Bank-firm relationship in default prediction models. An analysis on a sample of Italian firms.

Carmen Gallucci

Università degli Studi di Salerno

Vincenzo Formisano

Università degli Studi di Cassino e del Lazio Meridionale

Michele Modina

Università degli Studi del Molise

Filomena Pietrovito

Università degli Studi del Molise

Università degli Studi di Cassino e del Lazio Meridionale

Abstract

The aim of this paper is to evaluate the contribution of several aspects of the bank-firm relationships in anticipating the corporate default event at least one year before. Using a unique dataset on a sample of 113 co-operative credit banks and about 12,000 firms operating in Italy, between 2012 and 2014, this paper documents that checking accounts and long-term loans activities provide additional explanatory power respect to financial information gathered from balance sheets in predicting the likelihood of default. Estimating a probit model, it finds that overruns and usage of credit lines on checking accounts, as well as overruns and payments overdue on long-term loans increase the accuracy prediction of default events by about 10%.

1. Introduction

The corporate finance and banking literature has paid much attention on default prediction modeling. Traditional default prediction studies have examined various economic and financial aspects of a firm (leverage, cost of debt, working capital, asset turnover and profitability, among others) elaborated on the basis of information gathered from their balance sheets (Beaver, 1966; Altman, 1968; Altman and Saunders, 1998).[1]

A less investigated issue is the analysis of the impact of information obtained from the bank-firm relationships. By providing linked financial services, the bank can indeed have access to additional information, available in real time, concerning credit line usage, checking accounts and loans activities. This means that the bank can observe this information long before receiving the financial statements of a firm, which may later confirm that the borrower is in financial problems (Norden and Weber, 2010). This idea originates from two well-established facts in the banking literature: first, informational asymmetries and monitoring borrowers may be the primary reasons that intermediaries exist (Brealey et al., 1977; Diamond, 1984) and second, an important function of financial intermediation is the production of information (Campbell and Kracaw, 1980). Bank's monitoring efforts are therefore well repaid because it can use readily and costless information to forecast future changes in the borrower's ability to repay its loans. The ability of a bank to collect private information and thereby to produce a superior judgment of borrowers expected performance is of relevance not only for regulators and banks, but potentially also for the industrial organization of borrowers and for business cycle theory (Nakamura and Roszbach, 2016).[2]

The aim of this paper is to combine quantitative information gathered by balance sheets and the information on the bank-firm relationships, distinguishing the short-term (checking accounts) and the long term (loans) activities, in order to predict the default probability of a firm. The contribution to the literature is noteworthy. *First*, this paper investigates the combined activity on a firm's checking account, usage of credit lines and long-term loans in revealing significant information on its cash flow and therefore on its capacity of repaying loans. Using data on bank-firm relationships provides three main advantages: (*i*) it helps to ex-ante distinguishing default firms from non-defaulted ones; (*ii*) it is relevant to banks that want to anticipate firm's ability of paying for debts and (*iii*) it adds to the literature on the predictive ability of financial statement information. *Second*, this paper analyzes the marginal benefit of this bank-firm relationships information by comparing the default prediction accuracy of a model that incorporates only the financial information provided by balance sheets with that including also checking accounts and long-term loans. *Third*, while typical studies of firm default focus on large publicly traded companies, this study covers mainly limited liabilities firms, very different in size and type of activity.[3] This makes our study highly representative of the dynamics of Italian industry where firms are highly dependent on banks as relationship lenders and the banking industry has been deeply affected by the global financial crisis.

---

[1] The advantage of using balance sheets is threefold: first, balance sheets are publicly available and their content is verified by a third party; second, these data are readily available to many academics and practitioners (firms, banks, investors and regulatory agencies); and third, they provide a bunch of quantitative information, that helps to predict the default event at least one year before it occurs. However, using financial information alone is not highly informative since balance sheets are backward looking and they do not provide signals of financial distress in a timely manner.

[2] Monitoring the bank-firm relationship is especially informative about the current state of firm's leverage and cash flow (Norder and Weber, 2010).

[3] It should be recognized that Italian limited liability firms have balance sheets that are certified by a third part.

To this end, this paper adopts a unique dataset on a sample of 113 co-operative credit banks and about 12,000 firms operating in Italy, between 2012 and 2014, provided by *Centrale Rischi Finanziari* (Crif), an Italian rating agency, and by *Centro Servizi Direzionali* (CSD), an Italian consulting company of the co-operative banks. The full sample includes 19,748 firm-year observations representing 11,865 firm, while the share of defaulters is equal to 7% of the total sample (including 845 firms). The remaining 11,020 firms are non-defaulters over the sample period. The financial and economic information include the return on assets, the share of bank loans over total liabilities, the share of fixed assets over net equity and the share of net equity over net equity and inventories. In addition, the checking accounts activities include the share of credit line usage, the number of overrun days, the number of overrun checking accounts and the number of consecutive months of overrun. Finally, the long-term loans activities include the number of consecutive months of overrun and number of overdue payments.

First, our results show that the credit line usage, the depth of overruns on checking accounts as well as on long-term loans help the bank to predict the default probability of a borrower at least one year before and their impact is persistent over time. Second, they demonstrate that these aspects of bank-firm relationships help to increase the predictive power of balance sheets information by 10%. Third, predictors differ depending on firm characteristics, both in terms of size and sector of economic activity.

The rest of the paper is organized as follows. Section 2 briefly review the related literature. Section 3 presents the sources of data and gives and 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 sheets indicators and the full model, including bank-firm relationships indicators. Section 5 discusses the results for different sub-samples of firms and robustness checks and Section 6 concludes.

2. Literature review

This section briefly reviews the banking and corporate finance literature related to our study, that provides insights on the determinants of corporate default prediction.

