted firms with market data collected from their comparable publicly listed firms to proxy unlisted firms' asset price volatility. Specifically, our mapping approach relies on peer analysis provided by Refinitiv's Starmine Peers as peer groups are known for their influential nature as they are able to shape the decisions of members of the group. Specifically, the choice of the listed firms which are more closely associated and mapped to our sample of unlisted firms is generated using a proprietary algorithm that combines competitor lists provided in filings, analyst cross coverage, business classification, and revenue proximity. According to Refinitiv, this hierarchical approach produces very reasonable sets of peer companies for most securities, and thus, we construct eight portfolios of equally weighted daily returns for the market value of equity of our peers that represent the eight sectors of the equity market in which our sample of unlisted firms operate based on their domestic industry classification.

Our empirical results indeed show that the addition of market-based information to soft and hard information leads to a significant improvement in the prediction of unlisted firms' corporate defaults. This ability means that banks should recognise the value of market-based information derived from publicly traded industry peers when predicting credit risks. Next, we find that hardened soft information that materialise in codified and uncodified discretions has a significant effect on predicting firm defaults when they are integrated into the final ratings. Further, when we distinguish between the discretion to upgrade or downgrade a credit rating, we find that upgrades by loan officers can reduce the predictive probability of firm default. We also provide various robustness tests to further corroborate the validity of our main empirical findings.

The evidence provided in this study can be of interest to policymakers given the relevance of credit risk evaluations for the banking sector's stability.[5] One potential policy implication of our findings is to motivate banks to integrate market information into their hybrid credit rating methods to increase the precision of predicting unlisted firms' corporate defaults. This increased accuracy would therefore spur banks to produce enhanced borrowers' internal hybrid credit ratings that integrate hard and soft information with market information in the context of unlisted firms' credit risk assessment. Thus, the results of this study can be supportive to develop forward-looking frameworks on financial risk management (Breden, 2008; Rodriguez Gonzalez et al., 2018).

The rest of this paper is structured as follows. Section 2 discusses the related literature. Section 3 describes the bank's credit scoring process along with our banking and market variables. Section 4 describes our empirical method, while Section 5 presents our empirical findings. Section 6 provides several robustness checks. Section 7, finally, concludes the paper.

2 | RELATED LITERATURE

The need for reliable models for timely and accurately predicting business failures is important for banks to reach appropriate lending decisions. Current literature on modelling corporate credit risk has examined accounting- and structural market-based models. Thomas et al. (2002) defines accounting-based approaches, in the form of credit rating models, as a set of decision models and underlying techniques which support lenders in credit concession. Among the methods to assess credit quality, a distinction exists between purely quantitative and hybrid models.

A purely quantitative credit rating is based on hard information. This method does not account for qualitative soft information in the final ratings. Such methods include statistical techniques that quantify the financial information about a borrower as a numerical rating that reflects the credit quality and predictive power on the likelihood of borrower's default (Altman, 1968; Beaver, 1966; Louzada et al., 2016; Ohlson, 1980). In this context, several studies examine the effectiveness of credit ratings in predicting future defaults. Altman and Sabato (2008) study the most appropriate factors in predicting the quality of firm's future credit and build a model that predicts default. In a similar spirit, Altman et al. (2010) include non-financial information, such as creditors' legal action and audit reports, to predict the quality of unlisted firms' credit. Lin et al. (2012) test numerous accounting-based risk models to forecast UK-based small firms' defaults.

In some ways, the implementation by banks of the internal ratings-based approach (IRB) in the Basel II guidelines is likely to support this idea of accounting for not only quantitative, but also qualitative information when predicting borrowers' default probability (BCBS, 2001). In compliance with Basel II regulations, prudential regulators provide banks with the possibility to adopt IRB that allows them to develop their own model for credit rating to evaluate the credit quality of corporate borrowers. This provision has resulted in the widespread use of hybrid models of credit rating by banks that make use of both quantitative and qualitative components of borrowers' information. That is, such methods use qualitative soft information in adjunct to traditional quantitative hard information that is mostly collected from a borrower's financial statements or business plans. Specifically, hybrid rating models allow for the hardening of qualitative soft information in an attempt to combine hard and soft information about borrowers into numerical ratings that reflect their creditworthiness and repayment prospects that ultimately influence banks' lending decisions (Bertomeu & Marinovic, 2016; Brown et al., 2015, 2012; Casu et al., 2022; Filomeni et al., 2021; Gropp & Guettler, 2018; Liberti & Petersen, 2019; Roy & Shaw, 2021). Supervisory agencies also support this process of hardening soft information (BCBS, 2000; BCBS, 2005; Federal Reserve, 2011; OeNB, 2004).

Soft information refers to subjective impressions collected through repeated bank-firm personal interactions that, by assumption, are not quantifiable or verifiable in practice. To the contrary, hard information refers to the type of information that is easily observable and communicable inside the bank without dilution of its informative content. Thus, hybrid rating models allow banks to exploit all their private and soft information and harden it into a final rating that reflects a given borrower's creditworthiness. In this method, each rating is not automatically defined by hard information, but it must simultaneously account for both quantitative and qualitative information about the borrower. Soft information represents the bank's informational advantage that allows it to reach a more accurate credit assessment and, consequently, to determine the credit risk of firms (Acheampong & Elshandidy, 2021) and to improve financial efficiency (Edmans et al., 2016). This combined information is thus, injected into the credit rating model to produce internal credit ratings that reflects borrowers' creditworthiness. In this regard, the evidence shows that more opaque borrowers, such as SMEs, are dependent on local banks for soft information-intensive relationship lending (Berger & Udell, 1995; Cole, 1998; Gaudio et al., 2020; Petersen & Rajan, 1994; Xia et al., 2020).

Internal credit ratings, in turn, are likely to affect banks' lending decisions. While hybrid rating models could be virtually applied to all corporate segments, purely-quantitative models are mainly applied to small business lending (Asch, 1995; Eisenbeis, 1996; Mester, 1997). According to Feldman (1997), a purely quantitative credit rating alters lending to small firms in terms of lender-borrower collaboration, loan pricing, and credit availability. First, the credit rating allows lenders to grant credit and monitor disbursed loans without meeting borrowers (Hannan, 2003; Petersen & Rajan, 2002). This development is in contrast to the perceived relevance of a local bank-borrower relationship for small business lending (Berger & Udell, 1995; Petersen & Rajan, 1994). Lenders may therefore, enter markets without a pre-existing geographical presence. Second, the loan pricing for small firms has ambiguous theoretical predictions that depend on how a lender uses this technology (Berger et al., 2005). Third, the credit rating is likely to increase the availability of funds to small firms (Frame et al., 2001).

While there are several studies that investigate how a purely quantitative credit rating affects a bank's lending decisions, the evidence regarding hybrid credit ratings is still scarce. Volk (2012) estimates borrowers' probabilities of default using various models of default predictions and compares them to the classification of banks' credit ratings. They provide evidence that the probabilities of default estimated by purely quantitative information and banks' internal ratings give different measures of borrowers' credit quality which could be due to banks' dependence on additional qualitative information in assessing borrowers' creditworthiness. Qian et al. (2015) use loan-level data to show that delegation of authority improves the predictive power of internal ratings on Chinese borrowers' credit risks. Kraft (2015) also studies both quantitative adjustments to firms' financials reported in US GAAP financial statement and qualitative deviations to firms' credit ratings and the results indicate that both quantitative and qualitative adjustments to credit ratings allow to better capture credit risk consistent with effective collection and use of both hard and soft information.

In addition, several studies have also applied structural market-based models like Merton's model to predict firm defaults. Specifically, studies have extensively used Merton's (1974) DD measure based on observable equity market data to predict US listed firms' default risk (Bharath & Shumway, 2008; Byström, 2006; Vassalou & Xing, 2004). Further, some studies have also compared the corporate default predictive power of accounting and market-based models and have shown the incremental predictive power of Merton's (1974) model on corporate defaults as compared to alternative credit risk models which are based on accounting ratios, such as the Altman's (1968) Z-Score (Das et al., 2009; Doumpos et al., 2015). Doumpos et al. (2015) also find that prediction of default risk for a sample of European listed firms using the DD measure in addition to accounting-based financial ratios improves the predictive power. Chen and Wang (2023) construct a default model for micro- and small-sized firms that takes into account cash flow, default boundary, and actual cash flow's distribution in an incomplete information model. This model is suitable to dynamically forecast individual applicant and post-lending risks. However, these studies do not compare banks' internal ratings from the hybrid rating model with the market-based default models in predicting the defaults of unlisted firms.

3 | DATA

3.1 | The bank's environment and credit scoring

We address our research question by employing a unique and proprietary dataset collected from the Corporate and Investment Banking (CIB) Division of a major European banking player in the period September 2011 and 2012.[6] Such dataset comprises 437 loan applications, approved or denied. The multinational European banking group (which our data-providing bank belongs to) has total assets of about EUR 646 bn, market capitalization of about EUR 50 bn, and foreign subsidiaries operating in 12 European and Mediterranean countries. This European banking group owns 14 affiliated banks with more than 4000 branches and a market share of roughly 15% in the loans and deposits market in its home country of operations. Specifically, our data-providing bank is the lead of the aforementioned banking group and conducts its banking activity in its home country with around 1900 traditional bank branches spread out in 16 regions over the home country's territory. The CIB Division of the lead bank comprises 24 corporate branches disseminated in the 12 regions across the home country. The bank specifically designed the aforementioned corporate branches to reflect a new operational structure created ex novo with the objective to keep those credit relationship with mid- and large-sized companies separated from traditional retail banking related to the banking needs of families and small firms. Loan officers operating at corporate branches are responsible for initiating and managing the credit relationships with the applicant firms, for producing credit ratings to be submitted alongside the loan proposal to the appropriate hierarchical layer of the bank for approval. The mid-corporate segment includes enterprises that generate a yearly turnover ranging from EUR 150 million to EUR 1 billion and is generally less affected by issues of information opacity than SMEs. Due to this, mid-corporate lending should be less affected by frictions in transmitting information across the hierarchy of the bank. Micro-level information collected from each mid-corporate credit folder is very granular at the loan level and includes, among other information, detailed proprietary data on the rating process characterising a specific borrower, such as the final rating attributed and all intermediate rating components that are part of the rating process itself. In addition, each credit application discloses accounting information for the applicant firms.

Loan officers are compensated by a fixed salary plus an extra bonus awarded if the loan officer exceeds their allocated personal lending volume target assigned every year by the bank. Furthermore, the risk management department at the bank's headquarters closely monitors not only loan officers' activities and submitted loan applications, but also carefully examines advanced requests to override applicant firms' internal ratings. Such monitoring and assessment therefore takes place at the headquarters of the bank and ultimately determines the loan officers' career advancements. In light of their compensation scheme, loan officers may be incentivised to inflate internal credit ratings throughout the credit scoring process; in other words, they may find it convenient to 'game' the system with the objective to generate a greater volume of approved loans. However, loan officers also acknowledge that the success of approval and the post-disbursement performance of those loans whose internal ratings have been inflated or overridden are equally relevant for their career prospects. In this respect, if loan officers excessively inflate internal ratings, it can negatively affect their ability to get those 'inflated' loans approved which can expose them to severe reputational losses within the bank. As presented in detail in Sub-section 3.2, in the context of our paper, loan officers have the opportunity to inject qualitative soft information in two distinct discretionary moments during the applicant firm's credit scoring process. Indeed, injecting such soft information can materialise in both the mandatory qualitative questionnaire (i.e., codified discretion) and in the proposal to override the final credit rating produced by the credit scoring tool (i.e., uncodified discretion).

3.2 | Banking data

We use a banking proprietary dataset comprising of detailed information on the bank's internal ratings of mid-corporate borrowing firms as well as their different intermediate rating components. The bank implements a hybrid credit scoring methodology to rate borrowers' creditworthiness in which the final credit ratings are calculated on the basis of both quantitative and qualitative information. We now briefly present the hybrid credit scoring process adopted by our bank. To the purpose of this paper, this is relevant for two reasons. First, the bank's internal ratings reflect our measure of the bank's accuracy in predicting corporate defaults. Second, understanding the way internal ratings are generated is relevant to derive the underlying mechanism turning purely hard information-based statistical ratings into hybrid information-based final ratings simultaneously accounting for both hard and soft information.

