Small businesses funded by local banks:

The impact of bank-firm relationship on default risk

Michele Modina (corresponding author)

University of Molise

Via Francesco De Sanctis, 1, 86100 Campobasso (CB)

E-mail: michele.modina@unimol.it - Phone: +39 0874 404337

Filomena Pietrovito

University of Molise

Via Francesco De Sanctis, 1, 86100 Campobasso (CB)

E-mail: filomena.pietrovito@unimol.it

Carmen Gallucci

University of Salerno

Via Giovanni Paolo II, 132 – 84084 – Fisciano (SA)

E-mail: cgallucci@unisa.it

Vincenzo Formisano

University of Cassino and Lazio Meridionale

Loc. Folcara - 03043 Cassino (FR)

E-mail: v.formisano@unicas.it

Small businesses funded by local banks: The impact of bank-firm relationship on default risk

Abstract

This paper investigates the impact of the quantitative information on the bank-firm relationship on small-and-medium-sized enterprises (SMEs)' likelihood to default. It uses a unique dataset on over 12,700 Italian firms operating between 2012 and 2014 and 111 co-operative banks which have lent to them. The paper finds that, apart from the financial ratios obtained from the balance sheets, the use of credit lines, credit line violations, checking accounts performance and long-term loans overruns significantly affect the 1-year and 2-year probability of default (PD). Those additional variables also increase the model's discriminatory power, measured by the accuracy of default predictions. These results indicate that SMEs can improve their ability to repay loans, not only by focusing on their financial results, but also by correctly using credit lines and limiting credit violations in order to transmit information that serves to strengthen their access to banking credit.

Keywords: bank-firm relationship, credit risk, default prediction, firm default, SMEs

1 Introduction

SMEs significantly contribute to the economic development of the European industrial system, since they account for two-thirds of total EU-28 employment (66.4%) and slightly less than three-fifths (56.8%) of the value added generated by the non-financial sector (European Union, 2018). However, as far as the financial structure of European and Italian SMEs is concerned, academics and practitioners notice a bias in favor of bank loans. Compared to large firms, SMEs do not benefit from different sources of founding and therefore access to bank credit remains their main source of financing. Coverage of both short-term and medium-long-term financial requirements passes necessarily through the establishment of relations with the banking system. In the first case, checking accounts, as well as the self-liquidating credit lines, are the main technical forms; in the second case, long-term loans are the most common way of allowing the firm to sustain its financial needs or to better reshape the maturity of the debt by extending the time period. As a consequence, the construction of a solid relationship with the banking system, favoring the assignment of a good credit rating, is one of the cornerstones of the SMEs' financing policy. At the same time, operations with SMEs are the main credit activity of banks especially when the local credit market is concerned (Agostino *et al.*, 2011). Despite the banking concentration process, the recent regulatory changes and the novelties introduced by fintech (Barba Navaretti *et al.*, 2017), local bank lending continues to be an important channel for SMEs financing (Alessandrini *et al*, 2009).[1] Physical proximity and direct knowledge foster a closer credit relationship which might benefit both actors and the community in which they operate (Petersen and Rajan, 1994; Carpenter and Petersen, 2002; Udell, 2009; Agarwal and Hauswald, 2010). The granting of credit to firms constitutes the most consistent source of power of net interest margin. As a matter of fact, the activity of screening borrowers, the assignment of their creditworthiness and the subsequent monitoring of the bank-firm relationship constitute the fundamental steps of the credit process for banks. This has become even more important with the introduction of the Basel rules which, motivated by the intention to better determine the capital adequacy, have induced banks to develop accurate rating systems in order to connect capital requirements to the level of riskiness of corporate borrowers and to detect early warning signals in their loan portfolio (Altman and Sabato, 2007; Ciampi, 2015; Duarte *et al.*, 2018). In this framework, default prediction models have received not only the interest of firms, banks and practitioners, but also that of academics in supporting lending decisions, monitoring credit riskiness and pricing loans (Doumpos *et al.*, 2015; Duan *et al.*, 2017). The traditional literature has examined the contribution of

[1] In Italy cooperative credit banks are the most representative form of local banks (Bank of Italy, 2019). Present in 2,640 Italian municipalities (where in 620 of these they act as the sole intermediary), 59% of their assets are destined to loans to families and firms (6% points more than the other banks) and for the development of the territories (for each 100 euros of savings collected in the territory, 87 euros become credit to the real economy of that territory). During the years of the crisis (2008-2018), their market share increased and, in the period December 2012-December 2017, they introduced net loans of 8.2 billion euro into the economic circuits (compared to an overall reduction recorded in the credit market).

various economic and financial aspects of a firm (such as leverage, cost of debt, working capital, asset turnover, and profitability) in estimating default models, adopting accounting information extracted from balance sheets (Beaver, 1966; Altman, 1968). The use of accounting-based models in predicting the firm's insolvency risk is widely accepted in the banking industry and allows a higher level of risk-adjusted credit return (Agarwal and Taffler, 2008). However, using accounting information alone is not highly informative, for SMEs, for two main reasons. First, the balance sheets are backward looking and they do not provide timely signals on corporate difficulties in repaying bank loans.[2] Indeed, although the financial statements represent a rich source of information, they are not useful to detect cash difficulties of the firms and, therefore, their ability to repay the debts. Second, the opaqueness of information contained in financial indicators is usually higher for SMEs than in the case of large firms and therefore less effective to predict corporate default (Keasey and Watson, 2006; Altman and Sabato, 2007; Norden and Weber, 2010; Ciampi, 2015). Fortunately, as far as the investigation of the borrower's financial difficulties is concerned, relevant information can usually be observed by the banks long before receiving a firm's financial statements (Norden and Weber, 2010). This refers to readily available and costless private information, mainly related to specific services provided to the firm, that can be used to predict expected performance in repaying loans: credit and debit cards, credit lines, checking accounts, investment portfolio, saving accounts and long-term loans performance. Compared to quantitative balance sheet information, the content of these services is related to the specific bank-firm relationship, and therefore it is not publicly available. Although the role of credit information has acquired its relevance in the PD estimation banking models, the use of credit data in the bank-firm relationship has gone quite unnoticed among almost all literature relating to the estimate of corporate insolvency. Except for the seminal paper by Nakamura (1993) and for more recent contribution by Mester *et al.* (2007), Jiménez Lopez and

[2] In addition, 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 for instance, in terms of time, firms, and sector of economic activity (Grunert *et al.,* 2005).

Saurina (2009a, 2009b) and Norden and Weber (2010), a limited number of authors, both in the banking and business literature, study the link between bank-firm specific information and default performance. Therefore, additional analyses can be functional to strengthen the direct examination of the sources and types of information that financial institutions use to monitor credit (Matthias *et al.*, 2019) and 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). In this paper, 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. To this end, we adopt a unique and disaggregated dataset on a sample of 111 co-operative banks and 12,792 operating firms located in Italy over the period 2012-2014. The share of defaulters in the sample is about 9%, including 1,110 firms. A particular feature of our data is that the bank-firm relationship concerns many different banks operating in Italy 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. Firstly, we examine the impact of checking accounts on default rates using the share of the credit limit granted by the bank that is used by the firm, as well as information on credit limit violations, blank cheques and crediting operations. Secondly, we analyze the impact of long-term loans using the number of consecutive months of overruns and the share of overruns that are balanced by other bank-firm credit relationships. To the best of our knowledge, this paper is the first one dealing with this issue with a so large dataset, contributing to the literature in three ways. Firstly, it analyzes the marginal benefit of the bank-firm specific information, by comparing the default prediction accuracy of a model that incorporates the accounting information provided by balance sheets, controlling for firm size and system information, accounting for the share of loans granted by each co-operative bank, with that also including checking accounts and long-term loans performance. Secondly, this study covers a sample of SMEs, heterogeneous in terms of size and relational banking intensities.

This makes our study highly representative of firms mainly dependent on banks as relationship lenders and 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) and on relationship lending (e.g., duration and distance). Thirdly, provided that the data come from more than one hundred different local banks and that it is quite recent data (post crisis), we control for time varying bank characteristics, by including bank-time fixed effects. Two classes of results are derived. First, the use of the credit line and the depth of violations of the credit limit on checking accounts and long-term loans can help the bank to predict the probability of default of a borrower at least one year before and this impact is persistent over time. Second, the results also show that private information helps to increase the predictive power of balance sheets' information by 8% one year before the default event and by 7% two years earlier. These results indicate that SMEs must carefully monitor their creditworthiness by paying attention to their relationship with the bank, in terms of performance achieved on past short and long-term loans. In this perspective, financial planning plays a fundamental role in correctly managing short-term credit lines and in serving long-term debt. The rest of the paper is organized as follows. Section 2 briefly reviews the related literature. Section 3 presents the sources of data and gives an overview of their descriptive statistics. Section 4, after discussing the empirical model adopted, presents the main results distinguishing the baseline model, which includes the balance sheet indicators and the full model, including also bank-firm relationship indicators. Section 5 discusses the results for different sub-samples of firms, while section 6 concludes.

