Predicting Default Probability Using Delinquency: The Case of French Small Businesses[*1]*

By Asma Marouani

Small businesses may be expected to be more likely to fail because they are more volatile, have less power in negotiations with financial and social partners, are more credit-rationed by credit managers, are less likely to benefit from their experience or 'learning effects', compared to large firms, and often operate in small markets. From this point of view, financial ratios seem to be irrelevant when modelling their default probabilities. This current research is an attempt to fine tune variables and to find more dynamic information to include in a point-in-time probability of default model.

In this paper we explore the hypothesis that a firm's future default could be measured and explained solely by the historical data on the ability and willingness of a firm to pay its creditors. We use a credit scoring application to model default on a large data set of French Small and Medium-sized Enterprises (SMEs). We find that payment behavior data can be used to successfully predict SME bankruptcy in France and in a timely manner. New variables on late payment and delinquency are identified as alternatives to those usually used in failure models literature.

Introduction

For over 20 years, the prediction of firms' likelihood of bankruptcy appears to have been of paramount interest to the risk managers of banks and other non-financial lenders. Among the many important questions highlighted by advanced research on risk management are the features and the determinants of a company's failure. Most bankruptcy and recovery models have long used financial ratios as representations of the financial distress process (Altman 1968, Beaver 1966, and Edmister 1972). Based on Altman's earlier score (Altman 1968), the Zeta score uses seven ratios V1 = EBIT/total assets, V2 = normalized measure of standard error of estimate around a 10-year trend in V1, V3 = EBIT/total interest payments, V4 = retained earnings/total assets, V5 = current assets/current liabilities, V6 = five-year average equity market value/ total capitalization, V7 = total assets. Altman (1968) showed the accuracy of predicting future failure with only seven financial ratios. In general, models are based on liquidity and profitability ratios with other variables on activity and financial leverage. Banks were inspired by this research to build their models in assessing firms' likelihood of bankruptcy and still use them today. However, such models were essentially adapted to large and medium firms.

---

1 I would like to acknowledge the support of General Electric Capital France and Coface Services in providing their data . I would also like to thank Professor Annie Bellier and other professors from the department of Finance in the University of Cergy-Pontoise for their helpful comments and suggestions.

Asma Marouani is a PhD student in the Department of Finance, University of Cergy-Pontoise of Paris, France.

Correspondence address : Asma Marouani, 33 Boulevard du Port, 95011 Cergy-Pontoise Cedex, France. E-mail: asma.marouani87@gmail.com

Empirical studies have shown that SMEs, in comparison to large firms, have fewer liquid assets, are more prompt to rely on short-term debt, and have a volatile cash flow (Walker and Petty 1978). Given this characteristic, financial institutions are unable to efficiently evaluate the risks involved when lending to small firms. This is why SMEs are more likely to be credit rationed. In order to evaluate the financial differences between large and small firms, models for SMEs began to emerge. Other criticisms were levelled at financial-based models: financial ratios are static and do not take into account the dynamic aspects of risk factors. Failure models assume the distribution normality of input variables which is not often the case, especially because many of these ratios could not be negative.

Moreover, such predictive models tend to ignore macro-economic factors. Statistical tools held for the SME sector, such as Multiple Discriminant Analysis, are less likely to capture seasonality and cycles. Even though numerous solutions have been proposed to overcome such limits, these financial-based predictive models are still inaccurate (Dimitras et al. 1999). Access to financial funding for SMEs is still problematic. Small banks are reluctant to lend to SMEs and large banks demand more and more guarantees in return for a loan, with a higher interest rate in most cases. It is not surprising that small firms are particularly reliant on short-term sources of finance which is reflected in the extension of trade credit in several countries. Considered as a cash flow element, trade credit involves supplying goods and services with a possibility of paying after an agreed delay. The growing importance of trade credit for SMEs somehow mirrors a deteriorated banking relationship. SMEs use it as a complement or a substitute for financial resources (Brealey et al. 2010). They also perceive it as a strategic tool to generate profit and business by lending to their business partners (Petersen and Rajan 1997). In this way, managing late payment in a trade credit context appears to be the main interest of managers. Indeed, trade credit practices are not just a cash flow issue, but also an important way of signalling one's reputation and financial health (Paul and Wilson 2006; Wilson and Summers 2002). In some developed countries, trade credit, represented by accounts payable on the borrowers' balance sheets and by accounts receivable on the creditors' balance sheets, exceeds short-term bank credit and is an important way of financing a firm's working capital (Peel and Wilson 1996).

Several economic implications could be derived to justify our intent to pursue this path of research. A natural question emerges: could trade credit practices say more about a firm's financial situation and its credit risk than do financial accounts? Indeed, we believe that more dynamic late payment patterns derived from creditors reveal as much information about an SME's situation as does financial accounting. We can even talk about informational advantage for the benefit of non-financial lenders. Such informational advantage arises due to the fact that trade creditors are mostly engaged in the same non-financial transactions as the borrowers are. In many cases, access to financial data could be costly, whereas being in the same industry gives the trade creditors easier or cheaper access to information generated by their commercial relationships. Emery (1984) and Mian and Smith (1992) see trade credit as a more profitable short-term investment than marketable securities. Furthermore, it is difficult for banks to obtain detailed information from small firms, since the financial reports of small firms are mainly for tax purposes (Bhattacharya and Thakor 1993).

Despite the importance of trade credit, few studies have been conducted along this line of research, due to the unavailability of relevant published data and the reluctance of firms to communicate information regarding their trade credit practices. In the current study, we depart from the trade credit literature in at least two aspects: firstly, we extend the trade credit management literature by empirically quantifying cutoff points after which bankruptcy becomes an alarming possibility. Furthermore, we find evidence that trade credit could be more or less critical when other delinquency and late payment features are recorded for a given firm. So far, delinquency has referred to missing payments for consumer credit (including credit card loans and other consumer loans). Prior studies use the late payment data in addition to financial ratios to predict financial distress or impact on profitability, solvability, but never on a SME's bankruptcy. We investigate the most powerful explanatory variables reflecting payment patterns to predict default probabilities using a credit scoring method for French SME cases. Exploring delinquent behavior allows us to consider more dynamic aspects of the credit risk assessment. We believe that many risk factors remain to be identified when evaluating the risk of default of a SME. The lack of data has made SME credit risk an under-researched area in finance. We acknowledge that there are only a few studies on probability of defaultestimations specifically for SMEs.

Our paper is divided as follows: in the second section we explore the existing literature related to trade credit practices, delinquency patterns and the statistical tools generally used to predict default probability for SMEs. Section Three presents the methodology and variables we use. Section Four presents and discusses our results and Section Five concludes.

Review of Literature

Trade credit: its motives and determinants

Managing cash flow and working capital efficiently is done via good credit management practices. They have often been considered as pivotal to the health and performance of firms. Even for SMEs, dealing with working capital issues represents a great concern, particularly where small firms are growing and therefore need to finance increasing amounts and debtors. Research in recent years has focused on trade credit expansion as one essential element of cash flow management. Since Meltzer's paper (1960), illuminating statistics have revealed the importance of trade credit. Obviously, in industrialized economies, the volume of trade credit is higher than the short-term loans received from banks (De Blasio 2005) and it results from payment delays contractually agreed by non-financial companies. However, companies operating in countries having underdeveloped and/or inefficient legal and financial systems depend relatively more on trade credit (Rajan and Zingales 1995, Saito and Bandeira 2010).

Several reasons may be provided to explain the growing reliance on short-term sources of funding, such as trade credit. The literature usually refers to transactional motives and financial ones. First, trade credit is becoming an important form of credit when firms encounter credit rationing. Petersen and Rajan (1997) explain that large firms could play the role of intermediaries to credit rationed firms by granting longer payment delays in periods of monetary restrictions. Keasey and Watson (1992) conducted an empirical study on small firms in the UK and found a negative relationship between bank finance and trade credit, implying that trade credit is used as a substitute for other more traditional ways of financing. Secondly, firms with better access to credit agree to engage themselves in credit relationships with their suppliers, seeking informational advantages. Indeed, allowing for payment delays is a strategic way to get continuous information from borrowers (Frank and Maksimovic 2005). In a world of imperfect information, a supplier may learn about a firm's creditworthiness and future prospects in the course of their ongoing business relationship. Some borrowers intentionally tend to use trade credit as a tool to signal their financial situation (Cook 1999). Finally, trade credit serves also as price discrimination; the underlying hypothesis assumes that extending the credit period is synonymous to reducing prices. Some riskier borrowers may have been credit rationed. Consequently, this sector expresses its demand (Smith 1987) by buying higher quantities at lower prices. The total profit for suppliers increases even under a lower initial price and is essentially realized, thanks to price discrimination, via more flexible payment delays. From this perspective, the credit period can be an opportunity to reduce informational asymmetries about product quality and the seller's reputation, which makes trade credit a signal of product quality and seller reputation.

Trade credit choices may also differ from one firm to another, depending on several factors. Indeed, the company size is one of the most discriminatory factors when it comes to the financial choices of individual firms. In theory, it appears that large firms have relatively high bargaining power which results in longer payment intervals that may be due to the size of contracts and the confidence they inspire. At the same time, external funding sources available for the companies are more numerous, the greater their size. Numerous indicators have been used to measure the influence of the firm size factor in most of the empirical studies on trade credit. For Wilson and Summers (2002), the size criterion used is the amount of turnover. Emery et al. (1993) show that an increase in liquidity is more likely to cause a proportionate increase in trade receivables for a large firm than for a small firm. This result does not lead to the conclusion that larger firms are less liquidity-constrained. We are in line with these papers, as we introduce in our model the total turnover to reflect firms' size. In addition to that, small firms tend to extend their trade payables when their cash flows decrease.

Theories of agency (Jensen and Meckling 1976) and signal (Leland and Pyle 1977) presuppose the existence of a positive relationship between the company's maturity and the weight of debt. Conversely, the arguments of pecking order theory reflect the fact that older firms have more internal financing sources and rely less on debt (Myers and Majluf 1984). The firm's age is an approximation of capital information available to its borrowers. A relatively old business is generally considered to have a good reputation and thus gains trust from borrowers and easily establishes long lasting relationships with its bank lenders. Petersen and Rajan (1997) showed that the lifetime relationship formed by firms and financial institutions is highly correlated to the availability of bank loans. The degree of asymmetric information is assumed to be inversely proportional to the company's age. Previous research assumes the same logic in the trade credit context.

As explained above, under information asymmetry, the strength and duration of the ties between a business and its suppliers may play a role in the terms upon which trade credit is offered. Berger and Udell (1995) confirm this result and found that relationship measures are related to the availability and terms of credit from U.S. financial institutions. Another trade credit factor can be introduced. Indeed, more recent research has shown that ethnic and socio-cultural differences may impact the use of trade credit among small firms. Some empirical research has raised the relevance of ethnic relationships when it comes to providing payment delays for customers. In terms of trade credit, the feature has been particularly recorded for Hispanic and black-owned firms (Cavalluzzo and Cavalluzzo 1998). Proximity and neighborhood have been also mentioned in some research as being elements that determine the extent of trade credit. It appears that race/ethnicity and neighborhood are assimilated to proxies of credit networks that determine the extent of reliance on trade credit.

The trade credit determinants mentioned previously cannot be independently analyzed without taking into consideration the level of a country's financial development. Rajan and Zingales (1995) find that firms in industrial sectors with a greater need for external finance grow faster in countries with well-developed financial markets. These studies support the notion that a well-developed financial system can facilitate a country's economic growth. We question the fact that financial alternatives could be better developed in poorly developed counties to alleviate credit access problems. Love et al. (2007) examine the effect of financial crisis on trade credit in six emerging economies. They find that firms with weaker financial conditions are more likely to reduce trade credit after the crisis. In another paper, Fisman and Love (2003) examine the use of trade credit in different countries and find that industries with higher dependence on trade credit financing grow faster in countries with weaker financial institutions so that that it is used as a substitute for bank loans in countries with poor financial institutions.