The rating process starts with the loan officer who is the entry point of a given bank-borrower relationship. During this process, the loan officer performs due diligence by collecting borrowers' verifiable hard information from financial statements, pro-forma quarterly financials, and forward-looking business plans (if available). Moreover, interviews and on-site corporate visits give the opportunity to the loan officer to also gather borrowers' unverifiable soft information. Once the borrower's rating process is started by the loan officer, a 'statistical rating' is automatically generated by the rating tool adopted by the bank. Statistical ratings reflect the firm's likelihood of default purely based on its quantitative and hard information-based financials. Moreover, it does not consider market information for the assessment of unlisted firms. The loan officer's hardening of qualitative soft information can take place in the next two steps of the credit rating process.

Indeed, after the bank's model has automatically produced a statistical rating, the loan officer is called on to fill a mandatory questionnaire which gives them the opportunity through multiple choice questions to integrate quantitative information with their subjective judgement about several borrower's characteristics and market characteristics predefined by the bank.[7] The outcome from this process is the generation of an 'integrated rating' which can be equal to or deviate from the statistical rating previously produced. Any adjustments to statistical ratings, giving rise to an integrated rating that deviates from the latter, highlights the loan officer's execution of discretion. We label this type of loan officer's discretion as 'codified discretion' since the soft information injected at this stage of the credit scoring exercise is processed by the bank through a standard and codified numerical scale. Codified discretion is obligatory for loan officers, does not require any qualitative justification on the part of the loan officer, and is not subject to any validation process by the bank's headquarters.

In final step of the rating process, loan officers are given the opportunity to override integrated ratings, that is, to propose final ratings that deviate from integrated ratings. When submitting override proposals, loan officers have to provide a written explanation to senior bank

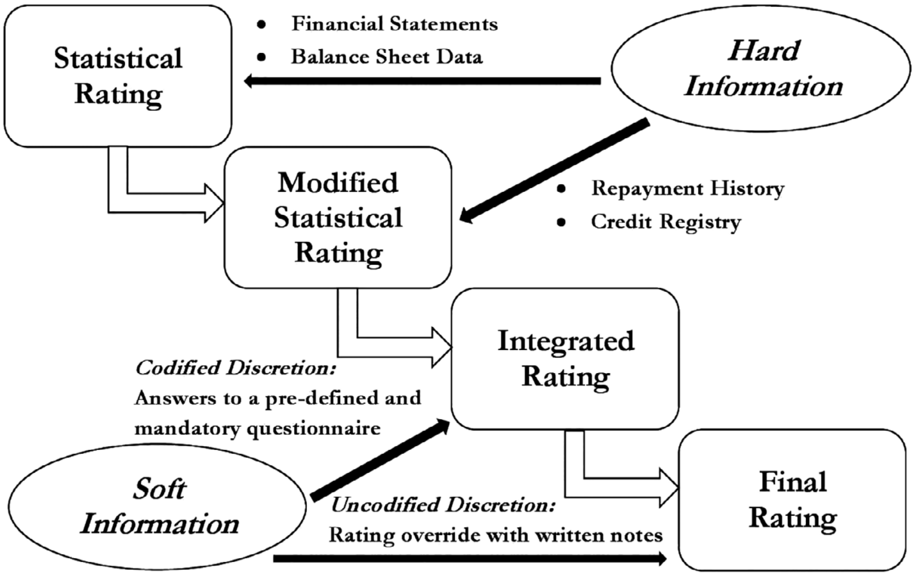

FIGURE 1 The bank's credit rating process. *Source*: Credit rating policy adopted by our data provider.

managers located at the bank's headquarters (i.e., the Rating Unit). Whatever the reasons, rating overrides cannot be quantified into a well-specified objective metric or codified into specific categorical statements. Instead, override proposals can only be transmitted within the banking organisation by providing detailed explanatory notes subject to inspections by the loan reviewers working at the bank's 'Rating Unit' with potential reputational consequences for the loan officer's career prospects within the banking organisation. We label this type of loan officer's discretion as 'uncodified discretion'. Relevant to the context of this study, separately analysing uncodified and codified discretion facilitates the investigation of the effects of different ways to harden soft information on the prediction of firm's default probability. Further, rating deviations in both codified and uncodified discretion can lead to either upgrading or downgrading overrides of integrated ratings. A graphical representation of the hybrid credit scoring process adopted by the bank is provided in Figure 1. The credit scoring process for a given borrowing firm terminates with the attribution of a 'final rating' that is composed of a maximum of 15 rating notches (subdivided into three macro-rating classes) represents the final output of this hybrid credit rating process, where the 15th notch reflects the most creditworthy rating notch and is comparable to the S&P's rating of AAA.

3.3 | Market data

The estimation of the DD encompasses several steps. First, we obtain the daily series for equity market prices and number of outstanding shares for the listed firms which are more closely associated with the sample of our unlisted firms. As previously mentioned, our mapping approach relies on peer analysis provided by Refinitiv's Starmine Peers based on a proprietary algorithm that provides accurate selection of comparable peers.[8] We therefore construct eight portfolios of equally weighted daily returns for the market value of equity of our peers that represent the eight sectors of the equity market in which our sample of unlisted firms operate based on their domestic industry classification. The detailed descriptions of the sectors of our peer group of publicly listed firms as well as the selected firms in that group are reported in the Appendix (Table A2). Second, we compute each individual firm's daily market value of equity as the number of outstanding shares multiplied by the share price and then computing its average market value of equity for our sample period. Third, we calculate the given portfolio's market value of equity by averaging the values for all firms that belong to the same portfolio (sector). Fourth, we estimate the equity volatility (σ_E) as the standard deviation of the daily percentage change (return) in the given portfolio's market value of equity for the year 2011. We then multiply the daily standard deviation by the square root of 252 to annualize the deviation.[9] Through this process, we can match the equity market volatility of our sample of unlisted mid-corporate firms with the volatility of equity values of our eight portfolios with the listed peers operating in the same industry. Definitions of the variables and their descriptive statistics are provided in Table 1.

Through this process, we estimate the DD that is unique for each unlisted firm, since the accounting-based

TABLE 1 Descriptive statistics.

Variable	Description	Obs	Mean	Median	SD	Min	Max
Dependent variables							
Default	Dummy variable equals 1 if the borrower defaults in the subsequent 2 years (Cox hazard model)	437	0.19	0	0.39	0	1
Financial Status_Ordered Probit	Categorical variable ranging from 0 to 7 that captures the intensity of the borrower's financial distress (ordered Probit model)	365	0.46	0	1.21	0	7
Credit risk measures							
Merton's DD	Merton's distance-to-default [logarithmic scale]	381	16.78 [2.43]	5.19 [1.96]	20.26 [0.95]	0 [0]	65.77 [4.21]
Alternative Merton's DD	Alternative Merton's distance-to-default [logarithmic scale]	351	12.85 [2.06]	2.80 [1.49]	16.24 [1.06]	0 [0]	60.11 [4.12]
Final rating	Final rating of a given borrowing firm	413	8.32	8	3.33	1.00	15.00
Integrated rating	Integrated rating of a given borrowing firm	408	7.73	8	3.35	1.00	15.00
Statistical rating	Statistical rating of a given borrowing firm	407	7.52	8	3.53	1.00	15.00
Soft information measures							
Soft information	Dummy variable equal to 1 if there is a difference between the final and the statistical ratings and 0 otherwise	407	0.46	0	0.50	0	1
Override	Dummy variable equal to 1 if there is a difference between the final and the integrated ratings and 0 otherwise	408	0.20	0	0.40	0	1
Questionnaire	Dummy variable equal to 1e if there is a difference between the integrated and the statistical ratings and 0 otherwise	407	0.37	0	0.48	0	1
Upgrade override	Dummy variable equal to 1 if there is an upgrading override decision and 0 otherwise	407	0.14	0	0.35	0	1
Downgrade override	Dummy variable equal to 1 if there is a downgrading override decision and 0 otherwise	407	0.06	0	0.23	0	1
Upgrade questionnaire	Dummy variable equal to 1 if there is an upgrading questionnaire decision and 0 otherwise	407	0.16	0	0.36	0	1
Downgrade questionnaire	Dummy variable equal to 1 if there is a downgrading questionnaire decision and 0 otherwise	407	0.21	0	0.41	0	1
Sentiment score	The sentimental tone of each loan application, where each positive word counts as +1 and each negative word as −1, divided by the number of words contained in the text section (after stopwords exclusion)	413	3.50	3.74	2.04	−6.43	9.06

TABLE 1 (Continued)

Variable	Description	Obs	Mean	Median	*SD*	Min	Max
Borrower's characteristics							
Borrower's size	Total assets of a given borrowing firm (expressed in logarithm in regressions)	400	190,000,000	84,000,000	313,000,000	224,000	2,380,000,000
Borrower's EBITDA	Ratio of a borrower's EBITDA over total assets	388	0.06	0.06	0.11	−1.25	0.60
Borrower's long-term debt	Ratio of a borrower's long-term debt over total assets	352	0.15	0.11	0.15	0.00	1.23
Borrower's liquidity	Ratio of a borrower's liquidity over total assets	400	0.07	0.04	0.10	0	0.93

Note: The table provides the descriptive statistics for all variables used in the empirical models.

assets and liability values used to compute this measure are individual at the firm-level, even if the equity volatility (σ_E) is calculated at the peers' portfolio-level for each given industry.

4 | ESTIMATED MODELS

4.1 | Bank's and market's predictive power for corporate defaults

To test the predictive power of bank's internal ratings and Merton's DD for firm defaults, we perform several specifications of the Cox proportional hazard model. Cox (1972) proposed a hazard regression model that many studies have widely used in medical research and in the analysis of survival data in finance (Buehler et al., 2006; Henebry, 1997; Lane et al., 1986). In this paper, we implement the Cox proportional hazard model to examine the association between the survival time of borrowing firms after the loan is disbursed and their relative hard, soft, and market information.[10] In this context, survival time refers to the number of days the borrower survives in the marketplace after it receives a loan from its lending bank. In this respect, the Cox proportional hazard model facilitates the simultaneous evaluation of the influence of several factors on firms' survival. The dependent variable in our survival-time analysis is the variable 'default' that is a binary variable equal to one if the given borrower defaults (at the end of year 2012 or 2013) and zero otherwise. We add the previous year (2011) values of all other covariates to our regressions except for the internal ratings that are revised during the annual renewal of the credit line that is performed by the bank. In this regard, a hazard model is more appropriate as it allows us to wipe out concerns about the vintage effect of firm's internal ratings.[11] Figures 2–4 show the Kaplan–Meier plots for the survivals of our sample of unlisted firms over time (Figure 2) that are based on final ratings (Figure 3) and Merton's DD (Figure 4). These plots show a series of declining horizontal steps which approaches the survival function of our sample.