Firstly, our analyses relates to the traditional models that use financial indicators gathered from balance-sheet to classify default and non-default firms or assign a default probability to firms in a given time horizon.[4] Academics in this field propose several techniques: (*i*) linear discriminant analysis (Beaver, 1966); (*ii*) multivariate discriminant analysis (Altman, 1968; Eisenbeis, 1977; Piesse and Wood, 1992; Altman et al., 1994; Foglia et al., 1998; Grice and Ingram, 2001); (*iii*) logistic regression (Ohlson, 1980; Altman and Sabato, 2007; Dainelli et al., 2013); and (*iv*) probit regression (Bottazzi et al., 2011). 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) a set of financial ratios (Foglia et al., 1998). The more recent logit and probit approaches instead aim to assign each firm a probability of default on the basis of several financial indicators. Although these analyses differ in methods adopted and datasets, they primarily identify five groups of financial indicators as most important in predicting the default probability of firms: leverage, liquidity, profitability, coverage, and activity.

[4] See Kumar and Ravi (2007), Rikkers and Thibeault (2009) for a review of the literature.

However, in the last decades it has been observed an increasing interest in embedding additional information in such models, in order to improve their prediction power. A first attempt has been made by including non-financial factors, at the firm-level as well as at the macroeconomic level, such as: credit information sharing (Jappelli and Pagano, 2002; Diekers et al., 2013), corporate governance (Ciampi, 2015; Grunert et al., 2005), corporate social responsibility (Sun and Cui, 2014) and economic determinants (Becchetti and Sierra, 2003; Bottazzi et al., 2011; Tinoco and Wilson, 2013).

Particularly instructive for this paper is a second strand of the literature attempting to capture signals to predict financial difficulties through the lenses of the bank-firm relationship. Several studies investigate the role of relationship lending in determining the borrower default probability. Fiordelisi et al. (2014), for instance, use credit-file data on a large sample of Italian manufacturing firms during the recession, finding that a longer relationship with lenders decreases the probability of default. Moreover, they demonstrate that the concentration of the bank-firm relationship, in terms of the number of lenders, and the geographical distance are positively associated with the incidence of firm defaults.

However, bank-firms relationships can be related to other aspects: scope (credit and debit cards, credit lines), depth (relationship length, credit line usage) and forms (checking and saving accounts, long-term loans, investment portfolio). Nakamura (1993), for instance, develops the "checking account hypothesis" stating that bank account activity is informative and is used to manage their relationship, especially with small-business borrowers whose banking relationship is usually with a single lender. The checking account of a small firm sheds a clear light on its revenues and expenses because the firm's cash flows are easily comprehensible and are typically completely documented within one account.[5] In the same vein, Mester et al. (2007) 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 terms, 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.

Norden and Weber (2010) 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. 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 business (and individuals). It is notable that not all previous studies document a positive effect of credit line utilization on the default probability. Bergeres et al. (2015), for instance, document that an increase in credit line utilization involves a decrease in default probability, providing evidence that borrowers use lines of credit to pay down their term loans in period of financial distress and that banks should manage both financial instruments simultaneously. The paper by Jiménez et al. (2009a), using the credit register database for Spanish firms, shows that defaulting firms have significantly higher credit line usage rates and line's 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.

[5] In other terms, the finances of a smaller business are more fully revealed to its banker through its bank account than those of a large business.

Taken together, the literature provides evidence that the bank takes an advantage in providing deposit-taking and lending services jointly. Information spillovers coming from past loans and checking accounts activities might help the bank to select the firms to grant the loan.

3. An overview of data and variables

In this paper, the data used for the empirical analysis include the bank-firm activities on checking accounts and long term loans, in addition to corporate balance sheets. These data are obtained from two influential sources: CRIF (*Centrale dei Rischi Finanziari*), that is a credit rating agency providing ratings on Italian firms, and CSD (*Centro Servizi Direzionali*), an Italian consulting company of the co-operative banks.

Firms in our sample belong to the following six macro-industries: agriculture, commerce, transports and hotels, manufacturing, building and services. Firms are segmented according to two main criteria: synthetic codes of economic activity (*SAE – codici sintetici di attività economica*) and legal form. The majority (99.5%) are productive firms, including limited companies, cooperative companies producing goods and non-financial services, and consortiums. Other included segments are holdings of non-financial groups and operating private holdings, managing and controlling a group of companies whose main activity is goods/services production. More than 80% of firms are limited liabilities companies (*società a responsabilità limitata*) and 5% of them are limited companies (*società per azioni*).

From the initial sample provided by CRIF and CSD, after checking for outliers, duplicates and missing values, and excluding financial services and the public administration sectors, the resulting sample consists of 11,865 firms and 19,748 firm-year observations.[6]

Regarding the financial factors, this study uses an initial long list of 60 indicators concerning several areas of firm's profile (leverage, profitability, liquidity and efficiency, among others).[7] Regarding bank-firm relationships in the short-term an initial set of 15 indicators on checking account activities (including overruns, usage of the line of credit, debit and credit accounts, among others) is adopted and for long-term loans performances 11 indicators (overdue payments and overruns, among others) are adopted.

These sets of indicators were reduced by a selection process made in two steps. An initial selection of 38 indicators was made by applying the variance inflation factor model (VIF) for each of the (potential) independent variables (Chatterjee and Hadi, 2012). This procedure allows to select only the variables that are not affected by multicollinearity concerns in the linear regression model in which the dependent variable is the default event.[8] Following Tinoco and Wilson (2013), only those ratios with VIF less than 5 are retained, thus reducing the set of financial indicators from 60 to 21, that of checking accounts from 15 to 11 and that of log-term loans activity from 11 to 6. This way, multicollinearity is

[6] Values higher than the 99th percentile of each indicator have been excluded. However, our results are also robust to the winsorizing procedure that consists in substituting extreme values of a variable with the value at the 99th percentile of the distribution.

[7] These ratios are constructed by CRIF from firms' balance sheets.