Overall, inter-firm credit appears to have many advantages for both suppliers and customers. This is still true for small firms that turn to short term funding to finance longer exploitation cycles. Yet trade credit practices also have disadvantages, which we enumerate in the following section.

Measuring trade credit risk and other late payment incidents: credit scoring models

Short-term sources of funding, such as trade credit, play a significant role in supporting the growth of firms and have numerous advantages for both non-financial lenders and firms' borrowers. However, some limits may arise, especially for small firms. One should not forget about costs related to extending payment delays and possible overdue trade credit. Overdue trade credit refers to trade credit that has expired but is not repaid. Firms are usually reluctant to have overdue trade credit because they may face significant late payment penalties, including the explicit cost of pecuniary penalties as well as the implicit costs of damaging long-term relationships with customers (Petersen and Rajan 1997). Moreover, trade credit is tied to the purchase of goods, which is less flexible than bank loans. Thus, even though trade credit appears to be relatively more attractive for financing purposes in the presence of constraints in bank loans, an effective formal financial system may be necessary to sustain a country's long run growth.

To alleviate late payment related to trade credit, policy makers have tried to impose numerous rules to manage the credit granted to firms' customers. In many countries, companies and government work together to establish an effective credit management policy aimed at preventing delayed payment, which is the major factor behind business failure (Wilson and Summers 2002). Credit policies are used internally to monitor firms' bad debt. In the UK, for instance, the debate still persists on the effectiveness of interest penalties on late payment in the trade credit context. In France, the 2008 LME law (Loi de Modernisation Economique) has been introduced to respond to late payment problems (Lorenzi and Kremp 2010). Despite all these attempts, payment delays are always considered critical due to a misunderstanding of the credit terms or failure to communicate the terms to the customers before the sale takes place. Consequently, legislation seems to do very little to deal with late payment problems. Indeed, companies often avoid adopting extreme penalties (charging interest on late payments, pursuing borrowers with overdue trade credit through the courts, to name a few) for several reasons. Firms may alter their relationship with their partners, especially large ones. There is evidence that a firm's size is positively correlated with the trade terms and claims conditions that allow the adoption of such extreme measures. It is worth noticing that larger firms have greater bargaining power. It appears obvious that small firms are by consequence reluctant to take action for fear of losing the loyalty of customers.

Many suggest, also, that late payment in trade credit can affect profitability. The incidence of credit period extended to customers may be useful to the credit managers for controlling the associated risk. In that sense, credit management becomes vital when a firm's performance may be altered in the case of longer and permanent late payment, especially when delayed payment by customers is often balanced in turn by delayed payment to their own suppliers. In addition to legal/regulatory action that could be taken to deal with late payment on trade credit, firms may consider other internal credit management policies such as setting credit limits and setting cash flow targets. But, due to their low cost, statistically derived credit scoring models have been proven to be reliable tools to predict delinquency for instance. Initially developed in consumer market (screening, pricing and monitoring consumer credit accounts), scoring models have been used worldwide in consumer lending for some time and their role has expanded internally among credit managers to address the risk profile of customers. Banks started to use these statistical techniques to moderate loan terms for credit card loans, mortgages and other consumer loans. Nowadays, non-financial lenders use them more and more for their internal purposes.

The most significant development in recent years has been the development of scores for small businesses. Adjusting for SMEs' specific characteristics in assessing risk credit is possible through objective and statistically validated models. The latter were commonly known worldwide in the 1990s when Fair Isaac Corporation introduced the Small Business Scoring Solution. The literature tends to distinguish two types of information generally used when applying credit scoring. First, we find hard information collected from credit bureaus or financial statements used for underwriting decisions (Berger and Frame 2007). The second type of information includes soft qualitative data gathered through the relationship with borrowers and lenders (Berger, Klapper and Udell 2001). Other purposes for the credit scoring systems have been identified in the literature, such as estimating the amount of profit an account is likely to generate, identifying applicants who may be candidates for other services, targeting prospective customers, predicting delinquencies for card loans, to name a few.

According to Berger and Frame (2007), Small Business Credit Scores increase small business credit availability in the following way: overall quantity of lending, lending to relatively opaque borrowers, lending within low-income as well as high-income areas and lending over greater distances. Other authors enumerate many other advantages for credit scoring: Ponicki (1996) finds that these techniques are simple and easy to manipulate. In addition to that, they can be used in a shorter timeframe.

Credit scoring is traditionally divided into two broad types (Lee and Chen 2005). The first application scoring is used at the time an application for credit is made and estimates an applicant's likelihood of default in a given time period. The data used for this task generally consists of financial and demographical information about a given sample of existing applicants. The second type of credit scoring, behavioral scoring, is used after credit has been granted as estimates, along with past data on credit worthiness, at some later date. Both types of credit scoring applications were extended to larger fields such as commercial credit, credit cards and trade credit. More generally, credit scoring and most recently behavioral scoring are the techniques that help organizations decide whether or not to grant credit to consumers who apply to them or to monitor future credit lines for existing customers. There is a substantiated tendency for lenders to buy delinquency data from credit bureaus as they have become aware of their utility in the credit scoring process. Obviously, the longer the payment is overdue, the more it will hurt your score. Estimating, then, the probability of default relies to a great extent on the historical payments of consumers in different fields (credit card consumers, loan consumers, trade credit). Credit analysts ultimately determined that the personal credit history of small business owners is highly predictive of the loan repayment prospects of the business.

To our knowledge, the extant evidence on the effects of small business credit scoring on small business credit is limited to two aspects. The first is related to the credit pricing. Technological progress has allowed banks to offer more or better services that may have raised costs, but customers were willing to pay more for these services, raising revenues by more than the increased costs. The second aspect focuses more on credit availability. A number of studies have found that large banks tend to devote lower proportions of their assets to small business lending than smaller institutions do. Our study is different from the existing literature. We are interested in credit scoring models for small businesses which take into account their peculiarities and heterogeneity. Secondly, we try to provide insights on how past credit behavior works to predict future bankruptcy. For the latter, we use dynamic patterns such as trade credit practices combined with other incidents.

Again, there has been little attention paid to credit scoring models which take into account the repayment behavior of small firms, especially incidents collected from government sources. This incites us to look into the variables that are relevant when predicting SME default. The next section describes the sample used for the purpose of the study and explains the statistical approach adopted.

Data Collection and Methodology

Data collection

Data on the payment behavior of a set of French companies is drawn from General Electric's factoring database in which several incidents of payment are recorded. In addition to late payment on trade credit, we can distinguish four other main payment incidents that will be used in our current study. Historical arrears on trade credit cover all clients of Factofrance, one of the major factors in the French market, which belongs to the General Electric Group. Factoring is a short term source of financing whereby a business sells its accounts receivable to a third party, called a factor, at a discount. It involves three parties: the seller of invoices who mitigates its risk on its clients (debtors) to the factor who becomes the sole owner of the receivables. Data on arrears are thus related to late payment of debtors (clients of Factofrance clients). They are recorded monthly, by firm. In our study, we only use frequencies and amounts of unpaid invoices that exceed one month, two months, three months, four months, five months and six months.

We were able to use historical data about unpaid trade bills on General Electric's clients.

The paper employs other different types of payment incident, all gathered by firm and collected from a French data provider, Coface Service. They fall into the following two categories: commercial litigation and debt to the French government (so-called "privilèges URSSAF" in France).

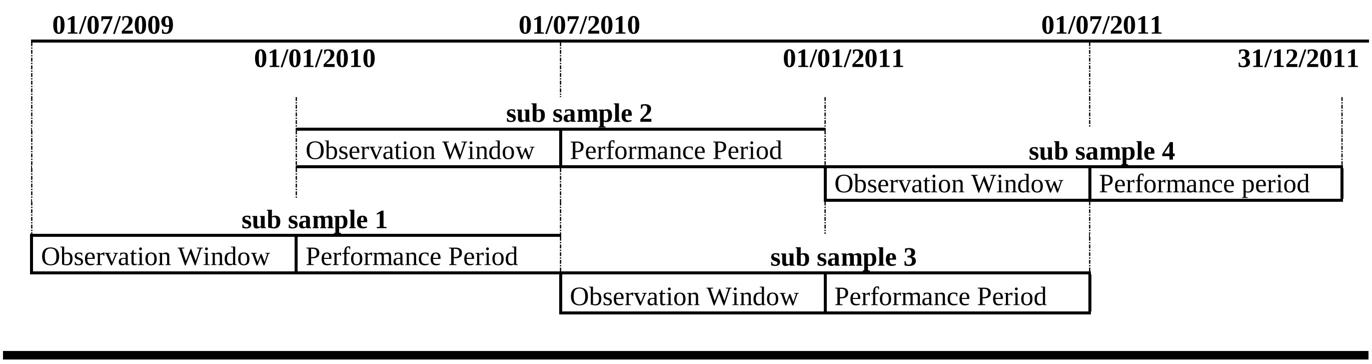
Finally, we add a firm's identity variables such as its age, geographical location, legal status and size, measured by the total turnover.

The initial sample accounts for 1,500,000 active commercial French companies at the beginning of July 2009. Public administration and insurance/financial activities were rejected from this sample. After cleaning files by controlling for the outliers or the missing values, the sample contains 973, 680 different French firms. The vast majority of the firms are small or medium sized and are representative of all sectors. We then define four snapshots taken at 01/07/2009, 01/01/2010, 01/07/2010 and 01/07/2011 as explained in Figure1. The classification into good or bad firms is made by means of a "default indicator". The latter is conducted by checking whether the firms go bankrupt six months after a given observation period, called an "outcome window". For example, for the group of active firms observed at 01/07/2011, the associated outcome window lasts until 31 December 2011 and aims to distinguish defaulted firms from non-defaulted firms. Selecting an appropriately sized outcome period requires careful consideration so as to capture a representative sample of defaulted firms with which to build a stable classifier model. To ensure that we obtain enough defaulted firms in our sample, we plot the cumulative default rate curve by time, measuring the bad cases of defaulted firms captured as the outcome window size increased. We find that a fixed period of six months is good enough to group representative defaulted firms. Recent studies have shown that the forecast accuracy of failure scores diminishes with the forecast horizon from one month up to six months. The credit scoring literature does not contain strong recommendations on how far forward into the future we should explore to make reliable predictions or how many months should be used to build the model. Apart from getting a better grip of default cases, the choice of six months for the outcome window is subordinate to the aim of our current study which is predicting short-run bankruptcy for French SMEs.

The period before the observation point is known as the performance period. In this period, we designed a set of explanatory variables and indicators to observe past payment behavior.
Again, it is crucial to select an adequate fixed period for the performance window to alleviate instability in making predictions rather than selecting a period arbitrarily. However, there is no standard way of defining the length of the performance and the outcome window. Kennedy et al.'s paper (2012) is an attempt to deal with these issues when modeling credit scoring for consumer loans. The authors compare the accuracy of scoring models that are built using different durations of historical customer repayment data (six months, 12 months and 18 months). Then they quantify the differences between varying outcome periods from which a customer's default status is defined (three months, six months, 12 months, 18 months and 24 months). Kennedy et al. (2012) consider that a 12-month performance window is best suited to the classification task, particularly when outcome window sizes of three months, six months and 12 months are specified. They emphasize the fact that seasonal trends could be ignored with a performance window of six months. However, we prefer to challenge their findings and carry out the research taking only six months of past payment data. The information held during this fixed period contains all we need to build a short-term bankruptcy prediction model. Our hypothesis is also supported by the fact that a shorter performance window, especially a six-month window, is enough to capture the highly dynamic events of payment incidents capable of signaling future distress and hence future bankruptcy. This might not be the case if we used accounting information over a longer horizon of time. Apart from these arguments, we believe that SMEs are more impacted by the short run fluctuations that may affect their capital. Rapid changes in the financial situation have an immediate impact on future bankruptcy for SMEs, compared to larger businesses.