2 Literature review

This section reviews the literature related to our study that provides insights on the determinants of corporate default prediction to be included in the empirical model. The issue analysed let this paper be partially close to the traditional models in corporate finance that use financial indicators, extracted from balance sheets, to classify default and non-default firms or to assign a default probability to firms in a given time horizon. Academics in this field propose several techniques: (i) linear discriminant analysis (Beaver, 1966); (ii) multivariate discriminant analysis (Altman, 1968; Altman *et al.*, 1994; Foglia *et al.*, 1998); (iii) binomial choice models (Altman and Sabato, 2007; Bottazzi *et al.,* 2011) and (iv) modern statistical learning techniques including neural networks, support vector machines (SVMs) and tree structure classifiers (Duffie and Singleton, 2003; Sun *et al.*, 2014; Jones *et al.*, 2017; Cheng *et al.*, 2018; Jiang and Jones, 2018; Jones and Wang, 2019). The first two models are based on the a priori assumption that there are two mutually exclusive groups of firms (defaulters and non-defaulters) and that the differences between them can be captured by observing (individually) several financial ratios (Foglia *et al.*, 1998). The more recent logit and probit approaches aim instead to assign each firm a probability of default considering the combined impact of several financial indicators. Although these analyses differ in the methods adopted and datasets, they primarily identify five groups of accounting indicators as most important in predicting the default probability of firms: leverage, liquidity, profitability, and financial and asset coverage (Altman and Sabato, 2007). Bonaccorsi di Patti *et al.* (2015), for instance, show the relevance of financial variables as determinants of corporate default. In particular, the weak capital structure, characterized by a high level of bank loans, increases the default rates and the sensitivity of the default probability to macroeconomic shocks. However, in the last decades, more interest has been observed in improving the prediction power of standard models of default prediction (Grunert *et al.*, 2005). Particularly instructive for this paper is the literature attempting to capture signals to predict financial difficulties through the information coming from the specific bank-firm relationship. This literature studies the informational content of transaction accounts and justifies the need to monitor revolving credit utilization simultaneously with borrowers' default probability (Bergerès *et al.*, 2015). As a matter of fact, Nakamura (1993) develops the 'checking account hypothesis', stating that bank account activity is informative and is used to manage the bank-firm relationship. In the same vein, Mester *et al.* (2007), using a Canadian dataset and both annual and monthly data over the period 1988-1992, document that changes in transaction accounts contain relevant information and reflect changes in accounts receivables, and that the number of prior borrowings in excess of collateral is an important predictor of credit downgrades and loan write-downs. In other words, by monitoring transaction accounts, the lender might obtain the timeliest information on cash flows of the business, and it can intensify monitoring as loans deteriorate. The checking accounts are typically offered as a bundle with the line of credit (Norden and Weber, 2010). In its broadest definition a line of credit can be defined as the maximum amount of financial resources which, over a given timespan, a bank commits to lend to a borrower. Borrowers can therefore borrow and repay funds up to the maximum amount authorized, while being charged only for the portion of the line that is used (Bergerès *et al.*, 2015). The paper by Jiménez *et al.* (2009a), for instance, using the credit register database for Spanish firms, shows that defaulting firms have significantly higher credit line usage rates and line exposure at default values (ratio between actual drawn amount and a fraction of undrawn amount) up to five years before the default event occurs. In a different contribution, Jiménez *et al.* (2009b) demonstrate that firms heading into default draw on their credit lines quite heavily. Even though they consider the impact of credit line history based on usage of the credit line, they do not consider the impact of other aspects related to the checking account relationship and aspects related to long-term loans. While these studies consider a sample period of five years pre-financial crisis, our paper uses more recent data focusing on the three years after the financial crisis. Every financial crisis has important repercussions on the bank-firm relationship. The general increase in uncertainty over economic growth and the expansion of asymmetry in information between banks and borrowers increases the risk of incorrect assessments in the selection of companies to be financed. Thus, it is worth verifying the accuracy of models based on pre-crisis data, which were constructed under stable economic conditions, using more recent post-crisis data. Our paper is closely related to Norden and

Weber (2010), who directly investigate whether information on credit line usage and checking account activity helps banks to monitor borrowers and how they use this information in managing their credit relationship with both firms and consumers. They adopt German data to demonstrate that credit line usage, limit violations, and cash inflows exhibit abnormal patterns 12 months before the default event, especially for small businesses (and individuals). Although the paper considers different borrowers' categories, it is based on data from one German bank and does 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 extension of the results to the banking system as a whole or to a sub-sample of it and does not allow to fully grasp the depth and intensity of the bank-firm relationship. Using a sample of Italian firms and data on specific credit history information, Dainelli *et al.* (2013), investigate the contribution of the usage level of short-term lines of credit to increasing the prediction power of a model including only financial information. However, their results cannot be extended to the Italian banking system since data come from only one leader bank and they do not consider bank level heterogeneity and other sources of hard information. It is notable that not all previous studies document a positive effect of credit line utilization on default probability. Bergerès *et al.* (2015), for instance, document that an increase in credit line utilization for consumers involves a decrease in default probability, providing evidence that they use lines of credit to pay their term loans in periods of financial distress and that banks should manage both financial instruments simultaneously. The evidence on long-term loans is indirect. Strahan (1999) notes that a typical deal between a firm and its bank should include both a term loan and a credit line, which could help the bank gain a better understanding of the borrower's credit quality. Information on both credit instruments should be combined so that credit line drawdown behavior can be connected to the default risk assessment of a term loan, and vice versa. Taken together, the literature provides evidence that the bank has an advantage in providing deposittaking and lending services jointly. In this way, the bank has the advantage of extracting hard information from several sources that are specific to the relationship with a given firm. Information spillovers coming from past loans and checking account activities might help the bank in the monitoring process of firms that apply for a loan. Our in-depth study of credit-related variables leads to improved forecasts in terms of bankruptcy and/or firms survival compared to studies that use only financial ratios.

3 Research methodology

3.1 Data and variables

Data used for the empirical analysis combine the bank-firm specific information on checking account activities and long-term loans with firm-level financial information gathered from balance sheets. These data have been 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.[3] In order to be included in our final sample, firms need to satisfy several requirements. First, we include firms with at least a checking account or a long-term loan relationship with one of the co-operative banks of the sample, over the period 2012-2013, and with reliable balance sheet ratios, to be sure that all information needed to estimate our empirical model is available. Second, we include firms with an average turnover in the range 5 thousand and 50 million euros in order to focus the analysis only on SMEs.[4] Third, following Arcuri and Levratto (2018), firms operating in the agricultural, financial and insurance, real estate, public administration, education, social and human health sectors have been excluded from the dataset because they may be subjected to particular failure regimes. Finally, we include multiple lending firms in the banking system to be able to

[3] The original sample includes anonymous firms identified by a unique code produced by CSD – the co-operative banks' provider. Each firm might be associated to one or more co-operative banks included in our sample, depending on whether it is involved in multiple lending or not. However, the same firm might have multiple relationships in the banking system.
[4] Within the European Union, SMEs are identified considering both the number of employees and turnover (or, alternatively, their total assets). Given the availability of information of employees for a limited sub-sample of firms, we adopt the average turnover over the sample period to identify SMEs, following Ciampi *et al.* (2018) and the Basel regulations.

include as control variable the share of loans granted by each co-operative bank of our sample with respect to the total amount of loans received by a specific firm from the banking system. We develop the empirical model from the analysis sample and test this model on the validation sample, composed of different SMEs.[5] To this end, from the resulting sample, we have drawn two different sub-samples: the analysis sample – including 70% of firms – and the validation sample – including the remaining 30% of firms. Firms were extracted without replacement using a proportional stratified random sampling method, based on four characteristics: default status (bonis and default, over the sample period), the province where firm operates, the Ateco 2 digits sector of economic activity and average turnover.[6] Through this method, final units of the sample are randomly and proportionally chosen from different non-overlapping strata, such that 70% come from each stratum to compose the analysis sample and 30% to compose the validation sample. Finally, we excluded influential observations, that are those for which values of each variable included in the analysis are higher than the 99th percentile. This procedure results in a final analysis sample of 12,792 firms and 22,602 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 default: one and two years before the event. Moreover, our sample includes only a small component of multiple lending firms in our sample of co-operative banks (9% for checking accounts and 2% for long-term loans), as well as a low number of multiple lending banks (the maximum number of banks for each firm is 5 for both checking accounts and long-term loans and the 99th percentile is 3 and 2, respectively). Concerning our event variable, we find 8.7% of default firms. Since we drop all firm-year observations with default in the previous year, in order to avoid last year's default, with the corresponding business information, being a mechanical predictor of this year's default (Dierkes *et al.*, 2013), this share refers to firms that became insolvent

