Quarterly Review of Economics and Finance 89 (2023) 254–268

Contents lists available at ScienceDirect

Quarterly Review of Economics and Finance

journal homepage: www.elsevier.com/locate/qref

Predicting SMEs' default risk: Evidence from bank-firm relationship data

Michele Modina [a], Filomena Pietrovito [b,*], Carmen Gallucci [c], Vincenzo Formisano [d]

[a] University of Molise, Italy
[b] University of Molise, Department of Economics, Via F. de Sanctis, snc, 86100 Campobasso, Italy
[c] University of Salerno, Italy
[d] University of Cassino e Lazio Meridionale, Italy

ARTICLE INFO

Article history:
Received 23 October 2022
Received in revised form 21 April 2023
Accepted 25 April 2023
Available online 28 April 2023

Keywords:
Bank-firm relationship
Credit risk
Default prediction
Lending process
SMEs

JEL classification:
G20
G21
G33

ABSTRACT

This paper uses a probit model on a unique dataset of 13,081 Italian firms and 111 co-operative banks involved in the lending process to provide empirical evidence suggesting that the use and violations of credit lines and long-term loan overruns predict one-year and two-year probability of default (PD). The analysis controls for balance sheet indicators and time varying bank characteristics, captured by bank-time fixed effects. When combined with accounting data, credit-related indicators obtained from private internal banking sources improve small and medium-sized enterprises' (SMEs) default prediction. The marginal benefit of the bank-firm specific information is also assessed by comparing the default prediction accuracy of a model that incorporates accounting information with that of a full model including private information. In terms of heterogeneity, the association between the balance sheet indicators and data on bank-firm relationships and default probability can vary across sectors and geographies. This highlights the importance for banks of specific analysis to better assess risk at the firm level.

1. Introduction

Small and medium-sized enterprises (SMEs) significantly contribute to the economic development of the European industrial system. They account for two-thirds (65 %) of the European Union (EU-27) employment. In terms of value added, in 2022, SMEs contributed to the European economy with approximately 3.9 trillion euros (European Union, 2022; Statista, 2022). Compared to large firms, SMEs do not benefit from multiple sources of funding and access to bank credit remains their main source of financing. As a consequence, the construction of a robust relationship with the banking system is one of the bases of SME financing policy (Berger & Udell, 1995, 2002; López-Espinosa, 2017; Petersen & Rajan, 1994a).

The importance of this relationship is reciprocal, since operations with SMEs are the main credit activity of banks, particularly in the local credit markets (Agostino et al., 2011).[1] The main determinant of a bank's net interest margin is the granting of credit to businesses. The fundamental steps in the lending process include screening borrowers, assessing their creditworthiness, and subsequent monitoring of the bank-firm relationship. In this framework, default prediction models have received interest from many parties, including firms, banks and practitioners, and academics investigating lending decisions, monitoring credit riskiness and pricing loans (Doumpos et al., 2015; Duan et al., 2018).

The traditional literature has examined the contribution of various economic and financial aspects of a firm (e.g., leverage, cost of debt, working capital, asset turnover and profitability) in estimating default models, adopting accounting information extracted from balance sheets (Altman, 1968; Beaver, 1966). However, using accounting information alone is not especially informative for SMEs. In

[1] In Italy, cooperative credit banks are the most representative form of local banks (Bank of Italy, 2021), operating in 2640 Italian municipalities (where, in 620 cases, they act as the sole intermediary). Over half (59 %) of their assets are destined for loans to families and firms (six percentage points higher than other banks) and the development of territories (for every 100 euros of savings collected in the territory, 87 become credit to the real economy of that territory). During the years of the crisis, their market shares increased and, in the period December 2012 - December 2017, they introduced net loans of 8.2 billion euros into economic circulation (compared to an overall reduction recorded in the credit market).

* Corresponding author.
E-mail address: filomena.pietrovito@unimol.it (F. Pietrovito).

https://doi.org/10.1016/j.qref.2023.04.008

addition to traditional balance sheet data and financial ratios (Ciampi et al., 2021), it is necessary to build default prediction models based on other information, such as data related to the bank relationship (Bartoli et al., 2013; Bergerès et al., 2015; Chen et al., 2013; Neuberger & Räthke-Döppner, 2015; Smondel, 2018). This refers to the private information mainly related to specific services provided to the firm, which can be used to predict expected performance in repaying loans: utilization of credit and debit cards, credit lines, checking accounts, investment portfolios, saving accounts and long-term loan performance. The relevance of this private information varies across institutional contexts but it becomes particularly relevant in national contexts (Aghabarari et al., 2021; Cororaton, 2019) such as Italy, where smaller banks mostly reap the informational benefits of relationship lending (Buffa et al., 2019), making their credit offering unique (Becchetti et al., 2016). These banks' clients mostly encompass financially marginalized households, cooperatives and SMEs in poorer regions (Wheelock & Wilson, 2011).

Although the role of credit information has acquired relevance in the default probability estimation of banking models, the use of credit data in the bank-firm relationship has been largely overlooked in the literature relating to the estimate of corporate insolvency. Notable work includes the seminal paper by Nakamura (1993) and more recent contributions by Mester et al. (2007), Jiménez et al. (2009a), (2009b), Norden and Weber (2010), Belaid et al. (2017), Yildirim (2020) and Moscatelli et al. (2020). Other than this work, a limited number of authors, both in the banking and business literature, have studied the link between bank-firm specific information and default performance. This research gap is important, particularly given the growing body of literature that recognizes the critical role soft information plays in relationship lending and, more broadly, in the information generation processes of banks (Álvarez-Botas & González, 2023; Filomeni et al., 2020, 2021; Zhao et al., 2021).

The present paper contributes to this debate by providing empirical evidence on how SMEs default prediction models can benefit from the addition of relationship lending variables, a topic that has received limited scholarly scrutiny (Ciampi et al., 2021). We deal with the use of bank-firm specific information in predicting corporate default in a model that includes firm-level balance sheet information and other firm-level characteristics. We adopt a unique and disaggregated dataset with a sample of 111 cooperative banks lending to 13,081 operating firms in Italy. The share of defaulters in the sample is 1128 firms (about 9 %). A particular feature of our data is that the bank-firm relationship concerns several banks and two different relationships: short-term contracts (checking accounts) and long-term contracts (long-term loans), typically offered in a bundle that also includes a line of credit. Our starting point is to examine the relationship between the checking accounts and default rates by looking at bank credit utilization, credit limit violations, blank cheques and credit transactions. Next, we analyse the association of the default with long-term loans, using the number of consecutive months of overruns and the amounts of overruns.

Two sets of results are obtained. First, the use of the credit line and the depth of violations of the credit limit on both checking accounts and long-term loans can help the bank to predict the probability of default of a borrower, at least one year before it occurs. and this relationship is persistent over time. Second, the results also show that private information helps to increase the predictive power of balance sheet information by 8 % one year and two years before the default event. These results suggest that SMEs should carefully monitor their creditworthiness by paying attention to their relationship with their bank in terms of performance achieved on past short- and long-term loans. From this perspective, financial planning plays a fundamental role in correctly managing short-term credit lines and in servicing long-term debt.

To the best of our knowledge, this paper is the first to examine this issue with a large dataset. The paper therefore contributes to the literature in three ways. First, it analyses the marginal benefit of bank-firm specific information by comparing the default prediction accuracy of a model that incorporates only accounting information provided by balance sheets, controlling for firm size and the share of loans granted by each co-operative bank, with that of a model also including checking accounts and long-term loan performance. Second, it covers a sample of SMEs, heterogeneous in size and strength of relational banking. This makes our study highly representative of firms that are mainly dependent on banks as relationship lenders. It also allows us to account for the potential heterogeneity of the effect of bank-firm information, depending on firm characteristics (e.g., firm size and age; sector of economic activity and geographical areas) and on relationship lending (e.g., length of the bank-firm relationship and distance). Third, since the data come from more than 100 different local banks and concern the post-financial crisis period, we control for time-varying bank characteristics by including bank-time fixed effects.

2. Literature review

Over the past decade, there has been renewed academic interest in assessing the creditworthiness of SMEs. The interest in the design of specific prediction models for SMEs, mainly based on accounting information (Dullmann & Koziol, 2013), has been driven by Basel II, with the introduction of banking capital requirements to better determine the capital adequacy and to detect early warning signals in their loan portfolio (Altman & Sabato, 2007; Ciampi, 2015; Duarte et al., 2018), and the effects of the financial crisis on firms' defaults.

Traditional empirical models for predicting default for SMEs use financial indicators, extracted from balance sheets, to classify defaulting and non-defaulting firms and to assign a probability of default to firms over a given time horizon. From an empirical point of view, models based on accounting factors are inherently constrained, since it is not clear how well they perform out-of-sample, e.g., in terms of time, firms and sector of economic activity (Grunert et al., 2005). Moreover, the opaqueness of information contained in financial indicators is generally higher for SMEs than for large firms and therefore less effective at predicting corporate default (Altman & Sabato, 2007; Ciampi, 2015; Keasey & Watson, 2006; Norden & Weber, 2010).

There is a wide range of evidence highlighting the ability of soft information to predict future business dynamics and prospects (Brunner et al., 2000; Berger and Udell, 2002; Grunert et al., 2005; Chen et al., 2013; Matthias et al., 2019). Some of this work has shown how the effectiveness of hard information (data and ratios inferable from financial statements) in predicting insolvency can vary significantly depending on soft information represented by environmental, social and governance aspects (Gabbi et al., 2020; Hasan et al., 2022). Bredart (2014), Ciampi (2015) and Ballester et al. (2020) indicate that the combination of financial data and corporate governance variables increases the capacity of bankruptcy prediction models.

Some studies have focused on the soft information that can be gleaned from bank-firm relationship or credit behavioural indicators (Jiménez et al., 2009a, 2009b; Mester et al., 2007; Moscatelli et al., 2020; Xia et al., 2020). Nakamura (1993) was one of the first to highlight the benefits of using bank transactions with particular reference to checking accounts, typically offered as a bundle with a line of credit. In its broadest definition, a credit line refers to the maximum amount of financial resources that a bank will lend to a borrower over a given period. Borrowers can use the loan up to the maximum amount by paying interest only on the sums actually used. The informative value of the checking account was confirmed by Mester et al.'s (2007) analysis of a set of annual and monthly data in a sample of Canadian companies during the period 1988–1992. The movement of bank accounts is correlated with the dynamics of receivables from customers; an anomalous use of credit lines is an anticipatory signal of cash tension which often leads to credit downgrades. Using the credit register database for Spanish companies, Jiménez et al. (2009a, 2009b) highlighted that defaulting firms experience more intense and rigid use of their credit line up to five years before the predefined event occurs. Analysing data from four German banks, Grunert et al. (2005) confirmed the hypothesis that the combined use of accounting data and bank-firm information improves the estimate of future insolvency events compared to individual use of each factor. Xia et al. (2020) investigated peer-to-peer lending using a soft information extraction technique based on keyword clustering. They showed that variables extracted from narrative data are powerful and that using narrative data significantly improves the predictability of default probability compared to using only hard information.

Therefore, additional analyses can strengthen the direct examination of the sources and types of information that financial institutions use to monitor credit (Matthias et al., 2019). It is also useful to run a comparative analysis of the combined use of financial variables and information on credit reports on a large sample of SMEs (Roggi et al., 2013). For financial institutions, soft information represents subjective knowledge accumulated over time during repeated interactions with borrowers. The accumulation of this soft information results in the conversion of this subjective information into a quantifiable input that can be used to evaluate the borrower in different dimensions. In sum, a growing body of literature recognizes soft information as a critical element in the information generation processes of banks (Filomeni et al., 2020, 2021; Zhao et al., 2021).