As detailed above, all information is aggregated at the level of the firm. One firm may have several payment incidents within the same month. Since we conducted monthly observations it was necessary to aggregate all the data at the level of the firm. We obtained a total of 3, 807, 598 observations. One firm may appear from 1 to 4 times in the whole sample, depending on whether it defaulted or not during a given performance period. It is worth noting that if a firm defaults during a given observation period, it is deleted from the following sub samples.

Figure 1
The Time Window of Analysis and The Constructions of The Sample



Design and refinement of variables

Behavioral scoring uses the characteristics of customers' recent behavior to predict whether or not firms are likely to default. Typical variables would be the average, maximum and minimum level. Other characteristics estimate the trends in payment or simply number of missing payments. We do not make any assumptions before data computation and statistical analysis. We would not suppose that some factors would affect the dependent variable in advance. The task of this phase is to design as many variables as possible. The stepwise process will retain the most significant and discriminant explanatory variables for our model. As a reminder, our goal is to verify the predictive power of late payment patterns of trade credit and cases in which this might not be sufficient. Also, we try to fine tune the variables set, to improve the model's performance. The detailed potential variables we explored are listed in Table A1 in the Appendix.

Dependent variable:

The dependent variable y_i is a binary set that indicates whether the firm becomes inactive during the following six months after a given observation date. We are in line with the Basel definition of legal default. A firm is considered at default if it goes bankrupt after a turnaround procedure or judicial liquidation.

According to the latter definition, we can recognize bad (y_i=1) and good firms ($y_i = 0$).

$$y_i = \begin{cases} 1 & \text{if firm i faces a judicial proceeding in performance period} \\ 0 & \text{if firm i didn't face any judicial proceedings in performce period} \end{cases}$$

Firms' identity

A firm's size, measured by the annual turnover, is supposed to play a central role in determining the level of credit risk parameters for default probabilities. Evidence from the literature generally supports the hypothesis that large firms are less likely to default because they have better access to various financing sources and they are less vulnerable to payment incidents. For the purposes of our paper, we consider a SME as a firm with fewer than 250 employees, which corresponds to 99 per cent of existing firms selected on 1st January 2010.

Historical payment behavior:

We first collect instances of commercial litigation, which are business disputes that might occur when one business partner sues the other partner for breaching a partnership agreement (defective goods, missed quantities, to name a few). They are considered as a severe payment incident and could be a sign of financial distress. 1.60% of the total sample had at least one incident of commercial litigation within the six months before a given observation date. The default rate for that 1.60 per cent of the total sample is equal to 7.45 percent, all else being equal, seven times higher when the firm didn't experience any commercial litigation during the preceding six months.

Secondly, we treat all data about sums due on trade credit, a term used to describe the time elapsed between the reception of a bill and the actual remittance of the payment that is due. The historical data covers 5.55 per cent of the total sample. The occurrence of at least one late payment during the preceding six months (from day 1 to six months of overdue trade credit) corresponds to a default rate equal to 1.36 per cent, all other things being equal, while 0.86 percent is the default rate of firms with no arrears on receivables during the preceding six months.

In addition, we investigate unpaid trade bills, collected from the Banque de France. The main reason for a default payment on trade bills is "Impossibility to pay" which occurs in many cases. From the official definition of "Incapacity to pay" by the Banque de France we enumerate these cases, to name but a few: holder dead, request for prorogation, shortage of funds, payment by subrogation, insufficient deposit, no order to pay, judicial decision, objection to payment on the account. Having at least one unpaid trade bill within the last six months provides a default rate of 7.35 per cent, all other things being equal, against 0.65 per cent for those firms with no unpaid trade bills. This variable seems to be very discriminant. This information covers 3.5 percent of the total sample.

Finally, we consider the unpaid social charges and taxes to State creditors (so-called Privilèges URSSAF) which correspond to a default in payment of legal liabilities (taxes and other liabilities) due to the Public Treasury and French Social Security. Firms should pay as a priority, on behalf of their employees, taxes and other liabilities to the State creditor. From this point of view, the existence of unpaid 'Privilège URSSAF' is indicative of serious and severe financial problems. 12.09 per cent is the default rate of 0.62 per cent of the total sample that has debts to state creditors. 1.29 per cent is the default rate of 7.75 per cent of the total sample for which no debt toward State creditors was recorded.

For each of these incidents of payment, we use a specific file provided by either Coface Services or General Electric Capital France. These files contain all the information about the frequency of each incident of payment during a given period, the amount and the date, all at the firm level. Table A1 in the Appendix shows other indicators and variables constructed from these files and used for the purpose of our paper.

Methodology

The underlying hypothesis is that a higher rate of late payment on trade credit will be associated with a lower default rate. To test for the latter hypothesis, we proceed as follows. We classify borrowers' firms into rating classes with respect to their default probability. The classification of firms into rating classes necessitates the finding of threshold values separating the rating classes. We aim at solving two problems: to distinguish the defaults from non-defaults and to put the firms in an order based on their payment behavior. To use a model to obtain the probability of default of each firm's receiver operating characteristics (ROC), analysis is employed to assess the distinguishing power of our model.

The logistic regression approach is used to identify short-run bankruptcy with the use of a default indicator. This statistical technique has long been considered as a powerful algorithm (Lee et al. 2006). Its specific form is as follows:

$$P\left(Y = 1 \middle| X_1(j_1), \dots, X_n(j_n)\right) = \frac{1}{1+\exp(\alpha_0 + \alpha_1(j_1) + \dots + \alpha_n(j_n))} \qquad (1)$$

The left side of equation (1) is the probability of default derived from a set of j_n explanatory variables of arrears, payment incidents and other variables as described above.

The transformation of the $\pi(x)$ logistic function is known as the logit transformation:

$$\mathrm{Ln}\, P\left(Y = 1 \middle| X_1(j_1), \dots, X_n(j_n)\right) = \mathrm{Ln}\left[\frac{P\left(Y=1 \middle| X_1(j_1),\dots,X_n(j_n)\right)}{1- P\left(Y=1 \middle| X_1(j_1),\dots,X_n(j_n)\right)}\right] \qquad (2)$$

To estimate the logistic parameters, we proceed by maximum likelihood estimations (Hosmer and Lemeshow 1989).

A common problem in regression analysis is that of variable selection. Often you have a large number of potential independent variables, and wish to select among them, perhaps to create a 'best' model. In order to reach this goal, some forms of automated procedure have been proposed, such as forward, backward or stepwise selection (Harell 2001). One common approach to select a subset of variables from a complex model is stepwise regression. A stepwise regression is a procedure to examine the impact of each variable to the model step by step. The variable that cannot contribute much to the variance explained would be thrown out. There are several versions of stepwise regression such as forward selection, backward elimination, and stepwise.

For the purposes of our article, we decide to apply a stepwise procedure with the logistic regression, which is a combination of the backward and the forward selection techniques. It differs in that variables already in the model do not necessarily stay there. As in the forward selection methods, variables are added one by one to the model according to its F-Statistic. After a variable is added, the stepwise method looks at all variables already included in the model and deletes those that do not hold an F-statistic significant to a chosen level. The iterations stop when none of the variables are significant following their F-Statistic.

To test for model robustness, we conduct several tests: we first test for multi-collinearity, then we look for variable significance, we verify if variables' signs from the logistic regression are as expected, and finally we undertake a ROC curve to validate the model's performance.

Results and Discussion

Descriptive Statistics and Preliminary Study

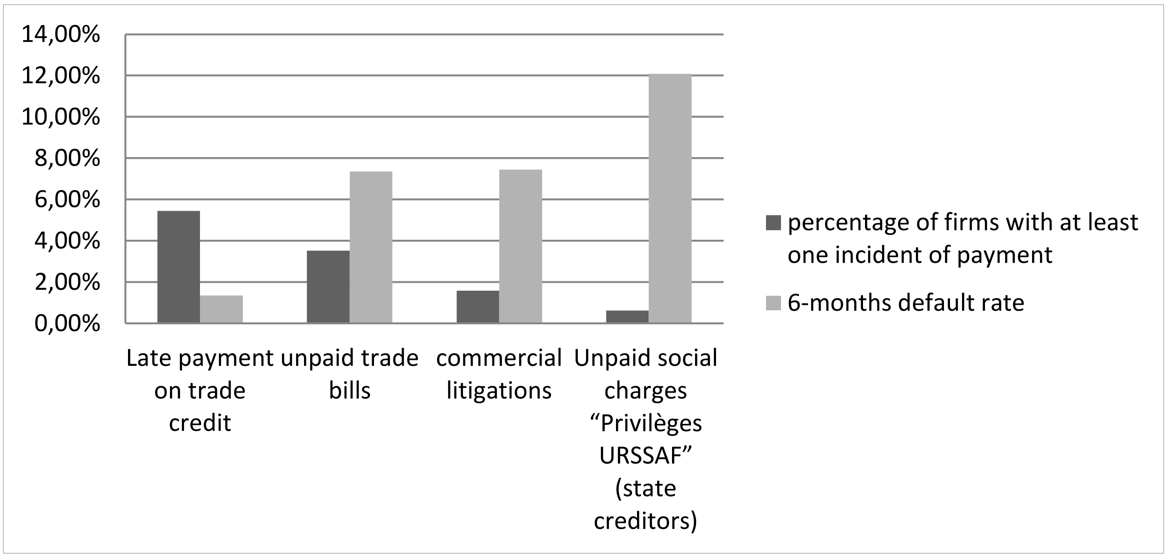
As a reminder, the database has a changing number of obligors from one observation date to another. There are some active obligors observed during a particular observation window that won't go into inactive status until December 2011. Others will default and will disappear at some point from our samples. The model works at the firm level. Therefore, each observation corresponds to a firm on a given observation date, that is, one firm might obtain different default statuses from different performance windows. The total size of the sample is 3, 807,598 observations. We observe the firms at four different dates and we calculate the six-month default rate following delinquencies collected. All tests and regressions are made on the pooled sample, that is, the four samples taken for different observation dates are treated together.

Figure 2 presents the six-month default rate of the pooled sample of firms. This a priori analysis is a suggestion that bankruptcy in a short horizon is affected by all sorts of delinquency in payment. It is worth noting that the default rate of the total sample is equal to 0.9%. The comparison between default rates of firms with any pattern of delinquency and those with no delinquency provides us with insights on the potential power of the explanatory variables that could be derived. Firms with arrears on trade credit have the lowest default rate. Short-term bankruptcy seems to be less affected by unpaid receivables than by the other listed incidents of payment. Despite the low proportion of firms with "privileges", the latter have the highest impact on default rates (12.09% as a six- month default rate). Indeed, firms can signal a critical financial situation when they start to not pay the government. Unpaid trade bills and commercial litigation have a similar default rate. However, our database contains more information about unpaid trade bills than commercial litigation.

All sectors are represented in the analysed sample with a slightly higher proportion of firms in the construction, trade and manufacturing industries (respectively 16.06%, 25.43% and 9.4% from the total sample). The results in Table A2 in the Appendix confirm again what is usually observed in the French market: firms operating in industries such as manufacturing or construction are more risky than other firms. Their six-month default rates are respectively equal to 1.06% and 1.49%.

