Predicting default probability using delinquency: the case of French SMEs

Asma Marouani

Université Cergy-Pontoise, 2013

Draft version

Abstract

In this paper we explore the hypothesis that probability of default (PD) could be measured and explained by the historical data on ability and willingness of a firm to pay its creditors. We report an application of credit scoring to model default on a large data set of French SMEs.

We find that payment behavior data can be used to predict successfully SME bankruptcy in a short horizon of 6 months. New variables on late payment and delinquency are identified as alternatives to what is usually know in failure models literature.

Keywords: bankruptcy, behavioral scoring, delinquency, late payment, trade credit, credit scoring, Logistic regression

Content

1.Introduction

For over 20 years, the prediction of firm bankruptcy appears to be of paramount interest of risk managers of banks and other non-financial lenders. Among the many important questions highlighted by advanced researches on risk management, one stressed the features and the determinant of a company failure. Most bankruptcy and recovery models have for long used financial ratios as representations of 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 7 financial ratios. In general, models are based on liquidity, profitability ratios with others variables on activity and financial leverage. Banks were inspired by these researches to build their models in assessing firms' bankruptcy and still use them nowadays. However such models were essentially adapted to large and medium firms.

Empirical studies have shown that SMEs, in comparison to large firms, are less liquid, 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 risks involved when lending to small firms. This is why SMEs are more likely to be credit rationing. In order to encounter financial differences between large and small firms, models for SMEs begin to emerge. Other critics were addressed to financial-based models: financial ratios are statics and do not look for dynamical 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 SME segment, i.e. Multiple Discriminant Analysis (MDA), are less likely to capture seasonality and cycles. Even if 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 SME is still problematic. Small banks are reluctant to lend to SME and large banks claim more and more guarantees in return with a higher interest rate in most cases. Not surprising that small firms are particularly relying in short-term sources of finance which was reflecting in the extension of trade credit in several countries. Considered among a cash flow element, trade credit involves supplying goods and services with a possibility to pay with an agreed delay. The growing importance of trade credit for SME reflects somehow a deteriorated banking relationship. In one hand, SMEs use it as a complement or a substitute of financial resources (Brealey et a.l, 2010). In the other hand, SMEs perceive it as a strategic tool to generate profit and business by lending to their business partners (Petersen and Rajan, 1997). In that way, managing late payment in a trade credit context appear to be on the top interest of managers. Indeed, trade credit practices are not just a cash flow issue, but also an important way to signal a one's reputation and financial health (Paul and Wilson, 2006; Wilson and Summers, 2002). In some developed countries, trade credit, represented by accounts payable in the borrowers' balance sheets and by accounts receivable in the creditors' balance sheets, exceeds short-term bank credit and is an important way of financing firm's working capital (Peel and Wilson, 1996).

Several economic implications could be derived justifying our intent to peruse this path of research. A natural question emerged: could trade credit practices say more about firm financial situation and their credit risk than do financial accountings? Indeed, we do believe that more dynamical late payment patterns derived from creditors enclose as such information about SMEs' situations as do financial accounting. We can even talk about informational advantage for the benefit of nonfinancial lenders. Such informational advantage arises due to the fact that trade creditors are mostly engaged in the same nonfinancial transactions that 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 that information held 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 in this path of research, due to the unavailability of relevant published data and the reluctance of firms to communicate about information regarding their trade credit practices. In the current study, we depart from the trade credit literature in at least two aspects: firstly, it extends trade credit management literature by empirically quantifying cutoffs under which it becomes alarming about bankruptcy. 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 refers to missing payment 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, etc. but never on SME's bankruptcy. We investigate on the most powerful explanatory variables reflecting payment patterns to predict default probabilities using a credit scoring for French SME cases. Exploring delinquency behavior allows us to consider for more dynamical aspects to the credit risk assessment. We believe that many risk factors remain to be identified when evaluating risk default of 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 PD estimation specifically for SMEs.

Our paper is divided as follow; 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 3 presents the methodology and variables we use. Section 4 presents and discusses our results and section 5 concludes.