Theoretical literature shows that banks have an advantage in jointly providing collection and lending services. This is because they can access factual information from different sources specific to their relationship with a given firm long before they receive the firm's financial statements (Norden & Weber, 2010; Nakamura & Roszbach, 2018). Information spillovers coming from loans and checking account activities might help a bank in the credit decision and monitoring process (Boyd & Prescott, 1986; Diamond, 1984, 1991; Ramakrishnan & Thakor, 1984). The literature emphasizes the role of information-intensive lending relationships between borrowers and lenders as a major aspect of reducing information asymmetries, restoring efficiency by establishing long-term implicit contracts between borrowers and lenders (Fischer, 2000; Greenbaum et al., 1989; Petersen & Rajan, 1994b, 1995; Sharpe, 1990).

Following the empirical work of Petersen and Rajan (1994b) and Berger and Udell (1995), a large body of literature has identified relationship variables such as length, distance, intensity or exclusivity to characterize relationship lending (Boot, 2000; Cole et al., 2004; López-Espinosa et al., 2017). Berger and Udell (2002) point out the importance of relationships in the lending process to SMEs and in assessing their creditworthiness. The importance of the relationship emerges most clearly in small banks (such as cooperative credit banks) which, given their proximity to the firm, are better able to process non-transparent or opaque information than large banks that interact more impersonally with borrowers (Bartoli et al., 2013; Hussain et al., 2021; McKillop et al., 2020; Stein, 2002). The informational advantage translates into favouring relational lending with a reduction in information asymmetries between lenders and borrowers and, consequently, better ability to monitor borrowers and improve the power of insolvency prediction models (Beck et al., 2006; Berger & Udell, 1995; Berger et al., 2014; Chen et al., 2013; Elsas, 2005; Howorth & Moro, 2012; Petersen and Rajan, 1994b). The main statistical tools in this field include linear and multiple discriminant analysis (Altman, 1968, 1994; Beaver, 1966), and logistic regression models (Altman & Sabato, 2007; Bottazzi et al., 2011). The default prediction methodologies that use discriminant analysis to predict corporate defaults are based on the a priori assumption that the differences between two mutually exclusive groups of firms (defaulters and non-defaulters) can be captured by observing individual financial ratios (Foglia et al., 1998) or by combining them (Altman & Sabato, 2007). Bonaccorsi di Patti et al. (2015) showed the relevance of financial variables as determinants of corporate default. In particular, weak capital structure, characterized by a high level of bank loans, increases default rates and the sensitivity of the default probability to macroeconomic shocks.

For our study, the binomial probit regression model is the most suitable tool to capture the signs of difficulty that come from accounting data and from the financial data collected by the banks from which SMEs request financing (so-called bank-firm information). Adding and combining variables, both of a qualitative (soft information) or financial nature, tends to improve the accuracy of forecasting models. However, collecting such data is not easy when dealing with SMEs and their reliability is often questioned (Luo et al., 2020). In particular, bank-firm information is kept by the banks who are very reluctant to grant external access. The uniqueness of our dataset, consisting of internal banking data on credit relationships between more than 100 community banks and thousands of SMEs, allows for a set of highly informative data of fundamental importance to offer a significant contribution to the existing literature.

Most of the previous work cited in the present paper has considered sample periods pre-financial crisis. Our paper uses post-crisis data focusing on the three years after the crisis. Every financial crisis has important repercussions for the bank-firm relationship. The general rise in uncertainty over economic growth and the expansion of information asymmetry between banks and borrowers increase the risk of incorrect assessments of companies to be financed. Thus, it is worth verifying the accuracy of models based on pre-crisis data, constructed under stable economic conditions, using post-crisis data.[2] Norden and Weber (2010) investigated whether, by monitoring transactions, a bank can obtain more timely information on a borrower's cash flow. They showed that monitoring the credit line and violations of its use up to 12 months before insolvency is a useful way to anticipate gradual deterioration of a borrower's credit rating. Using credit history data for a sample of Italian firms, Dainelli et al. (2013) confirmed the contribution of the utilization level of short-term lines of credit to increase the prediction power of a model including only financial information. However, although both of these papers consider different borrower categories, they are based on data from one bank and do not consider other contracts between the bank and the firm. The origin of the data from a single bank and the use of variables only related to checking accounts limit the generalizability of the results to the banking system as a whole, or to other samples, and do not fully elucidate the depth and intensity of the bank-firm relationship.

Not all previous studies have shown a positive effect of using the credit line to assess probability of default. Bergerès et al. (2015) found that an increase in the utilization of credit lines was associated with a decrease in probability of default. The use of credit lines can allow companies in periods of temporary liquidity shortage to honour their financial commitments. These works consider the impact of bank-firm information mainly on the basis of use of the credit line, neglecting the information contained in medium- or long-term loans. Strahan (1999) noted that the typical bank-firm relationship should include both a term loan and a short-term credit line in order to give a more complete picture of the borrower's

[2] Recent work has focused on the association between the COVID-19 pandemic and default probability. For example, Mirza et al. (2020) investigated the effect of COVID-19 on the solvency profile of firms in the EU member states, finding that the solvency profile deteriorated, particularly for firms in the manufacturing, mining and retail sectors.

financial behaviour. Information on both credit instruments should be combined to maximize the soundness of the default risk assessment. Our in-depth study of credit-related variables related to firms' banking transactions coming from private internal banking sources combined with accounting data leads to improved SME bankruptcy predictions.

3. Research methodology

3.1. Data and variables

The data analysed combine bank-firm specific information on checking account activities and long-term loans with firm-level financial information gathered from balance sheets. These data were provided by Centrale Rischi Finanziari (CRIF; an Italian credit rating agency) and have been produced by Centro Servizi Direzionali (CSD; a data provider of the co-operative banks). To be included in our sample, firms needed to satisfy three criteria. First, we include firms with an average turnover in the range of 5000–50,000,000 euros, having at least a checking account or a long-term loan over the period 2012–2013 and with reliable balance sheet ratios. Second, following Arcuri and Levratto (2018), firms operating in the financial and insurance, real estate, public administration, education, social and human health sectors were excluded because they may be subject to particular failure regimes.[3] Third, we included multiple lending firms in the banking system to be able to incorporate as a control variable the share of loans granted by each co-operative bank with respect to the total amount of loans received by a specific firm from the banking system.

We developed the empirical model from the analysis sample and tested this model on the validation sample, composed of different SMEs. Therefore, we developed two different sub-samples: the analysis sample, including 70 % of firms and the validation sample, including the remaining 30 %. Firms were extracted without replacement using a proportional stratified random sampling method based on four characteristics: default status (non-default and default, over the sample period), the province where firm operates, the sector of economic activity and average turnover.[4] Through this method, final units of the sample were randomly and proportionally chosen from different non-overlapping strata, so that 70 % came from each stratum to compose the analysis sample and 30 % to compose the validation sample. Finally, we excluded influential observations, which are those where values of each variable included in the analysis are higher than the 99th percentile. This procedure resulted in a final analysis sample of 13,081 firms and 23,205 firm-bank-year observations, with 56 % of firms observed both in 2012 and in 2013. This allows us to adopt two different time lags to predict defaults, one and two years before the event. For the event variable, 8.7 % of firms became insolvent during either 2013 or 2014. We used an initial long list of 60 financial indicators describing several areas of performance (e.g., leverage, profitability, liquidity and efficiency). Specific short-term, bank-firm information was contained in an initial set of 15 indicators on checking account activities (e.g., credit limit violations and utilization of the line of credit, debit and credit accounts). Long-term loan performance was described by 11 indicators (e.g., overdue payments and credit limit violations).

Since the initial set of indicators showed very high correlation coefficients, it was reduced by a two-step selection process (Ciampi, 2015). The first step derives from the variance inflation factor (VIF) model, which allows us to select only those variables unaffected by multicollinearity concerns in a linear regression model where the dependent variable is the default event.[5] Following Tinoco and Wilson (2013), only those ratios with a VIF of less than 5 are retained, thus reducing the set of financial indicators. The second step consists of estimating a stepwise regression on the probability of default, using the ratios selected in the first step as independent variables. The stepwise procedure helps to identify the best combination of significant explanatory variables in the regression and to include them in the empirical model (Shin & Lee, 2002; Shin et al., 2005). We adopted the backward selection method beginning with the model including all variables and iteratively eliminating non-significant variable. We used the 1 % level of significance for the addition of variables to the model and 5 % for their removal. At the end of the selection process, we obtained a model in which all explanatory variables proved to be jointly significant.

The final set of balance sheet ratios included three main measures and indicators: the return on assets (*roa*), defined as earnings before interests and taxes (EBIT) on total assets, as a measure of profitability; the share of owner's equity over owner's equity and inventories (*equity inventory coverage*), as an indicator of both capitalization and inventory incidence; the share of bank loans on total liabilities (*bank loans*), as a measure of the firm's debt. In addition, four attributes reflecting checking accounts and credit line performance were selected: the share of credit line received on a checking account that is used by the borrower in a month (*credit line utilization*); the number of months in a year of credit limit violations on a credit line (*credit limit violation*); the number of blank cheques that are not paid in the same day when they are shown to the bank for negotiation (*blank cheques*), signalling an initial potential liquidity difficulty; the number of crediting operations on the checking accounts (*crediting operations*). Finally, we selected two variables concerning long-term loans: the number of months of credit limit violations (*long-term overruns*) – i.e., the number of months in which there is a delay in payment of the loan instalment – and the share of limit violations on long-term credit granted in a quarter, balanced by other credit products of the firm with the same bank (*long-term balanced overruns*). *Turnover* refers to the annual amount of revenue, expressed in thousands. Summary statistics are reported in Table 1, distinguishing between defaulted and non-defaulted firms.

Significant differences were found between the two sub-samples on balance sheet indicators. The average *roa* indicator shows that profits are about 0.81 % of total assets, with a positive value for non-defaulted firms (0.95 %) and a negative average in the sub-sample of defaulted firms (−0.80 %). The average values of *equity inventory coverage* indicate that non-defaulted firms show a better coverage of inventories than defaulted firms (56 % vs. 43 %). On average, firms in the sample make large use of bank loans, which represent 35 % of their total liabilities, with higher values for defaulted firms than for non-defaulted firms (45 % vs. 34 %). These indicators are also statistically different between the two groups.[6]

[3] For example, credit cooperatives are very important lenders for small businesses in the agricultural sector in many advanced and emerging economies (e.g., in the rural parts of the US and several Latin American countries) (Aghabarari et al., 2021; Cororaton, 2019). This is also the case in Italy, as has been highlighted in previous literature (Galluzzo, 2012).

[4] We consider three different size groups: firms with an average turnover below 2,000,000 euros, firms with an average turnover between 2,000,000 and 5,000,000 euros and firms above 5,000,000 euros.

[5] The VIF method estimates how much the variance of an estimated regression coefficient is inflated because of linear dependence with other predictors. Computationally, it is defined as the reciprocal of tolerance – i.e., 1/(1−R2) (Ciampi, 2015). Its utility is that it indicates the magnitude of inflation in the standard errors associated with a beta weight that is due to multicollinearity. Other methods used to naturally overcome the multicollinearity problem include LASSO (Tian et al., 2015).