As we created many variables, the model was more likely to be complex and over-fitted. The more independent variables there are, the more probability the model had to carry mutually dependent and thus redundant predictors. Variance inflation factor (VIF) is a common way of detecting multi-collinearity. Mathematically speaking: VIF = 1/(1-R-square). As advanced in the literature (Janke and Tinslay 2005), if a VIF exceeds ten, the variable entry to the model becomes problematic. The definition of potential explanatory variables is listed in Table A1 with their VIF. Out of 56 variables tested in the model, 39 variables have less than ten for their VIF. In addition to multi-collinearity issues, correlation may affect the results. The general rule usually used in the literature is to keep variables with a Pearson coefficient of less than 0.7. As expected, variables of the amount of delinquency are highly correlated with variables of the number of incidents. We then get rid of either the first or the second variable. The model has 22 potential variables of delinquencies to explain and predict a firm's bankruptcy in a short horizon.

Figure 2
6-Months Default Rate and Proportions of Small Businesses With Incident of Payment



Univariate analysis and Segmentation

Univariate analysis is done to ensure that all default rates progress in the expected sense with the analyzed variable. Moreover, it enables us to identify classes for each independent variable representing a similar default rate. The process of fine classing allows us to determine which characteristics are worthy of consideration in the development of the model. Each characteristic is investigated to determine the underlying defaulter/non defaulter trends in the data at attribute level for discrete data and in small bands for continuous data. Once the trend has been identified, the attributes are grouped together into finer groups in order to smooth out fluctuations in continuous data and to combine attributes logically within discrete data. This process is aimed at determining whether or not the variable is able to distinguish between bad firms (defaulted) and good firms(non-defaulted firms).
Univariate analysis allows for segmentation. We notice that approximately 69% of firms don’t have any information about their delinquency (because they don’t have any or the relevant data is unavailable). We decide to create different subpopulations and conduct different scores. This alternative is recognized to provide us with better scores.
For our current study, we form four different segments in the following way:

1. Subpopulation with no delinquency features (no litigation, no unpaid trade bills and no unpaid “privileges”) but presents a positive outstanding. No late arrears have been recorded within the observation period. We expect to have a lower default rate for the latter subpopulation compared to 0.9%, the default rate of the whole sample.
2. Subpopulation with no delinquency information, with neither positive nor negative signals. We prefer to consider these firms separately because computing a delinquency model does not make sense operationally for firms with no delinquency characteristics.

A first logistic regression was run on the remaining subpopulations. The signs of some variables don't follow our expectations: indeed we found a negative relationship between probability of default and the increasing amount of commercial litigation. In addition to that, late payment on trade credit seems to be statistically insignificant and has a negative impact on default probability. We observe some fuzzy patterns related to firms that record late payment on trade credit. Firms with no sums due during a horizon window of six months have a short-term default rate of 6.3%, whereas firms that recorded at least once being a month overdue have a short run default rate equal to 1.4%. We also report a non-monotonic evolution of a six-month default rate when the number of delays over the total amount of outstanding bills becomes higher and higher. For instance, having more than 50% of the amount of outstanding bills in 180 delays is less risky than having no arrears in terms of default rate within the last six months Having no arrears within the last six months is much riskier than having about 25% of the outstanding bills owed after 90 delays. We suspect a selection bias at this stage of the analysis as the impact of late payment on short-run default is unclear.

To deal with this fuzzy pattern, we decide to divide again the remaining sample population into two other subpopulations:

1. Firms with late payment on trade credit incidents combined with other payment delinquencies (commercial litigation or unpaid trade bills or so-called "Privilèges URSSAF").
2. Firms with only late payment on trade credit payment incidents. (These firms do not appear on the data base of the other incidents of payment)

We discuss the results of scoring models conducted on the newly created subpopulations in the next section.

Scoring Results

For reasons explained in the previous section, we obtained four different sub-populations. Table 1 shows the numbers of total firms by each sub-population, the total number of defaulted firms and the corresponding default rate. The default rate of the group that we denote as G is highly driven by the occurrence of incidents of payment such as unpaid trade bills, or commercial litigation or unpaid "Privilèges URSSAF". The default rate is equal to 6.7%. According to the adopted segmentation, late payment on trade credit seems to have little effect on short term bankruptcy (0.6% of default rate for denoted group R consisting of firms with only late payment on trade credit). Also, the latter default rate is even equal to those of firms with no information about their late payment practices and less than the average rate of the total sample. We can say that low discriminatory power is associated with the late payment data in our possession.

Table 1
Construction of 4 Clusters of Firms

	Total number of firms	**Proportion of the total sample***	**numbers of defaulted firms**	**6-month default rate**
Firms with delinquencies, excluding late payment on trade credit	197190	5,20%	13276	6,70%
Firms with only late payment on trade credit	184192	4,90%	1125	0,60%
Firms with positive patterns of payment	867785	22,80%	4429	0,50%
Firms with no available information about their past payment behavior	2558431	67,10%	15080	0,60%

*The pooled sample counts for 3807598 observations at firm level. The average 6-month default rate of the total sample is 0.9%. Statistics are related to firms that have at least one kind incidents of payment.

For each sub-population, we follow the same process. We run univariate logistic regressions to estimate default probability, to check the accuracy ratio and to provide an idea of its predictive power. As we have already tested for variable correlations we run a final logit model on all variables we suspect to be potentially discriminant and significant. We apply a statistical stepwise selection procedure of the 22 initially selected variables. After checking for the slope of variables and its significance, we plot ROC curves of each model to gauge its performance. The evaluation of variables' predictive power is done by analyzing the different attributes with their corresponding default rate. In this sense significant differences among default rates for different values of the variables would suggest that such a variable is potentially relevant to the prediction of default.

Default Rate and Risk Categories for Firms of Group N:

Obviously, we can say nothing about firms with no delinquency data. Those firms are either good firms that pay on time or do not appear in our database. We decide to keep this group of firms. Indeed, their inclusion in the initial sample is essential to derive a good model that separates good and bad firms. For this group no logistic regression is computed. The average six-month default rate is however equal to 0.6% and this corresponds to 67% of the total population. The probability of default in this case is simply equal to the corresponding default rate. We thus have only one risk class.

Default Rate and Risk Categories for Firms of Group P:

This group contains firms with rather positive payment behavior. It represents 22% of the total population with an average default rate of 0.5%. We use two binary variables. The first variable takes 1 when a firm has a positive outstanding but displays no arrears during last 6 months. The second one is constructed from an internal confidential variable in General Electric. It is a specific variable used in-house by risk managers to identify firms that encounter severe incidents and judgments, based on an expert review of a firm's financial situation either by analyzing balance sheets or by looking for substantial information from several sources. This variable takes 0 to specify a firm that honors all its engagements. If, on the contrary, several alerts have been reported concerning payment behavior, delays or other incidents, it takes a positive number different from 0. We distinguish four different risk classes depending on the observed default rates.

As shown in Table 2, we notice that the default rate is driven downward when firms have positive outstanding but respect delayed payment terms (1.1% versus 2%) or when the internal variable is superior to 0 (0.2% versus 0.3%). Variables could be considered as indicators of good payment practices in favor of the firms.

Table 2
Default Rate and Risk Classes for Firms with Good Payment Behavior

Pivot table of two binary variables		Positive amount of outstanding with no past due	No outstanding recorded	Total
Expert Judgment deliver a good judgement about the firm	total number of firms	216882	524327	741209
	number of defaulted firms	387	1720	2107
	default rate	0,2	0,3	0,30%
Expert Judgment deliver a bad opinion about the firm	total number of firms	24924	101652	126576
	number of defaulted firms	272	2050	2327
	default rate	1,1	2,0	1,80%
Total	total number of firms	241806	625979	867785
	number of defaulted firms	659	3770	4429
	default rate	0,3	0,6	0,51%

* The Statistics concern the cluster of firms that have positive past payment behavior. The sample counts for 867785 observations at a firm level and represents 22.8% of the total pooled sample

Default rate and risk categories for firms of group R:

For the third subpopulation (4.6% of the population) dealing with the sole late payment on trade credit information we conduct a univariate logistic regression in order to evaluate the variables' power in predicting default. Between all remaining explanatory variables related to late payment on trade credit practices, those that are taken one month before a given observation date (the most recent data available on arrears are used) are the most significant and respond to the intuitive hypothesis that default rates are higher with high rates of late payment. According to the logistic regression results, the most powerful factors in terms of default prediction seem to be those computed from data one month before a given observation date. The latest information about late payment practices seems to matter more than past information of more than two months. The accuracy ratio for variables on late payment on trade credit up to one, two, three or six months back in the past is altered as we move further into the past. We decide to perform the logistic regression with a step option with only variables related to those of the most recent month. The stepwise selection process decides to keep two variables. However, it must be pointed out that the latter univariate regression results just give us a hint of an idea of potential powerful variables. Further research must be conducted in this area. When conducting a stepwise model, we obtain several negative signs obliging us to reject almost all variables. The Gini of the model is equal to 45.2%.

It is worth noticing from Table 3 that the average default rate on the current subpopulation is equal to 0.6% which is near to the default rate of the second population, corresponding to firms for which no additional delinquency information exists. This implies that late payment data that we have would not have as real a predictive power as expected; this may be explained by the fact that SME firms are more prompt to use late payments and this seems to be a frequent practice. Late payment is frequent in French industry and does not signal severe financial distress. We suspect a relatively low predictive power for late payment. To confirm this result, we compare the extent of late payment's discriminant power with the other incidents of payment for the following sub-population.

Table 3
Logistic Regression Results for Firms With Only Late Payment on Trade Credit

Explanatory Variables	Attributes	Degree of Freedom	Coefficient	Standard Error	Wald Chi-Square	Pr > Chisq
Intercept		1	-5,8743	0,0488	14516.0946	<.0001
1 month past due over the amount of outstanding	under 25% (ref)		0			
	between 25% and 75%	1	0,4219	0,09	21.99**	<.0001
	more than 75%	1	0,7147	0,0733	95.051**	<.0001
Expert Judgment on firm's future perspective	positive opinion (ref)		0			
	Negative opinion	1	1,8268	0,0607	905.0685**	<.0001
R-Square						0,0055
Max-rescaled R-Square						0,0762
Somers' D (Gini index)						0,452
AUROC						0,726

** denote confidence levels of 99%, 95% and 90% respectively. The missing value corresponds to the attrbues used as reference for the logistic regression

To construct segments of risk categories we first classify scores obtained by the previous logistic regression into deciles of the distribution of the score among all the firms of group R. We use the chi-square statistic to decide whether to combine adjacent deciles if their default rates are sufficiently similar. This technique is called "coarse classification" and widely used in the scorecard building process. Finally, we obtain six risk categories with an increasing default risk from segment R1 to segment R6 as mentioned in Table 4. We finally plot the ROC curve corresponding to this sub-population.

Table 4
6-month Default Rate by Risk classes for The Cluster of Firms With Only Late Payment on Trade Credit

Risk Categories	Number of non-defaulted firms	number of defaulted firms	Number of total firms	6-month Default rate
R1	123 107	342	123 449	0,3%
R2	14 857	65	14 922	0,4%
R3	16 242	96	16 338	0,6%
R4	20 721	366	21 087	1,7%
R5	3 355	88	3 443	2,6%
R6	4 785	168	4 953	3,4%
	183 067	1 125	184 192	0,6%

Default rate and risk categories for firms of group G:

We finally move to the last subpopulation which represents 5.4% of the total population with an average default rate of 6.4%. Firms of this group have several past incidents of payment such as commercial litigation, unpaid trade bills and "Privilèges URSSAF" combined with late payment on trade credit.