[5] Given the limited time span, we are not able to test the model over a different period (out-of-time) for the same (or for a different sample of) firms, as is common in the literature (see, for instance, Duarte *et al.*, 2018).
[6] We consider three different size groups: firms with an average turnover below 2 million euros, firms with an average turnover in the range between 2 million and 5 million euros and firms above 5 million euros.

either during 2013 or during 2014. This study uses an initial long list of 60 financial indicators, describing several areas of firms' profiles (leverage, profitability, liquidity and efficiency, among others).[7] Specific short term bank-firm information are instead contained in an initial set of 15 indicators on checking account activities (including credit limit violations, usage of the line of credit, debit and credit accounts, among others), whereas long-term loan performance are described by 11 indicators (overdue payments and credit limit violations, among others). Since the initial set of indicators showed very high correlation coefficients, it was reduced by a selection process conducted in two steps, that follows the literature (see, for instance, Ciampi, 2015). The initial selection derives from the variance inflation factor model (VIF), that allows us to select only those variables not affected by multicollinearity concerns in a linear regression model where the dependent variable is the default event.[8] 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 as independent variables the ratios selected in the previous step. 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 adopt the backward selection method, beginning with the model including all variables and iteratively eliminating non-significant ones. One per cent is used as the level of significance for the addition of variables to the model and 5% as the level for their removal. At the end of the selection process, we obtain a model in which all explanatory variables proved to be jointly significant. The final set of balance sheet ratios includes the return on assets ((*roa*)) defined as earnings before interests and taxes (EBIT) on total assets, as a measure profitability; the share of owner's equity over owner's equity and inventories (*equity inventory coverage*), as an indicator of

[7] These ratios have been constructed by CRIF from firms' balance sheets.
[8] 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). The utility of VIF is that it indicates the magnitude of the inflation in the standard errors associated with a beta weight that is due to multicollinearity (Ciampi, 2015).

both capitalization and inventory incidence; and the share of bank loans on total liabilities (*bank loans*), as a measure of firm's debt. In addition, four attributes reflecting checking accounts and credit lines performance are selected: the share of credit line received on a checking account that is used by the borrower in a month (*credit line usage*); the number of months of credit limit violations on a credit line, in a year (*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*)[9] and the number of crediting operations on the checking accounts (*crediting operations*), reflecting positive changes of firm's cash flows. Finally, concerning long-term loans, we selected two variables: the number of months of credit limit violations in a half year (*long-term overruns*) – i.e. the number of months in which there is a delay in payment of the loan installment – and the share of limit violations on long-term credit granted in a quarter that is balanced by other credit products of the firm with the same bank (*long-term balanced overruns*). Summary statistics are reported in Table 1, distinguishing between default and non-default firms. The table shows that our variables have a high variability between the minimum and the maximum value in both sub-samples. Concerning our balance sheet indicators, significant differences are found between the two subsamples. The average *roa* indicator shows that profits are about 0.83% of total assets, with a positive value for non-default firms (0.98%) and a negative average in the sub-sample of default firms (-0.83%), meaning that they are experiencing greater difficulties in obtaining positive economic performance. The average values of the share of net equity over net equity and inventories indicate that non-default firms show a better coverage of inventories than default firms (56% *vs* 43%), which, usually, constitutes a variable subject to greater volatility. It is interesting to note that, on average, firms in the sample make large use of bank loans, which represent on average 35% of their total liabilities, with higher values for default firms than for non-default ones (45% *vs* 34%). These indicators are also statistically different

[9] Interestingly, this indicator might signal an initial difficulty of the firm since not all blank cheques are then protested.

between the two groups. Looking at the short-term and long-term bank-firm specific contracts, similar results are obtained for the two subsamples. The credit line usage in a month is about 45% over the entire sample and, as expected, for default firms it is double (84%) than for non-default ones (42%). Borrowers that subsequently experience a default event exhibit an increasing need for liquidity, which could be due, for instance, to a decline in cash inflows because of a decline in sales (Norden and Weber 2010). However, in both subsamples, firms violate their credit lines, since the maximum credit line usage is about 190%. The average number of months of credit lines violation in checking accounts is about 1, with default firms showing three times the months of non-default firms. We found significant differences between the two subsamples also in terms of the number of blank cheques (0.14 for default firms and 0.02 for non-default ones). Finally, the number of crediting operations is higher for non-default firms (17) than for default firms (14). Concerning long-term overruns, they prove to be more problematic for default firms than for non-default. The number of months of overrun is significantly higher for default firms (0.49 *vs* 0.10), as well as the share of balanced overrun (0.06 *vs* 0.01). Concerning the control variables, firm size, measured by the *turnover*, shows average values that are statistically lower for default firms (4,717 thousand euros) than for non-default ones (4,896 thousand euros). System information are 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. Since this variable has a very skewed distribution, we choose to create three dummies for each tercile of the distribution of the share, and we use the two dummies for bank-firms in the second and third tercile as control variables. We argue that default firms show a lower share of loans.[10] Moreover, 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 subsamples have a checking account with

[10] This circumstance reasonably suggests that the cooperative credit banks included in our sample have only partially contributed to the phenomenon of "zombie lending" which tends to occur in geographical areas and economic sectors where loans are predominantly provided by weaker banks (Schivardi *et al.*, 2017).

the bank and about 47% of default firms and 39% of non-default firms have a long-term loan. Table 1 also reports descriptive statistics of variables used to split our sample of firms: the duration of the bank-firms relationship, measured in years, and the geographical distance between the bank and the firm, measured in minutes. For non-default firms the average duration is 10.46 years, significantly higher than that of default firms (10.03), and the average distance is 24 minutes, significantly lower than that of non-default firms (27 minutes).

[Insert Table 1 about here]

Table 2 reports the pairwise correlations between the six sets of variables. The figures reported confirm the evidence of Table 1. Default borrowers have 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 usage (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 and the long-term loan indicators: 0.20 for consecutive months of overruns and 0.17 with balanced overruns.

[Insert Table 2 about here]

Since these simple correlations do not consider the interrelationships between variables and the fact that some firms show industrial and localization specificities different from others, a multivariate analysis is conducted.

3.2 Econometric analysis

The present study is performed through a probit model to study which factors at time t_0 (and t_0-12) influence the probability of default at time t_{0+12}.[11] In a first set of analyses, we investigate whether

[11] The one-year prediction horizon is standard in the literature and is also required by the Basel regulations.

financial information gathered from balance sheets can be augmented by direct observations of bank-firm information to improve the prediction probability of default. We estimate 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 a second set of analyses, we investigate whether these results are different for sub-samples of firms and for different bank-firm relationship lending. We would like to assess, for instance, whether predictors exert a different impact on default probability depending on firm size, age, and on bank-firm distance and duration. Following several studies in the literature (Norden & Weber, 2010; Dierkes *et al.*, 2013; Duarte *et al.*, 2018), the econometric approach relies on a binomial probit model where the dependent variable, *default*, is a solvency condition at time t_{0+12} taking the value of 1 in case the bank *b* registers one of the following event for firm *f*: *unlikely-to-pay*, that is a temporary difficulty of the firm, that might be remedied; *bad loan*, that is a non-performing status that has not been ascertained; and a *forborne non-performing* loan for which the conditions (such as maturity and interest rate) have been re-negotiated with the firm.[12] The dependent variable assumes the value zero 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:

$$\Pr(default_{t_0+12}^{fb}=1)=\Phi(\beta_1 financial_{t_0}^{f}+\beta_2 size_{t_0}^{f}+\beta_3 system_{t_0}^{fb}+\eta^{p}+\lambda^{s}+v_{t_0}^{b}) \qquad (1)$$

where subscripts *f*, *b*, *p*, *s*, and *t*, are indicative of firm (*f* = 1,…., 12,792), bank (*b* = 1,…, 111), province (*p* = 1,…, 60), sector (*s* = 1,…, 60), 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*); and the share of net equity over net equity and inventories (*equity inventory coverage*), as described in section 3. Our specification allows to control for potential endogeneity problems as well as for omitted variables. First, the *default* indicator is equal to 1 for firm experiencing an event over