[6] Based on data from the ECB's Survey on the Access to Finance of Enterprises (SAFE), about 82 % of Eurozone SMEs use debt financing to fund their activities or investments. In the US, according to the latest available Annual Business Survey (ABS) data, 15 % of businesses formally applied for new credit in 2020, excluding applications for pandemic-related programmes. Data from the Federal Deposit Insurance Corporation show a 29 % year-on-year increase to $833 billion in 2020 from the $645 billion recorded in 2019 (Board of Governors, 2022). In this scenario, loans and lines of credit are still the most popular credit products (about 80 %). The percentage of

Table 1
Summary statistics.

	default firms				non-default firms				
	mean	cv	min	max	mean	cv	min	max	t-test
default indicators									
default (dummy)	0.67	0.70	0	1.00					
balance sheet indicators									
roa (share)	-0.80	-4.77	45.07	22.46	0.95	4.98	-62.24	25.39	18.46***
equity inventory coverage (share)	43.38	0.75	0.07	100	55.76	0.57	0.00	100	15.68***
bank loans (share)	44.69	0.40	0.02	95.02	34.27	0.56	0.00	95.28	-23.86***
firm characteristics									
turnover (thousands)	4983	1.41	5	48,700	4956	1.49	7	49,600	-0.15
system information									
share of loans (second tercile)	0.46	1.07	0	1	0.57	0.87	0	1	8.67***
share of loans (third tercile)	0.12	2.76	0	1	0.15	2.42	0	1	3.78***
credit line and checking account									
dummy checking account	0.98	0.13	0	1	0.99	0.11	0	1	1.04*
credit line utilization (share)	84.53	0.47	0	189.80	41.76	1.05	0	189.19	-43.80***
credit limit violation (months)	2.06	1.05	0	10	0.70	1.90	0	10	-26.46***
blank cheques (number)	0.14	3.56	0	3	0.30	7.65	0	3	-9.34***
crediting operations (number)	13.93	1.59	0	189	17.69	1.50	0	189	6.85***
long-term loan									
dummy long-term loan	0.45	1.10	0	1	0.39	1.25	0	1	-5.02***
long-term overruns (months)	0.48	2.18	0	5	0.11	3.99	0	5	-15.18***
long-term balanced overruns (share)	0.06	3.59	0	3.25	0.01	7.14	0	2.46	-10.08***
relationship lending									
relationship length (years)	9.80	1.06	0	113	10.57	1.05	0	114	2.90***
distance (minutes)	27.63	1.49	0	596.75	23.63	1.33	0	533.21	-3.97***

Notes: *t-test* indicates the value of the mean-difference test where H_0: mean (*non-default*) – mean (*default*) = 0. The approximate degrees of freedom for the *t-test* are obtained from Welch's formula (1947). * indicates significance at the 10 % level; * * indicates significance at the 5 % level; * ** indicates significance at the 1 % level.

In our analysis, monthly credit line utilization is at about 45 % across the entire sample and, as expected, for defaulted firms it is double that of non-defaulted firms (85 % vs. 42 %).[7] In both sub-samples, firms violate their credit lines, since the maximum credit line utilization is about 190 %. The latter values differ from those reported in the extant literature relative to the United States (US) context. For example, Acharya et al. (2013) provided evidence that the average percentage of unutilized credit lines over the overall corporate liquidity is about 45 % for US firms, equal to an average of credit line utilization of 55 %. Moreover, albeit broadly, the significant differences observed in the credit line usage of defaulted and non-defaulted firms corroborate the results of Acharya et al. (2020), suggesting that substantial drawdowns from credit lines precede covenant violations. This is a sign of the struggle experienced by a firm in its credit relationships with banks prior to a more severe default. Instead, our data on credit line usage for non-defaulted firms are mostly in line with those of Cingano et al. (2016), who reported an average ratio of 48.21 % between drawn and granted credit. However, the values presented here are significantly more subject to variability, as can be seen by comparing the maximum value of the overall credit line usage reported in Table 1 (189.19) with that of Cingano et al. (2016) (108.13).

The average number of months of credit line violation in checking accounts is about 1, with defaulted firms showing three times the value of non-defaulted firms. We found significant differences between the two sub-samples in terms of the number of blank cheques (0.14 for defaulted firms vs. 0.03 for non-defaulted firms). Finally, the number of crediting operations is higher for non-defaulted firms (18) than for defaulted firms (14). Concerning long-term overruns, the number of months of overrun is significantly higher for defaulted firms (0.49 vs. 0.11), as well as the share of balanced overrun (0.06 vs. 0.01). Concerning the control variables, firm size, measured by turnover, shows statistically higher average values for defaulted firms (4,983 thousand euros) than for non-defaulted firms (4,956 thousand euros).

Banking system information is contained in the share of loans granted by one of the co-operative banks of our sample to the average loans granted to the same firm by the banking system as a whole. Since this variable has a very skewed distribution, we chose to create three dummies for each tercile of the distribution of the share. We used the two dummies for bank-firms in the second and third tercile as control variables (see Fig. A1 in the Appendix). Our specification also controls for two dummies indicating whether the firm has a checking account and a credit line, and whether it has a long-term loan with the bank. The average values of these dummies indicate that about 99 % of firms in both sub-samples have a checking account with the bank, and about 45 % of defaulted and 39 % of non-defaulted firms have a long-term loan. Table 1 also shows descriptive statistics of variables used to split our sample of firms: the length of the bank-firm relationship, measured in years, and the geographical distance between the bank and the firm, measured in minutes. For non-defaulted firms, the average relationship length is 10.57 years, significantly longer than that of defaulted firms (9.80), and the average distance is 24 minutes, significantly closer than that of non-defaulted firms (28 minutes).

Table A1 in the Appendix reports the pairwise correlations between the six sets of variables. Defaulted borrowers have a lower return on assets and equity inventories coverage (−0.09), and higher bank loans (0.12). Concerning checking account activities, the probability of default is positively correlated with credit line utilization (0.22), credit limit violation months (0.23) and blank cheques (0.10), and negatively correlated with the dummy for the existence of a checking account (−0.01) and crediting operations (−0.04). Positive pairwise correlations are also found between the dummy default

(*footnote continued*)
applicants for loans or lines of credit in the US decreased between 2017 and 2021, while the percentage of applicants for credit cards, business credit and merchant cash advances remained stable or increased slightly (Board of Governors, 2022).

[7] Bank credit lines are a major source of financing for businesses and an important line of business for banks. For Italian firms, they account for an average of 45 % of corporate bank financing. They are defined as credit lines, bank overdrafts and credit card overdrafts and 26 % of SMEs used them in the past six months. Regarding the use of bank loans for 50 % of enterprises, they are a relevant form of financing and 16 % have used them in the past six months (European Commission, 2022). Looking at the short- and long-term bank-firm specific contracts, similar results are obtained for the two sub-samples.

and the long-term loan indicators: 0.20 for consecutive months of overruns and 0.16 with balanced overruns.

These simple correlations do not consider the interrelationships between variables and the fact that some firms show industrial and localization specificities different from others. We therefore developed our interrogation of the data with a multivariate analysis.

3.2. Econometric analysis

The present study used a probit model to study which factors at time t_0 (or t_0-12) influence the probability of default at time t_{0+12} (or t_0). In the first set of analyses, we investigated whether financial information gathered from balance sheets can be augmented by direct observations of bank-firm information to improve the prediction probability of default. We estimated a baseline pooled probit model, including only financial indicators as predictors, along with size and system information as controls, and two augmented models, including checking account and long-term loan performance. In the second set of analyses, we investigated whether these results are different for sub-samples of firms and for different bank-firm relationship lending. We aimed to assess, for instance, whether predictors exert a different relationship with default probability depending on firm size, age, sector and geographical area, and on bank-firm distance and relationship length. Our methodology followed several earlier studies (Dierkes et al., 2013; Duarte et al., 2018; Norden & Weber, 2010). The econometric approach relies on a binomial probit model where the dependent variable, *default*, is a solvency condition at time t_{0+12}. It assumes the value of 1 in cases where bank *b* registers one of three events for firm *f*: *unlikely-to-pay*, which is a temporary problem that could be remedied; *bad loan*, which is a non-performing status that has not been ascertained; a *forborne non-performing* loan for which the conditions (such as maturity and interest rate) have been re-negotiated with the firm. The dependent variable assumes the value of 0 when the firm does not present any anomaly in the service of the bank debt (the so-called *performing loan*).

The estimated *baseline model* can be summarized as follows:

$$\Pr(default^{fb}_{t_0+12}=1)=\Phi(\beta_1 financial^{f}_{t_0}+\beta_2 size^{f}_{t_0}+\beta_3 system^{fb}_{t_0}+\eta^p+\lambda^s+\nu^b_{t_0}) \quad (1)$$

where subscripts *f*, *b*, *p*, *s* and *t* indicate firm (*f* = 1, …., 13,081), bank (*b* = 1, …, 113), province (*p* = 1, …, 56), sector (*s* = 1, …, 61) and time (t_0 = 2012, 2013), respectively. The set of financial variables includes: the return on assets (*roa*); the share of bank loans in total liabilities (*bank loans*); the share of net equity over net equity and inventories (*equity inventory coverage*), as described in Section 3. Our specification allowed us to control for potential endogeneity problems and omitted variables.[8] First, the *default* indicator is equal to 1 for any firm experiencing an event over the period t_{0+12}, whereas the financial information refers to t_0, which is 2012 for defaults during 2013 and 2013 for defaults during 2014. We can also control for firm size (Altman et al., 2017; Dierkes et al., 2013; Fiordelisi et al., 2014), proxied by turnover, and include two dummies for the second and third tercile of the share of bank loans. Finally, the specification includes three sets of fixed effects: province fixed effects, to account for local context characteristics affecting firm performance; fixed effects at the sector level, to account for unobserved, time-invariant heterogeneity across sectors; bank-year fixed effects, to control for banks' time-varying unobserved characteristics that might affect borrower behaviour. Using bank-year fixed effects, we therefore compared how firms' default probability with the same bank changes relative to firms with different financial indicators. Provided that this *within bank* specification fully absorbs bank-year characteristics, the estimated default probability can be attributed to firm differences. Finally, standard errors are clustered by firms.

The second model (*full model*) includes checking accounts and long-term loan relationships, along with balance sheets indicators and control variables:

$$\begin{aligned}\Pr\left(default^{fb}_{to+12}=1\right)=\Phi(&\beta_2 financial^{f}_{to}+\beta_2 size^{f}_{to}+\beta_3 system^{f}_{to}\\&+\beta_4 flag\_ca^{fb}_{to}+\beta_5 bank\_firm\_ca^{fb}_{to}\\&+\beta_6 flag\_ltl^{fb}_{to}+\eta^p+\lambda^s+\nu^{b}_{to}\end{aligned} \quad (2)$$

where the checking account set includes credit line utilization, credit line violations months, blank cheques and crediting operations. Finally, the long-term loans set includes long-term overruns and long-term balanced overrun. We also control for dummies indicating whether the firm has a checking account and a long-term loan with a bank. All other variables and dummies are defined as in Eq. (1).

To evaluate the contribution of short- and long-term bank-firm hard information in predicting the probability of default, we perform receiver operating characteristics (ROC) analysis and compare the areas under the ROC curve (AUC) of the baseline and full models, as an index of accuracy. ROC analysis has many advantages over alternative methods of evaluating a wide range of diagnostic systems, including credit risk assessments (Irwin & Irwin, 2013): it is well suited to empirical data, relies on solid theoretical bases in setting the optimal threshold and does not depend on the probability of default.[9]

4. Empirical results

This section discusses the results obtained from probit regressions of the default indicator on predictor variables. The probability of default is estimated in the year prior to the observation of corporate financial distress (t_0) and two years prior to the financial distress event (t_0–12). Results are reported in Table 2, distinguishing between the *baseline model* (Eq. 1), including firm level balance sheets indicators and control variables, the *augmented model* and the *full model* (Eq. 2), which includes all additional information on bank-firm contracts. As required by the probit regression model, the dependent variable takes the values of 1 for firms classified as defaulters at time t_{0+12} and 0 for firms classified as non-defaulters.

Table 2 shows marginal effects calculated as the variation of default probability after a variation of predictors from the value at the 25th percentile to that at the 75th percentile (or 90th, depending on the distribution of variables).

Column (1) of Table 2 shows that *roa* has a negative and significant association with the probability of default, which is expected. Its marginal effect indicates that an increase in firm profitability from the value at the 25th percentile (−0.23 %) to that at the 75th percentile (1.93 %) reduces the probability that a firm is affected by a default event in the subsequent 12 months by 0.6 %. Considering that the average default rate in the sample is about 8 %, this relationship is also economically significant. Concerning the

[8] In addition, in the additional analysis section, we report the results obtained instrumenting credit line utilization with firm-level variables.