Table 5 shows the results of the logistic regression conducted on kept variables after a stepwise proceeding. All coefficients appear with the expected signs and are statistically significant. The distribution of default rate by risk classes is displayed in Table 6.

Analogically to group R, the risk categories are based on a firm's probability of default. The PD of an entity is computed via the logit transformation of a raw score given by the logistic regression. We finally obtain eight risk categories, shown in Table 5. The accuracy ratio is equal to 67.5% and the model works well with only seven variables. Our model is well fitted with the retained variables as default rates increase with high numbers/amount/recency of incidents of payment.

Table 5

Logistic Regression for the Cluster of Firms with Multiple Incident of Payments

	Attributes*	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-5.3654	0.1963	746.843	<.0001
Unpaid trade bills	less than 2 trade bills (ref)	1	0			
	2 unpaid trade bills or <=2000€	1	0.4188	0.0276	230.1658	<.0001
	from 3 to 5 unpaid trade bills	1	0.5879	0.0288	417.4179	<.0001
	Total amount of unpaid trade bills less than15000€	1	0.6811	0.0343	393.4447	<.0001
	Total amount of unpaid trade bills less than 45000€	1	0.8423	0.0355	561.9234	<.0001
	Total amount of unpaid trade bills more than 45000€	1	1.0565	0.0396	712.0532	<.0001
Total Number of Privilège URSSAF	no "Privilège URSSAF"	1	0.4169	0.0494	71.2711	<.0001
	one "Privilège URSSAF"	1	0.7459	0.0551	183.4736	<.0001
	Two "Privilège URSSAF"	1	0.9582	0.0367	683.2917	<.0001
	between three to nine "Privilège URSSAF"	1	1.0975	0.0383	822.0039	<.0001
	Ten or more "Privilège URSSAF"	1	0.9518	0.0395	581.1071	<.0001
Total amount of commercial litigations	No Commercial litigations	1	0			
	Total amount of Commercial Litigations less than 3000€	1	0.4729	0.0504	87.9979	<.0001
	Total amount of Commercial Litigations less than 5000€	1	0.551	0.0493	124.7833	<.0001
	Total amount of Commercial Litigations less than 9000€	1	0.7469	0.0503	220.4526	<.0001
	Total amount of Commercial Litigations less than 18000€	1	1.0558	0.0513	423.6483	<.0001
	Total amount of Commercial Litigations more than 18000€	1	1.1627	0.0548	450.9221	<.0001
Date of the most recent delinquency	in the first month before the observation window (ref)	1	0			
	in the second month of the observation window	1	0.1542	0.039	15.6506	<.0001
	in the 3rd month of the observation window	1	0.2183	0.0372	34.5251	<.0001
	In the 4th month of the observation window	1	0.315	0.033	91.2257	<.0001
	In the 5th and 6th months of the observation window	1	0.5775	0.0305	359.518	<.0001
Turnover	Less than 5 000 000€	1	1.9164	0.195	96.6216	<.0001
	Less than 15 000 000€	1	1.5462	0.1962	62.0814	<.0001
	Less than 30 000 000€	1	1.2937	0.2204	34.4432	<.0001
1 month Past Due Over The amount of Outstanding	Less than 50% (ref)					
	between 50% and 75%	1	0.2826	0.1049	7.2566	0.0071
	more than 75%	1	0.3828	0.0557	47.1891	<.0001
Expert Judgment about a firms' future perspective	Negative Expert Judgment (ref)	1	0.9569	0.087	435.867	<.0001
	Missing value	1	0			
	Positive Expert Judgment	1	0.8059	0.1768	294.456	<.0001

* All explanatpry variables related to aggregated amount or number of each incident of payment are calculated within an observation period of six Months.

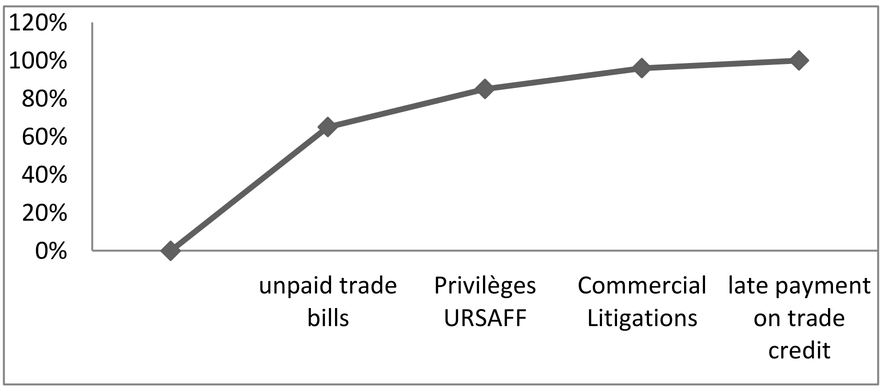
Table 6
Default Rate by Risk classes for Cluster of Firms With Multiple Incidents of Payment

Risk classes	Number of non-defaulted firms	number of defaulted firms	Number of total firms	6-month Default rate
G1	1 401	4	1 405	0,3%
G2	12 245	163	12 408	1,3%
G3	68 161	2 673	70 834	3,8%
G4	55 986	3 653	59 639	6,1%
G5	34 682	4 009	38 691	10,4%
G6	6 271	1 166	7 437	15,7%
G7	2 959	758	3 717	20,4%
G7	2 209	850	3 059	27,8%
Total	183 914	13 276	197 190	85,7%

We notice that other information about delinquencies (privilèges) has a better predictive power than late payment on trade credit. We succeeded by segmentation in resolving the anti-selection bias for late payment on trade credit variable data but we couldn't improve its predictive power. It seems that, among all variables, the latter have the lowest discriminant power to predict default in the short run.

We argue that the model is statistically robust and stable, although we acknowledge that the chosen cut-off for variables and rates is possibly not the optimal one. However, the main objective of the current study is to identify some alternative variables of firms' payment incidents to predict short term bankruptcy. Figure 3 shows that unpaid trade bills and "Privilèges URSSAF" are the most discriminant variables.

Figure 3
Marginal Contribution of each Incident of Payment on Model's Overall Performance for The Cluster of Firms With Multiple Delinquencies



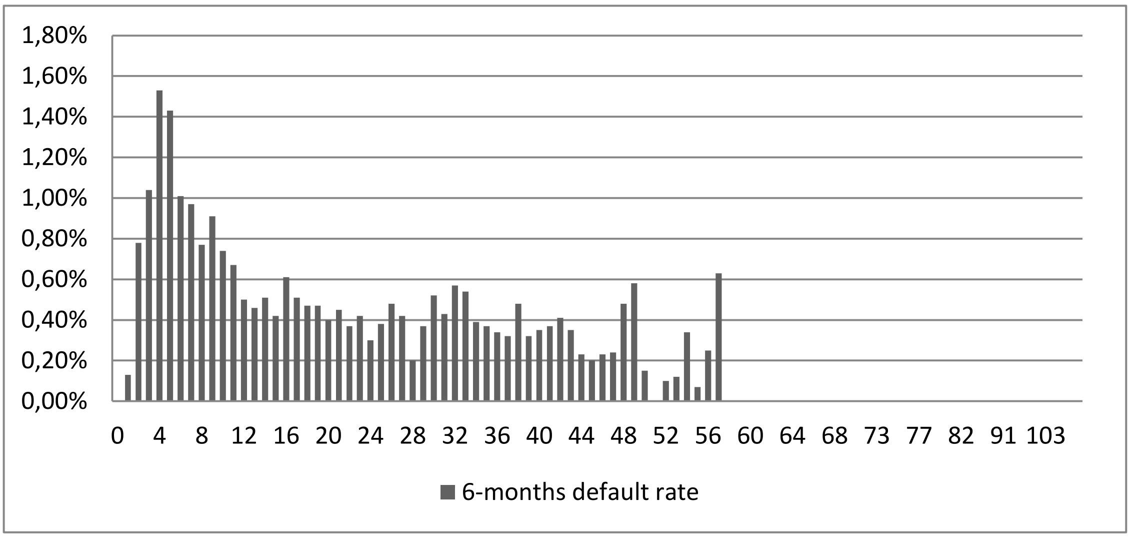
Impact of age, sector and geographic location for French SMEs

In this section, we aim to investigate the effects of location, age and industry on predicting default probabilities for SMEs in France. On the one hand, a mature firm is considered to be more experienced in managing its cash flows and account receivables. Hence delinquencies are less frequent or less important with the age of the firm. On the other hand, while industry and geographic location are relevant risk factors when estimating asset correlations (Duellman and Masschelein 2006), including these two variables in a PD model can increase the model's accuracy and stability.

Here is an attempt to improve the model by including other explanatory variables such as a firm's age, industry and location. We also verify whether these variables fit well with delinquency variables previously identified in the model using a stepwise logit transformation for both groups including delinquencies (group G with several types of incident of payment and group R with only trade credit variables)

Group R

figure 4
Distribution of The 6-Month Default Rate by Age



From Figure 4, we notice that the default rate increases considerably for firms aged from 1 to 5 years (from a default rate of 0.13% to 1.43% respectively). The trend is inversed for firms aged up to 57 years. The default rate is null for firms more than 60 years old.

When we apply the logistic regression to the model, it seems to reject the variable of the firm's age (after transformation into four attributes). The model is slightly improved but provided with non-significant coefficients.

The second factor we investigate in this section is geographic location. In order to test the hypothesis of the existence of location specific risk factors, we run a logit model relating the average default rate by 'département' to the credit event of default. As in Dietsche and Petey (2006), geographic location is identified though the French division into 'départements' which are 92 administrative regions of the French territory (we exclude overseas departments in the analysis). We rank departments into four classes of default rate, built by taking the quartiles of the distribution of departments' default rates. These four attributes are used as dependent variables. The logit transformation with the stepwise option transforms this variable under its discrete format into four attributes. The model is improved as demonstrated in Table 7. The Sommers'D vary from 45.2 (from the model without the location variable) to 51% and the coefficient is statistically significant. The Associated AUROC is consequently improved from 72.6% to 75.5%.

Table 7
Logistic Regression Results for Firms With Only Past Due on Trade Credit With the Inclusion of Location Variable

Explanatory variables	Attributes	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-6.5343	0.092	5046.182	<.0001
1 month Past Due over The amount of outstanding	less than 25% (ref)		0			
	between 25% and 75%	1	0.4232	0.0903	21.9533	<.0001
	more than 75%	1	0.7372	0.0736	100.2292	<.0001
Expert Judgment about a firm's future perspective	negative expert Judgment (ref)		0			
	Positive Expert Judgment	1	1.8357	0.0609	909.4864	<.0001
Geographic Location	1st quartile (ref)		0			
	2nd Quartile	1	0.4537	0.1069	17.9956	<.0001
	3rd Quartile	1	0.7804	0.0981	63.3006	<.0001
	4th Quartile	1	1.0759	0.0971	122.6693	<.0001

Association of Predicted Probabilities and Observed Responses

Percent Concordant	68.9	**Somers' D**	0.51
Percent Discordant	17.9	**Gamma**	0.587
Percent Tied	13.2	**Tau-a**	0.006
Pairs	204887412	**c**	0.755

Finally, we focus on the industry effect and its inclusion in the model. The industry effect has been examined in the literature of credit risk modeling. Chava and Jarrow (2004) show that industry groupings significantly affect both the intercept and slope coefficients in the forecasting equations. Our results are in line with previous work. Furthermore, we find similar results as in Dietsche and Petey's paper (2006): Industry and Services appear to be sectors with a relatively high risk, as expected. We notice that SMEs are less present among extractive industries as detailed in Table A3. SMEs are more concentrated in sectors close to the final consumer, such as hotels and services.
As shown in Table 8, the model was improved when we included the industry variables. The Gini is equal to 51% and the AUROC equal to 75.5%.