We estimate the reduced-form specifications of a Cox hazard model for the bank's internal credit ratings and the DD credit risk measure that takes the following form:

$$h(\text{default}_{i,t}) = h_0(t)x\exp\Big(\beta_0+\beta_2\text{Rating}_{i,t-1} + \beta_3\text{DD}_{i,t-1}+\beta_4\text{Size}_{i,t-1}+\beta_5\text{LTD}_{i,t-1} + \beta_6\text{EBITDA}_{i,t-1}+\beta_7\text{Liquidity}_{i,t-1} + \beta_8\text{GDPgrowth}+\delta D_{\text{industry}}+\varepsilon_{i,t}\Big) \quad (1)$$

Where the survival time t refers to the number of days the borrower i survives in the marketplace after the bank disburses the loan. 'Rating' denotes three different types of ratings used by banks in credit scoring such as statistical, integrated, and final ratings. These three rating components differ in terms of the informational content embedded into each one of them and allow to test the incremental value of each soft information component separately in predicting corporate defaults. More specifically, 'statistical ratings' represent the starting point of the rating procedure which are purely based on hard information and are automatically generated at the time when the loan officer starts the rating process for a given applicant firm. 'Integrated ratings' represent the second rating produced in credit scoring process and can be equal or different from previously generated statistical ratings based on a mandatory qualitative questionnaire

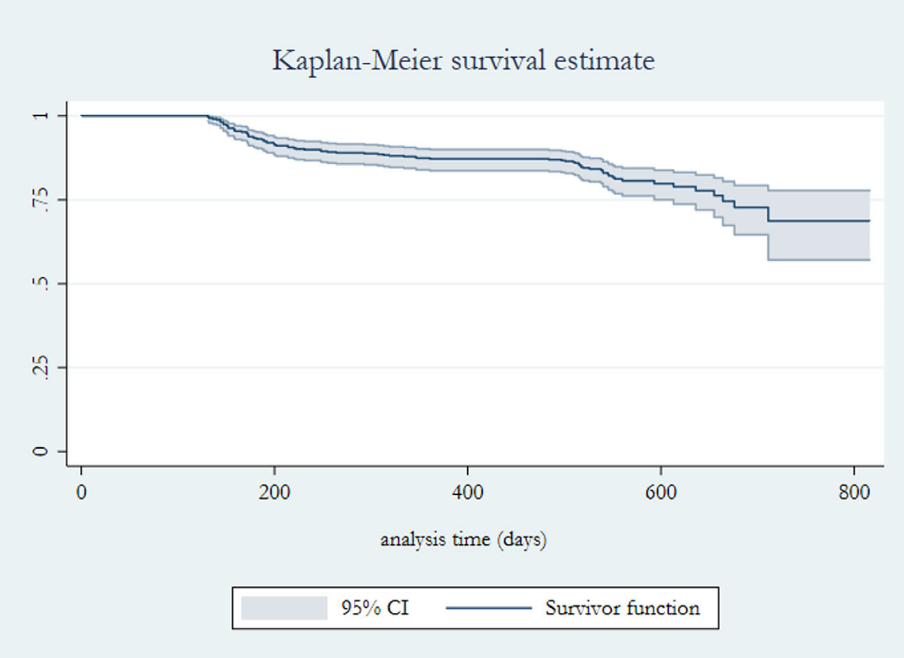

FIGURE 2 Kaplan–Meier plot for corporate survival. *Source*: Data from our data provider. [Colour figure can be viewed at wileyonlinelibrary.com]

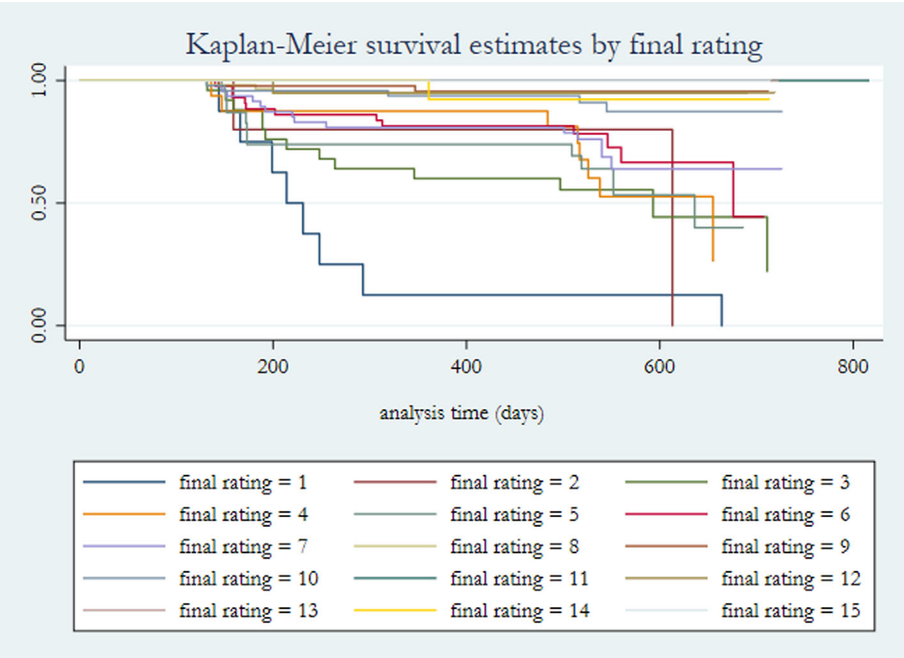

FIGURE 3 Kaplan–Meier plot for corporate survival by 'final rating'. *Source*: Data from our data provider. [Colour figure can be viewed at wileyonlinelibrary.com]

that the bank requires each loan officer to fill up as part of the borrower's credit scoring. Finally, 'final ratings' represent the last type of internal rating generated with respect to a given applicant (the one that is effectively attached to the given loan application) and can coincide or deviate from the previously produced integrated ratings based on the opportunity given to loan officer to override them at their discretion. Next, we study the impact of natural logarithm of the DD which is computed using the Merton's model as mentioned in Sections 3.3 and A1 of the Appendix in predicting corporate defaults. Further, we include the borrower-specific control variables such as size, long-term debt (LTD), EBITDA, and liquidity for each firm i at time $t-1$.[12] To address concerns of macroeconomic shocks, we also add the domestic annual GDP growth rate (GDPgrowth) of the geographic area in which the headquarters of the borrowing firm is located. $\delta D_{\text{industry}}$ allows to control for industry-specific fixed effects based on the domestic industry classification. $\varepsilon_{i,t}$ reflects the error term for borrower i at time t.

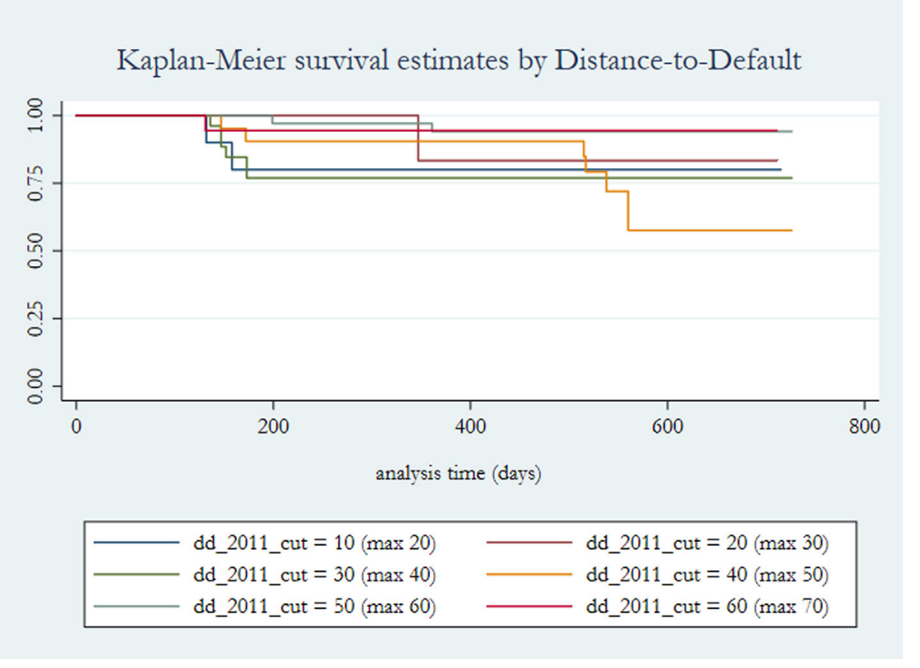

FIGURE 4 Kaplan–Meier plot for corporate survival by 'distance-to-default'. *Source*: Data from our data provider. [Colour figure can be viewed at wileyonlinelibrary.com]

4.2 | Bank's soft information and corporate defaults

We are now interested to investigate the predictive power of the bank's final ratings using different types of soft information on corporate defaults. We proceed by interacting the bank's final ratings with our measures of soft information. We first interact the final ratings with the variable 'soft information' that is a binary variable equal to one every time we observe a final rating that is different from the statistical rating and zero otherwise. We do this to test the effects of a change in the final rating based on the presence of overall soft information injected by the loan officer, without distinguishing between uncodified and codified discretion. Next, we interact the final ratings with the variable 'override' that is a dummy variable that equals one every time we observe a difference between the final and integrated ratings, and zero otherwise. In this regard, we test the effects of a change in the final ratings based on the presence of the purest form of soft information injected by the loan officer through rating overrides. Finally, we interact the final ratings with the variable 'questionnaire' that is a dummy variable that equals one every time we observe a difference between the integrated and the statistical ratings, and zero otherwise. In this regard, we test the effects of a change in the final ratings based on the deviations in the potential ratings that result from the qualitative questionnaire that the bank specifically requires the loan officer to complete as part of the credit rating process. Our Cox hazard estimations for the above-mentioned models are as follows:

$$\begin{aligned} h(\text{default}_{i,t}) = h_0(t) x \exp\Big(&\beta_0 + \beta_1 \text{Final Rating}_{i,t-1} \\ &+ \beta_2 \text{Final Rating}_{i,t-1} \\ &\times \text{Soft Information}_{i,t-1} \\ &+ \beta_3 \text{Soft Information}_{i,t-1} + \beta_4 \text{DD}_{i,t-1} \\ &+ \beta_5 \text{Size}_{i,t-1} + \beta_6 \text{LTD}_{i,t-1} \\ &+ \beta_7 \text{EBITDA}_{i,t-1} + \beta_8 \text{Liquidity}_{i,t-1} \\ &+ \beta_9 \text{GDPgrowth} + \delta D_{\text{industry}} + \varepsilon_{i,t}\Big) \end{aligned} \tag{2}$$

$$\begin{aligned} h(\text{default}_{i,t}) = h_0(t) x \exp\Big(&\beta_0 + \beta_1 \text{Final Rating}_{i,t-1} \\ &+ \beta_2 \text{Final Rating}_{i,t-1} \\ &\times \text{Override}_{i,t-1} + \beta_3 \text{Override}_{i,t-1} \\ &+ \beta_4 \text{DD}_{i,t-1} + \beta_5 \text{Size}_{i,t-1} + \beta_6 \text{LTD}_{i,t-1} \\ &+ \beta_7 \text{EBITDA}_{i,t-1} + \beta_8 \text{Liquidity}_{i,t-1} \\ &+ \beta_9 \text{GDPgrowth} + \delta D_{\text{industry}} + \varepsilon_{i,t}\Big) \end{aligned} \tag{3}$$

$$\begin{aligned} h(\text{default}_{i,t}) = h_0(t) x \exp\Big(&\beta_0 + \beta_1 \text{Final Rating}_{i,t-1} \\ &+ \beta_2 \text{Final Rating}_{i,t-1} \times \text{Questionnaire}_{i,t-1} \\ &+ \beta_3 \text{Questionnaire}_{i,t-1} + \beta_4 \text{DD}_{i,t-1} \\ &+ \beta_5 \text{Size}_{i,t-1} + \beta_6 \text{LTD}_{i,t-1} \\ &+ \beta_7 \text{EBITDA}_{i,t-1} + \beta_8 \text{Liquidity}_{i,t-1} \\ &+ \beta_9 \text{GDPgrowth} + \delta D_{\text{industry}} + \varepsilon_{i,t}\Big) \end{aligned} \tag{4}$$

TABLE 2 Baseline model of statistical, integrated, final rating and distance-to-default – survival analysis.