[9] The ROC curve is a graph of the sensitivity versus 1-specificity of the diagnostic test. The sensitivity is the fraction of positive cases that are correctly classified by the diagnostic test, whereas the specificity is the fraction of negative cases that are correctly classified. In other words, the ROC curve plots the sensitivity (the true-positive rate) against the specificity (the true-negative rate). If a test could perfectly discriminate, the AUC is equal to 1; the closer the ROC curve to this ideal point, the better its discriminating ability. A test with no discriminating ability will produce a curve that follows the diagonal of the grid (DeLong et al., 1988).

Table 2
Financial indicators and bank-firm relationships predicting the default probability.

	t_0			t_0 - 12		
	(1) Baseline model	(2) Augmented model	(3) Full model	(4) Baseline model	(5) Augmented model	(6) Full model
roa (share)	-0.006***	-0.004***	-0.003***	-0.005***	-0.003***	-0.003***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
equity inventory coverage (share)	-0.025***	-0.014***	-0.014***	-0.026***	-0.016***	-0.017***
	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)	(0.005)
bank loans (share)	0.026***	0.014***	0.012***	0.026***	0.017***	0.015***
	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)
turnover (thousand)	-0.000	0.012***	0.011***	0.007*	0.015***	0.014***
	(0.002)	(0.002)	(0.002)	(0.004)	(0.004)	(0.004)
share of loans (second tercile)	-0.021***	-0.014***	-0.017***	-0.025***	-0.021***	-0.023***
	(0.003)	(0.003)	(0.003)	(0.006)	(0.006)	(0.006)
share of loans (third tercile)	-0.022***	-0.017***	-0.021***	-0.019***	-0.019***	-0.024***
	(0.004)	(0.004)	(0.004)	(0.007)	(0.007)	(0.007)
dummy checking account		-0.106***	-0.052***		-0.078	-0.050
		(0.025)	(0.019)		(0.049)	(0.043)
credit line utilization (share)		0.055***	0.053***		0.043***	0.042***
		(0.003)	(0.003)		(0.005)	(0.005)
credit limit violation (months)		0.010***	0.009***		0.011***	0.010***
		(0.001)	(0.001)		(0.001)	(0.001)
blank cheques (number)		0.032***	0.028***		0.059***	0.052***
		(0.006)	(0.006)		(0.015)	(0.015)
crediting operations (number)		-0.010***	-0.009***		0.001	0.001
		(0.002)	(0.002)		(0.004)	(0.004)
dummy long-term loan			-0.001			0.004
			(0.003)			(0.005)
long-term overruns (months)			0.070***			0.031***
			(0.013)			(0.015)
long-term balanced overruns (share)			0.014***			0.014*
			(0.005)			(0.007)
bank-time fixed effects	yes	yes	yes	yes	yes	yes
industry fixed effects	yes	yes	yes	yes	yes	yes
province fixed effects	yes	yes	yes	yes	yes	yes
Observations	23,205	23,205	23,205	8231	8231	8231
R^2	0.120	0.231	0.254	0.140	8231	0.223

Notes: This table reports marginal effect obtained estimating Eq. (1) on the full sample of firms. Marginal effects are calculated as the variation of the default probability after a variation of the predictors from the value at the 25th percentile to that at the 75th percentile (or 90th, depending on the distribution of variables). Standard errors, clustered at firm level, are reported in parentheses; * indicates significance at the 10 % level; ** indicates significance at the 5 % level; *** indicates significance at the 1 % level.

financial structure, we observe a positive correlation between *bank loans* and default probability. This indicates that, other things being equal, if a firm becomes more dependent on bank loans, its probability of default increases by 2.6 %. Therefore, this result shows that the financial structure is an important determinant of SME ability to repay loans. A decrease of the *equity inventory coverage* ratio from the level at the 75th percentile (85 %) to that at the 25th percentile (26 %) predicts a 2.5 % increase in default probability. Within current assets, inventories are the item that suffers from superior volatility, because of their intrinsic nature and the effects of accounting techniques. Whatever the reason for their volatility, greater coverage of inventories with net equity contributes to strengthening a firm's financial soundness and produces positive effects on its creditworthiness. In addition, firm size, other things being equal, is not statistically significant. Finally, introducing system information does not change the results of the overall model and provides additional information. In particular, firms that are more exposed with a specific bank are less likely to suffer a negative event in the next 12 months than firms that are less exposed with the same bank. In summary, controlling for firm size, system information and for the fixed effects described in Section 3, all balance sheet indicators prove to be economically relevant and highly statistically significant (at the 1 % level) in predicting the default event 12 months before it occurs. This result is consistent with the empirical literature summarized above.

Additionally, if we include the short-term relationship indicators in the baseline model (column 2), we obtain three relevant results. First, the correlation between balance sheet indicators and default probability is not affected by the bank-firm relationship, at least in terms of sign and significance, while in some cases the magnitude is reduced. On the contrary, the coefficient of firm size becomes positive and statistically significant in this specification, indicating that larger firms are more likely to experience a default event. In particular, when moving from firms in the 25th percentile (with an average turnover of 886,000 euros) to firms in the 75th percentile (with an average turnover of 5,395,000 euros) the default probability increases by 1.1 %.

Second, the correlation of all five checking account factors in the model (*dummy checking account*, *credit line utilization*, *credit limit violation*, *blank cheques* and *crediting operations*) and default is highly significant (at the 1 % level). Specifically, our results indicate that SMEs with a checking account show a probability of default 10.6 % lower than that of those that do not have a checking account. Moreover, increasing credit line utilization from the value at the 25th percentile (0 %) to that at the 75th percentile (90 %) produces an increase of about 5.5 % in default risk. In other words, a firm that is close to reaching its permitted credit limit is about 6 % more likely to experience difficulties in repaying its loan on a one-year horizon than a firm that does not use at all its credit line. Looking at the months of credit limit violations on checking accounts, we obtain a positive association with default probability: when increasing the number of consecutive credit limit violation months from 0 (the 25th percentile) to 1 (the 75th percentile), the probability of default increases by about 1 %. When the number of blank cheques changes from the 25th percentile (0) to the 99th percentile (1), the default probability increases by 3.2 %. Finally, crediting operations reduce

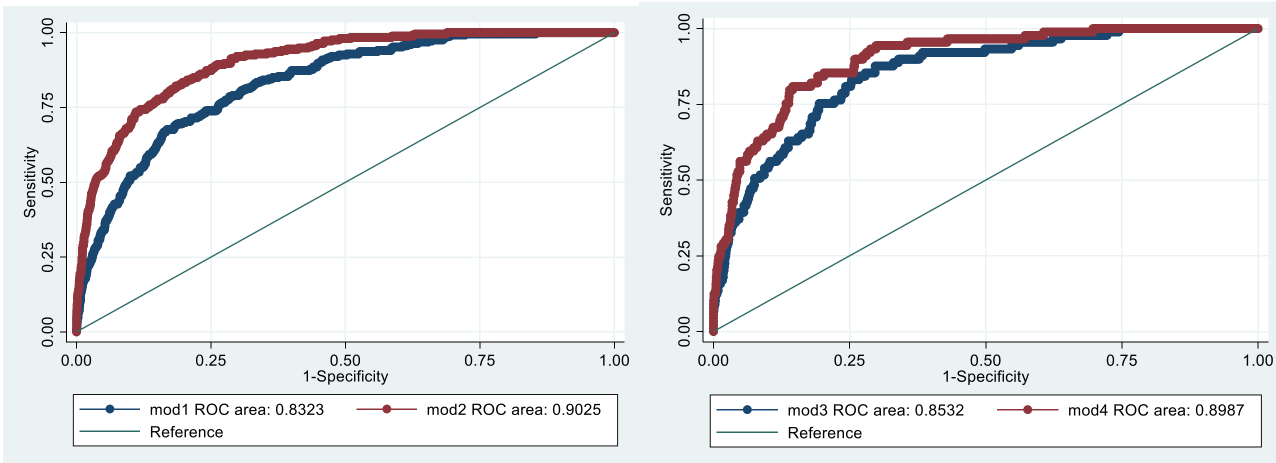

Fig. 1. ROC curve comparing baseline model and full model at time t_0 and t_0-*12*. *Notes*: This figure reports the comparison of areas under the receiver operating characteristics (ROC) curve of the *baseline model* and *full model* estimated at time t_0 (left hand side) and t_0-*12* (right hand side) on the *validation sample*. The figure plots the area under the ROC curve (AUC). The comparison is performed using the non-parametric method suggested by DeLong et al. (1988). Baseline model AUC is equal to 0.83 and full model AUC is equal to 0.90, at time t_0. Baseline model AUC is equal to 0.85 and full model AUC is equal to 0.90, at time t_0-*12*. *Sensitivity* is the fraction of positive cases that is correctly classified by the diagnostic test (true-positive rate), whereas *specificity* is the fraction of negative cases that is correctly classified (true-negative rate).

default probability by about 1 % when its value changes from 0 to 20. As a whole, these findings indicate that when controlling for quantitative information obtained from balance sheets, checking accounts can help a bank predict future SME defaults, consistent with the findings of Norden and Weber (2010).

Third, it should be noted that considering checking account activities helps to improve the accuracy of default predictions compared to a model that only includes balance sheets. Specifically, the R^2 values more than quadruple when additional checking account information factors are added to the baseline model. This means that a bank can obtain additional information on a firm's ability to repay loans when it considers checking account activity in addition to balance sheets indicators.

Column (3) of Table 2 shows the results obtained augmenting the previous specification with long-term loan factors. SMEs that obtain a long-term loan show a 0.1 % higher probability of default than those that do not obtain long-term loans, but this marginal effect less precise and not statistically significant. A higher number of consecutive months of overruns in a half year (from 0 to 3) increases the probability of default by 7.0 %, while the effect of the share of balanced overruns is lower, increasing the probability by 1.4 % (for a change from 0 % to 0.37 %). The last result indicates that even if a firm has long-term loan overruns offset by other relationships with the bank, its probability of defaulting is higher. This is because despite compensation, failure to repay a long-term loan is a sign that a firm is struggling to generate sufficient operating cashflow to service its debts. Other coefficients remain unchanged in terms of their economic effect and significance. In this case, the R^2 value increases to 25 %, demonstrating that the full model has an explanatory power that is about four times that of the baseline model.

We also considered a different prediction horizon for factors at time t_0-*12* influencing default at t_0, shown in columns 4–6 of Table 2. Interestingly, even though the number of observations reduces by about 70 %, our key finding — that the measures of checking accounts and long-term loans have a significant relationship with probability of default in addition to balance sheet factors, size and system information — remains true. This indicates that our key results are persistent over time. Reassuringly, marginal effects and significance remain almost unchanged in the full model.

AUC analysis offers a direct measure of the real performance of a probit model. Fig. 1 shows the comparison of AUCs of the baseline and full models. The baseline model at time t_0 provides an accuracy ratio of 83 %, while the full model stands at 90 %. The difference corresponds to the area between the two cumulative accuracy profiles (Dierkes et al., 2013). The same comparison provides an overall *p*-value of 0.000, which rejects the null hypothesis of equal AUCs. In other words, the small *p*-value suggests that the two areas are statistically significant. This finding supports our expectation that bank-firm specific hard information helps to improve the accuracy prediction of balance sheet indicators. The AUC also indicates that the baseline model at t_0-*12* gives an accuracy ratio of 85 % and the full model 90 %. The *p*-value equal to 0.000 also suggests in this case that the two areas are statistically different.[10]

5. Robustness and heterogeneity analysis

5.1. Firm size, relationship length and distance from the bank

In the previous section, we showed that the inclusion of bank-firm variables in the baseline model not only provides additional information that a bank can use in the short term to monitor and predict default events of SMEs, but also increases prediction accuracy. In the present section, we investigate whether the previous results hold in the sub-samples of firms and relational banking intensities. Norden and Weber (2010) argued that "the complexity of cash flows in the checking accounts, the likelihood of having multiple banking relationships, and the mechanism of default differ considerably across these borrower types". For instance, larger firms might benefit from the use of different sources of funding, whereas SMEs rely primarily on bank loans. Consequently, although information in financial statements may be more reliable for larger companies, information in the bank-company relationship is more accurate in the case of small borrowers.