To summarize the logistic results derived from the model with the inclusion of three different variables of firms' age, industry and geographic location: we find that age doesn't enhance the accuracy of the model. Age doesn't seem to bring any additional information and explanation for short term default for firms with only missed payment on their account receivables. Late payment on trade credit doesn't seem to follow a specific trend and it concerns all firms. Past dues on trade credit are les controlled than other incidents of payment (see the next section for the same analysis on group G of firms with multiple incidents of payment). However, the model is considerably improved when we include variables of both location and industry. The new model has a Gini of 53.6% compared to the initial one which has a Gini equal to 45.2%.

Table 8
Logistic Regression Results for Firms with Only Incidents of Paiement on Trade Credit With The Inclusion of The Industry Variable

Explanatory variables	attributes	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-7.3809	0.411	322.467	<.0001
1 month Past Due Over The Amount of Oustanding	less than 25% (ref)					
	between 26% and 75%	1	0.4138	0.0901	21.1196	<.0001
	more than 75%	1	0.6932	0.0735	88.9671	<.0001
Expert Judgment about a firm's future perspective	negative Expert Judgment (ref)	1	0			
	positive Expert Judgment	1	1.835	0.0608	910.778 7	<.0001
Industry	1st Quartile (ref)	1	0			
	2nd Quartile	1	1.1398	0.413	7.6155	0.0058
	3rd Quartile	1	1.5642	0.4121	14.4064	0.0001
	4th Quartile	1	1.8912	0.4119	21.0837	<.0001

Association of Predicted Probabilities and Observed Responses

Percent Concordant	69.6	**Somers' D**	0.51
Percent Discordant	18.6	**Gamma**	0.579
Percent Tied	11.8	**Tau-a**	0.006
Pairs	2059503 75	**c**	0.755

Group G

The previous analysis is conducted on the sample of firms with multiple incidents of payment. We start by transforming the continuous variables on discrete attributes. In Table 9 we find that all of the new included variables improve the model. The Somers'D rises from 36.6% to 42.8% and the AUROC from 68.3% to 71.4%.

It is interesting to notice that for this group of firms, age matters. The default rate decreases with age when multiple incidents of payment are considered. In other words, firms learn to better manage their cash flows that might severely impact the firm's capital. The industries included in each attribute differ slightly from those established for group R: they migrate to the next quartile or the previous quartile as shown in TableA4. But globally, Construction and Services are still deemed sectors of high risk.

Table 9

Logistic Regression With the Inclusion of Variables of Age, Industry and Location of Firms

	Attributes*	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-7.118	0.2169	1076.48	<.0001
Unpaid trade bills	less than 2 trade bills (ref)	1	0			
	2 unpaid trade bills or <=2000€	1	0.2737	0.028	95.4354	<.0001
	from 3 to 5 unpaid trade bills	1	0.3727	0.0294	160.5235	<.0001
	Total amount of unpaid trade bills less than15000€	1	0.4853	0.035	192.0408	<.0001
	Total amount of unpaid trade bills less than 45000€	1	0.6371	0.0363	307.3028	<.0001
	Total amount of unpaid trade bills more than 45000€	1	0.8395	0.0407	425.4961	<.0001
Total Number of Privilège URSSAF	no "Privilège URSSAF"	1	0			
	one "Privilège URSSAF"	1	0.2729	0.0503	29.4725	<.0001
	Two "Privilège URSSAF"	1	0.5652	0.0559	102.0801	<.0001
	between three to nine "Privilège URSSAF"	1	0.7677	0.0375	418.1611	<.0001
	Ten or more "Privilège URSSAF"	1	0.9518	0.0395	581.1071	<.0001
Total amount of commercial litigations	No Commercial litigations	1	0			
	Total amount of Commercial Litigations less than 3000€	1	0.5541	0.0511	117.7034	<.0001
	Total amount of Commercial Litigations less than 5000€	1	0.6279	0.0499	158.0705	<.0001
	Total amount of Commercial Litigations less than 9000€	1	0.8135	0.0512	252.6153	<.0001
	Total amount of Commercial Litigations less than 18000€	1	1.1201	0.0525	455.6266	<.0001
	Total amount of Commercial Litigations more than 18000€	1	1.2176	0.0562	469.2563	<.0001
Date of the most recent delinquency	in the first month before the observation window (ref)	1	0			
	in the second month of the observation window	1	0.1442	0.0392	13.5465	0.0002
	in the 3rd month of the observation window	1	0.2223	0.0374	35.3571	<.0001
	In the 4th month of the observation window	1	0.2979	0.0332	80.4638	<.0001
	In the 5th and 6th months of the observation window	1	0.5876	0.0307	367.3516	<.0001
Turnover	Less than 5 000 000€	1	1.1747	0.1981	35.1592	<.0001
	Less than 15 000 000€	1	1.19	0.2026	34.4984	<.0001
	Less than 30 000 000€	1	0.9937	0.2231	19.8458	<.0001
1 month Past Due Over The amount of Outstanding	Less than 50% (ref)					
	between 50% and 75%	1	0.2635	0.1072	6.0439	0.014
	more than 75%	1	0.3562	0.0575	38.3767	<.0001
Expert Judgment about a firms' future perspective	Negative Expert Judgment (ref)	1	0			
	Missing value	1	0.7765	0.0419	342.8707	<.0001
	Positive Expert Judgment	1	0.9531	0.0426	500.0835	<.0001
Geographic location	1st Quartile	1	0			
	2nd Quartile	1	0.2264	0.0257	77.7126	<.0001
	3rd Quartile	1	0.3444	0.0273	159.5274	<.0001
	4th Quartile	1	0.4693	0.0262	320.8949	<.0001
Age of the firm	less than 3 old years	1	0			
	03 to 06 old years	1	1.1357	0.0468	588.5776	<.0001
	07 to 10 old years	1	0.9359	0.0497	355.0397	<.0001

	11 to 15 old years	1	0.7116	0.0525	183.945	<.0001
	more than 16 old years	1	0.5942	0.0502	140.1984	<.0001
	1st Quartile					
	2nd Quartile	1	0.4494	0.0804	31.2452	<.0001
Industry	3rd Quartile	1	0.5259	0.0731	51.7094	<.0001
	4th Quartile	1	0.8964	0.0727	151.9378	<.0001

* All explanatpry variables related to aggregated amount or number of each incident of payment are calculated within an observation period of six Months.

Association of Predicted Probabilities and Observed Responses

Percent Concordant	70.8	**Somers' D**	0.428
Percent Discordant	28	**Gamma**	0.433
Percent Tied	1.1	**Tau-a**	0.054
Pairs	2441642264	**c**	0.714

Validation of the final model

Stability of the models

We assess the stability of the model by observing the accuracy ratios for our tests in the training set and the validation set and also by observing the size, signs and significance of the coefficients for individual variables. We follow the same discretization's variables and form four groups of firms depending on whether they have encountered delinquency in the past.

The Validation set contains 943, 675 new active firms on 1st June 2011. We show here the accuracy results of logistic regressions on the training set for the two R groups. Table 10 shows comparisons of descriptive statistics for the validation data set before clustering into four groups of firms.

Table 10
Descriptive Statistics for the Training Set and Validation Set

	validation set			training set		
Cluster of Firms	**Frequency**	**Percent**	**6 months Default rate**	**Frequency**	**Percent**	**6 months Default rate**
G : firms witrh multiple incidents of payment	52816	5.65	7.3	197190	5.18	6.7
N: Firms with no delinquency information	613332	65.62	0.7	2550187	66.98	0.6
P: Firms with positive past payment behavior	219118	23.44	0.5	876029	23.01	0.5
R : Firms with only late payment on trade credit	49391	5.28	0.6	184192	4.84	0.6

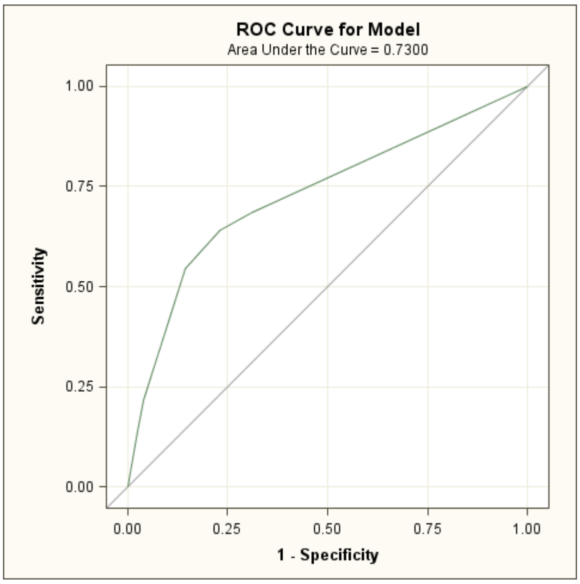
Table 11
Prediction Accuracy Rates for Clusters of Firms R and G

	Group R	Group G
	training set	
Sommer' D	45.20%	36.60%
AUROC	72.68%	68.3%
	validation set	
Sommer's D	45.40%	35.6%
AUROC	73.00%	67.8%

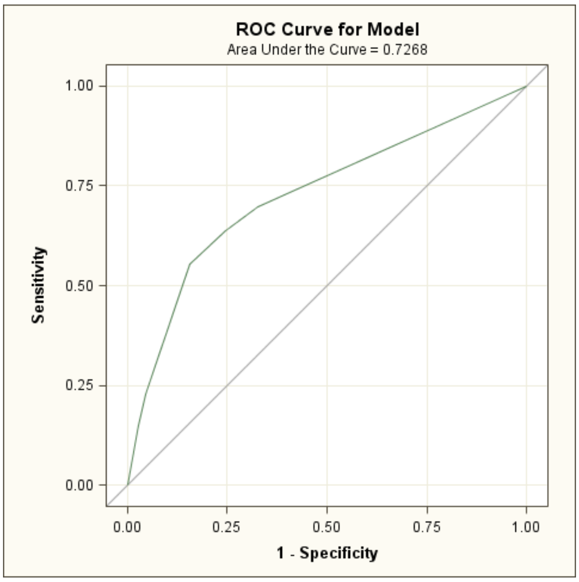
The accuracy ratios were very similar for the two sample periods and the coefficients and significance tests were extremely close. Table 11 highlights this point. The results aren't materially altered from the training set and the validation set. Furthermore, the ROC curves in Figure 5 provide reassuring shapes.