	Default				
Dependent variable	(1)	(2)	(3)	(4)	(5)
Statistical rating	**−0.375*** **				
	(0.086)				
Integrated rating		**−0.396*** **			
		(0.089)			
Final rating			**−0.425*** **		**−0.373*** **
			(0.070)		(0.081)
Merton's DD				**−1.088*** **	**−0.517*** **
				(0.221)	(0.191)
Borrower's size	0.183**	0.207**	0.160	0.029	0.105
	(0.086)	(0.096)	(0.126)	(0.093)	(0.114)
Borrower's long-term debt	−4.550***	−4.162***	−3.447***	−4.962***	−4.623***
	(1.476)	(1.170)	(0.841)	(1.228)	(0.706)
Borrower's EBITDA	2.992	1.863	2.512	−8.194**	1.571
	(2.871)	(2.588)	(2.167)	(3.717)	(3.177)
Borrower's liquidity	−2.185	−1.785	−1.258	−5.559	−1.546
	(4.330)	(3.576)	(3.319)	(5.435)	(3.314)
GDP growth rate	−1.627	−9.499	−12.204	−41.377	−20.804
	(35.422)	(37.028)	(34.863)	(46.040)	(32.439)
Observations	304	304	304	304	304
Industry FE	YES	YES	YES	YES	YES
Area FE	YES	YES	YES	YES	YES

Note: The bold value indicates to highlight the most important and statistically significant coefficients. The table presents the results of the Cox hazard model for the predictive power of bank's internal ratings and Merton's distance-to-default (DD) on corporate defaults. The dependent variable 'Default' is a binary variable equal to one if the borrower defaults, and zero otherwise. The independent variables are: (i) Merton's DD; (ii) statistical rating; (iii) integrated rating; (iv) final rating; (v) size; (vi) long-term debt; (vii) EBITDA; (viii) liquidity; (ix) GDP growth rate. All variables are defined in Table 1. Area and industry fixed effects are incorporated in all regressions (not reported). Robust errors reported in parentheses are clustered at the industry level. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$, respectively.

4.3 | Bank's upgrades and downgrades in predicting corporate defaults

Following our previous analysis of Sub-section 4.2, we now investigate the predictive power of a bank's final ratings with soft information on corporate defaults by distinguishing between upgrades and downgrades. Our data facilitates to differentiate between two steps of injecting soft information into borrowers' internal credit ratings: the qualitative questionnaire and the rating overrides. We start by interacting the bank's final ratings with the rating overrides in the form of both upgrades and downgrades that is performed in the 'uncodified' stage of the rating process. To this purpose, we build the variables 'upgrade override' and 'downgrade override' which are dummies that equal one in the case of an upgrade or downgrade override, respectively, and zero otherwise. Further, we then proceed to interact the bank's final ratings with loan officers' upgrades and downgrades performed through the mandatory qualitative questionnaire required in the 'codified' stage of the rating process. To this purpose, we build the variables 'upgrade questionnaire' and 'downgrade questionnaire' which are dummies that equal one in case when the loan officer makes a decision to upgrade or downgrade based on the questionnaire respectively, and zero otherwise. Our Cox hazard estimations for the above-mentioned models are as follows:

$$\begin{aligned} h(\text{default}_{i,t}) = h_0(t) x \exp\Big(&\beta_0 + \beta_1 \text{Final Rating}_{i,t-1} \\ &+ \beta_2 \text{Final Rating}_{i,t-1} \times \text{Upgrade Override}_{i,t-1} \\ &+ \beta_3 \text{Final Rating}_{i,t-1} \times \text{Downgrade Override}_{i,t-1} \\ &+ \beta_4 \text{Upgrade Override}_{i,t-1} \\ &+ \beta_5 \text{Downgrade Override}_{i,t-1} \\ &+ \beta_6 \text{DD}_{i,t-1} + \beta_7 \text{Size}_{i,t-1} + \beta_8 \text{LTD}_{i,t-1} \\ &+ \beta_9 \text{EBITDA}_{i,t-1} + \beta_{10} \text{Liquidity}_{i,t-1} \\ &+ \beta_{11} \text{GDPgrowth} + \delta D_{\text{industry}} + \varepsilon_{i,t}\Big) \end{aligned} \tag{5}$$

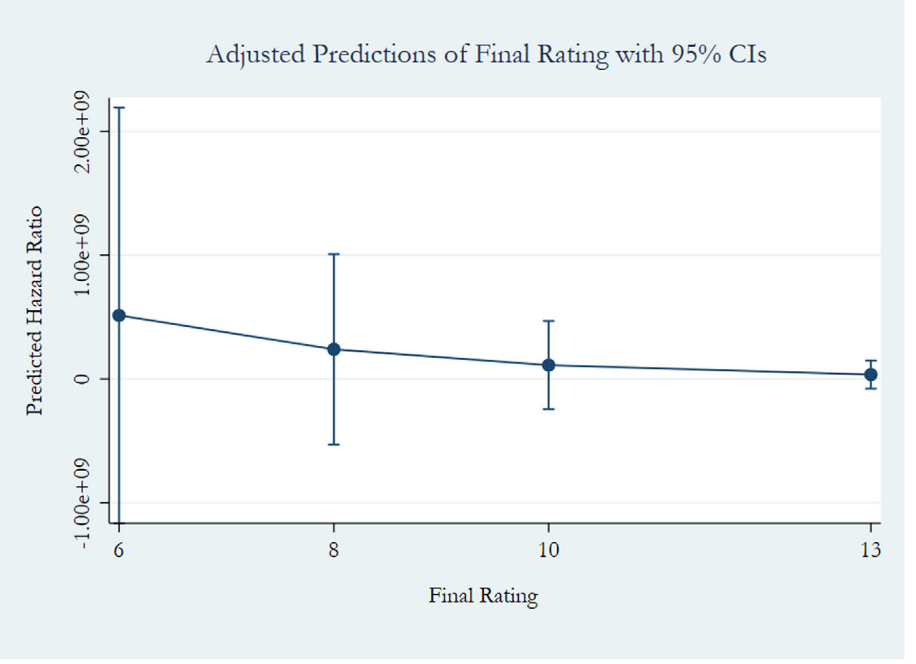

FIGURE 5 Predicted hazard ratio at percentiles of 'final rating'. On the x-axis are reported the 25th, 50th, 75th, and 90th percentiles of 'final rating' distribution. [Colour figure can be viewed at wileyonlinelibrary.com]

$$\begin{aligned} h(\text{default}_{i,t}) = h_0(t)x\exp\Big(&\beta_0 + \beta_1\text{Final Rating}_{i,t-1} \\ &+ \beta_2\text{Final Rating}_{i,t-1} \\ &\times\text{Upgrade Questionnaire}_{i,t-1} \\ &+ \beta_3\text{Final Rating}_{i,t-1} \\ &\times\text{Downgrade Questionnaire}_{i,t-1} \\ &+ \beta_4\text{Upgrade Questionnaire}_{i,t-1} \\ &+ \beta_5\text{Downgrade Questionnaire}_{i,t-1} \\ &+ \beta_6\text{DD}_{i,t-1} + \beta_7\text{Size}_{i,t-1} + \beta_8\text{LTD}_{i,t-1} \\ &+ \beta_9\text{EBITDA}_{i,t-1} + \beta_{10}\text{Liquidity}_{i,t-1} \\ &+ \beta_{11}\text{GDPgrowth} + \delta D_{\text{industry}} + \varepsilon_{i,t}\Big) \end{aligned} \tag{6}$$

5 | RESULTS

5.1 | The predictive power of banks and markets for corporate defaults

In this section, the bank's internal ratings that are attributed to a given borrowing firm are disentangled into intermediate and final components, and we estimate the models discussed in Equation (1). Our results are provided in Table 2. In columns (1)–(4), we investigate the effects of the bank's statistical rating, integrated rating, final rating, and the DD on the predictive probability of defaults, respectively. In the last column (5), we jointly examine the effect of 'final rating' and 'DD' on the prediction of corporate defaults.

In columns (1)–(2), we find that intermediate ratings have a negative and statistically significant effect that indicates a borrower's higher credit quality is associated with a lower likelihood of default. Further, the overall final rating (in column 3) shows a negative and significant influence with a greater magnitude effect on predicting corporate defaults as it contains the informational content of all intermediate ratings and represents the definitive rating evaluated by the bank when granting credit to a given borrowing firm. In terms of the economic magnitude, we find that a one standard deviation increase in statistical, integrated, and final ratings is associated with a reduced default probability by 22.22%, 25.92%, and 29.03%, respectively.[13]

Next, our findings in column (4) show a negative and statistically significant coefficient for the DD that indicates a smaller distance-to-default is associated with a rising likelihood of default for unlisted firms. In terms of the economic magnitude, we find that a one standard deviation increase in 'DD' is associated with lower probability of defaults by 103.56%. This probability confirms that market-based information contains significant predictive power on the incidence of default for unlisted firms. Further, in column (5), the results confirm the statistical significance of 'final rating' and 'DD' on predicting corporate defaults that in addition to hard and soft information, market-related information improves the predictive power of the credit rating assessments by banks for corporate defaults for unlisted corporate borrowers. We find that a one standard deviation increase in the 'final rating' and the 'DD' improves the predictive ability which helps to reduce the probability of defaults

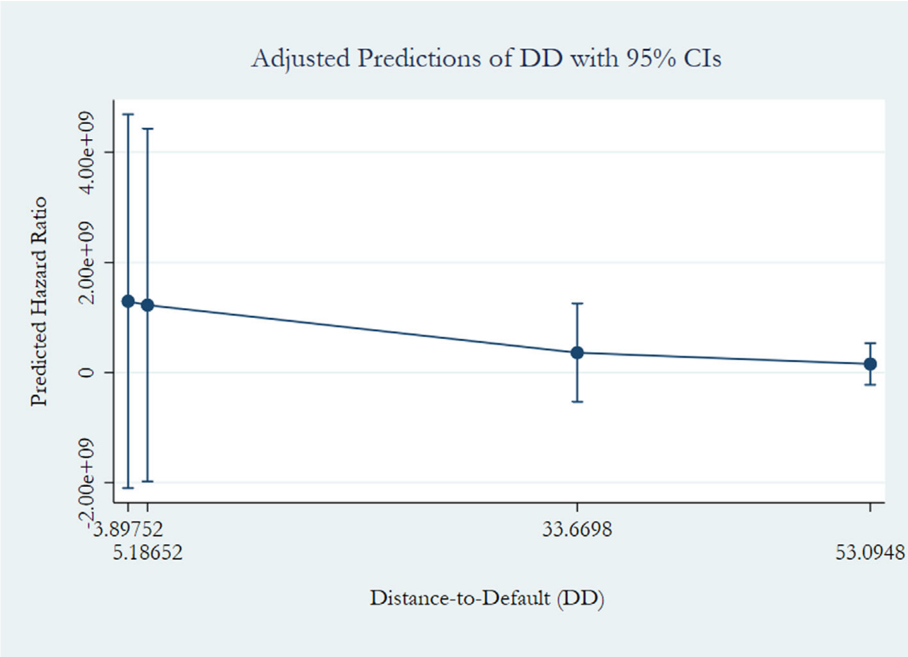

FIGURE 6 Predicted hazard ratio at percentiles of 'DD'. On the x-axis are reported the 25th, 50th, 75th, and 90th percentiles of 'DD' distribution. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Out-of-sample prediction.

Observations of the training set				**Observations of the test set**
243				61
	Mean	***SD***	**Min**	**Max**
Standard error of the default prediction: final rating	2.523369	0.273676	1.703571	3.19157
Standard error of the default prediction: final rating and market information	1.914626	0.314176	1.126758	2.545858

Note: The table presents the results of out-of-sample prediction where we assign 80% of the data points to the training set (243 observations) and the remaining 20% to the test set (61 observations). We train the model using the training set and then apply the model to the test set. Specifically, specification (1) gives the standard error of the default prediction of the model generated by Equation (1) with only the final rating, while specification (2) gives the standard error of the default prediction of the model generated by Equation (1) with both the final rating and market information.

by 24.03% and 76.19%, respectively. The graphical evidence provided in Figures 5 and 6 confirm the predictive margins of the hazard ratio for different percentiles of 'final rating' and 'DD' that show the predicted hazard ratio increases with the borrower's final rating (where a higher rating corresponds to worse rating classes) and decreases with the DD (where higher DD values reflect a firm's higher distance-to-default), which is in line with our empirical findings.