There is a substantial body of literature investigating factors that affect the probability of SMEs defaulting according to size (Álvarez et al., 2023; Agostino et al., 2011; Berger & Udell, 1995; El Kalak &

[10] In unreported regressions, available on request, we estimate a baseline model including the bank-firm relationships information and then the model augmented with balance sheet information, demonstrating that the latter provides only limited additional information. The accuracy ratio, in this case, increases by 2 %, one and two years beforehand. This strengthens our previous results that balance sheet information is not fully informative and requires additional bank-firm specific information to predict future corporate default events.

Table 3
Sample splits by firm size and bank-firm relationships, one year before default.

	small firms	medium firms	short relationship length	long relationship length	small distance	high distance
	(1)	(2)	(3)	(4)	(5)	(6)
roa (share)	-0.002 ***	-0.011 ***	-0.004 ***	-0.004 ***	-0.004 ***	-0.004
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.003)
equity inventory coverage (share)	-0.019 ***	-0.007	-0.017 ***	-0.009 **	-0.014 ***	-0.013
	(0.004)	(0.006)	(0.005)	(0.004)	(0.003)	(0.020)
bank loans (share)	0.010 ***	0.025 ***	0.009 ***	0.018 ***	0.013 ***	0.013
	(0.002)	(0.005)	(0.003)	(0.003)	(0.002)	(0.012)
turnover (thousand)	0.005 **	0.006	0.013 ***	0.013 ***	0.012 ***	0.012
	(0.002)	(0.005)	(0.003)	(0.003)	(0.002)	(0.011)
share of loans (second tercile)	-0.014 ***	-0.018 ***	-0.015 ***	-0.022 ***	-0.018 ***	-0.015
	(0.004)	(0.007)	(0.005)	(0.005)	(0.003)	(0.018)
share of loans (third tercile)	-0.019 ***	-0.006	-0.016 **	-0.027 ***	-0.019 ***	-0.066 ***
	(0.005)	(0.017)	(0.007)	(0.005)	(0.004)	(0.017)
dummy checking account	-0.089 ***	0.027	-0.063 **	-0.039	-0.034 *	-0.273 **
	(0.026)	(0.018)	(0.027)	(0.027)	(0.019)	(0.108)
credit line utilization (share)	0.055 ***	0.047 ***	0.060 ***	0.052 ***	0.053 ***	0.101 ***
	(0.003)	(0.005)	(0.004)	(0.004)	(0.003)	(0.018)
credit limit violation (months)	0.009 ***	0.013 ***	0.011 ***	0.009 ***	0.009 ***	0.016 ***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.004)
blank cheques (number)	0.052 ***	0.041 **	0.051 ***	0.033 ***	0.027 ***	0.060
	(0.013)	(0.017)	(0.016)	(0.010)	(0.006)	(0.055)
crediting operations (number)	-0.015 ***	-0.003	-0.010 ***	-0.010 ***	-0.009 ***	-0.008
	(0.003)	(0.004)	(0.003)	(0.003)	(0.002)	(0.015)
dummy long-term loan	-0.003	0.002	-0.010 **	0.010 **	-0.001	0.006
	(0.004)	(0.007)	(0.005)	(0.005)	(0.003)	(0.020)
long-term overruns (months)	0.072 ***	0.053 **	0.088 ***	0.075 ***	0.069 ***	0.157 *
	(0.015)	(0.022)	(0.023)	(0.019)	(0.014)	(0.095)
long-term overruns (share)	0.013 **	0.026 *	0.019 **	0.009	0.011 **	0.106 **
	(0.006)	(0.015)	(0.009)	(0.007)	(0.005)	(0.052)
bank-time fixed effects	yes	yes	yes	yes	yes	yes
industry fixed effects	yes	yes	yes	yes	yes	yes
province fixed effects	yes	yes	yes	yes	yes	yes
Observations	16,700	5073	11,327	10,187	20,928	1244
R^2	0.266	0.321	0.252	0.317	0.260	0.378

Notes: This table reports marginal effect obtained estimating Eq. (1) on sub-samples of firms. Marginal effects are calculated as the variation of the default probability after a variation of the predictors from the value at the 25th percentile to that at the 75th percentile (or 90th, depending on the distribution of variables). Standard errors, clustered at firm level, are reported in parentheses; * indicates significance at the 10 % level; ** indicates significance at the 5 % level; *** indicates significance at the 1 % level. Standard errors, clustered at firm level, are reported in parenthesis; * indicates significance at the 10 % level; ** indicates significance at the 5 % level; *** indicates significance at the 1 % level.

Hudson, 2016; Gupta et al., 2018; Hasan et al., 2022). Analysing a sample of US firms, Gupta et al. (2018) documented that the determinants of bankruptcy vary between the different dimensional categories of SMEs. Following this literature and to explore these differences, we divided firms into two categories according to their average turnover over the observation period: small firms (average turnover of less than 5,000 thousand euros) and medium-sized firms (over 5,000 thousand euros). Table 3 shows some differences between the two categories. The default probability of small firms is affected by the same factors that are found to be economically and statistically significant for the entire sample, whereas that of medium firms is affected by a smaller number of factors. In terms of balance sheet indicators, the default probability of medium firms reduces by 1.1 % if *roa* increases, which is five times the correlation of the same change for small firms. Similarly, the *share of bank loans on total loans* increases the probability of default by 2.5 % in medium firms and by 1 % in small firms. *Equity inventory coverage* exerts a negative and significant marginal effect (1.9 %) on default probability only for small firms. These results are consistent with the hypothesis that the information extracted from balance sheets is more reliable for medium than for small firms, and therefore more useful in monitoring their future payment behaviour.

In a model controlling for balance indicators, system information and fixed effects, among checking account activities the utilization of credit line, months of credit limit violation and blank cheques are relevant for both small and medium firms. The marginal effects in both sub-samples are in line with those of the entire sample. It is notable that the marginal effects of the dummies of the share of bank loans respect to the total banking loans display a negative and highly significant effect on the default probability of small firms. A plausible explanation for this result is that a strong tie with an individual bank represents a way for a small firm to amplify its bargaining power and generate greater availability of credit and lower costs (Grunert & Norden, 2012). *Crediting operations* are found to be economically significant only for small firms, reducing default probability by 1.5 %.

All results on the relevance of long-term loan activities are confirmed for both small and medium firms. Consecutive months of *long-term loans overruns* increase default probability by 7 % and 5 %, respectively, in the two sub-samples. The share of long-term loans balanced overruns increases the default probability of small firms by 1.3 % and of medium firms by 2.6 %. In keeping with Gupta et al. (2018), these results seem to confirm the importance that predictive models appropriately consider the different characteristics of the SMEs linked to their size and do not refer to them as an indistinct whole.

In addition to firm-specific characteristics, our previous results might be sensitive to the intensity of the bank-firm relationship. The classic way in which this intensity is modelled is through the length of lending relationships, as past payment history might give important information to the bank, along with the share of finance provided by each borrower, the number of lending banks and the distance between the local bank and the firm (Foglia et al., 1998; Fiordelisi et al., 2014; Norden & Weber, 2010; Omiccioli & Carmignani, 2007; Refait-Alexandre & Serve, 2020). In the bank-firm relationship context, variables that capture the relationship intensity and length are found to affect credit supply (Agarwal & Hauswald, 2010), cost of borrowing (Bharath et al., 2011) and small business liquidity (Han et al., 2017) which in turn may affect the probability of default.

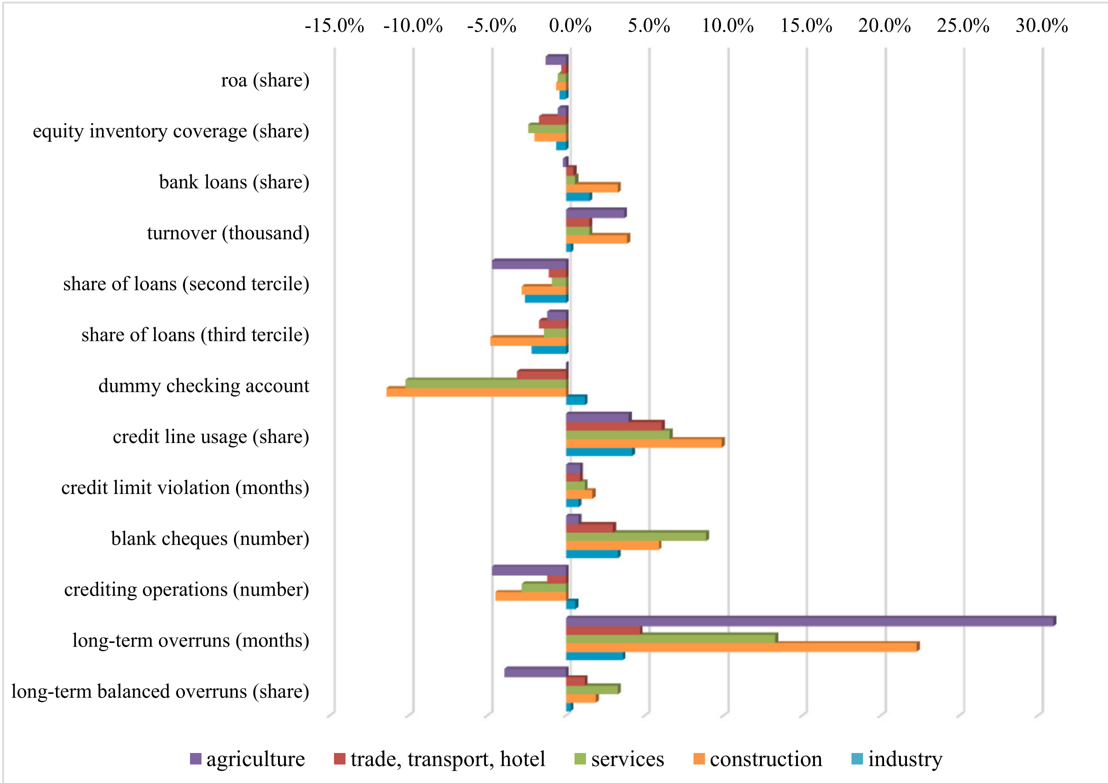

Fig. 2. Sample splits by sector. *Notes*: This figure reports only significant marginal effects obtained estimating Eq. (1) on sub-samples of firms belonging to different sectors.

To better understand the correlation of the intensity of the relationship between bank and firm and default risk, our work investigates the intensity of the relationship by measuring it in two ways: (a) the geographical distance between the local bank and the firm; (b) the length of their relationship. While geographic proximity improves the quality of private information (Agarwal & Hauswald, 2010), the closeness of the business relationship helps to capture signals that predict financial difficulties through the lens of the lending relationship. From this perspective, we divided our sample of firms into two groups, depending on the median physical distance between the bank and the firm (one hour) and on the median relationship length (eight years). The results for short and long bank-firm relationships are shown in columns 3 and 4, respectively, of Table 3. The magnitude of the *roa* marginal effect remains at 0.4 % over the length of the relationship and the marginal effect of *bank loans* increases from 0.9 % to 1.8 % in absolute value as we move from short to long relationships. The share of loans of a bank with respect to the banking system reduces default probability in both sub-samples, but their association is higher in cases of long bank-firm relationships. With the exception of crediting operations, the relation with checking account activities reduces as we move from short to long relationships. This result indicates that as the intensity of the bank-borrower relationship increases, the increase in credit line utilization, credit limit violations and blank cheques have a positive but lower relationship with default probability. The correlation with long-term loans indicators, both in terms of months of overruns and balanced overruns, is positive and very similar in both bank-firm sub-samples of short and long relationship length.