[8] Initial indicators were very highly correlated. The VIF methods 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. $1/(1 - R^2)$. The utility of VIF is that it indicates the magnitude of the inflation in the standard errors associated with a particular beta weight that is due to multicollinearity (Ciampi, 2015).

not present in the model and the levels of coefficients obtained are reliable. The second step consists in 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 and Lee, 2002; Shin *et al.*, 2005). In particular, the *backward selection* method is adopted, beginning with the model including all variables and iteratively eliminating non-significant ones. 1% 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, convergence towards a model in which 10 explanatory variables are jointly significant is obtained.

The final set of balance sheet factors includes: the return on asset defined as profits (or losses) over total assets (*roa*) to measure profitability; the share of bank loans on total liabilities (*bank loans*) as indebtedness; the share of fixed assets over net equity (*fixed assets*) as financial riskiness; and the share of net equity over net equity and inventories (*equity inventory coverage*) as financial balance. In addition, three attributes reflecting checking accounts relationships are selected: the share of credit line given in a month on a checking account that is used by the borrower (*credit line usage*); the number of overrun days on checking account in a month (*overrun days*), the number of overrun checking accounts with the same bank (*overrun relationships*) and the number of consecutive months of overrun in a year (*consecutive months c.a.*). Finally, long-term loans performances include the number of consecutive months of overrun in a semester (*consecutive months l.t.l*) and the number of overdue payments in a month (*overdue payments*).

Variables and summary statistics are described in Table 1, distinguishing between default and non-default firms. It shows that our variables are normally distributed and all of them have a high variability between the minimum and the maximum value in both samples. Concerning our balance sheets indicators, significant differences between the two subsamples are found. The average return on assets indicator shows that profits are about 0.8% of total assets, with positive value for non-default firms (0.93%) and a negative average in the sub-sample of default firms (-1.26%), meaning that they are experiencing losses. It is interesting to notice that on average, firms in the sample make a large usage of bank loans which represent on average 37% of their total liabilities, with higher values for default firms than for non-default ones (46% *vs* 36%). The average share of fixed assets over net equity shows that in both cases firms make use of external resources to finance fixed assets, but the higher share is found for default firms (284% *vs* 266%), even though the difference between the two subsamples is significantly different at 5% level. The average values of the share of net equity over net equity and inventories indicate that non-default firms are more efficient than default ones (60% *vs* 45%) and this difference is highly significant (at the 1% level).

Looking at the short-term and long-term bank-firm relationships, similar results for the two subsamples are obtained. The credit line usage in a month is about 50% over the entire sample and, as expected, it is higher for default firms, since it reaches almost 100%, than for non-default ones (47%). Borrowers that subsequently default exhibit an increasing need for liquidity, which is due to a decrease in credits such as sales (Norder and Weber, 2010). In both cases, the credit line usage shows that the checking account has a debit balance, meaning that the credit line is completely used. The average number of overrun days in checking accounts is 3.5, with default firms showing 8 more days than non-default firms. Significant differences between the two subsamples are find also in terms of overrun relationships and consecutive months of overruns in a year. The first performance indicator is below 1 in both subsamples, while the second indicator shows that non-default firms have one month of overrun in a year, and default ones about 3 consecutive months. The same reasoning applies to the number of consecutive months of overrun in a semester and overdue payments for long-term loans, but with slightly lower values for both subsamples (0.32 *vs* 1.40 and 0.08 *vs* 0.75, respectively).

Table 1 – Summary statistics

	Non-defaut firms					Default firms					
	mean	**cv**	**min**	**max**	**obs**	**mean**	**cv**	**min**	**max**	**obs**	**ttest**
balance sheets factors											
roa	0.93	5.20	-62.85	24.48	11,020	-1.26	-4.18	-64.05	17.47	845	11.74***
bank loans	36.02	0.53	0.00	99.16	11,020	45.70	0.38	0.02	99.17	845	-15.67***
fixed assets	265.75	1.13	0.01	1,542.86	11,020	283.85	1.07	0	1,536.36	845	-1.67**
equity inventory coverage	60.04	0.54	0.04	100	11,020	44.85	0.73	0.51	100	845	13.01***
checking accounts activities											
credit line usage	47.50	0.92	0	289.66	11,020	91.65	0.43	0	266.37	845	-31.30***
overrun days	3.24	2.13	0	51	11,020	11.57	1.00	0	55	845	-20.58***
overrun relationships	0.17	2.18	0	4	11,020	0.61	0.99	0	5	845	-20.97***
consecutive months c.a.	1.01	1.66	0	12	11,020	2.93	0.91	0	12	845	-20.66***
long-term loans activities											
consecutive months l.t.l.	0.32	2.34	0	6	11,020	1.40	1.16	0	6	845	-19.09***
overdue payments	0.08	4.47	0	9	11,020	0.75	2.41	0	22	845	-10.83***

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 reports the pairwise correlations between the three sets of variables. The figures reported confirm the evidence of Table 1. Default borrowers have: lower return on assets and equity inventories coverage (-0.115 and -0.119); higher bank loans and share of fixed assets over net equity (0.131 and 0.016, respectively). It is interesting to remark that these pairwise correlations indicate a very low impact of fixed assets and a similar impact on default probability of all other ratios. Concerning checking accounts activities, the probability of default is positively correlated to all indicators: 0.253 for credit line usage, 0.28 for overrun days and relationships, and 0.27 for consecutive months of overrun in a year. The highest pairwise correlations are found with the long-term loans indicators: 0312 for consecutive months of overrun in a semester and 0.286 for overdue payments.

Since this simple correlations do not take into account the interrelations between factors and the fact that some firms show industrial and localization specificities different from others, a multivariate analysis is conducted.