[12] It should be noticed that our definition does not include 90 days past-due.

the period t_{0+12}, whereas the financial information refers to t_0, that is to 2012 for defaults over 2013 and to 2013 for defaults over 2014. Moreover, the specification controls for firm size (Dierkes *et al.*, 2013; Fiordelisi *et al.*, 2014), proxied by the turnover, and includes two dummies for the second and third tercile of the share of bank loans. For Altman *et al.* (2017), the size of the firm contributes differently to the determination of failed and non-failed companies. Finally, to rule out omitted variable issues, the specification includes three sets of fixed effects: province fixed effects, to account for local context characteristics affecting firm performance; sector fixed effects, to account for firms unobserved, time-varying, loan demand characteristics and quality shocks; and bank-year fixed effects, to control for banks' time varying unobserved characteristics that might affect the borrower behavior. Using bank-year fixed effects we are comparing how firms' default probability, in 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:[13]

$$\begin{aligned}\Pr(default_{t_0+12}^{fb}=1)=\Phi(&\beta_1 financial_{t_0}^{f}+\beta_2 size_{t_0}^{f}+\beta_3 system_{t_0}^{f}+\beta_4 flag\_ca_{t_0}^{fb}+\\&+\beta_5 bank\_firm\_ca_{t_0}^{fb}+\beta_6 flag\_ltl_{t_0}^{fb}+\beta_7 bank\_firm\_ltl_{t_0}^{fb}+\eta^{p}+\lambda^{s}+\nu_{t_0}^{b})\end{aligned} \tag{2}$$

where the checking account set includes *credit line usage*, *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 Equation (1). To evaluate the contribution of short-term and long-term bank-firm hard information in predicting the probability of default, we perform the receiver operating characteristics (ROC)

[13] We also estimate an augmented model that augments the baseline model including checking account factors.

analysis and compare the areas under the ROC curve (AUC) of the baseline and full models, as an index of accuracy. As stated by DeLong *et al.* (1988), indeed, the overall value of a test based on an observed variable that lies on a continuous or graded scale can be made using a ROC curve. Over alternative methods of evaluating a wide range of diagnostic systems, including credit risk assessments, the ROC has many advantages (Irwin and Irwin, 2013): it is well suited to empirical data, which in turn may reflect the nature of the underlying rating process; it relies on solid theoretical bases in setting the optimal threshold; it does not depend on the probability of default. 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 (that is, the true-positive rate) against the specificity (that is, the true-negative rate). If a test could perfectly discriminate, the area under the ROC curve 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).

4 Empirical results

This section discusses the results obtained from probit regressions of the default indicator on predictor variables. Accordingly, the probability of default is estimated in the year prior to the observation of corporate financial distress (t_0) as well as two years prior to the financial distress event (t_0–12). Results are reported in Table 3, distinguishing the *baseline model* (Equation 1), including firm level balance sheets indicators along with control variables, the *augmented model*, and the *full model* (Equation 2), which includes all additional information on the bank-firm contracts. As required by the probit regression model, the dependent variable takes the value of 1 for firms classified as defaulters at time t_{0+12} and the value of 0 for firms classified as non-defaulters. Table 3 reports marginal effects 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). Column (1) of Table 3 shows that return on assets has a negative and significant impact on the probability of default, as expected. Its marginal effect indicates that an increase in firm profitability from the value at the 25th percentile (-0.22%) to that at the 75th percentile (1.97%) 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 impact is also economically significant. Concerning the financial structure, we observe a positive impact of *bank loans* on the default probability, indicating that, other things equal, if a firm becomes more dependent on bank loans, its probability of default increases by 2.7%. Therefore, consistent with previous studies, this result shows that the financial structure is an important determinant of SMEs ability to repay loans. However, a decrease of the *equity inventory coverage* ratio from the level at the 75th percentile (85%) to that at the 25th percentile (26%) determines an increase of the default probability of 2.5%. Within current assets, inventories are the item that suffers from superior volatility due to both its intrinsic nature and the effects of accounting techniques. Whatever the reason for its volatility, greater coverage of inventories with net equity contributes to strengthening the company's financial soundness and produces positive effects on its creditworthiness. In addition, the results on firm size indicate, that for the same bank and other things equal, firm size is not statistically significant. Finally, introducing the 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 sum, controlling for firm size, system information, and for the fixed effects described in section 3, all balance sheet indicators prove not only to be economically relevant, but also highly statistically significant (at the 1% level) in predicting the default event 12 months before. This result is consistent with the empirical contributions summarized in the to the literature review. Additionally, if we include the short-term relationship indicators in the baseline model (column 2), we obtain three relevant results. Firstly, the impact of balance sheet indicators on the default probability is not affected by the bank-firm relationship, at least in terms of sign and significance, while the magnitude is reduced in some cases. 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, moving from firms in the 25th percentile (with an average turnover of 880 thousand euros) to firms in the 75th percentile (with an average turnover of 5,290 thousand euros) the default probability increases by 1.1%. Secondly, the impact of all five checking account factors in the model (*dummy checking account*, *credit line usage*, *credit limit violation*, *blank cheques* and *crediting operations*) is highly significant (at the 1% level). Specifically, our results indicate that SMEs that have a checking account show a probability of default that is 12% lower compared to that of SMEs that do not have a checking. Moreover, increasing the credit line usage from the value at the 25th percentile (0%) to that at the 75th percentile (90%) produces an increase in the default risk of about 5.5%. In other words, a firm that is close to reaching the credit limit allowed by the bank 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 positive impact on the default probability. Increasing the number of consecutive credit limit violation months from zero (the 25th percentile) to one (the 75th percentile), the probability of default increases by about 1%. A higher impact is found when the number of blank cheques changes from the 25th percentile (zero) to the 99th percentile (1): the default probability increases indeed by 3.5%. Finally, crediting operations, that reflect cash inflows, reduce the default probability by about 1% when its number changes from 0 to 20. As a whole, these findings indicate that, controlling for quantitative information obtained from balance sheets, checking accounts help the bank to predict future SMEs defaults, consistent with the findings of Norden and Weber (2010). Thirdly, it should be noticed that considering checking account activities helps to improve the accuracy of default predictions with respect to a model that includes balance sheets alone. The McFadden R^2 indeed more than quadruple when we consider additional checking account information factors to the baseline model. This means that the bank obtains additional information on the firm's ability to repay loans when it considers the checking account activities, other than the balance sheets indicators. Column (3) of Table 3 reports the results obtained augmenting the previous specification with long-term loan factors. SMEs that obtain a long-term loan show a probability of default that is 0.3% higher than that of SMEs that do not obtain long-term loans. However, this marginal effect is estimated with less precision and is not statistically significant. Moreover, a higher number of consecutive months of overruns in half year (from zero to 3) increases the probability of default by 6.5%, while the impact of the share of balanced overruns is lower, increasing the probability by 1.7% (for a change from zero to 0.37%). The last result indicates that even if the firm has long-term loan overruns offset by other relationships with the bank, its probability of default is higher. The reason lies in the fact that, despite the compensation, the failure to pay the long-term loan is a sign of the difficulty of the firm to generate sufficient operating cash flows to serve the repayment of the debts incurred. Other coefficients remain unchanged in terms of their economic impact and significance. In this case the McFadden R^2 increases to 16%, demonstrating that the full model has an explanatory power that is about four times that of the baseline model.[14] Furthermore, we consider a different prediction horizon and we report the results for factors at time $t_0\text{-}12$ influencing default at t_0+12 in columns 4–6 of Table 3. 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 impact on the probability of default that goes beyond the balance sheet factors, size and system information—remains confirmed. This indicates that our key results are persistent over time. Reassuringly, marginal effects and the significance of our factors remain almost unchanged in the full

[14] Very similar results have been obtained by estimating the augmented model on the restricted subsample of firms with both a checking account and a long-term loan relationship with the bank. Results are available on request.

model. However, considering a longer horizon, the flag for checking accounts and the crediting operations are estimated with less precision and are not economically and statistically significant. A more appropriate and direct measure of the real performance of a probit model is the area under the ROC curve. Figure 1 reports the comparison of AUCs of the baseline and full models. The baseline model at time t_0 provides an accuracy ratio of 84%, while the full model stands at 92%. The difference in the accuracy ratio between the full model and the baseline model 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 indicates that the null hypothesis of equality of areas under the ROC curve can be rejected. In other words, the small *p*-value suggests that the two areas are statistically significant. This finding offers support to our expectation that bank-firm specific hard information helps to improve the accuracy prediction of balance sheet indicators. The AUC indicates also that the baseline model at t_0*-12* gives an accuracy ratio of 81% and the full model 88%. The p-value, equal to 0.000, also suggests in this case that the two areas are statistically different.[15]