The control variables show that higher long-term debt is significantly associated with a lower likelihood of default, consistent with the idea that higher long-term debt supports business continuity by allowing firms to invest in profitable projects. Moreover, we add the domestic annual GDP growth rate to control for macroeconomic shocks. In a similar vein, we control for the country's area and the industry in which the given borrower operates according to the domestic industry classification.[14]

To strengthen our claim that market related information improves the banks' predictive power for defaults, we perform an out-of-sample prediction and demonstrate that adding a certain type of information causes the model's predictive power to improve. To divide the modelling dataset into the training set and the testing set, we assign 80% of the data points to the training set (243 observations) and the other 20% to the test set (61 observations).[15] We then use the model with the training set and apply it to the test set in order to evaluate the performance of Equation (1). We gather predictions

TABLE 4 Predicting corporate default by using bank's final rating with soft information – survival analysis.

	Default		
Dependent variable	**(1)**	**(2)**	**(3)**
Final rating	**−0.311*** **	**−0.351*** **	**−0.303*** **
	(0.113)	(0.109)	(0.109)
Merton's DD	**−0.634** **	**−0.501*** **	**−0.666*** **
	(0.293)	(0.203)	(0.270)
Final rating × Soft information	**−0.168** **		
	(0.102)		
Final rating × Override		**−0.706* **	
		(0.436)	
Final rating × Questionnaire			**−0.164** **
			(0.090)
Soft information	1.048		
	(0.790)		
Override		1.820	
		(2.295)	
Questionnaire			1.303**
			(0.604)
Borrower's size	0.092	0.042	0.082
	(0.112)	(0.128)	(0.102)
Borrower's long-term debt	−5.187***	−3.667***	−5.109***
	(0.977)	(0.666)	(0.829)
Borrower's ebitda	1.135	2.441	0.128
	(4.003)	(3.970)	(3.976)
Borrower's liquidity	−1.813	−1.831	−1.799
	(3.971)	(3.965)	(3.922)
GDP growth rate	−25.814	−18.827	−36.049
	(36.198)	(32.030)	(38.988)
Observations	304	304	304
Industry FE	YES	YES	YES
Area FE	YES	YES	YES

Note: The bold value indicates to highlight the most important and statistically significant coefficients. The table presents the results of the Cox hazard model for the predictive power of bank's internal ratings and Merton's distance-to-default (DD) on corporate defaults. The dependent variable 'Default' is a binary variable equal to one if the borrower defaulted, and zero otherwise. The independent variables are: (i) Merton's DD; (ii) statistical rating; (iii) integrated rating; (iv) final rating; (v) soft information; (vi) override; (vii) questionnaire; (viii) long-term debt; (ix) EBITDA; (x) liquidity; (xi) GDP growth rate. All variables are defined in Table 1. Area and industry fixed effects are incorporated in all regressions (not reported). Robust errors reported in parentheses are clustered at the industry level. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$, respectively.

from the training set and compare them to the withheld output values of the test set. This comparison allows us to evaluate how accurate these predictions are and to determine whether the statistics of their errors are similar to those in the fitted model. We report these test results in Table 3. Specifically, the out-of-sample validation results show that the standard error of the default prediction is lower for a credit risk model that includes both banking and market information on firms. This error validates the use of market information as a better predictor of defaults by banks for unlisted firms.

Further, to evaluate the goodness of fit of the risk models in the survival analysis, we use Harrell's C-index, also referred to as the concordance index (Harrell et al., 1982).[16] In our study, it represents the probability that a randomly selected defaulting company will have a higher predicted default risk than a randomly selected non-defaulting company. If the value of the Harrell's C-index is close to 0.5, it indicates that the predicted risk is not accurate in determining whether the borrowers will go bankrupt. If the values are close to zero, it means that the risk ratings are highly misleading. Further, if the values are close to one, they indicate that the risk ratings are good in determining which of the borrowing firms will default. In our analysis, the Harrell's C-index is 0.8116 which is very close to the optimal value of one that indicates a high degree of accuracy for our survival analysis. This evidence is further corroborated by a value of Somers' D that is equal to 0.6233, since large values (tending towards −1 or 1) show that the model has good predictive ability (Somers, 1962).[17]

TABLE 5 Predicting corporate default by using bank's upgrades versus downgrades – survival analysis.

	Default	
Dependent variable	**(1)**	**(2)**
Final rating	−0.351***	−0.307***
	(0.119)	(0.107)
Merton's DD	−0.443**	−0.790***
	(0.225)	(0.283)
Final rating × Upgrade override	**−0.329*** **	
	(0.114)	
Final rating × Downgrade override	−0.497	
	(0.425)	
Upgrade override	−43.522***	
	(1.175)	
Downgrade override	1.586	
	(2.234)	
Final rating × Upgrade questionnaire		**−0.480*** **
		(0.117)
Final rating × Downgrade questionnaire		−0.044
		(0.080)
Upgrade questionnaire		3.396***
		(0.615)
Downgrade questionnaire		0.543
		(0.629)
Borrower's size	0.044	0.093
	(0.132)	(0.119)
Borrower's long-term debt	−2.987***	−5.592***
	(0.889)	(0.888)
Borrower's EBITDA	2.990	0.890
	(4.568)	(4.611)
Borrower's liquidity	−2.017	−1.555
	(4.030)	(3.273)
GDP growth rate	−30.445	−48.945
	(34.572)	(38.864)
Observations	304	304
Industry FE	YES	YES
Area FE	YES	YES

Note: The bold value indicates to highlight the most important and statistically significant coefficients. The table presents the results of the Cox hazard model for the predictive power of bank's internal ratings interacted with upgrading and downgrading decisions and Merton's distance-to-default (DD) on corporate defaults. The dependent variable 'Default' is a binary variable equal to one if the borrower defaults, and zero otherwise. The independent variables are: (i) Merton's DD; (ii) statistical rating; (iii) integrated rating; (iv) final rating; (v) upgrade override; (vi) downgrade override; (vii) upgrade questionnaire; (viii) downgrade questionnaire; (ix) long-term debt; (x) EBITDA; (xi) liquidity; (xii) GDP growth rate. All variables are defined in Table 1. Area and industry fixed effects are incorporated in all regressions (not reported). Robust errors reported in parentheses are clustered at the industry level. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$, respectively.

5.2 | Bank's soft information and corporate defaults

In Table 4, we show our findings for the different types of soft information in lending decisions as per Equations (2)–(4). To this purpose, we represent soft information as the potential change in a borrower's internal intermediate ratings that could occur throughout the credit rating process. Specifically, according to the hybrid rating process, the loan officer has the possibility to inject soft information into the process in two different discretionary moments. The first moment is a step specifically required by the bank where the loan officer who is in-charge of the banking relationship has to fill out a qualitative questionnaire about several borrowers' characteristics. This first step in the credit rating procedure turns a statistical rating into an integrated one which could be equal to or deviate from the purely hard information-based statistical rating initially generated by the credit scoring tool of the bank. The second moment grants to the loan officer the opportunity to propose a rating override by proposing a final rating which deviates from the previously produced integrated rating. We argue that any rating amendment on the part of the loan officer reflects the use of discretion that stems

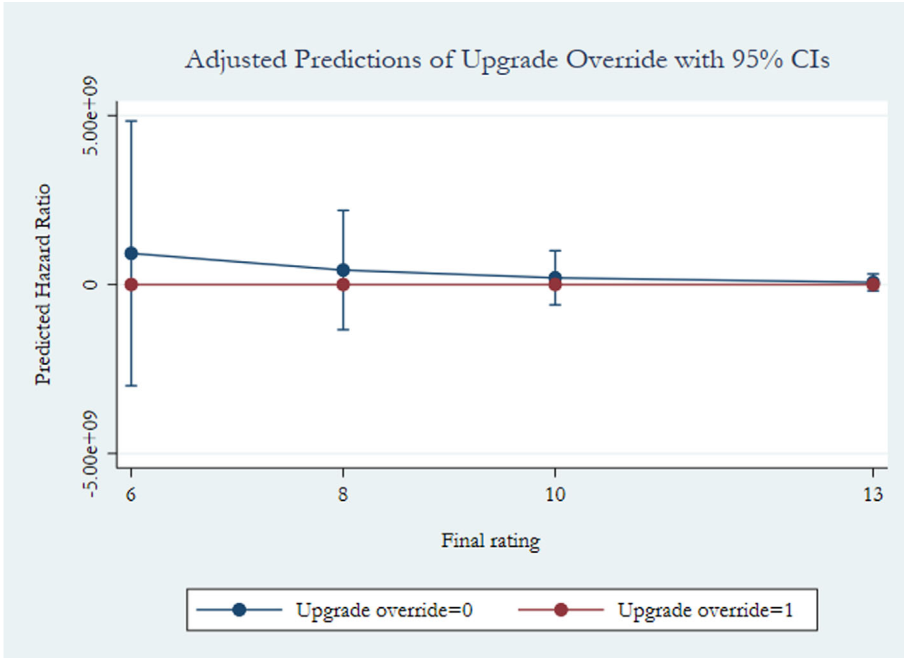

FIGURE 7 Predicted hazard ratio at percentiles of 'final rating' based on 'override' upgrades. On the x-axis are reported the 25th, 50th, 75th, and 90th percentiles of 'final rating' distribution. [Colour figure can be viewed at wileyonlinelibrary.com]

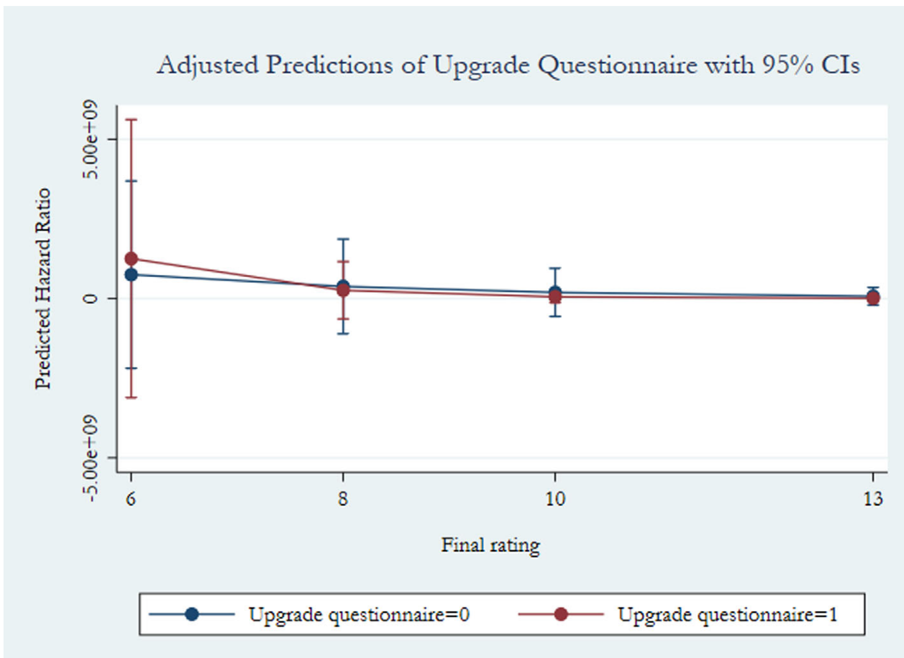

FIGURE 8 Predicted hazard ratio at percentiles of 'final rating' based on 'questionnaire' upgrades. On the x-axis are reported the 25th, 50th, 75th, and 90th percentiles of 'final rating' distribution. [Colour figure can be viewed at wileyonlinelibrary.com]

from accumulated soft information throughout repeated interactions with the same borrowing firm. While soft information injected into the qualitative questionnaire is performed through answers given to a set of pre-determined questions required by the bank, a proposed rating override represents a proactive behaviour on the part of the loan officer. Therefore, in Table 4 we investigate three different models based on the nature of the soft information injected into the credit scoring process.