Table 2 – Correlation matrix

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1)	flag default	1										
(2)	roa	-0.115	1									
(3)	bank loans	0.131	-0.183	1								
(4)	fixed assets	0.016	-0.133	0.328	1							
(5)	equity inventory coverage	-0.119	0.167	-0.181	-0.050	1						
(6)	credit line usage	0.253	-0.214	0.300	0.119	-0.160	1					
(7)	overrun days	0.280	-0.115	0.121	0.049	-0.143	0.507	1				
(8)	overrun relationships	0.282	-0.116	0.167	0.065	-0.108	0.606	0.567	1			
(9)	consecutive months c.a.	0.270	-0.151	0.173	0.083	-0.150	0.527	0.600	0.590	1		
(10)	consecutive months l.t.l.	0.312	-0.079	0.152	0.059	-0.067	0.326	0.373	0.369	0.443	1	
(11)	overdue payments	0.286	-0.047	0.102	0.016	-0.046	0.221	0.271	0.266	0.233	0.544	1

Our entire sample of 19,748 firm-year observations shows an average default rate of 7% that slightly resembles that of Italian firms, with a higher percentage in 2013 than in 2014 (Table 3). Approximately 5% of the firms in the sample were distressed at the end of the sample period. Table 3 also provides the distribution of default borrowers separating them depending on size, geographical area, and sector of economic activity. Most of them are small firms (83%) and lower shares are medium and large firms (16% and 1% respectively). Firms are mainly located in the Centre-North of Italy (73%) and only 27% of them is operating in the South. The highest percentage of default is in the construction sector (12%), as expected, while other sectors show an average default rate ranging between 6% and 7%.

Therefore the probability of a default event decreases with increasing size and is higher for small firms. The default event strongly involves firms located in the Centre of Italy and operating in the construction sector.

Table 3 – Distribution of default and non-default firms

sample	share of default firms	share of non-default firms
year		
2013	0.08	0.92
2014	0.05	0.95
size		
very small	0.07	0.93
small	0.08	0.92
medium	0.08	0.92
large	0.03	0.97
sector		
agriculture	0.06	0.94
other services	0.06	0.94
trade, transport and hotels	0.06	0.94
construction	0.12	0.88
manufacturing	0.07	0.93
geographic area		
North	0.07	0.93
Centre	0.09	0.91
South	0.07	0.93

Note: *very small* firms show a turnover lower that 2 millions of euros; *small* firms are those with a turnover between 2 and 5 millions of euros; *medium* firms are between 5 and 50 millions of euros; and *large* firms are above 50 millions of euros.

4. Econometric analysis

4.1 The econometric specification

We now estimate a probit model to study which factors at time t_0-1 (and t_0-2) influence the probability of default at time t_0.[9] In a first set of analyses, we investigate whether financial information gathered from balance sheets could be augmented by direct observations of bank-firm relationships to improve the prediction probability of default. We estimate a baseline pooled probit model, including only financial indicators as predictors, and two augmented models, including checking accounts and long-term loans activities in the bank-firm relationship. In a second set of analyses, we investigate whether these results are different for sub-samples of firms. We would like to assess whether indicators suitable to predict the default of manufacturing firms differ from those predicting the default of firms operating in the construction or service sectors and whether predictors are different depending on firm size.

The econometric approach relies on a binomial probit model where the endogenous variable, *default*, is a solvency condition at time t_0 taking the value of 1 in case the firm experiences one of the following events in the next 12 months: impaired loans, renovated loans or non-performing loans. The estimated *baseline model* can be summarized as follows:

$$\begin{aligned} \Pr(default_{t_0}^{ijhtp}) = \Phi(\beta \sum FinancialVariables_{t_0-1}^{i} + du\_size^{i} + du\_\sec tor^{h} + \\ + du\_bank^{j} + du\_year^{t_0} + du\_province^{p}) \end{aligned} \tag{1}$$

where subscripts *i*, *j*, *h*, *t* and *p*, are indicative of firm (i = 1,…., 11,865), bank (j = 1,…, 113) , sector (h = 1,…, 84), time (t_0 = 2012, 2013) and province (p = 1,…, 50), respectively.[10] The set of financial variables includes the return on asset (*roa*); the share of bank loans on total liabilities (*bank loans*); the share of fixed assets over net equity (*fixed assets*); and the share of net equity over net equity and inventories (*equity inventory coverage*), described in Section 3. *default* refers to t_0, 2012 and 2013, and identifies a firm that has payment difficulties in the performance period (i.e. the next 12 months), through 2013 and 2014, respectively. Financial information refers instead to the observation period, that is established in 12 months before the referring dates t_0. Following Fiordelisi et al. (2014) and Diekers et al. (2013), the specification controls for other firm-specific variables. Three dummies are added to distinguish the size of a firm (small, medium and large) and the specification is saturated with a bank fixed effect to control for banks unobserved characteristics that might affect borrower behavior. Further, to account for firms unobserved, time-varying loan demand and quality shocks, fixed effects for industry are included. Finally, the specification controls for time invariant fixed effects and for local context.

The second model (*full model*) includes checking accounts and long-term loans relationships between the firm and the bank.[11] The model is the following:

$$\begin{aligned} \Pr(default_{t_0}^{ijhtp}) = \Phi(\beta_1 \sum FinancialVariables_{t_0-1}^{i} + \beta_2 \sum CheckingAccounts_{t_0-1}^{ij} + \beta_3 \sum LongTermLoans_{t_0-1}^{ij} + \\ du\_size^{i} + du\_\sec tor^{h} + du\_bank^{j} + du\_year^{t_0} + du\_province^{p}) \end{aligned} \tag{2}$$

---

[9] The one year prediction horizon is standard in the literature and is also required by the Basel II and Basel III Regulations. In the internal rating system the probability of default (PD) estimates must be a long-run average of one-year realized default rater for borrowers in the grade.

[10] The list of provinces included in the analysis is provided in Appendix 1B.