[Insert Table 3 about here]

[Insert Fig. 1 about here]

5 Sample split

In previous section, we have demonstrated that including bank-firm variables in the baseline model not only provides additional information that the bank can use in the short-run to monitor and to predict the default events of SMEs, but also increases its prediction accuracy. However, in this section we investigate whether the previous results hold in the sub-samples of firms and relational banking intensities. As a matter of fact, Norden and Weber (2010) argue that "the complexity of cash flows in the checking accounts, the likelihood of having multiple banking relationships, and the

[15] 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 a limited additional information. The accuracy ratio, in this case, increases by 2%, one and two years before. This strengthens our previous results that balance sheet information are not fully informative and they require additional bank-firm specific information to predict future corporate default events.

mechanism of default differ considerably across these borrower types". For instance, larger firms might benefit from the use of different sources of funding and a higher number of bank relationships, whereas SMEs rely primarily on bank loans as a source of funding, and this relationship is usually exclusive. As a result, despite the information contained in the balance sheets might be more reliable for larger firms, the information content of the bank-firm relationship is more accurate in the case of small business borrowers. However, the factors that affect the probability of SMEs default according to their size have not been extensively investigated. Analyzing a sample of US SMEs, Gupta *et al.* (2018) document that SMEs are not all equal and the determinants of bankruptcy are different between the different categories of SMEs (micro, small and medium size). Furthermore, the magnitude of the coefficients of the most significant variables in predicting the probability of failure of micro, small or medium enterprises respectively show significant differences in the models. In order to explore the differences amongst the sub-categories of SMEs, we divide firms according to their average turnover over the observation period. Small firms are defined as those with an average turnover of less than 5 million euros and medium-sized firms are those with an average turnover of over 5 million euros. The results, reported in Table 4, highlight some differences between the two categories. The sub-sample of small firms accounts for 16,258 observations, while that of large firms includes about one third of firm-year observations. 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 1.2% if *roa* increases, which is four times the impact 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 case of medium firms and by 1% for small firms. The equity inventory coverage exerts a negative and significant impact (1.7%) on the default probability only for small firms. These results are consistent with the hypothesis that the information extracted from balance sheets are more reliable for medium than for small firms and therefore are more useful in monitoring their future payment behavior. Among checking account activities, the usage of credit line, the months of credit limit violation and blank cheques are relevant for both small and medium firms in a model controlling for balance indicators, for system information, and for fixed effects. The marginal effects in both sub-samples are in line with those of the entire sample. It is worth noting 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 impact on default probability of small firms. A plausible explanation of this result is that a strong tie with and individual bank represents for a small firm a way to amplify the bargaining power, and as a result, have greater availability of credit and lower costs (Grunert and Norden, 2012). Crediting operations are found to be economically significant only for small firms, reducing the default probability by 1.4%. 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 the default probability by 7% and 5%, respectively in the two sub-samples. The share of long-term loans balanced overruns increases the probability of small firms by 1.7% and that of medium firms by 2.5%. In line with the conclusions of Gupta *et al.* (2018), these results seem to confirm the importance that the 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 modeled is through the duration of lending relationships, as the 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; Omiccioli and Carmignani, 2007; Norden and Weber, 2010; Fiordelisi *et al.*, 2014). In the bank-firm relationship context, variables that capture the intensity and duration of lending relationship are found to affect the credit supply

(Agarwal and Hauswald 2010), the cost of borrowing (Bharath *et al.*, 2011) and the small business liquidity (Han *et al.*, 2017) which, in turn, may affect the probability of default. To better understand the impact of the intensity of the relationship between bank and firm on default risk, our work, in addition to the specification linked to the dimensional characteristics of the firm, investigates the intensity of the relationship measuring it in two ways: a) the geographical distance between the local bank and the firm, and b) the duration of their relationship. While the geographic proximity improves the quality of private information (Agarwal and Hauswald, 2010), the closeness in the business relationship helps to capture signals to predict financial difficulties through the lenses of the lending relationship. In this perspective, our sample of firms has been divided into two groups, depending on the median physical distance between the bank and the firm (1 hour) and on the median duration (8 years). The results for short duration and long duration bank-firm relationships are displayed in columns 3 and 4, respectively, of Table 4. It is worth noting that the magnitude of *roa* and *bank loans* marginal effect increases, in absolute value, as we move from short to long duration bank-firm relationships, moving, respectively, from 0.4% to 0.5% and from 1% to 2%. The share of loans of a bank with respect to the banking system reduces the default probability in both subsamples, but its impact is higher in case of long bank-firm relationships. On the contrary, the impact of checking account activities reduces as we move from short to long duration, with the exception of crediting operations. This result indicates that as the intensity of the bank-borrower relationship increases, the increase in credit line usage, credit limit violations and blank cheques exert a positive but lower impact on default probability. The impact of long-term loans indicators, both in terms of months of overruns and balanced overruns, is positive and is very similar in both bank-firm subsamples of short and long duration. Considering distance as an alternative measure of the intensity of the bank-firm relationship, we notice that for small distance relationships all the results obtained on the whole sample are confirmed. However, the balance sheet indicators are not statistically significant in predicting the default probability of large distance firms. A higher distance not only reduces the availability of soft information, as pointed out by the literature, but it also weakens the positive effects of a relational approach between bank and firm that allows to overcome the problems of information opacity and to mitigate the procyclical behavior of banks (Grunert *et al.*, 2005). Checking account activities exert a significant impact on the default probability of both small and large distance bank-firm that is higher in the case of large distance. The impact of credit line usage and blank cheques on the default probability of large distance firms is about twice that of small distance ones (9% vs 5.3% and 7% vs 3%, respectively). Similarly, credit limit months of violation increases by 1.4% the default of large distance firms and by 0.8% in the small distance ones.

[Insert Table 4 about here]

6 Discussion and Conclusions

For small and medium-sized enterprises the main source of financing 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. This paper provides 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 including only accounting information. This analysis is based on a unique dataset of a sample of 111 co-operative credit banks and more than 12,700 firms operating in Italy between 2012 and 2014.[16] This paper provides several marginal contributions to the literature that observes the links between the probability of business insolvency and the short and medium-term credit line utilization rates (Nakamura, 1993; Mester *et al.*, 2007; Jiménez *et al.*, 2009a, 2009b; Norden and Weber, 2010; Nakamura and Roszbach, 2018). First, it adds additional explanatory and discriminatory power of

[16] Continuing to investigate the bank-firm relationship is important in light of the crucial role of the SMEs in the global economy, of the recent Basel regulatory changes, of the still latent effects of the financial crisis and of new challenges, driven by technological evolution, which will affect the two players in the near future.

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, very different in size and type of activity, and highly dependent on banks. Third, thanks to the depth of the data contained in the database, the paper is able to check for time-varying characteristics of the banking relationship (e.g. duration and distance of the relationship), including the bank-year fixed effects. Finally, it uses a complete post-crisis data set that eliminates the selection bias of previous studies. By estimating a probit model, we obtain two main results. First, bank-firm specific information that is obtained from checking account activities and long-term loan performance have a significant impact on the probability of default that goes beyond the balance sheet indicators. The use of the credit line and the depth of credit limit violations on current accounts and long-term loans raise the bank's ability to predict the probability of a borrower's insolvency at least a year before. The impact of additional private information 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 8% in a one-year horizon and by 7% in a two-year horizon. The combination of credit line usage, checking account activity and long-term loans performance effectively gives the lender a real-time window into the borrower's cash flows, confirming and extending the results of Norden and Weber (2010). These results provide useful implications for both SMEs and banks. SMEs must closely monitor their creditworthiness, paying attention to the management of the relationship with the bank, with particular focus on overruns and past due. The knowledge of variables that influence their rating, requires firms to increase the accuracy of short- and medium-term financial planning and to improve the financial communication with banks. At the same time, the combination of the use of bank-firm relationships is confirmed to be a valid tool for banks to intercept the borrower's ability to properly serve the debt. The spillovers of information from past loans and current account activities is a concrete advantage in better managing their lending relationships with SMEs. However, this work has some limitations. The major constraint is that the paper doesn't capture much time dynamics. For the 1-year default prediction, it only has data for two-time spots, i.e., firm data in 2012 for default prediction through 2013, and firm data in 2013 for default prediction through 2014. Without much training with 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 circumstance could limit the extension of the results to different industrial and financial systems. However, the source of data from several banks and the large sample of companies mitigates the aforementioned limit and makes it possible to generalize the results to those economies where the relationship between SMEs and local banks is significant. Ultimately, there might be other important variables that could influence the probability of corporate default. For instance, local and institutional conditions could impact the bank-firm relationship context which, in turn, may affect the probability of default.