First, in column (1) in Table 4, we investigate the differential effect of the final rating based on the presence (or absence) of overall soft information by interacting it with the borrower's final rating (final rating × soft information). Our results show that the coefficient of the interaction term is negative and statistically significant that means that when the overall soft information is added to the bank's final ratings, it better predicts the probability of corporate defaults. In terms of the economic magnitude, we find that one standard deviation increase in soft information in the final ratings improves the predictability by reducing the default probabilities by 67.91%. Further, we find a positive but insignificant coefficient for soft information.

TABLE 6 Robustness: Alternative measure of soft information.

Dependent variable	Default (1)
Final rating	−0.384**
	(0.175)
Merton's DD	−1.119**
	(0.526)
Final rating × Sentiment score	−0.027***
	(0.010)
Sentiment score	−0.156***
	(0.056)
Borrower's size	0.028
	(0.168)
Borrower's long-term debt	−9.543***
	(2.751)
Borrower's EBITDA	−4.108
	(6.117)
Borrower's liquidity	−18.434**
	(8.101)
GDP growth rate	7.084
	(23.941)
Observations	304
Industry FE	YES
Area FE	YES

Note: The table presents the results of the Cox hazard model for the predictive power of bank's internal ratings interacted with an alternative measure of soft information and Merton's distance-to-default (DD) on corporate defaults. The dependent variable 'Default' is a binary variable equal to one if the borrower defaults, and zero otherwise. The independent variables are: (i) Merton's DD; (ii) statistical rating; (iii) integrated rating; (iv) final rating; (v) sentiment score; (vi) upward sentiment score; (vii) downward sentiment score; (viii) size; (ix) long-term debt; (x) ebitda; (xi) liquidity; (xii) GDP growth rate. All variables are defined in Table 1. Area and industry fixed effects are incorporated in all regressions (not reported). Robust errors reported in parentheses are clustered at the industry level. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$, respectively.

In columns (2) and (3) of Table 4, we investigate the influence of soft information using override and questionnaire on predicting corporate defaults and interact it with the final rating of the borrower. Our results show a negative and statistically significant coefficient of the interaction terms 'final rating × override' and 'final rating × questionnaire' indicating that when soft information is injected through uncodified and codified discretions into the final rating, they help to better predict the probability of corporate defaults. In terms of the economic magnitudes, we find that one standard deviation increase in uncodified and codified discretions in the final ratings predicts lower probability of defaults by 94.13% and 68.59%, respectively. The economic magnitudes show a larger effect of uncodified discretion as compared to codified discretion. Next, we provide evidence of a positive and significant coefficient only for codified discretion (questionnaire) which means that for the least creditworthy borrowers, loan officers might be incentivised to inflate credit scores to game the system that results in higher corporate defaults. This is because codified discretion is neither validated nor monitored by the bank's headquarters and does not expose loan officers to any reputational kickbacks. The results for all other control variables are as inferred.

5.3 | Bank's upgrades and downgrades in predicting corporate defaults

In Table 5, we further investigate the potential differential effects on firm defaults of upgrading and downgrading discretionary decisions performed by the loan officer in both uncodified step, that is, override, and codified step of the rating process, that is, questionnaire.

In column (1) of Table 5, we interact the variables of upgrade and downgrade overrides with a borrower's final rating to create the terms of 'final rating × upgrade override' and 'final rating × downgrade override', respectively. On the one hand, the coefficient of the interaction term 'final rating × upgrade override' shows a negative and statistically significant effect on corporate defaults that indicates an increase in positive information in the 'uncodified' step of the rating process predicts a lower probability of experiencing corporate defaults. On the other hand, the coefficient of the interaction term 'final rating × downgrade override' is not statistically significant even though the coefficient has a positive sign. In terms of the economic magnitude, we find that one standard deviation increase in the upgrade override of the final rating predicts lower probability of defaults by 84.45%. Confirmatory graphical evidence of the effect of upgrading overrides on the predictive margins of hazard ratio for different percentiles of 'final rating' is provided in Figure 7. The figure shows that loan officers' decisions on upgrade overrides are associated with lower predicted hazard ratios.

Next, in column (2) of Table 5, we create the two interaction terms 'final rating × upgrade questionnaire' and 'final rating × downgrade questionnaire'. Our results show that the coefficient of 'final rating × upgrade questionnaire' has a negative and statistically significant effect on corporate defaults prediction, while the coefficient of 'final rating × downgrade questionnaire' is not statistically significant. These results indicate that the upgrades that result from mandatory qualitative questionnaire taken

TABLE 7 Robustness: Baseline model – ordered probit model.

Dependent variable	Financial status_ordered probit (1)	(2)	(3)
Final rating	**−0.354*** **		**−0.337*** **
	(0.037)		(0.045)
Merton's DD		**−0.541*** **	**−0.240*** **
		(0.137)	(0.092)
Borrower's size	0.162***	0.056	0.134***
	(0.024)	(0.050)	(0.019)
Borrower's long-term debt	−4.847***	−4.435**	−5.738***
	(1.421)	(1.866)	(1.609)
Borrower's EBITDA	4.099***	−4.122	4.095**
	(1.419)	(2.554)	(1.788)
Borrower's liquidity	−1.158	−2.198	−1.264
	(1.284)	(2.190)	(1.189)
GDP growth rate	10.836	−15.614	7.430
	(12.799)	(21.856)	(12.388)
Observations	297	297	297
Industry FE	YES	YES	YES
Area FE	YES	YES	YES

Note: The bold value indicates to highlight the most important and statistically significant coefficients. The table presents the results of the ordered Probit model for the predictive power of bank's internal ratings and Merton's distance-to-default (DD) on corporate defaults. The dependent variable 'Financial Status_Ordered Probit' is a categorical variable reflecting the borrower's financial status. The independent variables are: (i) Merton's DD; (ii) statistical rating; (iii) integrated rating; (iv) final rating; (v) size; (vi) long-term debt; (vii) EBITDA; (viii) liquidity; (ix) GDP growth rate. All variables are defined in Table 1. Area and industry fixed effects are incorporated in all regressions (not reported). Robust errors reported in parentheses are clustered at the industry level.
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$, respectively.

by the loan officers show a lower probability of corporate defaults. In terms of the economic magnitude, we find that one standard deviation increase in the upgrade questionnaire for the final rating predicts lower probability of defaults by 89.14%. Confirmatory graphical evidence of the effects of upgrades on the predictive margins of the hazard ratio for the different percentiles of 'final rating' is provided in Figure 8. The figure shows that loan officers' questionnaire upgrades are associated with higher predicted hazard ratios, particularly for the medium-to-worst rating classes, thus corroborating the idea that loan officers tend to inflate the ratings of relatively risky borrowers through the exercise of codified discretion that does not expose them to any reputational kickbacks.

6 | ROBUSTNESS CHECKS

We now perform several robustness checks to further confirm our main empirical results.[18]

6.1 | Alternative measure of soft information

In line with Campbell et al. (2019), we now consider an alternative measure of soft information based on the keywords used by loan officers in the written notes of the loan applications and measured as the difference between positive and negative keywords divided by the total number of words contained in each loan application. We follow Li's (2010) recommendation of performing a textual analysis on the basis of dictionaries specifically developed to analyse soft information. To implement our dictionary-based approach, we adapt appropriate lists of positive and negative words using the Loughran McDonald (LM) dictionary that is a list of words that reflect the LM sentiment by category (Loughran & Mcdonald, 2011). For our purposes, we focus on positive and negative keywords. The negative and positive keywords are translated and individually checked for potential spelling and construction mismatching and depurated from stopwords as a cleaning procedure. Moreover, dictionaries of positive and negative words are extended to enrich the

TABLE 8 Robustness: Predicting corporate default by using bank's final rating with soft information – ordered probit model.

Dependent variable	Financial status_ordered probit (1)	(2)	(3)
Final rating	**−0.305*****	**−0.322*****	**−0.296*****
	(0.062)	(0.048)	(0.055)
Merton's DD	**−0.303*****	**−0.233*****	**−0.324*****
	(0.119)	(0.091)	(0.109)
Final rating × Soft information	**−0.110***		
	(0.078)		
Final rating × Override		**−0.538****	
		(0.279)	
Final rating × Questionnaire			**−0.111***
			(0.062)
Soft information	0.929		
	(0.643)		
Override		2.153	
		(1.402)	
Questionnaire			1.065**
			(0.433)
Borrower's size	0.125***	0.098***	0.112***
	(0.029)	(0.033)	(0.027)
Borrower's long-term debt	−6.509***	−5.533***	−6.427***
	(1.355)	(1.308)	(1.637)
Borrower's EBITDA	4.340**	4.372***	3.668*
	(1.774)	(1.557)	(2.071)
Borrower's liquidity	−1.253	−1.616	−1.193
	(1.283)	(1.315)	(1.268)
GDP growth rate	3.631	6.395	−0.444
	(11.348)	(14.029)	(9.615)
Observations	297	297	297
Industry FE	YES	YES	YES
Area FE	YES	YES	YES

Note: The bold value indicates to highlight the most important and statistically significant coefficients. The table presents the results of the ordered Probit model for the predictive power of bank's internal ratings and Merton's distance-to-default (DD) on corporate defaults. The dependent variable 'Financial Status_Ordered Probit' is a categorical variable reflecting the borrower's financial status. The independent variables are: (i) Merton's DD; (ii) statistical rating; (iii) integrated rating; (iv) final rating; (v) size; (vi) long-term debt; (vii) EBITDA; (viii) liquidity; (ix) GDP growth rate. All variables are defined in Table 1. Area and industry fixed effects are incorporated in all regressions (not reported). Robust errors reported in parentheses are clustered at the industry level. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$, respectively.

algorithm capability to detect and score words into the positive and negative categories in the context of different languages (Chen & Skiena, 2014).

The 'sentiment score' is therefore a continuous variable extracted from the text of the individual loan applications, where each positive word counts as +1 and each negative word as −1. The sentiment score for each document has then been divided by the number of words contained in the text section (after stopwords exclusion) to avoid the dependence of the score to the number of words contained in the given loan application. Loan officers' notes with greater sentiment scores reflect a more positive sentimental tone of the given loan application.

Table 6 provides the results of these estimations. The results corroborate a significant effect of hard, soft, and

TABLE 9 Robustness: Predicting corporate default by using bank's upgrades versus downgrades – ordered probit model.

Dependent variable	Financial status_ordered probit (1)	(2)
Final rating	−0.322***	−0.283***
	(0.047)	(0.062)
Merton's DD	−0.205*	−0.395***
	(0.115)	(0.121)
Final rating × Upgrade override	**−0.126****	
	(0.051)	
Final rating × Downgrade override	**−0.393***	
	(0.204)	
Upgrade override	−5.180***	
	(0.334)	
Downgrade override	1.923	
	(1.177)	
Final rating × Upgrade questionnaire		**−0.306*****
		(0.097)
Final rating × Downgrade questionnaire		−0.071
		(0.062)
Upgrade questionnaire		2.063***
		(0.599)
Downgrade questionnaire		0.974*
		(0.550)
Borrower's size	0.097**	0.131***
	(0.039)	(0.026)
Borrower's long-term debt	−5.058***	−6.580***
	(1.589)	(1.415)
Borrower's EBITDA	4.537***	3.290
	(1.552)	(2.536)
Borrower's liquidity	−1.656	−1.099
	(1.362)	(1.131)
GDP growth rate	1.915	−11.168
	(10.720)	(8.726)
Observations	297	297
Industry FE	YES	YES
Area FE	YES	YES

Note: The bold value indicates to highlight the most important and statistically significant coefficients. The table presents the results of the ordered Probit model for the predictive power of bank's internal ratings and Merton's distance-to-default (DD) on corporate defaults. The dependent variable 'Financial Status Ordered Probit' is a categorical variable reflecting the borrower's financial status. The independent variables are: (i) Merton's DD; (ii) statistical rating; (iii) integrated rating; (iv) final rating; (v) size; (vi) long-term debt; (vii) EBITDA; (viii) liquidity; (ix) GDP growth rate. All variables are defined in Table 1. Area and industry fixed effects are incorporated in all regressions (not reported). Robust errors reported in parentheses are clustered at the industry level. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$, respectively.

TABLE 10 Robustness: Alternative DD measure – survival analysis.