[11] We also estimate an augmented model that contains all factors of the baseline model and checking accounts factors.

where the checking account set includes the share of credit line that is used by the borrower (*credit line usage*); the number of overrun days (*overrun days*), the number of overrun checking accounts with the same bank (*overrun relationships*) and the number of consecutive months of overrun in a year (*consecutive months c.a.*). Finally, the long-term loans performances include the number of consecutive months of overrun in a semester (*consecutive months l.t.l*) and the number of overdue payments (*overdue payments*). All other variables and dummies are defined as in Equation (1). Checking accounts and long-term loans activities refer to the observation period, as defined above.

In order to evaluate the contribution of short-term and long-term bank-firm relationships in predicting the probability of default, two standard methods are adopted in this paper. The first consists in comparing the goodness-of-fit of the baseline and full model through the McFadden R^2 adjusted for the number of regressors, following Norden and Weber (2010) and Diekers et al. (2013). The second consists in performing the receiver operating characteristics (ROC) analysis and in comparing the areas under the ROC curve (AUC) of baseline and full model 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 through the use of a ROC curve (Hanley and McNeil, 1983; Metz, 1978). 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 terms, 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 an 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).

4.2 The econometric evidence

This section discusses the results obtained from probit regressions of the default indicator on the predictor variables. Accordingly, the probability of default is estimated in the year prior to the observation of corporate financial distress ($t_0 - 1$) as well as two years prior to the financial distress event ($t_0 - 2$). Results are reported in Table 4, distinguishing the *baseline model* (equation 1) including balance sheets indicators alone, the *augmented model* and the *full model* (equation 2), which includes additional information on the bank-firm relationships. As required by the probit regression model, the dependent variable takes the value of 1 for firms classified as defaulters at time t_0 and the value of 0 for firms classified as non-defaulters. Table 4 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 (or 90th).

Column (1) of Table 4 shows that return on assets has a negative and significant impact on the probability of default, as expected. Return on asset is a profitability ratio indicating whether the firm is producing gains (or losses) and what is its share over total assets. A higher (positive) value this ratio means that the gains are high and the probability of default is expected to be low. Its marginal effect indicates that an increase of this ratio from the value at the 25th percentile(-0.36%) to that at the 75th percentile (2.04%) reduces the probability that a firm is affected by a default event in the next year by 0.7%. Considering that the average default rate in the sample is 7%, this impact is also economically significant. Considering the financial structure, we observe a positive impact of *bank loans* on the default probability: a higher resort to banks as external financiers is likely to increase the probability that a firm incurs in a default event. Looking at the marginal effect, it is interesting to notice that the variation of bank loans determines a higher probability of default: other thinks equal, if a firm becomes more dependent on bank loans its probability of default increases by 2.2%. Therefore, consistent with previous studies this result shows that the financial structure of a firm is an important determinant of its ability to repay loans. On the contrary, the *fixed assets* ratio does not seem to affect the default probability on a one year horizon, indicating that firm's solidity is not relevant for the probability that it repays loans. The most relevant impact on default is exerted instead by the *equity inventory coverage* ratio: a decrease of this ratio from the level at the 75th (93%) to that at the 25th (30%) percentile determines an increase of the default probability of 2.6%.

In addition, considering firm size, the results show that small and medium sized firms have a higher probability of default respect to very small ones (the excluded category), while large firms do not seem to have a different behavior. This result can be due to the limited size of the subsample of large firms.

In sum, controlling for firm size and for fixed effects described in Section 3, all balance sheet indicators (excluding *fixed assets*) prove to be not only economically relevant, but also statistically significant (at the 1% level) in predicting the default event one year before. This result is consistent with the traditional literature summarized in Section 2. More interestingly, if we include the short-term relationship indicators in the baseline model (column 2) we obtain three key results.

First, the impact of balance sheets 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. Second, the impact of all four checking account factors in the model (*credit line usage*, *overrun days*, *overrun relationships* and *consecutive months c.a.*) are positive and highly significant (at the 1% level). Specifically, increasing the credit line usage from the value at the 25th (0%) to that at the 75th percentile (93%) suggests an increase in the default risk of about 2.6%. In other terms, a firm that is close to reach the credit limit allowed by the bank is about 3% 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 overruns on checking accounts, we obtain positive impact on the default probability, regardless of the indicator used. For instance, increasing the number of overrun days from 0 (the 25th percentile) to 1 (the 75th percentile) the probability of default increases by 1%. The same impact is found for the number of consecutive months of overrun in a year on a checking account. The number of overrun relationships with the bank increases the default probability by 1.4% if it increases from the 25th to the 95th percentile. These findings indicates that checking account information helps to predict future firm defaults and confirm the evidence of Norden and Weber (2010) found for German firms. Third, and more important, the checking account activities help improve the accuracy of default predictions. The McFadden R^2 adjusted for the number of regressors more than triples when we consider additional checking account information factors to the baseline balance sheet factors: it varies from 6% to 19%. This means that the bank obtains additional information on the firm ability to repay loans when it considers the checking account activities, other than the balance sheets indicators.

Column (3) of Table 4 includes the long-term loans factors and shows that a higher number of consecutive months of overruns (from the value at the 25th percentile to that of the 90th) increases the probability of default by 1.7% while the number of overrun payments has an even larger impact increasing the probability by 2.5% (for a change between the 25th and the 95th percentile). Other coefficients remain unchanged in terms of economic impact and significance. The most interesting result is that the McFadden adjusted R^2 increases to 22%, demonstrating that the full model has an explanatory power that is about four times that of the baseline model.

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 areas under the curve of the baseline and full model. The baseline model at time t_0-*1* provides an accuracy ratio of 77%, while the full model of 88%. The difference in the accuracy ratio between the full model and the baseline one corresponds to the area between the two cumulative accuracy profiles (Diekers et al., 2013). The same comparison provides an overall *p*-value of 0 that indicates that the null hypothesis of equality of areas under the ROC curve can be rejected. In other terms, the small *p*-value suggests that the two areas are statistically significant. This finding provides support to our expectation that bank-firm relationships help to improve the accuracy prediction of balance sheets indicators.