When considering distance as an alternative measure of the intensity of the bank-firm relationship, we see that for small-distance relationships, the results obtained for the whole sample are all confirmed. However, the balance sheet indicators do not statistically significantly predict the default probability of firms in a long-distance relationship. A higher distance reduces the availability of soft information (as shown in the literature) and weakens the positive effects of a relational approach between bank and firm that mitigates the problems of information opacity and procyclical behaviour of banks (Grunert et al., 2005). Checking account activities show a significant relationship with default probability in contexts of both short and long-distance relationships, which is higher in cases of greater distance. The marginal effect of credit line utilization and blank cheques on default probability of long-distance firms is about twice that of short-distance firms (10 % vs. 5.3 % and 6 % vs. 2.7 %, respectively). Similarly, credit limit months of violation increases default probability of long-distance firms by 1.6 % and of short-distance firms by 0.9 %.

5.2. Sector of economic activity and geographical area

To account for heterogeneity, we next split the sample according to: (i) the sector of economic activity (agriculture; trade, transport, hotel; services; construction; industry); (ii) the geographical areas (North, Central, South). The results in Fig. 2 show that in terms of balance sheet indicators, an increase in *roa* reduces the probability of default by 1.3 %, 0.3 % and 0.6 % for the agriculture, trade, transport and hotel, and construction sectors, respectively. *Equity inventory coverage* exerts a negative and significant association (1.7 %) with default probability for the trade, transport and hotel sector and −2.4 % for the services sector. Overall, these results suggest that improving profitability and equity levels, and enhancing inventory management, can have positive effects on reducing the probability of default. However, the specific correlations with particular balance sheet indicators may differ across sectors, highlighting the importance for banks of sector-specific analyses in assessing firm-level risk.

Consistent with previous results, the share of *credit line utilization* increases the probability of default by about 6 % in the trade, transport and hotel and the services sectors, by 9.9 % in the construction sector,

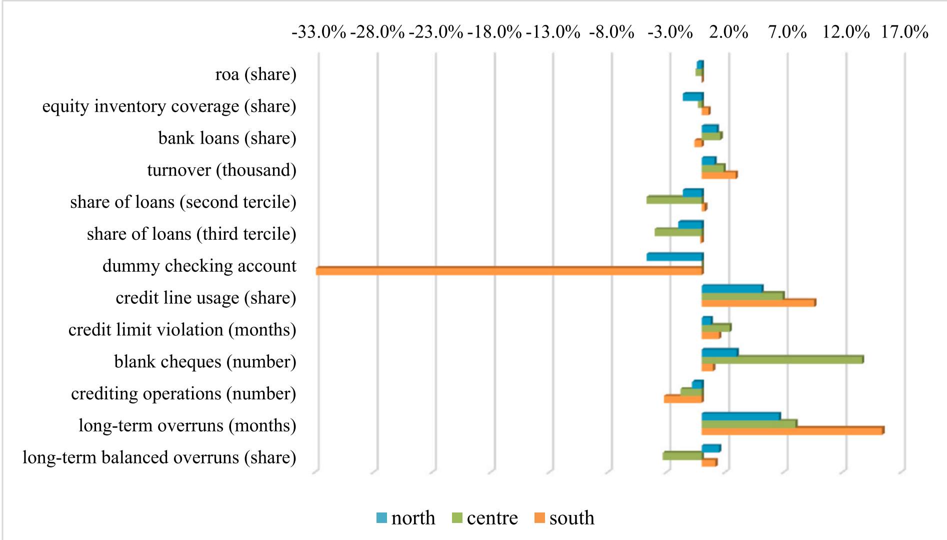

Fig. 3. Sample splits by geographical area. *Notes*: this figure reports only significant marginal effects obtained estimating Eq. (1) on sub-samples of firms operating in different areas.

and by about 4 % in the agricultural and industry sectors. Months of *credit limit violations* are significant for the trade, transport and hotel and the construction sectors (0.9% and 1.7 %, respectively), whereas *blank cheques* are significant in the industry sector (3.3 %). In keeping with previous results, increasing the number of credit operations reduces the probability of default by 4.7 % in the agriculture sector, 1.2 % in the trade, transport and hotel sector, and 4.5 % in the construction sector. *Long-term overruns* significantly increase the probability of default in the trade, transport and hotel sector, the construction sector and in the industry sector (by 4.7 %, 22.3 % and 3.6 %, respectively). Analysis of behaviour with the banking system shows that in some sectors (particularly the construction, agriculture and trade, transport and hotel sectors), there is a greater correlation with certain indicators (e.g., overruns and credit line utilization). This is due to the ineffective management of financial dynamics that may be determined by both the modest level of financial skills found within the sector and the greater complexity and relationship length of the operating cycle. The mismatch between the lower predictability of financial requirements and available financing instruments may be another explanation for the greater association with financial anomalies in some specific economic sectors.

Fig. 3 shows the results of the split by geographic area. In terms of balance sheet indicators, the probability of default decreases by 0.4 % and 0.5 % with an increase in *roa* for the North and the Centre, respectively. The share of *equity inventory coverage* has a significant relationship with default only on northern firms, with a reduction in probability of default of 1.6 %. The *checking account* activities significantly predict default probability for firms in the South (amounting to a 32.9 % reduction), while *utilization of credit line* and months of *credit line violations* significantly affect firms in the North and Centre (raising the probability of default by 5.1 % and 0.8 % in the North and by 6.9 % and 2.4 % in the Centre, respectively). The relevance of long-term lending activities is confirmed for firms in the North. Consecutive months of *long-term loan overruns* increase the probability of default by 6.6 %, while the share of *long-term balanced overruns* increases the probability by 1.5 %. Overall, southern areas show less virtuous behaviour in using bank credit lines, as evidenced by longer overdraft periods and more intensive use of credit.[11] However, when a firm located in the South uses checking accounts in addition to loans, the propensity to default tends to decrease dramatically. Checking accounts provide a flexible source of financing, helping the firm to manage cashflow more effectively and meet financial obligations more punctually. In addition, having a checking account helps to build a stronger relationship with the bank, which in turn benefits from more information to monitor the behaviour of the firm it finances.

5.3. Instrumental variables

Thus far, we have analysed the relationship between banks and their borrowers in the context of asymmetric information between lender and borrower in which borrowers have more information on their default risk than banks. Borrowers are small firms that are usually private or unlisted, so the information is more asymmetric than for public firms, since public shares provide information about financial conditions. However, borrowers can engage in other activities with the bank that provide some information on default risk. One such activity is the use of credit lines, as shown by the results in previous section. In the previous analysis, we considered the use of a credit line as a purely exogenous variable in a probit analysis to predict default risk. In other words, we observe the behaviour of defaulting firms before they default, which should exclude the fact that default risk variation affects use of credit lines.

However, it may be argued that the effect of credit line utilization on default is ambiguous (Bergerès et al., 2015). On the one hand, a borrower in financial distress might use the liquidity provided by the credit line to pay back a loan, effectively reducing its default probability. On the other hand, if higher credit line utilization is associated with bad risk borrowers, it may correlate positively with default probability. Additionally, Jiménez et al. (2009b) found that credit lines are drawn down more by firms that eventually default on these lines than by firms that do not default. Therefore, introducing the use of credit lines means introducing a potential endogenous action from the borrower that is not independent from that of defaulting: the use of credit lines affects default probability and, conversely, the default risk variation might affect the use of credit lines.[12]

Therefore, following the contribution of Bergerès et al. (2015), we adopted an instrumental variable approach to deal with the potential issue of endogeneity of credit line utilization. In particular, we estimated a two-stage regression model where in the first stage, credit line utilization is instrumented by a dummy indicating whether the firm is

[11] Results for firms in the Centre and South are estimated on a smaller number of observations than those for northern firms.

[12] We would thank an anonymous referee for highlighting this issue.

Table 4
Using instrumental variables for credit line utilization.

	t_0			t_0 - 12		
	(1)	(2)	(3)	(4)	(5)	(6)
credit line utilization (share)	0.069*** (0.019)	0.047** (0.022)	0.036* (0.022)	0.121*** (0.042)	0.077* (0.040)	0.063* (0.036)
balance sheet indicators	yes	yes	yes	yes	yes	yes
firm characteristics	yes	yes	yes	yes	yes	yes
system information	yes	yes	yes	yes	yes	yes
credit line and checking account		yes	yes		yes	yes
long-term loans			yes			yes
bank-time fixed effects	yes	yes	yes	yes	yes	yes
industry fixed effects	yes	yes	yes	yes	yes	yes
province fixed effects	yes	yes	yes	yes	yes	yes
Kleibergen-Paap test	135.57 (0.000)	106.10 (0.000)	111.04 (0.000)	80.39 (0.000)	58.72 (0.000)	62.86 (0.000)
Hansen test	2.077 (0.149)	0.035 (0.852)	0.274 (0.600)	1.060 (0.303)	0.477 (0.490)	0.529 (0.467)
Observations	23,079	23,079	23,079	8493	8, 493	8, 493
R^2	0.234	0.366	0.367	0.237	0.362	0.364

Notes: this table reports the second stage marginal effect of credit line utilization obtained estimating an instrumental variable model. Credit line utilization is instrumented by a dummy indicating whether the firm is a collateralized counterpart and by the number of months since the first contract between the bank and the firm. Kleibergen–Paap first-stage F-statistic (with the p-value) is the value of the F-statistic (and p-value) for the hypothesis that instruments have jointly zero coefficients in the first-stage regression. Over-identifying restrictions statistic (p-value) is the value of the Hansen statistic (with the p-value). Standard errors, clustered at firm level, are reported in parenthesis; * indicates significance at the 10 % level; ** indicates significance at the 5 % level; *** indicates significance at the 1 % level.

a collateralized counterpart and by the number of months since the first contract between the bank and the firm. The second stage of our model is Eq. (1), where credit line utilization is instrumented.

The results are provided in Table 4, together with the tests of validity of instruments that are statistically shown to overcome endogeneity. The results mirror those in Table 2, since credit line utilization positively predicts default probability in all models and both one and two years beforehand. In particular, increasing credit line utilization from 0 % (25th percentile) to 90 % (75th percentile) predicts an increase in the default probability of 3.6 % one year beforehand, in the full model (column 3). The same increase in credit line utilization predicts a higher increase in default probability two years beforehand (6.3 % in the full model). Therefore, our results support the hypothesis that higher credit line utilization is associated with bad risk borrowers and is positively correlated with default probability.

6. Discussion and conclusions

The main source of financing for SMEs is the banking channel. From the supply side, financing small companies is one of the main activities of local banks, contributing significantly to the net interest margin. The present paper has provided a direct examination of the contribution of bank-firm specific information, gathered from checking account activities and long-term loan performance, in improving the discriminatory power of default prediction models that only include accounting information. The analysis is based on a unique dataset of a sample of 111 co-operative credit banks and 13,081 firms operating in Italy between 2012 and 2014.

Our work provides several incremental contributions to the literature observing the links between the probability of business insolvency and short- and medium-term credit line utilization rates. First, it demonstrates the additional explanatory and discriminatory power of bank-firm relationships, highlighting the role of aspects related to the checking account relationship and to long term-loans in two different time horizons. Second, it focuses on SMEs that are very different in size and type of activity, and that are highly dependent on banks. Third, the depth of the dataset means that we were able to check for time-varying characteristics of the banking relationship (e.g., relationship length and distance), including the bank-year fixed effects. Finally, our study used a complete post-crisis dataset that eliminates the selection bias found in previous studies.

The paper also contributes to the current academic debate on improving the prediction of SME defaults by employing relationship lending variables. This topic has been subject to limited scholarly scrutiny (Ciampi et al., 2021). This represents an important gap in the research, particularly given the growing body of literature that recognizes the critical role played by soft information in relationship lending and, more broadly, in the information generation processes of banks (Álvarez-Botas & González, 2023; Filomeni et al., 2020, 2021; Zhao et al., 2021).