References

Agarwal, S., and R. Hauswald (2010), 'Distance and private information in lending', *The Review of Financial Studies*, Vol. 23, No. 7, pp. 2757-88.

Agarwal, V., and R. Taffler (2008), 'Comparing the Performance of Market-Based and Accounting-Based Bankruptcy Prediction Models', *Journal of Banking and Finance*, Vol. 32, No. 8, pp. 1541-51.

Agostino, M., F. Gagliardi and F. Trivieri (2011), Bank competition, lending relationships and firms' default risk. An investigation on Italian SMEs', *International Small Business Journal*, Vol. 30, No. 8, pp. 907-43.

Alessandrini, P., A. F. Presbitero and A. Zazzaro (2009), 'Geographical organization of banking systems and innovation diffusion', in P. Alessandrini, M. Fratianni and A. Zazzaro (ed.), The Changing Geography of Banking and Finance, Springer, New York, NY, pp. 75-108.

Altman, E. I. (1968), 'Financial ratios, discriminant analysis and the prediction of corporate bankruptcy', *The Journal of Finance*, Vol. 23, No. 4, pp. 589-609.

Altman, E. I., and G. Sabato (2007), 'Modelling credit risk for SMEs: Evidence from the US market', *Abacus*, Vol. 43, No. 3, pp. 332-57.

Altman, E. I., M. Iwanicz-Drozdowska, E. K. Laitinen and A. Suvas (2017), 'Financial Distress Prediction in an International Context: A Review and Empirical Analysis of Altman's Z-Score Model', *Journal of International Financial Management & Accounting*, Vol. 28, No. 2, 131-71.

Altman, E. I., G. Marco and F. Varetto (1994), 'Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience)', *Journal of Banking and Finance*, Vol. 18, No. 3, pp. 505-29.

Arcuri, G., and N. Levratto (2018). 'Early stage SME bankruptcy: does the local banking market matter?', *Small Business Economics*, https://doi.org/10.1007/s11187-018-0042-4

Bank of Italy (2019), 'The 2018 Annual Report', *Bank of Italy*, 31 May, retrieved from https://www.bancaditalia.it/pubblicazioni/relazione-annuale/2018/

Barba Navaretti, G., G. Calzolari and A. Pozzolo (2017). 'Fintech and Banks: friends or foes?', *European Economy, Banks, Regulation, and the Real Sector*, No. 2, pp. 9-50.

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

Bergerès, A. S., P. d'Astous and G. Dionne (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*, Vol. 33, pp. 276-86.

Bharath, S. T., S. Dahiya, A. Saunders and A. Srinivasan (2011), 'Lending Relationships and Loan Contract Terms', *The Review of Financial Studies*, Vol. 24, No. 4, pp. 1141-203.

Bonaccorsi di Patti, E., A. D'Ignazio, M. Gallo and G. Lopes Micucci (2015), 'The Role of Leverage in Firm Solvency: Evidence from Bank Loans', *Italian Economic Journal*, Vol. 1, No. 2, pp. 253-86.

Bottazzi, G., M. Grazzi, A. Secchi and F. Tamagni (2011), 'Financial and economic determinants of firm default', *Journal of Evolutionary Economics*, Vol. 21, No. 3, pp. 373-406.

Carpenter, R., and B. Petersen (2002), 'Capital Market Imperfections, High-Tech Investment, and New Equity Financing', *The Economic Journal*, Vol. 112, No. 477, pp. F54-F72.

Cheng, C., S. Jones and W. Moser (2018), 'Abnormal trading behavior of specific types of shareholders before US firm bankruptcy and its implications for firm bankruptcy prediction', *Journal of Business Finance and Accounting*, Vol. 45, No. 9-10, pp. 1100-1138.

Ciampi, F. (2015), 'Corporate governance characteristics and default prediction modeling for small enterprises. An empirical analysis of Italian firms', *Journal of Business Research*, Vol. 68, No. 5, pp. 1012-25.

Ciampi, F., V. Cillo and F. Fiano (2018), 'Combining Kohonen maps and prior payment behavior for small enterprise default prediction', *Small Business Economics*, https://doi.org/10.1007/s11187-018-0117-2.

Dainelli, F., F. Giunta and F. Cipollini (2013), 'Determinants of SME credit worthiness under Basel rules: The value of credit history information', *PSL Quarterly Review*, Vol. 66, No. 264, pp. 21-47.

DeLong, E. R., D. M. DeLong and D.L. Clarke-Pearson (1988), 'Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach', *Biometrics*, Vol. 44, No. 3, pp. 837-45.

Dierkes, M., C. Erner, T. Langer and L. Norden (2013), 'Business credit information sharing and default risk of private firms', *Journal of Banking and Finance*, Vol. 37, No. 8, pp. 2867-78.

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

Duan, J. C., B. Kim, W. Kim and D. Shin (2018), 'Default probabilities of privately held firms', *Journal of Banking and Finance*, Vol. 94, pp. 235-50.

Duarte, F. D., A. P. M. Gama and M. A. Gulamhussen (2018), 'Defaults in bank loans to SMEs during the financial crisis', *Small Business Economics*, Vol. 51, No. 3, pp. 591-608.

Duffie, D., and K. J. Singleton (2003), *Credit Risk, Pricing, Measurements, and Management*, Princeton University Press, Princeton, NY.

European Union, (2018), 'Annual Report on European SMEs 2017/2018', *EU publications*, 14 November, retrieved from https://publications.europa.eu/en/publication-detail/-/publication/a435b6ed-e888-11e8-b690-01aa75ed71a1/language-en.

Fiordelisi, F., S. Monferrà and G. Sampagnaro (2014), 'Relationship lending and credit quality', *Journal of Financial Services Research*, Vol. 46, No. 3, pp. 295-315.

Foglia, A., S. Laviola and P. M. Reedtz (1998), 'Multiple banking relationships and the fragility of corporate borrowers', *Journal of Banking and Finance*, Vol. 22, No. 10-11, pp. 1441-56.

Grunert, J., and L. Norden (2012), 'Bargaining power and information in SME lending', *Small Business Economics*, Vol. 39, No. 2, pp. 401-17.

Grunert, J., L. Norden and M. Weber (2005), 'The role of non-financial factors in internal credit ratings', *Journal of Banking and Finance*, Vol. 29, No. 2, pp. 509-31.

Gupta, J., M. Barzotto and A. Khorasgani (2018), 'Does size matter in predicting SMEs failure?', *International Journal of Finance & Economics*, Vol. 23, No. 4, pp. 571-605.

Han, L., S. Zhang and F. J. Greene (2017), 'Bank market concentration, relationship lending, and small business liquidity', *International Small Business Journal*, Vol. 35, No. 4, pp. 365-84.

Irwin, R. J., and T. C. Irwin (2013), 'Appraising credit ratings: Does the CAP fit better than the ROC?', *International Journal of Finance & Economics*, Vol. 18, No. 4, pp. 396-408.

Jiang, Y. and S. Jones (2018), 'Corporate Distress Prediction in China: A Machine Learning Approach', *Accounting and Finance*, Vol. 58, No. 4, pp. 1063-1109.

Jiménez, G., J. A. Lopez and J. Saurina (2009a), 'Calibrating exposure at default for corporate credit lines', *Journal of Risk Management in Financial Institutions*, Vol. 2, No. 2, pp. 121-29.

Jiménez, G., J. A. Lopez and J. Saurina (2009b), ‘Empirical analysis of corporate credit lines’, *Review of Financial Studies*, Vol. 22, No. 12, pp. 5069-98.

Jones, S., D. Johnstone and R. Wilson (2017), ‘Predicting Corporate Bankruptcy: An Evaluation of Alternative Statistical Frameworks’, *Journal of Business Finance & Accounting*, Vol. 44, No. 1-2, pp. 3-34.

Jones, S., and G. Wang (2019), ‘Predicting private company failure: A multi-class analysis’, *Journal of International Financial Markets, Institutions and Money*, Vol. 61, Issue C, pp. 161-188.

Keasey, K., and R. Watson (2006), ‘Non-Financial Symptoms and the Prediction of Small Company Failure: A Test of Argenti’s Hypotheses’, *Journal of Business Finance & Accounting*, Vol. 14, No. 3, pp. 335-54.