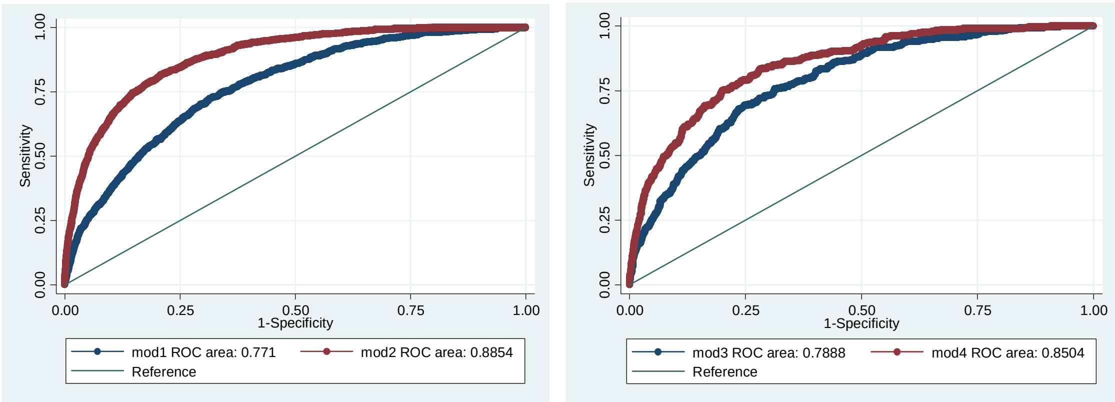
Furthermore, we consider different prediction horizons and we report the results for factors at time t_0-*2* in columns 4-6 of Table 4. Interestingly, even though the number of observations reduces by about 70%, our key findings that the measures of checking accounts and long-term loans have a significant impact on the probability of default that go beyond the balance sheets factors remain confirmed. This indicates that our key results do not depend on sample. Reassuringly, marginal effects and significance of our factors in the full model remains almost unchanged. However, considering a longer horizon the solidity of total assets becomes statistically significant with a negative sign: an increase of fixed assets over net equity from the value at the 25th percentile (71%) to that at the 75th (330%) reduces the probability of default by about 1%. In addition, overdue payments on long-term loans factor is not economically and statistically significant. Looking at the accuracy of the prediction, we notice that considering a longer time horizon reduces the R^2 both in the baseline and in the augmented and full models. This is confirmed by the AUC indicating that the baseline model gives an accuracy ratio of 79%, while the full model 85%. The *p*-value equal to 0.000 suggests also in this case that the two areas are statistically significant.

Table 4 – Probit estimation results for the entire sample, one and two years before default

	t_0-1			t_0-2		
	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline model	*Augmented model*	*Full model*	*Baseline model*	*Augmented model*	*Full model*
roa	-0.007***	-0.005***	-0.005***	-0.006***	-0.004***	-0.004***
	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)
bank loans	0.022***	0.012***	0.010***	0.025***	0.016***	0.015***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
fixed assets	-0.002	-0.002	-0.001	-0.007***	-0.007***	-0.007***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
equity inventory coverage	-0.026***	-0.013***	-0.013***	-0.033**	-0.022***	-0.022***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
credit line usage		0.026***	0.023***		0.022***	0.021***
		(0.00)	(0.00)		(0.00)	(0.00)
overrun days		0.010***	0.007***		0.012***	0.01***
		(0.02)	(0.02)		(0.03)	(0.03)
overrun relationships		0.014***	0.007*		0.01	0.007***
		(0.07)	(0.07)		(0.12)	(0.12)
consecutive months c.a.		0.015***	0.011***		0.017***	0.014***
		(0.04)	(0.04)		(0.06)	(0.06)
consecutive months l.t.l.			0.017***			0.022***
			(0.05)			(0.08)
overdue payments			0.025***			0.001
			(0.06)			(0.13)
small	0.012***	0.021***	0.021***	0.012	0.022***	0.022***
	(0.05)	(0.05)	(0.05)	(0.08)	(0.08)	(0.08)
medium	0.019***	0.040***	0.039***	0.015*	0.034***	0.034***
	(0.05)	(0.06)	(0.056)	(0.08)	(0.09)	(0.0)
large	-0.016	0.016	0.016	-0.011	0.019	0.021
	(0.20)	(0.23)	(0.24)	(0.31)	(0.32)	(0.32)
year fixed effects	yes	yes	yes	yes	yes	yes
bank 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	19,748	19,748	19,748	6,307	6,307	6,307
McFadden R^2	0.12	0.25	0.29	0.14	0.22	0.23
McFadden adjusted R^2	0.06	0.19	0.22	0.01	0.09	0.09

This table reports results of the pooled probit regressions with the firm default indicator, *default,* at time t_0 as dependent variables. Model (1) is the *baseline model* including balance sheets indicators at time t_0*-1* (or t_0*-2*) and (unreported) bank, industry, time, province fixed effects. Model (2) is the *augmented model* including the checking accounts indicators at time t_0*-1* (or t_0*-2*). Model (3) is the *full model* with the long-term loans indicators at time t_0*-1* (or t_0*-2*). *overrun days*, *overrun relationships*, *consecutive months c.a.*, *consecutive months l.t.l.* and *overdue payments* are expressed in logarithm. The sample includes data on 11,865 firms over the period 2012-2014. 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 firm are reported in parenthesis. * indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicate significance at the 1% level.

Figure 1 – ROC curve comparing baseline model and full model at time t_0-1 and t_0-2



This figure reports the comparison of areas under the receiver operating characteristics (ROC) curve of *baseline model* and *full model* estimated at time t_0*-1* (left hand side) and t_0*-2* (right hand side). 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 0.77 and full model AUC is equal to 0.88, at time t_0*-1*. Baseline model AUC is equal to 0.79 and full model AUC is equal to 0.85, at time t_0*-2*. S*ensitivity* is the fraction of positive cases that are correctly classified by the diagnostic test (true-positive rate), whereas *specificity* is the fraction of negative cases that are correctly classified (true-negative rate).