By estimating a probit model, the paper offers two key findings. First, bank-firm specific information obtained from checking account activities and long-term loan performance has a significant association with the probability of default that goes beyond balance sheet indicators. Use of credit line and the depth of credit limit violations on current accounts and long-term loans increase a bank's ability to predict the probability of a borrower's insolvency at least a year before it occurs. The association of additional private information with default, with respect to financial indicators, is persistent over time. Second, a comparison of the goodness-of-fit of the models shows that using bank-firm specific measures leads to a better fit than that obtained using balance sheets alone. The accuracy ratio, considering both the true-positive rate and the true-negative rate, increases by 7 % in a one-year horizon and by 5 % in a two-year horizon. The combination of credit line utilization, checking account activity and long-term loan performance effectively gives the lender a real-time window into the borrower's cashflow, confirming and extending the results of Norden and Weber (2010).

These results have useful implications for both SMEs and banks. SMEs must closely monitor their creditworthiness, paying attention to the management of their relationship with the bank, with particular focus on overruns and past due. Knowledge of variables that influence their rating requires SMEs to increase the accuracy of short- and medium-term financial planning and to improve their financial communication with banks. The combination of the use of bank-firm relationships with accounting data is also a valid tool for banks to ascertain a borrower's ability to properly service their debt. The spillover of information from past loans and current account activities is a concrete advantage in better managing their lending relationships with SMEs.

This work has some limitations. The major constraint is that the paper is lacking in time dynamics. For the one-year default prediction, it only has data for two-time spots, i.e., firm data in 2012 for default predictions during 2013 and firm data in 2013 for default predictions during 2014. Without much training using historical data, the model may be less informative for predicting future defaults. In addition, our sample includes firms and banks operating in Italy and this could limit the generalization of results to different industrial and financial systems. However, the source of data from several banks and the large sample of companies mitigates these limitations and makes it possible to generalize the results to economies where the relationship between SMEs and local banks is significant.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was conducted within the Observatory on Ethics, Finance, Banks and Firm (*Osservatorio Etica, Finanza, Banca e Impresa*) of the University of Cassino and Lazio Meridionale (Italy). We would like to aknowledge Alberto Pozzolo for his valuable suggestions and comments. We also would like to acknowledge the comments of participants at the EFiC Conference on Banking and Finance held at the University of Essex (Colchester - UK) on 4–6 July 2019, at European Financial Management Association 2021 Conference held at University of Leeds (Leeds – UK) on 30 June-3 July 2021, at the World Finance & Banking Symposium held at Florida International University (Miami – USA) on 16–17 December 2022, and seminar participants at the University of Molise (Italy).

Appendix

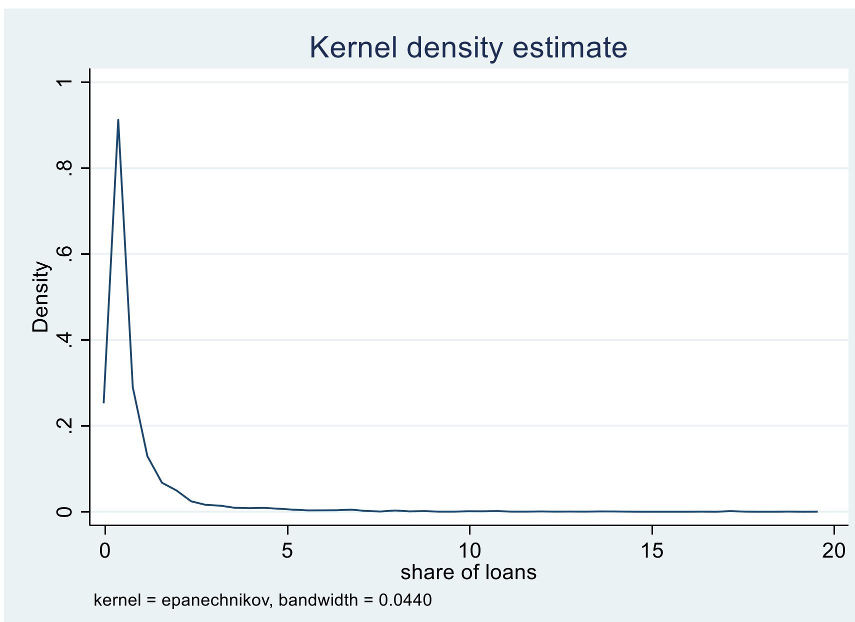

Fig. A1. Kernel density of the variable *Share of loans*.

Table A1
Correlation matrix.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
(1)	default (dummy)	1														
(2)	roa (share)	-0.09	1													
(3)	equity inventory coverage (share)	-0.09	0.20	1												
(4)	bank loans (share)	0.12	-0.22	-0.24	1											
(5)	turnover (thousand)	-0.02	0.07	0.00	-0.09	1										
(6)	share of loans (second tercile)	-0.04	0.03	-0.01	-0.09	-0.16	1									
(7)	share of loans (third tercile)	-0.01	0.02	0.06	-0.01	-0.16	-0.46	1								
(8)	dummy checking account	-0.01	0.01	-0.05	-0.05	0.03	0.04	0.00	1							
(9)	credit line utilization (share)	0.22	-0.22	-0.16	0.31	-0.16	-0.03	0.02	0.11	1						
(10)	credit limit violation (months)	0.23	-0.15	-0.12	0.18	-0.11	-0.01	0.05	0.06	0.45	1					
(11)	blank cheques (number)	0.10	-0.03	-0.04	0.02	-0.04	0.01	0.02	0.02	0.15	0.18	1				
(12)	crediting operations (number)	-0.04	0.01	-0.03	-0.07	0.14	0.04	0.13	0.08	-0.05	-0.01	0.04	1			
(13)	dummy long-term loan	0.03	-0.03	0.00	0.18	-0.02	0.06	0.18	-0.14	0.03	0.07	0.03	0.13	1		
(14)	long-term overruns (months)	0.20	-0.06	-0.04	0.13	-0.04	0.00	0.08	-0.17	0.15	0.26	0.08	-0.01	0.32	1	
(15)	long-term balanced overruns (share)	0.16	-0.04	-0.01	0.09	-0.02	0.00	0.03	-0.19	0.10	0.15	0.04	-0.03	0.18	0.62	1

References

Acharya, V., Almeida, H., & Campello, M. (2013). Aggregate risk and the choice between cash and lines of credit. *The Journal of Finance, 68*(5), 2059–2116.

Acharya, V., Almeida, H., Ippolito, F., & Perez Orive, A. (2020). Bank lines of credit as contingent liquidity: Covenant violations and their implications. *Journal of Financial Intermediation, 44*, Article 100817.

Agarwal, S., & Hauswald, R. (2010). Distance and private information in lending. *The Review of Financial Studies, 23*(7), 2757–2788.

Aghabarari, L., Guettler, A., Naeem, M., & Van Doornik, B. (2021). Is there help indeed, if there is help in need? The case of credit unions during the global financial crisis. *Economic Inquiry, 59*(3), 1215–1233.

Agostino, M., Gagliardi, F., & Trivieri, F. (2011). Bank competition, lending relationships and firms' default risk. An investigation on Italian SMEs. *International Small Business Journal, 30*(8), 907–943.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance, 23*(4), 589–609.

Altman, E. I., & Sabato, G. (2007). Modelling credit risk for SMEs: Evidence from the US market. *Abacus, 43*(3), 332–357.

Altman, E. I., Iwanicz-Drozdowska, M., Laitinen, E. K., & Suvas, A. (2017). Financial distress prediction in an international context: A review and empirical analysis of Altman's Z-score model. *Journal of International Financial Management and Accounting, 28*(2), 131–171.

Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking and Finance, 18*(3), 505–529.

Álvarez, L., García-Posada, M., & Mayordomo, S. (2023). Distressed firms, zombie firms and zombie lending: A taxonomy. *Journal of Banking and Finance, 149*, Article 106762.

Álvarez-Botas, C., & González, V. M. (2023). The value of relationship banking: International evidence. *Research in International Business and Finance, 64*, Article 101822 . https://doi.org/10.1016/j.ribaf.2022.101822

Arcuri, G., & Levratto, N. (2018). "Early stage SME bankruptcy: does the local banking market matter? *Small Business Economics, 54*(2), 421–436.

Ballester, L., González-Urteaga, A., & Martínez, B. (2020). The role of internal corporate governance mechanisms on default risk: A systematic review for different institutional settings. *Research in International Business and Finance, 54*, Article 101293.

Bank of Italy (2021), *The 2020 annual report, Bank of Italy*, Retrieved from https://www.bancaditalia.it/pubblicazioni/relazione-annuale/2021/index.html. (Accessed 31 May2021).

Bartoli, F., Ferri, G., Murro, P., & Rotondi, Z. (2013). Bank–firm relations and the role of Mutual Guarantee Institutions at the peak of the crisis. *Journal of Financial Stability, 9*, 90–104.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research, 4*, 7–111.

Becchetti, L., Ciciretti, R., & Paolantonio, A. (2016). The cooperative bank difference before and after the global financial crisis. *Journal of International Money and Finance, 69*, 224–246.

Beck, T., Degryse, H., De Haas, R., & Van Horen, N. (2006). When arm's length is too far: Relationship banking over the credit cycle. *Journal of Financial Economics, 127*(1), 174–196.

Belaid, F., Boussaada, R., & Belguith, H. (2017). Bank-firm relationship and credit risk: An analysis on Tunisian firms. *Research in International Business and Finance, 42*, 532–543.

Berger, A., Goulding, W., & Rice, T. (2014). Do small businesses still prefer community banks? *Journal of Banking and Finance, 44*, 264–278.

Berger, A. N., & Udell, G. F. (1995). Relationship lending and lines of credit in small firm finance. *Journal of Business, 68*, 351–382.

Berger, G. F., & Udell (2002). Small business credit availability and relationship lending: The importance of bank organisational structure. *The Economic Journal, 112*, F32–F53.

Bergerès, A. S., d'Astous, P., & Dionne, G. (2015). Is there any dependence between consumer credit line utilization and default probability on a term loan? Evidence from bank-customer data. *Journal of Empirical Finance, 33*, 276–286.

Bharath, S. T., Dahiya, S., Saunders, A., & Srinivasan, A. (2011). Lending relationships and loan contract terms. *The Review of Financial Studies, 24*(4), 1141–1203.

Board of Governors (2022). *Availability of credit to small businesses. federal reserve system*, Washington, DC.

Bredart, X. (2014). Financial distress and corporate governance: The impact of board configuration. *International Business Research, 7*(3), 72–80.

Brunner, A., Krahnen, J., & Weber, M. (2000). Information production in credit relationships: On the role of internal ratings in commercial banking. *Centre for Financial Studies Working*(Paper No. 2000/10).

Bonaccorsi di Patti, E., D'Ignazio, A., Gallo, M., & Lopes Micucci, G. (2015). The role of leverage in firm solvency: evidence from bank loans. *Italian Economic Journal, 1*(2,), 253–286.

Boot, A. W. A. (2000). Relationship banking: What do we know? *Journal of Financial Intermediation, 9*, 7–25.

Bottazzi, G., Grazzi, M., Secchi, A., & Tamagni, F. (2011). Financial and economic determinants of firm default. *Journal of Evolutionary Economics, 21*(3), 373–406.

Boyd, J., & Prescott, E. C. (1986). Financial intermediary-coalitions. *Journal of Economic Theory, 38* 211-32.

Buffa, F., Franch, M., Martini, U., & Tamanini, A. (2019). Scelte strategiche e pratiche ambientali nelle Pmi alberghiere. Verso un nuovo modello di business sostenibile. *Sinergie Italian Journal of Management, 37*(1), 85–106.

Chen, Y., Huang, R. J., Tsai, J., & Tzeng, L. Y. (2013). Soft information and small business lending. *Journal of Financial Services Research, 47*(1), 115–133.