Dependent variable	Default (1)	(2)
Alternative Merton's DD	**−0.943*****	**−0.449*****
	(0.206)	(0.184)
Final rating		−0.377***
		(0.081)
Borrower's size	0.046	0.124
	(0.100)	(0.116)
Borrower's long-term debt	−4.424***	−4.349***
	(1.070)	(0.628)
Borrower's ebitda	−8.033**	1.781
	(3.734)	(3.099)
Borrower's liquidity	−5.595	−1.461
	(5.280)	(3.219)
GDP growth rate	−42.491	−20.821
	(46.954)	(32.958)
Observations	303	303
Industry FE	YES	YES
Area FE	YES	YES

Note: The bold value indicates to highlight the most important and statistically significant coefficients. The table presents the results of the Cox hazard model for the predictive power of bank's internal ratings and the alternative measure of Merton's distance-to-default (Alternative Merton's DD) on corporate defaults. The dependent variable 'Default' is a binary variable equal to one if the borrower defaulted, and zero otherwise. The independent variables are: (i) alternative Merton's DD; (ii) statistical rating; (iii) integrated rating; (iv) final rating; (v) size; (vi) long-term debt; (vii) EBITDA; (viii) liquidity; (ix) GDP growth rate. All variables are defined in Table 1. Area and industry fixed effects are incorporated in all regressions (not reported). Robust errors reported in parentheses are clustered at the industry level. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$, respectively.

market information in reducing the likelihood of experiencing corporate defaults. Specifically, our results confirm that a more positive sentimental tone is associated with a decreased corporate default probability. Thus, our empirical findings prove to be robust to a different measure of soft information represented by the sentiment score.

6.2 | Alternative estimation method

We re-estimate all our models using an ordered probit estimation method. In these models we define the dependent variable of 'financial status_ordered probit' as a categorical variable which ranges from a minimum value of zero to a maximum value of seven that is

TABLE 11 Robustness: Type I and type II errors.

Riskiness buckets (final rating)	% of defaulted firms	Default rate
Low risk		
Defaulted	1.22%	3.13%
Medium risk		
Defaulted	65.85%	15.88%
High risk		
Defaulted	31.71%	60.47%

Riskiness buckets (bank's PD)	% of defaulted firms	Default rate
Low risk		
Defaulted	9.76%	5.59%
Medium risk		
Defaulted	39.02%	16.16%
High risk		
Defaulted	50.00%	55.41%

Note: The table presents the percentage of firms that eventually default and the default rate that from ranking the firms into default risk buckets based on the borrower's final rating and probability of default (computed by the bank's algorithm). The results reported in this table help to address any type I and type II errors in the data.

based on the severity of the borrower's financial distress as provided in Appendix A3.[19] Specifically, a value of zero represents a borrower who is in good financial health, while a value of seven represents a borrower who has defaulted.

The results are provided in Tables 7–9. We continue to find a significant effect of hard, soft, and market information in predicting firms' default in Table 7, while in Table 8 we confirm the relevance of hardened soft information in the form of codified and uncodified discretions that improve banks' ability to predict corporate defaults in addition to hard and market information. Table 9 shows that the decisions of upgrading overrides and upgrading questionnaires better predict corporate default. Thus, our main results prove to be robust even to this alternative estimation method.

6.3 | Alternative measure of DD

As a further robustness test, we now construct an alternative measure of the Merton's DD. We now estimate the asset volatility for each publicly traded firm by first taking the average of these individual asset volatilities characterising our peers, and then inserting the computed averaged volatilities into Equation (vii) as described in Appendix A1. Our results are provided in Table 10. We continue to find a highly significant and negative effect of the DD in predicting corporate default. Thus, our main results remain unaffected even when using this alternative measure in our empirical specifications.

6.4 | Eliminating type I and II errors

In default prediction studies, it is useful to present metrics to help evaluate the extent of type I and type II errors in the data. Hence, outside of the regression analysis, we rank firms into default risk buckets based on the borrower's final rating and probability of default (computed by the bank's algorithm) and evaluate the percentage of firms that eventually default and the default rate that lie in each default risk buckets (Shumway, 2001; Beaver et al., 2005).[20] We provide the results of the three bankruptcy risk buckets in Table 11. The table shows that both percentage of the firms that eventually default and default rates are increasing in credit risk. Thus, these results help to address any potential type I and type II errors in the data.

7 | CONCLUSIONS

In this study, we provide a novel contribution to the current empirical literature by testing the influence of soft, hard, and market information in predicting corporate defaults. We employ a unique proprietary dataset that contains detailed information on internal credit ratings and on the different intermediate credit ratings for European unlisted mid-sized corporations. Using the semiparametric Cox proportional hazard model, we evaluate the predictive power of hard (quantitative), soft (qualitative), and market information on corporate default probabilities. We proxy asset price volatility of our sample of unlisted firms by computing their Merton's DD with market data collected from comparable publicly listed firms.

We find that the simultaneous use of hard, soft, and market information in the bank's credit rating process significantly improves its predictive ability of firm defaults. Further, we show that in addition to hard and market information, codified and uncodified discretions have significant effects on improving the bank's ability to predict corporate defaults. Finally, when distinguishing between rating upgrades and downgrades performed in both the codified and uncodified stages of the credit scoring process, we find that it is precisely the upgrading discretion that is significantly associated with a reduced probability of a firm defaulting.

Despite the granularity and the depth of our proprietary data that make our empirical setting unique, we acknowledge that the empirical analysis of this study covers a limited time-period. Nonetheless, our empirical framework is consistent with other studies using very similar empirical settings, such as Liberti and Mian (2009), Filomeni et al. (2020, 2021), among others. Hence, despite the limited time-period, the conclusions derived from this study are relevant to policy and indicate that integrating market information into banks' hybrid credit rating models can significantly improve the accuracy of credit risk assessments of unlisted firms. Our results are indeed of crucial relevance to policy makers for the design of forward-looking financial risk management frameworks to safeguard the soundness of the banking sector playing a crucial intermediation role in the economy.

ACKNOWLEDGEMENTS

We are grateful to the Editor, Associate Editor, and the anonymous Reviewers for their insightful comments. We also thank Jerry Coakley, Claudia Girardone, Michele Modina, and participants at the 2019 Annual EFiC Conference in Banking and Corporate Finance held at Essex Business School (Colchester, UK) for providing constructive suggestions. Any remaining errors are our own.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study cannot be disclosed due to confidentiality reasons.

ORCID

Stefano Filomeni https://orcid.org/0000-0002-0838-5510
Udichibarna Bose https://orcid.org/0000-0002-7550-9974
Athanasios Triantafyllou https://orcid.org/0000-0002-0711-9851

ENDNOTES

[1] Market information is generally added where market signals specific to firms are available, and it can be incorporated via Merton's DD model, or by adding variables like market cap, excess stock returns, and firm volatility in the hazard model as implemented in Shumway (2001).

[2] Shumway (2001) shows that hazard models exhibit statistical superiority over static models which do not consider that a company is exposed to insolvency risk over multiple periods (i.e., 2 years after loan disbursement). Indeed, static models do not account for the fact that companies change over time, and, therefore, generate biased default probabilities.

[3] Credit scoring process and rating process are used interchangeably throughout this paper and refer to the same credit score/rating assignment process.

[4] We use the terms 'overrides' and 'uncodified discretion', and 'questionnaires' and 'codified discretion' interchangeably throughout the paper.

[5] A recent study by Naili and Lahrichi (2022) provide a structured review of the recent empirical and theoretical literature on bank's credit risk.

[6] We acknowledge that the empirical analysis of this study covers a limited time period; however, the literature has extensively used similar granular proprietary data with a limited time-period (e.g., Filomeni et al., 2020, 2021; Liberti & Mian, 2009).

[7] Unlike rating overrides, filling up the qualitative questionnaire does not require loan officers to adopt a proactive behaviour as it is a mandatory requirement by the bank during the rating process. See Appendix A4 for an example of the qualitative questionnaire provided by the bank.

[8] This proprietary algorithm combines competitor lists provided in filings, analyst cross coverage, business classification, and revenue proximity to produce the comparable peers. According to Refinitiv, this hierarchical approach produces very reasonable sets of peer companies for most securities.

[9] The equity volatilities of our eight portfolios of listed firms remain nearly the same even if we use the market value weighted average (instead of the simple average) of our portfolios' equity values.

[10] For detailed discussion, please refer to Kalbfleisch and Prentice (2002) or Cox and Oakes (1984).

[11] In fact, it seems unrealistic that all the ratings occur at the same time; presumably as the firm's credit quality falters, the internal ratings are revised more frequently by the bank.

[12] We drop firms with negative earnings and negative equity from the sample and the results are unaltered.

[13] Following Claessens and Schmukler (2007), the economic magnitude is computed as the log of estimated coefficients times one standard deviation of the covariates. For example, the standard deviation of 'statistical rating' (3.53) is multiplied by the log of its estimated coefficient (0.375), and the exponential of the resulting number represents the percentage change in the probability of corporate defaults (22.23%).

[14] Domestic firms account for 87.5% of our sample of 437 European non-listed firms; for this reason, in our regressions we control for possible macroeconomic and industry shocks at the domestic country level. In unreported regressions, we find that even if we exclude foreign borrowers from the analysis, our results remain qualitatively unaltered.

[15] Our out-of-sample results hold even after assigning 70% of the data points to the training set and the other 30% to the test set.

[16] Harrell's concordance-index is referred as the proportion of observations that the model can order correctly in terms of survival times. Harrel's C is called the AUC (Area Under the Curve) and is a useful way to evaluate the quality of the prediction model.

[17] Somers' D is referred to as the difference between the number of concordant pairs and the number of discordant pairs divided by the total number of pairs not tied on the independent variable.

[18] In unreported robustness tests, we restrict our sample based on borrower's size to test if the smallest or largest borrowers drive our results. In addition, we remove foreign borrowers from our sample to test if the inclusion of foreign borrowers drives our results. In both cases, our estimation results do not change much both qualitatively and quantitatively from the main estimation results.

[19] Appendix A3 shows that while some elements of the borrower's financial distress rely on objective criteria, others rely on subjective assessments. These categories of financial default are not determined by the same loan officer who produces the soft information, but by a different department of the organisation that works independently.

[20] As for the probability of default, firms with a PD below the 25th percentile of the PD distribution are assigned to the low-risk bucket, firms with a PD falling in between the 25th and 75th percentiles of the PD distributions are assigned to the medium-risk bucket, and companies with a PD above the 75th percentile of the PD distribution are assigned to the high-risk bucket. As for the borrowers' internal credit ratings, the low-, medium-, and high-risk buckets are determined by the given rating class to which the borrower belongs: investment, mezzanine, and speculative grade, respectively.

REFERENCES

Acheampong, A., & Elshandidy, T. (2021). Does soft information determine credit risk? Text-based evidence from European banks. *Journal of International Financial Markets, Institutions and Money*, *75*, 101303.

Agarwal, V., & Taffler, R. (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking and Finance*, *32*, 1541–1551.

Allen, F., Carletti, E., & Gu, X. (2015). The roles of banks in financial systems. In A. Berger, P. Molyneux, & J. Wilson (Eds.), *The Oxford handbook of banking* (pp. 292–311). Oxford University Press.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, *23*, 589–609.

Altman, E. I., & Sabato, G. (2008). Modeling credit risk for SMEs: Evidence from the US market. *Abacus*, *43*(3), 332–357.

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*, 33–127.

Andrikopolous, P., & Khorasgani, A. (2018). Predicting unlisted SMEs' default: Incorporating market information on accounting-based models for improved accuracy. *British Accounting Review*, *50*(5), 559–573.

Asch, L. (1995). How the RMA/fair, Isaac credit-scoring model was built. *Journal of Commercial Lending*, *77*(10).

Basel Committee on Banking Supervision (BCBS). (2000). Range of practice in banks' internal ratings systems. BCBS Discussion Paper.

Basel Committee on Banking Supervision (BCBS). (2001). The internal ratings-based approach, Bank for International Settlements.