5. Results by firm size and sector of economic activity

In Section 4 we have found that dummy variables capturing firm size are highly related to borrowers default probability, demonstrating that small firms are less likely to experience repayment difficulties than medium and large firms. As such, in this section we investigate whether the previous results in terms of impact of balance sheets, checking account and long-term loans factors, hold if we differentiate by small firms and large firms. As stated by Norden and Weber (2010), '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, large firms benefit from the use of different sources of funding and a higher number of bank relationships. Small Italian firms rely primarily on bank loans as a source of funding and this relationship is usually exclusive.

In Table 5 we report the results obtained splitting the sample in two subsamples, according to borrowers' average turnover over the sample period. Small firms are defined as those with average turnover below 5 million euros and large firms are those with average turnover above 5 million euros. The results mainly confirm our previous findings on the full sample that balance sheet factors and bank-firm relationships are relevant for default. However, the analysis highlights differences across categories. The sample of small firms accounts for about 16,000 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. In terms of balance sheets indicators, the default probability of large firms is affected by the return on assets and by the share of bank loans over total liabilities. It is interesting to notice that the impact of these two indicators is about three times that of the same indicators for small firms. Return on assets decreases the probability of default by 1.4%, whereas bank loans increases this probability by 2.8%. Differently from small firms, large firms are not affected by the equity inventory coverage. The marginal impact (1.2%) is indeed not significant. Among checking account activities, the usage of credit line and the number of overrun days are relevant in a model controlling for balance indicators and for fixed effects. However, while the marginal effect of credit line usage is higher than that of the entire sample and of small firms (3%), that of overrun days is very low. Finally, all results on the relevance of long-term loans activities are confirmed for large firms. Consecutive months of overrun and overdue payments increase the default probability by 1.7% and 4.4%, respectively.

Taking advantage of the fact that our entire sample covers 84 Ateco-2 digits sector of economic activity, we also investigate whether the predictive power of balance sheets and bank-relationships factors are robust across industries. For this purpose, we split the sample into six subsamples: agriculture, manufacturing, construction, trade, transport and hotels and other services. We find that firms in all industries, except construction, show a higher probability of default after one year the lower its return on assets. The impact of this ratio ranges between 4% (for other services) and 15% (for agriculture). On the contrary, the impact of the share of bank loans is significant (at the 5% level) only for firms in the construction industry. Equity inventory coverage is negative and significant for all sectors (except for agriculture) with marginal impacts ranging between 5% (for manufacturing) and 4.1% (for construction). Among checking accounts, credit line usage is a predictor of default for firms in all sectors of economic activity, consistent with our results for the entire sample and with other findings in the literature (Norden and Weber, 2010). Banks that detect an abnormal increase of credit line usage should consider that the risk of default is high for that borrower, with a very high impact on the probability of default. The number of overrun days is highly significant for all sectors except for construction and other services, with the highest impact in trade, transport and hotels industry (1.3%). Considering the number of overrun relationships we obtain positive impact on the probability of default for construction (2.1%) and other services (1.5%), whereas a (surprising) negative impact on agriculture. The depth of overrun, both on checking accounts and on long-term loans, measured by the number of consecutive months determines an increase in the probability of default for all sectors, with the highest impacts in the construction sector. Finally, looking at overdue payments on long-term loans the bank can anticipate the default event in a one year horizon for firm in all sectors.

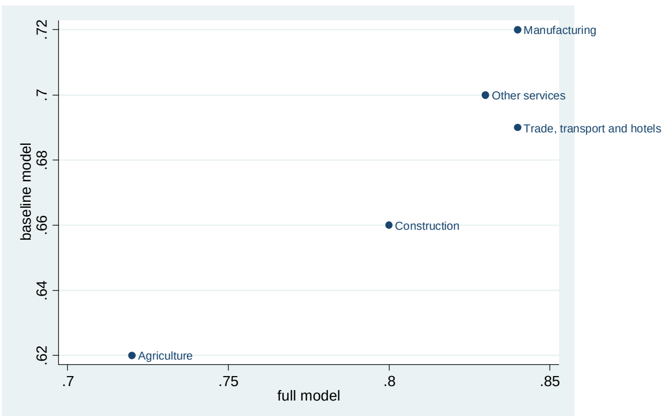
To evaluate the accuracy prediction of our model in different sectors, we adopt the same tests as before and we notice that the McFadden adjusted R^2 ranges between 6% for agriculture and 21% for construction. We also investigate whether the accuracy prediction is robust across industry groups, following Diekers et al. (2013). Figure 2 plots the AUC for baseline model and for the full model for each industry separately. It shows that the baseline model provides an accuracy ratio of at least 60%, whereas the full model increases it by at least 10%. The manufacturing sector shows the highest performance in identifying both the true positive cases and the true-negative cases (72% and 84%, respectively).

Table 5 – Probit estimation results for different types of firms, one year before default