2. Review of Literature

2.1 Trade credit: its motives and determinants

Managing cash flow and working capital efficiently go through a good credit management practices. They have often been considered as pivotal to the health and performance of firms. Even for SME, dealing with capital working issues represents a great concern particularly where small firms are growing and therefore need to finance increasing amount and debtors. Researches in recent years have focused on trade credit expansion as one essential element of cash flow management. Since Meltzer (1960)'s paper, enlightening statistics have recognized the importance of trade credit. Obviously, in industrialized economies, the volume of trade credit is higher than short-term loans received from banks (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 system 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. Literature usually refers to transactional motives and financial ones. First, trade credit is becoming an important form of credit when firms encounter credit rationing problem. Petersen and Rajan (1997) explains 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 from UK and found a negative relationship between bank finance and trade credit, implying that Trade Credit is used as substitute to other more traditional way of financing. Secondly, firms with a better access to credit agree to engage themselves in credit relationship with their suppliers seeking for 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 signaling tool to 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 segment expresses its demand (Smith, 1987) by buying higher quantities at lower prices. Total profit for suppliers increases even under a lower initial price and is essentially realized thanks to price discrimination into a more flexible payment delays. From this perspective, the credit period can give the opportunity to reduce informational asymmetries about product quality and the seller reputation, which makes trade credit a signal of product quality and seller reputation.

Trade credit choices may differ from one firm to another depending on several factors. Indeed, the company size is one of the most discriminating factors when it comes to financial choices of individual firms. In theory, it appears that large firms have a relatively high bargaining power which results to larger payment intervals that may be due to the importance of contracts and the confidence they inspire. At the same time, external funding sources available for the companies are more numerous as its size is larger. Numerous indicators have been used to measure the influence of the firm size factor in most of the empirical works 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 do not stand for the conclusion that larger firm 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 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 researches assume 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 researches have shown that ethnic and sociocultural 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 provide 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 researches 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 in somehow extent of reliance on trade credit.

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 questioned the fact that financial alternatives could be better developed in poorly developed counties to encounter credit access problems. Love et al. (2005) examines the effect of financial crisis on trade credit in six merging economies. They found 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 substitute for bank loans in countries with poor financial institutions.

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

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

Short-term sources of funding, i.e. trade credit play a significant role to support the growth of firms and have numerous advantage for both nonfinancial lenders and firms borrowers. However, some limits may rise especially for small firms. One should not forget about cost 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 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 constraint 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 tried to settle numbers of rules to manage the credit granted to firms' customers. In many countries, companies and government worked together to establish an effective credit policy management aimed to prevent from delayed payment which is the major factor behind the business failure (Wilson and Summers, 2002). Credit policies are used internally to monitor firms' bad debt. In UK, for instance, the debate still persists on the effectiveness of interest penalties on late payment in trade credit context. In France, 2008's LME law (Loi de Modernisation Economique) have 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 written to the customers before the sale takes place. Consequently, legislation seems to do very little to deal with late payment problems. Indeed, companies often ovoid to adopt extreme penalties (charging interest on late payment, pursuing borrowers with overdue trade credit through courts, etc.) for several reasons. Firms may alter their relationship with their partners, especially large ones. There is evidence that firm's size is positively correlated with the trade terms and claims conditions that allow adopting such extreme measures. It is worthy to notice that larger firms have bigger bargaining power. It appears obvious that small firms are by consequence reluctant to take actions 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 risk associated. In that sense, the credit management becomes vital when firm's performance may be altered in 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 actions that could be taken to face late payment on trade credit, firms may consider other internal credit management policies such as setting credit limits and setting cash flows target. But, due to their low cost, statistically derived credit scoring models have been proven to be reliable tools to predicted delinquency for instance. Initially developed in consumer market (screening, pricing and monitoring consumer credit accounts), models of scoring have been used worldwide in consumer lending for some time and their role has expanded internally among credit managers to address risk profile of customers. Banks started to use these statistical techniques to moderate terms loans as credit card loans mortgages and other consumer loans. Nowadays, non-financial lenders use it more and more for their internal purposes.