Matthias, M., M. Giammarino and G. Gabbi (2019), ‘Modeling hard and soft facts for SMEs: Some international evidence’, *Journal of International Financial Management & Accounting*, Vol. 30, No. 3, pp. 203-22.

Mester, L. J., L. I. Nakamura and M. Renault (2007), ‘Transactions accounts and loan monitoring’, *Review of Financial Studies*, Vol. 20, No. 3, pp. 529-56.

Nakamura, L. (1993), ‘Commercial bank information: Implications for the structure of banking’, in M. Klausner and L. J. White (ed.), Structural Change in Banking, Business One/Irwin, Homewood, IL, pp. 131-60.

Nakamura, L. I., and K. Roszbach (2018), ‘Credit ratings, private information, and bank monitoring ability’, *Journal of Financial Intermediation*, Vol. 36, pp. 58-73.

Norden, L., and M. Weber (2010), ‘Credit line usage, checking account activity, and default risk of bank borrowers’, *Review of Financial Studies*, Vol. 23, No. 10, pp. 3665-99.

Omiccioli, M., and A. Carmignani (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., and R. G. Rajan (1994), ‘The benefits of lending relationship’, *The Journal of Finance*, Vol. 49, No. 1, pp. 3-37.

Roggi, O., E. Altman, G. Sabato and A. Giannozzi (2013), ‘Building SME rating: is it necessary for lenders to monitor financial statements of the borrowers?’, *Bancaria*, Vol. 10, pp. 54-71.

Schivardi, F., E. Sette and G. Tabellini (2017), ‘Credit misallocation during the European financial crisis’, *Bank for International Settlement*, retrieved from https://www.bis.org/publ/work669.pdf.

Shin, K. S., and Y. J. Lee (2002), ‘A genetic algorithm application in bankruptcy prediction modeling’, *Expert Systems with Applications*, Vol. 23, No. 2, pp. 321-28.

Shin, K. S., T. S. Lee and H. J. Kim (2005), ‘An application of support vector machines in bankruptcy prediction model’, *Expert Systems with Applications*, Vol. 28, No. 1, pp. 127-35.

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.

Sun, J., H. Li, Q-H. Huang and K-Y. He (2014), ‘Predicting financial distress and corporate failure: a review from the state-of-the-art definitions, modeling, sampling, and featuring approaches’, *Knowledge-Based Systems*, Vol. 57, pp. 41-56.

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

Udell, G. F. (2009), ‘Financial innovation, organizations and small business lending’, in P. Alessandrini, M. Fratianni and A. Zazzaro (ed.), The Changing Geography of Banking and Finance, Springer, Boston, MA, pp. 15-26.

Welch, B. L. (1947), ‘The generalization of ‘student’s’ problem when several different population variances are involved’, *Biometrika*, Vol. 34, No. 1/2, pp 28–35.

Table 1 Summary statistics

	default firms				non-default firms				
	mean	cv	min	max	mean	cv	min	max	t-test
default									
default (dummy)	0.67	0.71	0	1					
balance sheet indicators									
roa (share)	-0.83	-4.70	-45.07	22	0.98	4.80	-59.73	25.27	18.54***
equity inventory coverage (share)	42.62	0.77	0	100	55.70	0.57	0	100	16.35***
bank loans (share)	45.14	0.40	0	94.89	34	1	0	95	-24.67***
firm characteristics									
turnover (thousands)	4,717	1	5	49,400	4,896	1.50	5	49,700	1.08
system information									
share of loans (second tercile)	0.48	1.05	0	1	0.58	0.85	0	1	8.27***
share of loans (third tercile)	0.12	2.77	0	1	0.14	2.43	0	1	3.67***
credit line and checking account									
dummy checking account	0.98	0.13	0	1	0.99	0.11	0	1	1.63*
credit line usage (share)	84.14	0.48	0	189.80	41.68	1.05	0	190.33	-42.84***
credit limit violation (months)	2.05	1.06	0	10	0.71	1.90	0	10	-25.86***
blank cheques (number)	0.14	3.50	0	3	0.03	7.70	0	3	-9.56***
crediting operations (number)	13.89	1.56	0	189	17.83	1.49	0	189	7.27***
long-term loan									
dummy long-term loan	0.47	1.07	0	1	0.39	1.25	0	1	-6.30***
long-term overruns (months)	0.49	2.17	0	5	0.11	4.05	0	5	-15.23***
long-term balanced overruns (share)	0.06	3.52	0	3.25	0.01	7.06	0	2.29	-10.22***
relationship lending									
duration (years)	10.03	1.10	1	113	10.46	1.05	0	114	1.61**
distance (minutes)	26.65	1.51	0	596.75	23.59	1.33	0	533.21	-3.05***

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.

Table 2 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	0.05	-0.01	-0.16	-0.47	1								
(8)	dummy checking account	-0.01	0.01	-0.05	-0.06	0.00	0.04	0.00	1							
(9)	credit line usage (share)	0.22	-0.22	-0.17	0.31	-0.17	-0.04	0.02	0.11	1						
(10)	credit limit violation (months)	0.23	-0.15	-0.13	0.19	-0.11	-0.03	0.06	0.06	0.45	1					
(11)	blank cheques (number)	0.10	-0.03	-0.04	0.02	-0.03	0.01	0.01	0.02	0.15	0.18	1				
(12)	crediting operations (number)	-0.04	0.01	-0.03	-0.07	0.14	0.04	0.12	0.08	-0.05	-0.01	0.04	1			
(13)	dummy long term loan	0.04	-0.03	0.00	0.19	-0.02	0.06	0.18	-0.14	0.03	0.07	0.03	0.12	1		
(14)	long-term overruns (months)	0.20	-0.06	-0.04	0.14	-0.04	-0.01	0.08	-0.17	0.15	0.25	0.08	-0.01	0.32	1	
(15)	long-term balanced overruns (share)	0.17	-0.04	-0.01	0.10	-0.02	-0.01	0.04	-0.18	0.10	0.16	0.03	-0.03	0.19	0.63	1

Table 3 Impact of financial indicators and bank-firm relationships on the default probability

	t_0			t_0 - 12		
	(1)	(2)	(3)	(4)	(5)	(6)
roa (share)	-0.006 ***	-0.004 ***	-0.004 ***	-0.007 ***	-0.005 ***	-0.005 ***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
equity inventory coverage (share)	-0.025 ***	-0.014 ***	-0.014 ***	-0.026 ***	-0.014 **	-0.015 ***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
bank loans (share)	0.027 ***	0.015 ***	0.013 ***	0.030 ***	0.021 ***	0.020 ***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
turnover (thousand)	-0.001	0.012 ***	0.011 ***	0.008 **	0.015 ***	0.014 ***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
share of loans (second tercile)	-0.022 ***	-0.013 ***	-0.016 ***	-0.025 ***	-0.021 ***	-0.024 ***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
share of loans (third tercile)	-0.025 ***	-0.020 ***	-0.026 ***	-0.019 ***	-0-022 ***	-0.027 ***
	(0.00)	(0.00)	-0.010	(0.00)	(0.00)	-0.010
dummy checking account		-0.119 ***	-0.058 ***		-0.059 *	-0.032
		(0.03)	(0.02)		(0.05)	(0.04)
credit line usage (share)		0.055 ***	0.053 ***		0.039 ***	0.038 ***
		(0.00)	(0.00)		(0.01)	(0.01)
credit limit violation (months)		0.010 ***	0.009 ***		0.012 ***	0.011 ***
		(0.00)	(0.00)		(0.00)	(0.00)
blank cheques (number)		0.035 ***	0.032 ***		0.078 ***	0.072 ***
		(0.01)	(0.01)		(0.02)	(0.02)
crediting operations (number)		-0.010 ***	-0.009 ***		-0.001	0.001
		(0.00)	(0.00)		(0.00)	(0.00)
dummy long-term loan			0.003			0.006
			(0.00)			(0.01)
long-term overruns (months)			0.065 ***			0.027 **
			(0.01)			(0.01)
long-term balanced overruns (share)			0.017 ***			0.019 ***
			(0.01)			(0.01)
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	22,602	22,602	22,602	7,886	7,886	7,886
McFadden R^2	0.118	0.236	0.254	0.143	0.222	0.235