Ciampi, F. (2015). Corporate governance characteristics and default prediction modeling for small enterprises. An empirical analysis of Italian firms. *Journal of Business Research, 68*(5), 1012–1025.

Ciampi, F., Giannozzi, A., Marzi, G., & Altman, E. I. (2021). Rethinking SME default prediction: A systematic literature review and future perspectives. *Scientometrics, 126*, 2141–2188.

Cingano, F., Manaresi, F., & Sette, E. (2016). Does credit crunch investment down? New evidence on the real effects of the bank-lending channel. *The Review of Financial Studies, 29*(10), 2737–2773.

Cole, R. A., Goldberg, L. G., & White, L. J. (2004). Cookie-cutter versus character: The micro structure of small business lending by large and small banks. *Journal of Financial and Quantitative Analysis*.

Cororaton, A., (2019). *Banking on the firm objective*. Retrieved from SSRN: https://ssrn.com/abstract=3231614.

Dainelli, F., Giunta, F., & Cipollini, F. (2013). Determinants of SME credit worthiness under Basel rules: The value of credit history information. *PSL Quarterly Review, 66*(264), 21–47.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics, 44*(3), 837–845.

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

Diamond, D. W. (1991). Monitoring and reputation: The choice between bank loans and directly placed debt. *Journal of Political Economy, 99*, 688–721.

Dierkes, M., Erner, C., Langer, T., & Norden, L. (2013). Business credit information sharing and default risk of private firms. *Journal of Banking and Finance, 37*(8), 2867–2878.

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.

Duan, J. C., Kim, B., Kim, W., & Shin, D. (2018). Default probabilities of privately held firms. *Journal of Banking and Finance, 94*, 235–250.

Duarte, F. D., Gama, A. P. M., & Gulamhussen, M. A. (2018). Defaults in bank loans to SMEs during the financial crisis. *Small Business Economics, 51*(3), 591–608.

Dullmann, K., & Koziol, P. (2013). *Evaluation of minimum capital requirements for bank loans to SMEs*. Deutsche Bundesbank22.

El Kalak, I., & Hudson, R. (2016). The effect of size on the failure probabilities of SMEs: An empirical study on the US market using discrete hazard model. *International Review of Financial Analysis, 43*, 135–145.

Elsas, R. (2005). Empirical determinants of relationship lending. *Journal of Financial Intermediation, 14*, 32–57.

European Commission (2022). Survey on the access to finance of enterprises (SAFE). *Analytical report 2022*. Retrieved from https://single-market-economy.ec.europa.eu/access-finance/data-and-surveys-safe_en.

European Union, (2022), *Annual report on European SMEs 2021/2022*, EU publications. Retrieved from https://le-europe.eu/publication/2021–2022-annual-report-on-european-smes-june-2022/. (Accessed June 2022).

Fischer, K. H. (2000). Acquisition of information in loan markets and bank market power – an empirical investigation. *Mimeo Johann Wolfgang Goethe University Frankfurt*.

Filomeni, S., Udell, G. F., & Zazzaro, A. (2020). Communication frictions in banking organizations: Evidence from credit score lending. *Economic Letters, 195*, Article 109412 . https://doi.org/10.1016/j.econlet.2020.109412

Filomeni, S., Udell, G. F., & Zazzaro, A. (2021). Hardening soft information: Does organizational distance matter. *The European Journal of Finance, 27*(9), 897–927. https://doi.org/10.1080/1351847X.2020.1857812

Fiordelisi, F., Monferrà, S., & Sampagnaro, G. (2014). Relationship lending and credit quality. *Journal of Financial Services Research, 46*(3), 295–315.

Foglia, A., Laviola, S., & Reedtz, P. M. (1998). Multiple banking relationships and the fragility of corporate borrowers. *Journal of Banking and Finance, 22*(10–11), 1441–1456.

Gabbi, G.; Giammarino, M.; Matthias, M. (2020), *Die hard: Probability of default and soft information. risks*, 8, 46.

Galluzzo, N. (2012). Multifunctionality in agriculture and development of cooperative credit banks: A comparative analysis in Italy and Ireland. *International Journal of Cooperative Studies, 1*(1), 1–8.

Greenbaum, S. I., Kanatas, G., & Venezia, I. (1989). Equilibrium loan pricing under the bank-client relationship. *Journal of Banking and Finance, 13* 221-35.

Grunert, J., & Norden, L. (2012). Bargaining power and information in SME lending. *Small Business Economics, 39*(2), 401–417.

Grunert, J., Norden, L., & Weber, M. (2005). The role of non-financial factors in internal credit ratings. *Journal of Banking and Finance, 29*(2), 509–531.

Gupta, J., Barzotto, M., & Khorasgani, A. (2018). Does size matter in predicting SMEs failure? *International Journal of Finance and Economics, 23*(4), 571–605.

Hasan, I., Lynch, D.K. , Siddique, A.R. (2022) *Usefulness of hard information and soft information: case of default risk and ESG factors*. Retrieved from SSRN: https://ssrn.com/abstract=4015878 or http://dx.doi.org/10.2139/ssrn.4015878.

Han, L., Zhang, S., & Greene, F. J. (2017). Bank market concentration, relationship lending, and small business liquidity. *International Small Business Journal, 35*(4), 365–384.

Howorth, C., & Moro, A. (2012). Trustworthiness and interest rates: An empirical study of Italian SMEs. *Small Business Economics, 39*, 161–177.

Hussain, I., Durand, R. B., & Harris, M. N. (2021). Relationship lending: A source of support or a means of exploitation? *Global Finance Journal, 48*(C), Article 100549.

Irwin, R. J., & Irwin, T. C. (2013). Appraising credit ratings: Does the CAP fit better than the ROC? *International Journal of Finance and Economics, 18*(4), 396–408.

Jiménez, G., Lopez, J. A., & Saurina, J. (2009a). Calibrating exposure at default for corporate credit lines. *Journal of Risk Management in Financial Institutions, 2*(2), 121–129.

Jiménez, G., Lopez, J. A., & Saurina, J. (2009b). Empirical analysis of corporate credit lines. *Review of Financial Studies, 22*(12), 5069–5098.

Keasey, K., & Watson, R. (2006). Non-financial symptoms and the prediction of small company failure: A test of Argenti's hypotheses. *Journal of Business Finance and Accounting, 14*(3), 335–354.

López-Espinosa, J., Mayordomo, S., & Moreno, A. (2017). When does relationship lending start to pay? *Journal of Financial Intermediation, 31*, 16–29.

Luo, Z., Hsu, P., & Xu, N. (2020). SME default prediction framework with the effective use of external public credit data. *Sustainability, 12*(18), 7575.

Matthias, M., Giammarino, M., & Gabbi, G. (2019). Modeling hard and soft facts for SMEs: Some international evidence. *Journal of International Financial Management and Accounting, 30*(3), 203–222.

McKillop, D., French, D., Quinn, B., Sobiech, A. L., & Wilson, J. O. S. (2020). Cooperative financial institutions: A review of the literature. *International Review of Financial Analysis, 71*, 1–11.

Mirza, N., Rahat, B., Naqvi, B., & Rizvi, S. K. A. (2020). Impact of Covid-19 on corporate solvency and possible policy responses in the EU. *The Quarterly Review of Economics and Finance*. https://doi.org/10.1016/j.qref.2020.09.002

Mester, L. J., Nakamura, L. I., & Renault, M. (2007). Transactions accounts and loan monitoring. *Review of Financial Studies, 20*(3), 529–556.

Moscatelli, M., Parlapiano, F., Narizzano, S., & Viggiano, G. (2020). Corporate default forecasting with machine learning. *Expert Systems with Applications, 161*, Article 113567.

Nakamura, L. (1993), Commercial bank information: Implications for the structure of banking, Klausner M. & White, L.J. (Ed.s), *Structural change in banking*. Irwin, Homewood, IL: Business One(pp. 131–160).

Nakamura, L. I., & Roszbach, K. (2018). Credit ratings, private information, and bank monitoring ability. *Journal of Financial Intermediation, 36*, 58–73.

Neuberger, D., & Räthke-Döppner, S. (2015). The role of demographics in small business loan pricing. *Small Business Economics, 44*, 411–424. https://doi.org/10.1007/s11187-014-9602-4

Norden, L., & Weber, M. (2010). Credit line utilization, checking account activity, and default risk of bank borrowers. *Review of Financial Studies, 23*(10), 3665–3699.

Omiccioli, M., & Carmignani, A. (2007), *Costs and benefits of creditor concentration: An empirical approach*, Banca d'Italia. Retrieved from https://www.bancaditalia.it/pubblicazioni/temi-discussione/2007/2007–0645/en_tema_645.pdf.

Petersen, M., & Rajan, R. G. (1994a). The benefits of lending relationship. *The Journal of Finance, 49*(1), 3–37.

Petersen, M.A., & Rajan, R.G. (1994b). The effect of credit market competition on firm-creditor relationships. *Working paper*. Chicago University of Chicago, February.

Petersen, M. A., & Rajan, R. G. (1995). The effect of credit market competition on lending relationships. *The Quarterly Journal of Economics, 110*(2), 407–443.

Ramakrishnan, S., & Thakor, A. V. (1984). Information reliability and a theory of financial intermediation. *Review of Economic Studies, 51* 415-32.

Refait-Alexandre, C., & Serve, S. (2020). Multiple banking relationships: Do SMEs mistrust their banks? *Research in International Business and Finance, 51*.

Roggi, O., Altman, E., Sabato, G., & Giannozzi, A. (2013). "Building SME rating: is it necessary for lenders to monitor financial statements of the borrowers?". *Bancaria, 10*, 54–71.

Sharpe, S. A. (1990). Asymmetric information, bank lending, and implicit contracts: A stylized model of customer relationships. *Journal of Finance, 45*(1), 69–87.

Shin, K. S., & Lee, Y. J. (2002). A genetic algorithm application in bankruptcy prediction modeling. *Expert Systems with Applications, 23*(2), 321–328.

Shin, K. S., Lee, T. S., & Kim, H. J. (2005). An application of support vector machines in bankruptcy prediction model. *Expert Systems with Applications, 28*(1), 127–135.

Smondel, A. (2018). SMEs' soft information and credit rationing in France. *Human Systems Management, 37*(2), 169–180.

Strahan, P.E. (1999), *Borrower risk and the price and non-price terms of bank loans*. Federal Reserve Bank of New York, October, Retrieved from https://www.newyorkfed.org/ medialibrary/media/research/staffreports/sr90.html.

Statista (2022), Value added by small and medium-sized enterprises (SMEs) in the European Union (EU27) from 2008 to 2022, by enterprise size. Retrieved from https://www.statista.com/statistics/936386/value-added-by-smes-in-eu-memberstates/#:~:text=Small%20and%20medium%2Dsized%20enterprises,trillion %20euros%20of%20this%20amount. (Accessed August 2022).

Stein, J. C. (2002). Information production and capital allocation: Decentralized versus hierarchical firms. *Journal of Finance, 57*, 1891–1921.

Tian, S., Yu, Y., & Guo, H. (2015). Variable selection and corporate bankruptcy forecasts. *Journal of Banking and Finance, 52*, 89–100.

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.

Welch, B. L. (1947). The generalization of 'student's' problem when several different population variances are involved. *Biometrika, 34*(1/2), 28–35.

Wheelock, D. C., & Wilson, P. W. (2011). Are credit unions too small? *Review of Economics and Statistics, 93*(4), 1343–1359.

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*.

Yildirim, A. (2020). The effect of relationship banking on firm efficiency and default risk. *Journal of Corporate Finance, 65*, Article 101500.

Zhao, T., Luintel, K. B., & Matthews, K. (2021). Soft information and the geography of SME bank lending. *Regional Studies, 55*(4), 679–692. https://doi.org/10.1080/00343404.2020.1851024