The most significant development in the last years has been the development of scores for small business. Adjusting for SMEs's 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. Literature tends to distinguish two types of information generally used when applying credit scoring. First we find hard information collected from credit bureau or financial statements used for underwriting decisions (Berger and Frame, 2007). The second type of information includes soft qualitative data gathered throughout the relationship with borrowers and lenders (Berger, Klapper and Udell, 2001). Other purposes for the credit scoring systems were identified in the literature such as estimating 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 few names.
According to Berger and Frame (2007), Small Business Credit Scores increases 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) find 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 model holding 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 credit worthiness at some late date. Both types of credit scoring applications were extended to larger fields as commercial credit, credit card, 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 an accurate tendency for lenders to buy delinquency data from credit bureau as they became aware of their utility in the credit scoring process. Obviously, the longer the payment is past due, the more it will hurt your score. Estimating, then, probability of default relies in great part on payment historical 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 allowed banks to offer more or better services that may have raised costs, but that customers were willing to pay more for these services, raising revenues by more than the cost increases .The second focus rather on the credit availability. A number of studies found that large banks tend to devote lower proportions of their assets to small business lending than smaller institutions. Our study is different from the existing literature. We are rather interested in credit scoring models for small business to take into accounts their peculiarities and heterogeneity. Second, we try to provide insights on how credit past behavior works to predict future bankruptcy. For the latter, we use dynamical patterns such as trade credit practices combined with other incidents.
Again there was little attention toward credit scoring models integrating the repayment behavior of Small firms, especially incidents collected from public administrations. This incites us to look into the variables that are relevant when predicting SME’s default. The next section describes the sample used for the purpose of the study and explains the statistical approach adopted.

3. Data collection and methodology

3.1. Data collection

Data on 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 (LP), we can distinguish four main other payment incidents that will be used in our current study. Historical arrears on trade credit cover all clients of Factofrance, one of the major factor in the French market and belongs to General Electric Group. The 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 involved 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 at a firm level. In our study, we only use frequencies and amounts of unpaid invoices that exceed 1 month, 2 months, 3 months, 4 months, 5 months and 6months.

For in-house modeling purposes, General Electric uses historical data about unpaid trade bills on its clients. We were able to use the same data but, for confidential concerns, we won't divulge detailed results on this part of data.
The paper employs three other different types of payment incidents, all gathered at a firm level and collected from a French data provider Coface Service. They fall into the two following categories: commercial litigations, debt to French government (so-called "privilèges URSSAF" in France).
Finally, we add firm's identity variables such as the age, the department, the legal status and the firm's size measured by the total turnover.

The initial sample counts for 1 500 000 active commercial French company at beginning of July 2009. Public administrations and insurance/financial activities have been 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 observe the active firms from the beginning of July 2009 and set the "default indicator" to verify whether a firm goes bankrupt by 31st of December 2011. We create 4 different sub-samples for which we split the time into two phases, i.e. the "observation window" of 6 month and the "performance observation" for 6 months. The period after the observation point is known as the outcome window or performance period. In the observation period, we designed set of explanatory variables and indicators to observe the past payment behavior. We then define 4 snapshots taken at 01/07/2009, 01/01/2010, 01/07/2010 and at 01/07/2011 (see figure 1).The choice of 6 month for the performance window is subordinated to the aim of our study witch is predicting a short-run bankruptcy taking into account the sole historical payment variables. However, there is an interesting area of research dealing with the length of the performance and outcome window. The recommendations in the literature typically range from 6 to 24 month (Gestel and Baesens, 2009).

As detailed above, all information is aggregated at the firm level. One firm may have several payment incidents within the same month. Since we conduct monthly observations it was necessary to aggregate the all data at the firm level. We obtain a total of 3 807 598 observations. One firm may exist from 1 to 4 times in the whole sample, depending whether she defaulted or not during a given performance period. It is worthy to notice that if a firm default during a given observation period, it is deleted from the followings sub samples.