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Small firms	Large firms	Agriculture	Manufacturing	Construction	Trade, transport and hotels	Other services
roa	-0.004***	-0.014***	-0.149***	-0.007***	-0.003	-0.006***	-0.004***
	(0.00)	(0.01)	(0.08)	(0.01)	(0.01)	(0.01)	(0.01)
bank loans	0.009***	0.028***	0.000	0.019	0.019**	0.006	0.007
	(0.00)	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)
fixed assets	-0.001	-0.005	0.077***	-0.003	0.004	-0.005**	-0.001
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
equity inventory coverage	-0.014***	-0.012	0.000***	-0.005***	-0.041***	-0.011**	-0.012***
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
credit line usage	0.021***	0.030***	0.168***	0.016***	0.035**	0.032***	0.029***
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
overrun days	0.008***	0.000***	0.009***	0.008***	0.006	0.013***	0.000
	(0.02)	(0.05)	(0.51)	(0.03)	(0.05)	(0.03)	(0.05)
overrun relationships	0.010***	-0.004	-0.227***	0.002	0.021***	0.005	0.015*
	(0.08)	(0.19)	(0.92)	(0.14)	(0.18)	(0.14)	(0.18)
consecutive months c.a.	0.019***	0.008	0.079***	0.011***	0.051***	0.020***	0.008**
	(0.04)	(0.11)	(0.52)	(0.07)	(0.10)	(0.07)	(0.09)
consecutive months l.t.l.	0.018***	0.017***	0.477	0.016***	0.027***	0.019***	0.023***
	(0.05)	(0.14)	(.)	(0.09)	(0.13)	(0.09)	(0.11)
overdue payments	0.024***	0.044***	-0.182***	0.032***	0.040***	0.037***	0.014*
	(0.07)	(0.19)	(0.48)	(0.12)	(0.14)	(0.12)	(0.15)
small				0.003	0.051	0.031***	0.047***
				(0.09)	(0.14)	(0.10)	(0.16)
medium			0.471***	0.025	0.113	0.030***	0.069***
			(0.98)	(0.09)	(0.15)	(0.10)	(0.18)
large				0.030	0.194		
				(0.30)	(0.45)		
year fixed effects	yes	yes	yes	yes	yes	yes	yes
bank fixed effects	yes	yes	yes	yes	yes	yes	yes
industry fidex effects	yes	yes					
province fixed effects	yes	yes	yes	yes	yes	yes	yes
Observations	15,834	2,789	44	5,782	2,164	5,500	2,975
McFadden R^2	0.29	0.35	1.00	0.30	0.38	0.30	0.33
McFadden adjusted R^2	0.22	0.16	0.06	0.18	0.21	0.18	0.15

This table reports results of the pooled probit regression by firms size and firm sector of economic activity. Small firms are defined as those with an average turnover below 5 million euros and large firms are those with an average turnover above 5 million euros. Results refer to the *full model* including all indicators at time t_0-*1* and (unreported) bank, industry, time, 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 25th percentile to that at the 75th (or 90th) percentile. Standard errors clustered within firm are reported in parenthesis. * indicates significance at the 10% level, ** indicates significance at the 5% level and *** indicate significance at the 1% level.

Figure 2 – Baseline model and full model AUC for each industry



6. Conclusions

(to be done)

References

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., & Saunders, A. (1998). Credit risk measurement: Developments over the last 20 years. *Journal of Banking and Finance*, *21*(11/12).

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 & finance*, *18*(3), 505-529.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of accounting research*, 71-111.

Becchetti, L., & Sierra, J. (2003). Bankruptcy risk and productive efficiency in manufacturing firms. *Journal of banking & finance*, *27*(11), 2099-2120.

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.

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

Brealey, R., Leland, H. E., & Pyle, D. H. (1977). Informational asymmetries, financial structure, and financial intermediation. *The Journal of Finance*, *32*(2), 371-387.

Campbel, T. S., & Kracaw, W. A. (1980). Information production, market signalling, and the theory of financial intermediation. *The Journal of Finance*, *35*(4), 863-882.

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.

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., 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* 44: 837–845.

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

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

Eisenbeis, R. A. (1977). Pitfalls in the application of discriminant analysis in business, finance, and economics. *The Journal of Finance*, *32*(3), 875-900.

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 & Finance*, *22*(10), 1441-1456.

Grice, J. S., & Ingram, R. W. (2001). Tests of the generalizability of Altman's bankruptcy prediction model. *Journal of Business Research*, *54*(1), 53-61.

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

Hanley, J. A. & McNeil, B. J. (1983). A method of comparing the area under two ROC curves derived from the same cases. *Radiology* 148, 839-843.

Jappelli, T., & Pagano, M. (2002). Information sharing, lending and defaults: Cross-country evidence. *Journal of Banking & Finance*, *26*(10), 2017-2045.

Jiménez, G., & Saurina, J. (2004). Collateral, type of lender and relationship banking as determinants of credit risk. *Journal of banking & Finance*, *28*(9), 2191-2212.

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.

Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques–A review. *European journal of operational research*, *180*(1), 1-28.

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

Metz, C. E. (1978). Basic principles of ROC analysis. Seminars in Nuclear Medicine 8, 283-298.

Nakamura, L. (1993). Commercial Bank Information: Implications for the Structure of Banking. In Michael Klausner, and Lawrence J. White (eds.), Structural Change in Banking. Homewood, IL: Business One/Irwin.

Nakamura, L. I., & Roszbach, K. (2016). Credit Ratings, Private Information, and Bank Monitoring Ability. Federal Reserve Bank of Philadelphia, Working Paper n.16-14.

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

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of accounting research*, 109-131.

Piesse, J., & Wood, D. (1992). Issues in assessing MDA models of corporate failure: A research note. *The British Accounting Review*, *24*(1), 33-42.

Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques–A review. *European journal of operational research*, *180*(1), 1-28.

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

Shin, K. S., & Lee, Y. J. (2002). A genetic algorithm application in bankruptcy prediction modeling. *Expert Systems with Applications*, *23*(3), 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.

Sun, W., & Cui, K. (2014). Linking corporate social responsibility to firm default risk. *European Management Journal*, *32*(2), 275-287.

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, 28–35.

Appendix 1A – List of provinces

Ancona	Como	Milano	Ragusa	Trapani
Avellino	Cosenza	Modena	Ravenna	Trento
Bari	Crotone	Napoli	Rimini	Treviso
Belluno	Enna	Novara	Roma	Trieste
Benevento	Ferrara	Padova	Rovigo	Udine
Bergamo	Firenze	Palermo	Salerno	Varese
Bologna	Forlì-Cesena	Perugia	Sassari	Venezia
Bolzano	Frosinone	Pordenone	Siena	Verona
Brescia	Gorizia	Potenza	Taranto	Vicenza
Catanzaro	Latina	Prato	Torino	Viterbo