Figure 5
Model Performance of training set and validation set

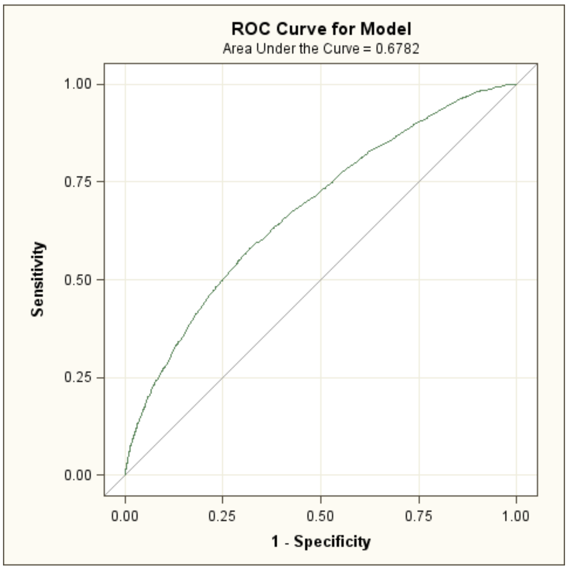
ROC Curve on a Validation Set of the Cluster of Firms with Only past Due on Late Credit



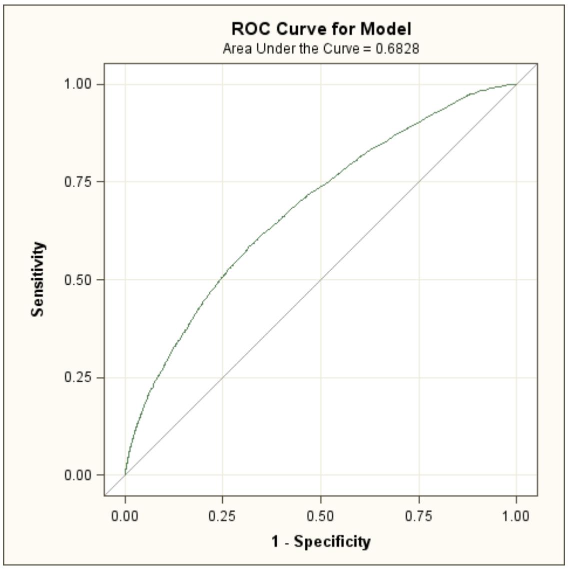
ROC Curve on a training Set of the Cluster of Firms with Only past Due on Late Credit



ROC Curve On Validation Set For The Cluster of Firms with Multiple Delinquencies



ROC Curve On Training Set For The Cluster of Firms with Multiple Delinquencies



Selection of the final model

In this section we merge all the risk categories to identify the risk categories for the final model. We obtain 18 risk categories with increasing default rates from 0.2% to 33.2% (We exclude firms with no available information about their past payment behavior, the cluster of firms G, to obtain the latter results and to conduct the ROC curve of the final model). The corresponding accuracy rate is equal to 71.7%. Figure 6 plots the short default rate by risk category.

Figure 7
ROC Curve for The Final Model

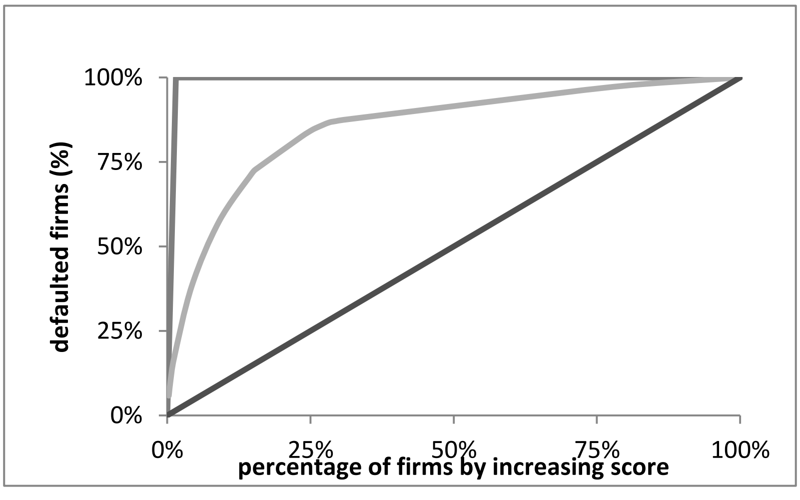
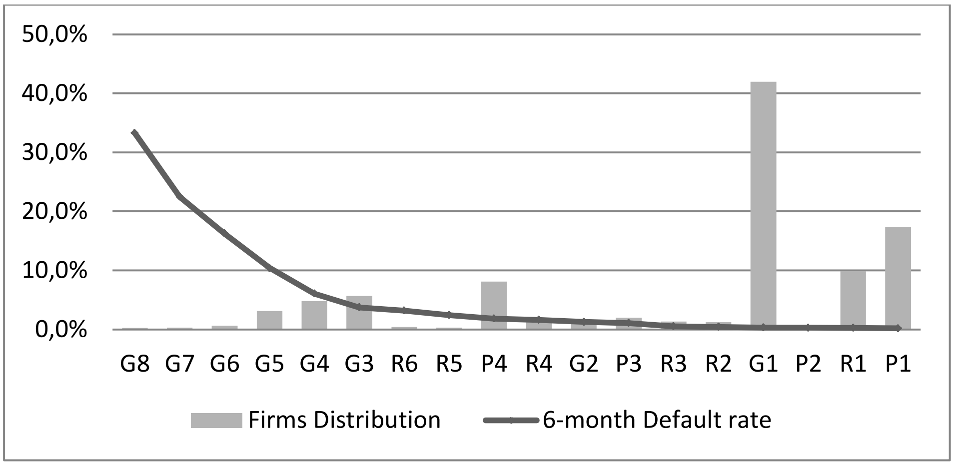


Figure 6
Short term Default rate by risk categories for firms with several incidents of payment



Conclusion

Not much attention has been paid so far to modelling the credit risk of Small and Medium-sized Enterprises over time, although SMEs' exposure is relatively important for European banks. SMEs have specific characteristics that influence the development and use of credit risk models. SMEs are sensitive to the state of the economy. They may be expected to be more likely to fail, because they have less power in negotiations with financial and social partners, are less likely to benefit from their experience or 'learning effects', compared to large firms, and often operate in small markets. Due to the lack of product and market diversification, SMEs face high uncertainty about their future cash flow levels and timing. This leads to inconsistent and volatile financial statement data over time. The financial data of one year can be totally inconsistent with the data of the next year. Until recently, banks still applied to SMEs the same procedures and criteria they used in lending to large firms. But since the 90s, the use of adequate credit scoring models for SMEs has begun to emerge.

This paper aimed to propose another view of modeling the default probability for SMEs when only past information about payment behavior is used. We propose alternatives to traditional scoring models which rely more on financial information. By doing this, we obtain insights into a second dimension that could be taken into consideration when predicting SME bankruptcy. No additional financial ratios or accounting statements are added to the model to avoid jeopardizing existing scoring. Notice that this SME scoring could be seen as an application to use in business-to-business trade credit loans: a lender's clients could benefit from it by reducing their losses from bad trade debts. Risk managers can refer to it as a complementary tool, not a substitute for other credit scoring that relies more on financial information. Apart from looking at the financial situation of business partners, credit managers could be interested in checking whether their client is capable of honoring its future financial commitments.

The originality of our study lies in both the very large size of the dataset we use and the diversity of the data we handle. Beyond qualitative information about French firms, we have access to very precise information on the dates of payment incidents, on trade bills concerning each firm and other incidents of payment. Delinquency variables are widely used in modeling behavioral scores for consumer loans. However this is not the case for corporate credit risk modeling that, to our knowledge, has marginally used this information as a complement to other accounting ratios. This is the first paper that tries to explain corporate default for SMEs with only variables on delinquencies. Our study was initially motivated by a few papers on consumer credit. Our paper provides us with three different conclusions. First, we propose a model to predict short-term default probability based solely on information about the past payment behavior of corporates. With aggregate variables on frequencies, amount and the last date of incident which occurred during a window of just six months, we were able to propose a scoring model with a good accuracy ratio. The result indicates the ability of delinquencies to serve as precursors to potential default. Delinquencies provide signals and information to users in a fashion similar to credit ratings. They are considered a sensitive barometer of a firm's financial condition, updated daily. They have the practical advantage of reflecting the current situation of corporates and can be interpreted as early warning signals of a short-run bankruptcy. Unlike accounting statements that are only available quarterly, this is an attempt to model in a more timely manner and to capture more dynamic features of credit risk. Secondly, the impact of payment incidents on future default differs from one incident to another. Missing payments to the French government (not paying the URSSAF) has the highest explanatory power. The latter reveals that such indicators could forecast real financial distress for the scored firm at an earlier time. Finally, late payment on trade credit has the lowest impact on firms' default in the short run when firms encounter other incidents of payment. The effect of late payment on trade credit becomes less pronounced when no additional delinquencies occurred simultaneously.

It could be interesting to analyze other types of incident combined with demographic and qualitative information. Future research along these lines is surely warranted and can be expected to refine our understanding of the credit risk modeling of SMEs. Besides, it is worth studying the change in credit quality for SMEs. A more through-the-cycle model could be addressed to analyze the ratings dynamics of SMEs under stressed scenarios.

References

Altman, E. I. (1968). "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," *The Journal of Finance* 23(4), 589-609.

Beaver, W. H. (1966). "Financial Ratios as Predictors of Failure," in *Empirical Research in Accounting, Selected Studies*. University of Chicago, 71-111.

Berger, A. N., and W. S. Frame (2007). "Small Business Credit Scoring and Credit Availability," *Journal of Small Business Management* 45(1), 5-22.

Berger, A. N., L. F. Klapper, and G. F. Udell (2001). "The Ability of Banks to Lend to Informationally Opaque Small Businesses," *Journal of Banking and Finance* 25(12), 2127-2167.

Berger, A. N., and G. F. Udell (1995). "Relationship Lending and Lines of Credit in Small Firm Finance," *Journal of Business* 68(3), 351-81.

Bhattacharya, S., and A. V. Thakor (1993). "Contemporary Banking Theory," *Journal of Financial Intermediation* 3, 2-50.

Brealey, R., S. C. Myers, and F. Allen (2010). *Principles of Corpoarte Finance*. MC Graw Hill 10th Edition, New York.

Cavalluzzo, K S., and L. C. Cavalluzzo (1998). "Market Structure and Discrimination: The Case of Small Businesses," *Journal of Money, Credit, and Banking* 68(3), 351-81.

Chava, S., and R. A. Jarrow (2004). "Bankruptcy Prediction with Industry Effect," Review of Finance 8(3), 537-569.

Cook, L. D. (1999). "Trade Credit and Bank Finance: Financing Small Firms in Russia," *Journal of Business Venturing* 14, 493-518.

De Blasio, G. (2005). "Does Trade Credit Substitute for Bank Credit? Evidence from Firm-Level Data," in *Economic Notes Review of Banking, Finance and Monetary Economics* 4(1), 85-112.

Dietsche, M., and J. Petey (2006). "The Impact of Size, Sector and Location on Credit Risk in SME LoansPortfolios," Working paper.

Dimitras, A. I., R. Slowinski, R. Susmaga, and C. Zopounidis (1999). "Business Failure Prediction Using Rough Sets," *European Journal of Operational Research* 114, 263-280.

Duellmann K., and N. Masschelein (2006). "Sector Concentration in Loan Portfolios and Economic Capital," working paper presented at the Journee de l'AFSE, Strasbourg, May.

Edmister, R. O. (1972). "An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction," *The Journal of Financial and Quantitative Analysis* 7, 1477-1493.

Emery, G. W. (1984). "A Pure Financial Explanation for Trade Credit," *Journal of Financial and Quantitative Analysis* 19, 271-285.

Frank, M.Z., and V. Maksimovic (2005). "Trade Credit, Collateral, and Adverse Selection," in Banca D'Italia, June, 553.

Fisman, R., and I. Love (2003). "Trade Credit, Financial Development and Industry Growth," *Journal of Finance* 58(1), 353-373.

Harell, F.E. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis,* Springer Series in Statistic.

Hosmer, D. W., and S. Lemeshow (1989). *Applied Logistic Regression,* New York: John Wiley Sons, Inc.

Janke, S.J., and F. Tinsley (2005). *Introduction to Linear Models and Statistical Inference*, Wiley-Interscience.