Basel Committee on Banking Supervision (BCBS). (2005). Studies on the validation of internal rating systems. BCBS Working Paper No. 14.

Bauer, J., & Agarwal, V. (2014). Are hazard models superior to traditional bankruptcy prediction approaches? A comprehensive test. *Journal of Banking and Finance*, *40*, 432–442.

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

Beaver, W. H., McNichols, M. F., & Rhie, J. (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.

Berger, A. N., Miller, N. H., Petersen, M. A., Rajan, R. G., & Stein, J. C. (2005). Does function follow organizational form? Evidence from the lending practices of large and small banks. *Journal of Financial Economics*, *76*, 237–269.

Berger, A. N., & Udell, G. F. (1995). Relationship lending and lines of credit in small firm finance. *The Journal of Business*, *68*, 351–381.

Bertomeu, J., & Marinovic, I. (2016). A theory of hard and soft information. *The Accounting Review*, *91*(1), 1–20.

Bharath, S. T., & Shumway, T. (2008). Forecasting default with the Merton distance to default model. *Review Financial Studies*, *21*, 1339–1369.

Breden, D. (2008). Monitoring the operational risk environment effectively. *Journal of Risk Management in Financial Institutions*, *1*, 156–164.

Brown, M., Schaller, M., Westerfeld, S., & Heusler, M. (2015). Internal control and strategic communication within firms – Evidence from bank lending. University of St. Gallen, School of Finance Research Paper No. 2015/04.

Brown, M., Westerfeld, S., Schaller, M., & Heusler, M. (2012). Information or insurance? On the role of loan officer discretion in credit assessment. MoFiR Working Paper No 67.

Buehler, S., Kaiser, C., & Jaeger, F. (2006). Merge or fail? The determinants of mergers and bankruptcies in Switzerland, 1995–2000. *Economics Letters*, *90*, 88–95.

Byström, H. N. E. (2006). Merton unraveled: A flexible way of modeling default risk. *The Journal of Alternative Investments*, *8*, 39–47.

Campbell, D., Loumioti, M., & Wittenberg-Moerman, R. (2019). Making sense of soft information: Interpretation bias and loan quality. *Journal of Accounting and Economics*, *68*, 101240.

Casu, B., Chiaramonte, L., Croci, E., & Filomeni, S. (2022). Access to credit in a market downturn. *Journal of Financial Services Research*. https://doi.org/10.1007/s10693-022-00388-x

Chen, T., & Wang, S. (2023). Incomplete information model of credit default of micro and small enterprises. *International Journal of Finance and Economics*, *28*(3), 2956–2974. https://doi.org/10.1002/ijfe.2577

Chen, Y., & Skiena, S. (2014). Building sentiment lexicons for all major languages. Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 383–389.

Claessens, S., & Schmukler, S. L. (2007). International financial integration through equity markets: Which firms from which countries go global? *Journal of International Money and Finance*, *26*(5), 788–813.

Cole, R. A. (1998). The importance of relationships to the availability of credit. *Journal of Banking and Finance*, *22*, 959–977.

Cox, D. R. (1972). Regression models and life tables. *Journal of the Royal Statistical Society, Series B*, *34*, 187–220.

Cox, D., & Oakes, D. (1984). *Analysis of Survival Data*. London: Chapman and Hall.

Das, S. R., Hanouna, P., & Sarin, A. (2009). Accounting-based versus market-based cross-sectional models of CDS spreads. *Journal of Banking and Finance*, *33*, 719–730.

Demetriades, P. O., & Devereux, M. P. (2000). Investment and financial restraints: Theory and evidence. *International Journal of Finance and Economics*, *5*, 285–296.

Diamond, D. (1984). Financial intermediation and delegated monitoring. *Review of Economic Studies*, *51*(3), 393–414.

Diamond, D. (1991). Monitoring and reputation: The choice between bank loans and directly placed debt. *Journal of Political Economy*, *99*, 689–721.

Doumpos, M., Niklis, D., Zopounidis, C., & Andriosopoulos, K. (2015). Combining accounting data and a structural model for predicting credit ratings: Empirical evidence from European listed firms. *Journal of Banking and Finance*, *50*, 599–607.

Ederington, L., & Goh, J. (1998). Bond rating agencies and stock analysts: Who knows what when? *Journal of Financial and Quantitative Analysis*, *33*, 569–585.

Edmans, A., Heinle, M. S., & Huang, C. (2016). The real costs of financial efficiency when some information is soft. *Review of Finance*, *20*(6), 2151–2182.

Eisenbeis, R. A. (1996). Recent developments in the application of credit-scoring technique to the evaliation of commercial loan. *IMA Journal of Management Mathematics*, *7*, 271–290.

Faulkender, M., & Petersen, M. (2006). Does the source of capital affect capital structure? *Review of Financial Studies*, *19*, 45–79.

Fazzari, S. M., Hubbard, R. G., & Petersen, B. C. (1988). Investment and finance reconsidered. *The Brookings Papers on Economic Activity*, *19*(1), 141–195.

Federal Reserve. (2011). Supervisory guidance on model risk management. Board of Governors of the Federal Reserve System, Office of the Comptroller of the Currency, SR Letter 11-7.

Feldman, R. J. (1997). Small business loans, small banks and a big change in technology called credit. *The Region*, *11*, 19–25.

Filomeni, S., Modina, M., & Tabacco, E. (2023). Trade credit and firm investments: Empirical evidence from Italian cooperative banks. *Review of Quantitative Finance and Accounting*, *60*, 1099–1141. https://doi.org/10.1007/s11156-022-01122-3

Filomeni, S., Udell, G. F., & Zazzaro, A. (2020). Communication frictions in banking organizations: Evidence from credit score lending. *Economics Letters*, *195C*, 109412.

Filomeni, S., Udell, G. F., & Zazzaro, A. (2021). Hardening soft information: Does organizational distance matter? *The European Journal of Finance*, *27*(9), 897–927.

Frame, W. S., Srinivasan, A., & Woosley, L. (2001). The effect of credit scoring on small-business lending. *Journal of Money, Credit and Banking*, *33*, 813–825.

Gadenne, D., & Iselin, E. R. (2000). Properties of accounting and finance information and their effects on the performance of bankers and models in predicting company failure. *Journal of Business Finance and Accounting*, *27*(1–2), 155–193.

Gaudio, B. L. D., Griffiths, M. D., & Sampagnaro, G. (2020). Soft information production in SME lending. *The Journal of Financial Research*, *43*(1), 121–151.

Gropp, R., & Guettler, A. (2018). Hidden gems and borrowers with dirty little secrets: Investment in soft information, borrower self-selection and competition. *Journal of Banking and Finance*, *87*, 26–39.

Hannan, T. H. (2003). Changes in non-local lending to small business. *Journal of Financial Services Research*, *24*, 31–46.

Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. *Jama*, *247*(18), 2543–2546.

Henebry, K. L. (1997). A test of the temporal stability of proportional hazards models for predicting bank failure. *Journal of Financial and Strategic Decisions*, *10*(3), 1–11.

Hillegeist, S., Keating, E., Cram, D., & Lundstedt, K. (2004). Assessing the probability of bankruptcy. *Review of Accounting Studies*, *9*(1), 5–34.

Kalbfleisch, J. D., & Prentice, R. L. (2002). *The Statistical Analysis of Failure Time Data* (2d ed.). New York: John Wiley & Sons.

Kraft, P. (2015). Rating agency adjustments to GAAP financial statements and their effect on ratings and credit spreads. *The Accounting Review*, *90*(2), 641–674.

Lane, W. R., Looney, S. W., & Wansley, J. W. (1986). An application of the cox proportional hazards model to bank failure. *Journal of Banking and Finance*, *20*, 835–852.

Li, F. (2010). Textual analysis of corporate disclosures: A survey of the literature. *Journal of Accounting Literature*, *29*, 143–165.

Liberti, J. M., & Mian, A. R. (2009). Estimating the effect of hierarchies on information use. *Review Financial Studies*, *22*, 4057–4090.

Liberti, J. M., & Petersen, M. A. (2019). Information: Hard and soft. *The Review of Corporate Finance Studies*, *8*(1), 1–41.

Lin, S.-M., Ansell, J., & Andreeva, G. (2012). Predicting default of a small business using different definitions of financial distress. *The Journal of the Operational Research Society*, *63*, 539–548.

Loughran, T., & Mcdonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance*, *66*(1), 35–65.

Louzada, F., Ara, A., & Fernandes, G. B. (2016). Classification methods applied to credit scoring: Systematic review and overall comparison. *Surveys in Operations Research and Management Science*, *21*, 117–134.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, *29*, 449–470.

Mester, L. (1997). What's the point of credit scoring? *Federal Reserve Bank of Philadelphia Business Review*, 3–16.

Myers, S. C., & Majluf, N. S. (1984). Corporate financing and investment decisions when firms have information that investors do not have. *Journal of Financial Economics*, *13*, 187–221.

Naili, M., & Lahrichi, Y. (2022). The determinants of banks' credit risk: Review of the literature and future research agenda. *International Journal of Finance and Economics*, *27*, 334–360.

Oesterreichische Nationalbank (OeNB). (2004). *Guidelines on credit risk management: Rating models and validation*. Cooperation with the Financial Market Authority (FMA).

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, *18*, 109–131.

Petersen, M. A., & Rajan, R. G. (1994). The benefits of lending relationships: Evidence from small business data. *Journal of Finance*, *49*, 3–37.

Petersen, M. A., & Rajan, R. G. (2002). Does distance still matter? The information revolution in small business lending. *Journal of Finance*, *57*, 2533–2570.

Qian, J., Strahan, P. E., & Yang, Z. (2015). The impact of incentives and communication costs on information production and use: Evidence from bank lending. *Journal of Finance*, *70*, 1457–1493.

Ramakrishnan, R. T. S., & Thakor, A. (1984). Information reliability and a theory of financial intermediation. *Review of Economic Studies*, *51*, 415–432.

Rikkers, F., & Thibeault, A. E. (2009). A structural form default prediction model for SMEs, evidence from the Dutch market. *Multinational Finance Journal*, *13*(3/4), 229–264.

Rodriguez Gonzalez, M., Basse, T., Kunze, F., & Vornholz, G. (2018). Early warning indicator systems for real estate investments: Empirical evidence and some thoughts from the perspective of financial risk management. *ZVersWiss*, *107*, 387–403.

Roy, P. K., & Shaw, K. (2021). A credit scoring model for SMEs using AHP and TOPSIS. *International Journal of Finance and Economics*, *2021*, 1–20.

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business*, *74*, 101–124.

Somers, R. H. (1962). A new asymmetric measure of association for ordinal variables. *American Sociological Review*, *27*(6), 799–811.

Thomas, L., Edelman, D., & Crook, J. (2002). Credit scoring and its applications. In *Mathematical modeling and computation*. Society for Industrial and Applied Mathematics.

Tinoco, M. H., & Wilson, N. (2013). Financial distress and bankruptcy prediction among listed companies using accounting, market and macroeconomic variables. *International Review of Financial Analysis*, *30*, 394–419.

Vassalou, M., & Xing, Y. (2004). Default risk in equity returns. *Journal of Finance*, *59*, 831–868.

Volk, M. (2012). Estimating probability of default and comparing it to credit rating classification by banks. *Economic and Business Review*, *14*, 299–320.

Xia, Y., He, L., Li, Y., Liu, N., & Ding, Y. (2020). Predicting loan default in peer-to-peer lending using narrative data. *Journal of Forecasting*, *39*, 260–280.

Xu, D., Zhang, X., & Geng, H. (2018). Generalized fuzzy soft sets theory-based novel hybrid ensemble credit scoring model. *International Journal of Finance and Economics*, *24*(2), 903–921.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Filomeni, S., Bose, U., Megaritis, A., & Triantafyllou, A. (2024). Can market information outperform hard and soft information in predicting corporate defaults? *International Journal of Finance & Economics*, *29*(3), 3567–3592. https://doi.org/10.1002/ijfe.2840