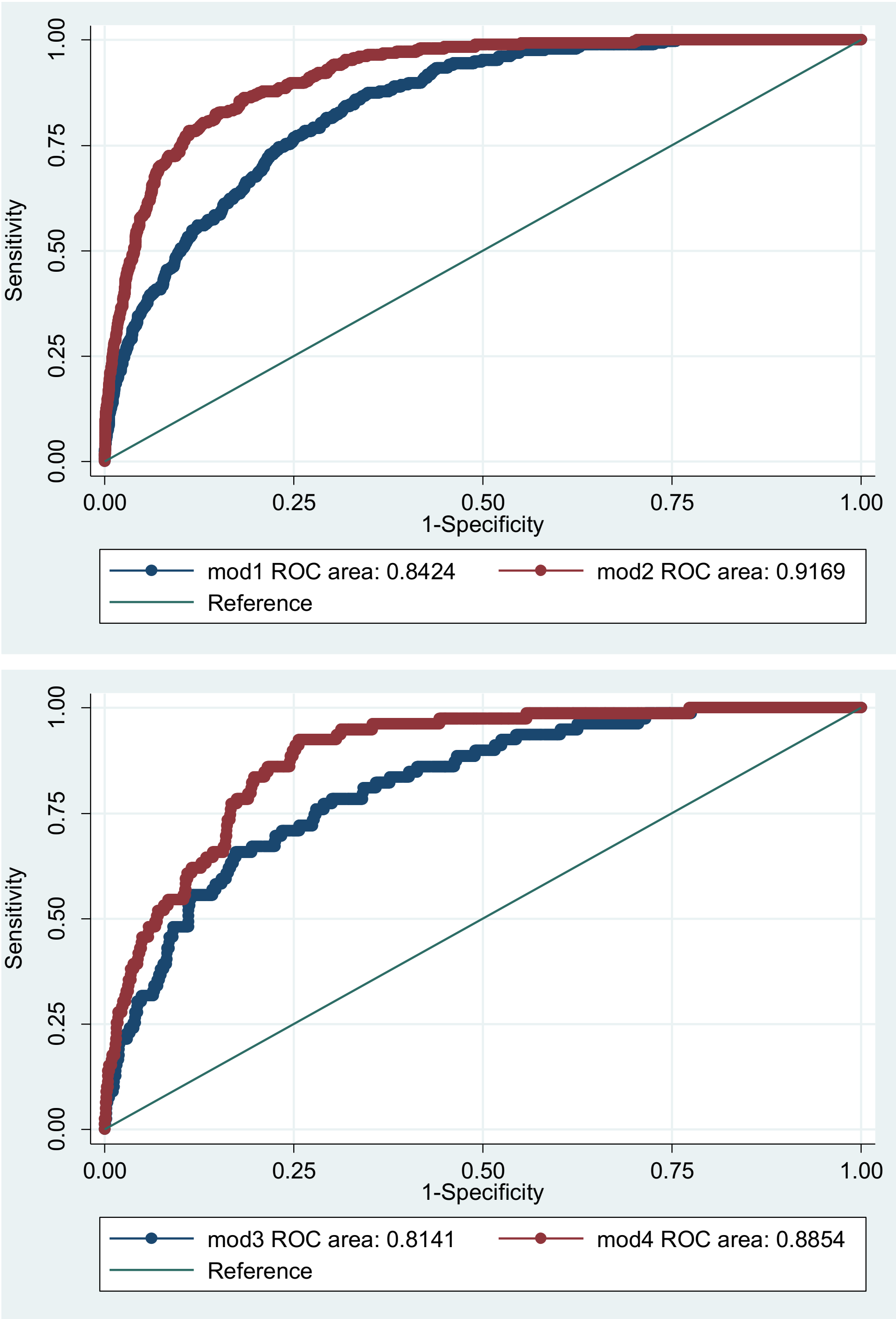
This table reports the results of the pooled probit regressions estimated on the analysis sample. The firm default indicator, default, at time t0+12 is the dependent variable. Model (1) is the baseline model, including balance sheet indicators at time t0 (or t0-12) and (unreported) bank-year, industry and province fixed effects. Model (2) is the augmented model, including the checking account indicators at time t0 (or t0-12). Model (3) is the full model, with the long term-loan indicators at time t0 (or t0-12); turnover, blank cheques and crediting operations are in logarithms. The table reports marginal effects calculated as variations in the default probability after a change in the independent variable from the value at the 25th percentile to that at the 75th (or 90th) percentile. Standard errors, clustered within firms, are reported in parentheses; * indicates significance at the 10% level; ** indicates significance at the 5% level; *** indicates significance at the 1% level.

Table 4 Probit estimation results for different types of firm and bank-firm relationship, one year before default

	small firms		medium firms		short duration		long duration		small distance		large distance	
	(1)		(2)		(3)		(4)		(5)		(6)	
roa (share)	-0.003	***	-0.012	***	-0.004	***	-0.005	***	-0.004	***	-0.004	
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)	
equity inventory coverage (share)	-0.017	***	-0.009		-0.018	***	-0.007	*	-0.013	***	-0.025	
	(0.01)		(0.01)		(0.00)		(0.00)		(0.00)		(0.02)	
bank loans (share)	0.010	***	0.025	***	0.009	***	0.019	***	0.012	***	0.011	
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.01)	
turnover (thousand)	0.005	**	0.003		0.011	***	0.013	***	0.011	***	0.014	
	(0.00)		(0.01)		(0.00)		(0.00)		(0.00)		(0.01)	
share of loans (second tercile)	-0.018	***	-0.012	*	-0.012	**	-0.025	***	-0.018	***	-0.016	
	(0.01)		(0.01)		(0.00)		(0.01)		(0.00)		(0.02)	
share of loans (third tercile)	-0.026	***	-0.016		-0.021	***	-0.034	***	-0.025	***	-0.047	*
	(0.01)		(0.02)		(0.01)		(0.00)		(0.00)		(0.02)	
dummy checking account	-0.075	***	0.007		-0.087	***	-0.018		-0.055	***	-0.174	*
	(0.03)		(0.03)		(0.03)		(0.03)		(0.02)		(0.13)	
credit line usage (share)	0.054	***	0.051	***	0.058	***	0.052	***	0.053	***	0.090	***
	(0.02)		(0.01)		(0.00)		(0.00)		(0.00)		(0.02)	
credit limit violation (months)	0.009	***	0.011	***	0.010	***	0.008	***	0.008	***	0.014	***
	(0.00)		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)	
blank cheques (number)	0.058	***	0.044	***	0.055	***	0.040	***	0.030	***	0.070	**
	(0.02)		(0.02)		(0.02)		(0.01)		(0.01)		(0.04)	
crediting operations (number)	-0.014	***	-0.003		-0.009	***	-0.011	***	-0.009	***	-0.016	
	(0.01)		(0.00)		(0.00)		(0.00)		(0.00)		(0.02)	
dummy long-term loan	0.005		-0.003		-0.005		0.012	**	0.004		-0.003	
	(0.00)		(0.01)		(0.00)		(0.00)		(0.00)		(0.02)	
long-term violations (months)	0.067	***	0.049	***	0.079	***	0.073	***	0.062	***	0.120	
	(0.02)		(0.02)		(0.02)		(0.02)		(0.01)		(0.10)	
long-term violation (share)	0.017	***	0.025	**	0.019	**	0.013	*	0.013	***	0.184	**
	(0.01)		(0.01)		(0.01)		(0.01)		(0.01)		(0.10)	
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	16,258		4,790		11,198		10,006		20,408		1,164	
McFadden R^2	0.270		0.307		0.257		0.310		0.258		0.366	

This table reports results of the pooled probit regression estimated on the analysis sample divided by firms' size duration of the bank-firm relationship and distance between the bank and the firm. Small firms are defined as those with an average turnover below 5 million euros and medium firms are those with an average turnover above 5 million euros. Short/long duration and small/large distance sub-samples of firms are defined depending on the median duration of the bank-firm relationship (8 years) and on the median physical distance between the bank and the firm (1 hour). Results refer to the full model, including all indicators at time t0-1 and (unreported) bank-year, industry and province fixed effects. The table reports marginal effects calculated as variations in the default probability after a change in the independent variable from the value at the 25thpercentile to that at the 75th (or 90th) percentile. Standard errors, clustered within firms, are reported in parenthesis; * indicates significance at the 10% level; ** indicates significance at the 5% level; *** indicates significance at the 1% level.

Fig. 1 ROC curve comparing baseline model and full model at time t_0 and t_0-*12*



This figure reports the comparison of areas under the receiver operating characteristics (ROC) curve of the baseline model and full model estimated at time t0 (left hand side) and t0-12 (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 0.84 and full model AUC is equal to 0.92, at time t0. Baseline model AUC is equal to 0.81 and full model AUC is equal to 0.88, at time t0-12. 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).