Jensen, M., and W. Meckling (1976). "Theory of the Firm: Managerial Behavior, Agency Costs and Capital Structure," *Journal of Financial Economic* 3(4), 305-360.

Johnson, S., J. McMillan, and C. Woodruf (2002). "Courts and Relational Contracts," *Journal of Law, Economics and Organization* 18, 221-277.

Keasey K., and R. Watson (1992). *Investment and Financing Decisions and the Performance of Small Firms,* Robert Watson: Books.

Kennedy K., B. MacNamee, S.J. Delany, M. O'Sullivan, and N. Watson (2012). "A Window of Opportunity: Assessing Behavioral Scoring," *Expert Systems with Applications* 40(4), 1372-1380.

Lee, T., and I. Chen (2005). "A Two-Stage Hybrid Credit Scoring Model Using Artificial Neural Networks and Multivariate Adaptive Regression Splines," *Expert Systems with applications* 28, 743-752.

Lee, T. S., C .C. Chiu, Y. C. Chou, and C. J. Lu (2006). "Mining the Customer Credit Using Classification and Regression, Tree and Multivariate Adaptive Regression Splines," *Computational Statistics and Data Analysis* 50, 1113-1130.

Leland, H., and D. Pyle (1977). "Informational Asymmetries, Financial Structure and Financial Intermediation," *Journal of Finance* 32(2), 257-270.

Lorenzi, J. H, and E. Kremp (2010). *Rapport annuel de l'Observatoire des délais de paiement,* Banque de France.

Love, I., L. A. Preve, and V. Sarria-Allende (2007). "Trade Credit and Bank Credit: Evidence from Recent Financial Crises," *Journal of Financial Economics* 83(2), 453-469.

Meltzer, A. H. (1960). "Mercantile Credit Monetary Policy, and Size of Firms," *The Review of Economics and Statistics* 42(4), 429-437.

Mian, S., and C. Smith, (1992). "Accounts Receivable Management Policy: Theory and Evidence," *Journal of Finance* 47(1), 169-200.

Myers, S. C., and N. S. Majluf (1984). "Corporate Financing and Investment Decisions When Firms Have Information That Investors Do Not Have," *Journal of Financial Economics* 13(2), 187-221.

Paul, S. Y., and N. Wilson (2006). "Trade Credit Supply: An Empirical Investigation of Companies Level Data," *Journal of Accounting Business and Management* 13, 85-113.

Peel, M. J., and N. Wilson (1996). "Working Capital and Financial Management Practices in the Small Firm Sector," *International Small Business Journal* 14(2), 52-68.

Petersen, M. A., and R. G. Rajan (1997). "Trade Credit: Theories and Evidence," *Review of Financial Studies* 10(3), 661-691.

Ponicki, C. (1996). "Case Study: Improving the Efficiency of Small Business Lending at First National Bank of Chicago," *Commercial Lending Review* 11, 51-60.

Rajan R. G., and L. Zingales (1995). "What Do We Know About Capital Structure? Some Evidence From International Data," *Journal of Finance* 50(5), 1421-146.

Saito R., and M. L. Bandeira (2010). "Empirical Evidence of Trade Credit Uses of Brazilian Publicly-Listed Companies," *Brazilian Administration Review* 7(3), 242-259.

Smith, J. K. (1987). "Trade Credit and Informational Asymmetry," *Journal of Finance* 42(4), 863-872.

Van Gestel, T., and B. Baesens (2009). *Credit risk management: Basic concepts,* USA: Oxford University Press.

Walker, E., and W. Petty (1978). "Financial Differences Between Large and Small Firms," *Financial Management* 7(4), 61-68.

Wilson, N., and B. summers (2002). "Trade Credit Terms Offered by Small Firms: Survey Evidence and Empirical Analysis," *Journal of Business Finance and Accounting* 29, 317-351.

Table A1
List of Potential Explanatory Variables

Description	VIF	Description	VIF
cumulative amount of unpaid trade bills within the observation window	1.12	the maximum amount of LP of more than 120 days within the observation window	35.56
cumulative number of unpaid trade bills within the observation window	22.19	the maximum amount of LP of more than 150 days within the observation window	1.14
date of last unpaid trade bills	3.23	the maximum amount of LP of more than 180 days within the last 6 months	3.79
average ratio of the cumulative amount of unpaid trade bill within the observation window by the turnover of the same year	1.04	the average amount of arrears of more than 30 days within the observation window	2.52
average ratio of the cumulative amount of unpaid trade bill within the observation window by the total amount of account receivables of the same year	1.04	the average amount of LP of more than 60 days within the observation window	1.66
cumulative amount of so-called 'Privilèges URSAFF' within the observation window	2.82	the average amount of LP of more than 90 days within the observation window	2.54
cumulative number of 'Privilèges URSAFF' within the observation window	2.7	the average amount of LP of more than 120 days within the observation window	9.94
date of last 'Privilèges URSAFF'	1.03	the average amount of LP of more than 150 days within the observation window	1.03
cumulative amount of commercial litigations within the observation window	1.22	the average amount of LP of more than 180 days within the observation window	2.30
cumulative number of commercial litigations within the observation window	19.76	the percentage of LP of more than 30 days over the amount of outstanding taken1 month before a given observation date	7.89
date of last commercial litigations	1.42	the percentage of LP of more than 60 days over the amount of outstanding taken1 month before a given observation date	1.17
the maximum amount of LP on trade credit of mroe than 30 days within the observation window	1.32	the percentage of LP of more than 90 days over the amount of outstanding taken1 month before a given observation date	3.23
the maximum amount of arrears of more than 60 days within the observation window	9.28	the percentage of LP of more than 120 days over the amount of outstanding taken1 month before a given observation date	1.4
the maximum amount of LP of more than 90 days within the observation window	1.22	the percentage of LP of more than 150 days over the amount of outstanding taken1 month before a given observation date	1.0
the percentage of LP of more than 180 days over the amount of outstanding taken1 month before a given observation date	9.9	the ratio of cumulative amount of LP of more than 150 days divided by the total LP of trade credit in the observation window	89.4
Dummy variable that takes 1 if the total amount of LP of more than 30 days arises between the beginning and the end of the observation window	2.5	the ratio of cumulative amount of LP of more than 180 days divided by the total LP of trade credit in the observation window	83.6
the maximum amount of outstanding within the observation window	2.56	the cumulative amount of all incident of payment (minus LP on trade credit) in the observation window divided by the last turnover recorded	41.2
the minimum amount of outstanding within the observation window	16.61	the percentage of outstanding over LP of more than 30 days 3 month before a given observation date	25.9
turnover in euros	4.02	the percentage of outstanding over LP of more than 60 days 3 month before a given observation date	9.0
Dummy variable that takes 1if the outstanding is positive but no LP recorded during the observation window	4.41	the percentage of outstanding over LP of more than 90 days 3 month before a given observation date	3.79
The ratio of late payment of at least 30 days by the cumulative amount of arrears of more than 1 month before a given observation date	2.22	the percentage of outstanding over LP of more than 120 days 3 month before a given observation date	25.0
the number of incident of payment in the observation window	9.92	the percentage of outstanding over LP of more than 150 days 3 month before a given observation date	1.08

the date of the last incident of payment	3.25	the percentage of outstanding i LP of more than 180 days 3 month before a given observation date	7.40
the ratio of cumulative amount of LP of more than 30 days divided by the total LP of trade credit in the observation window	10.83	the percentage of outstanding over LP of more than 30 days 6 month before a given observation date	14.87
the ratio of cumulative amount of LP of more than 60 days divided by the total LP of trade credit in the observation window	19.74	the percentage of outstanding over LP of more than 60 days 6 month before a given observation date	1.81
the ratio of cumulative amount of LP of more than 90 days divided by the total LP of trade credit in the observation window	20.17	the percentage of outstanding over LP of more than 30 days 6 month before a given observation date	14.87
the ratio of cumulative amount of LP of more than 120 days divided by the total LP of trade credit in the observation window	1.08	the percentage of outstanding over LP of more than 60 days 6 month before a given observation date	1.81
the percentage of LP of more than 180 days over the amount of outstanding taken1 month before a given observation date	9.9	the ratio of cumulative amount of LP of more than 150 days divided by the total LP of trade credit in the observation window	89.4
Dummy variable that takes 1 if the total amount of LP of more than 30 days arises between the beginning and the end of the observation window	2.5	the ratio of cumulative amount of LP of more than 180 days divided by the total LP of trade credit in the observation window	83.6
the maximum amount of outstanding within the observation window	2.56	the cumulative amount of all incident of payment (minus LP on trade credit) in the observation window divided by the last turnover recorded	41.2
the minimum amount of outstanding within the observation window	16.61	the percentage of outstanding over LP of more than 30 days 3 month before a given observation date	25.9
the percentage of outstanding over LP of more than 90 days 6 month before a given observation date	1.09	the percentage of outstanding over LP of more than 150 days 6 month before a given observation date	20.17
the percentage of outstanding over LP of more than 120 days 6 month before a given observation date	12.28	the percentage of outstanding over LP of more than 180 days 6 month before a given observation date	37.4

Table A2
6-months Default rate distribution by firm's industry*

	Proportion	6-months Default rate
Agriculture, forestry and fishing	1,08%	0,42%
Manufacturing	9,40%	1,06%
Production and distribution of electricity, gas, steam and air conditioning	0,55%	0,09%
Production and distribution of water, sanitation, waste management and remediation activities	0,35%	0,05%
Construction	16,06%	1,49%
Trade, repair of motor vehicles and motorcycles	25,43%	0,87%
Transportation and storage	3,12%	1,22%
Accommodation and food	7,93%	0,80%
Information and communication	4,39%	0,69%
Real estate activities	6,82%	0,36%
Administrative activities and support services	4,74%	0,94%
Education	1,25%	0,71%
Human health and social work	1,65%	0,31%
Arts, entertainment and recreation	1,03%	0,93%
Other service activities	3,32%	0,87%
Extractive industries	0,16%	0,33%

* The pooled sample count for 3807598 observations. Financial activities and public administrations were excluded from the sample. There is 215 802 observations with missing information about their industry. The 6 months Default rate of the pooled sample is equal to 0.9%

Table A3
defaut rate by quartiles of French industry

department quartiles	industry	number of firms	6 months default rate
1st quartile	D: Production and distribution of electricity, gas, steam and air conditioning L: Real estate activities B: Extractive industries	4187	0.14%
2nd Quartile	P:Education Q:Human health and social work G: Trade, repair of motor vehicles and motorcycles E: Production and distribution of water, sanitation, waste management and remediation activities I: Accommodation and food	72365	0.42%
3rd Quartile	N: Administrative activities and support services R: Arts, entertainment and recreation J: Information and communication C: Manufacturing	59819	0.66%
4th Quartile	M: management and remediation activities A: Agriculture, forestry and fishing H: Transportation and storage F: Construction S: Other service activities	47821	0.88%

TableA4
6-month Default Rate by Quartiles of French Industry for Cluster of firms With Multiple Incident of Payment

department quartiles	industry	number of firms	6 months default rate
Industry_Q1	L: Real estate activities Q:Human health and social work E: Production and distribution of water, sanitation, waste management and remediation activities	7321	2.83%
Industry_Q2	A: Agriculture, forestry and fishing D: Production and distribution of electricity, gas, steam and air conditioning B: Extractive industries	16580	4.84%
Industry_Q3	I: Accommodation and food S: Other service activities R: Arts, entertainment and recreation J: Information and communication G: Trade, repair of motor vehicles and motorcycles M: management andremediation activities	88078	5.75%
Industry_Q4	P:Education N: Administrative activities and support services H: Transportation and storage F: Construction C: Manufacturing	85211	8.45%