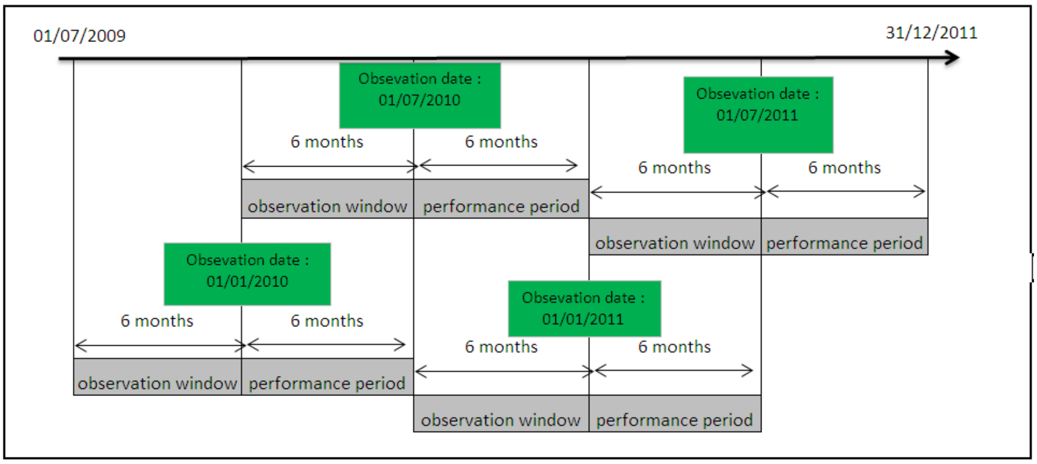

Figure 1: The time window of analysis

3.2. Variable design and refinement

Behavioral scoring uses characteristics of customer's 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 trend in payment or simply numbers of missing payment, etc. We do not any assumption 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 fine tune the variables set to improve models performance. The detailed variables are listed in Appendix 1.

Dependent variable:

The dependent variable y_i used is a binary set that indicates whether the firms become inactive during the following 6 months after a given observation date. We are in line with the Basel definition of legal default, i.e firms is considered at default if it goes bankrupt after turnaround procedure or judicial liquidation.
According to the latter definition, we can recognize bad (y_i=1) and good firm ($y_i = 0$).

$$y_i = \begin{cases} 1 & \text{if firm i face a juridical proceeding in performance period} \\ 0 & \text{if firm i didn'tface any juridical proceeding in performce period} \end{cases}$$

Firms' identity

Firm's size: we generally observe that large firms are less likely to default because they have better access to various financing sources and they are less vulnerable to payment incident.

For the purpose of our paper, we consider a SME as firm with less than 250 employees which corresponds to 99% of existing firms selected at 1st January 2010.

Historical payment behavior:

Commercial litigations considered as a severe payment incident and could be a sign to a financial distress. 1.60% of the total population has at least one commercial litigations within the last past 6 months before a given observation date. The default rate for that 1.60% of total population is equal to 7.45%, all else equal, seven times much higher when the firm didn't experience commercial litigations during the last 6 month.

Late payment on trade credit: The historical data covers 5,55% of total population. The occurrence of at least one late payment during the last 6 months (from 1 day to 6 month of overdue trade credit) correspond to a default rate equal to 1,36%, all else equal, while 0,86% is the default rate of firms with no arrears on receivables during the last 6 months.

Unpaid trade bills: having at least one unpaid trade bills within the last 6 months provide a default rate of 7.35%, all else equal, against 0,65% for those firms with no unpaid trade bills. This variable seems to be very discriminant. This information covers 3,5% of total population.

Default in payment to State creditors (so-called Privilèges URSSAF) which corresponds to default in payment of legal liabilities (taxes and other liabilities) due to Public Treasury and Social Security System. 12,09% is the default rate of 0.62% of total population that has deb to state creditors. 1,29% is the default rate of 7,75% of total population for which no debt toward State creditors was recorded.

3.3. Methodology

The underlying hypothesis is that higher late payment on trade credit will be associated with lower default rate. To test for the latter hypothesis, we process as follow. We classify borrowers firms into rating classes with respect to their default probability. The classification of firms into rating classes necessitates the finding of thresholds values separating the rating classes. We aim at solving two problems: to distinguish the default from non-defaults and to put the firms in an order based on their payment behavior. For using a model to obtain the probability of default of each firm receiver operating characteristics (ROC) analysis is employed to assess the distinction power of our model.
The logistic regression approach is used to identify short-run bankruptcy with the use of default indicator. This statistical technique has been considered for long as a powerful algorithm (Lee et al, 2006). Its specific form is at follow:

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

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 estimation (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 purpose 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: i) we first test for multicolinearity, ii) then we look for variable significance, iii) we verify if variables signs from the logistic regression are as expected, iv) and finally we undertake a ROC curve to validate the model performance.

4. Results and discussion

4.1 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 a firm level. Therefore, each observation corresponds to a firm of a given observation date, i.e. one firm might obtain different default status from different performance window. The total size of the sample is 3 807,598 observations. We observe the firms at 4 different date and we calculate the 6-month default rate (See table 1) following delinquencies collected. All tests and regressions are made on the pooled sample, i.e. the 4 samples taken for different observation date are treated together.

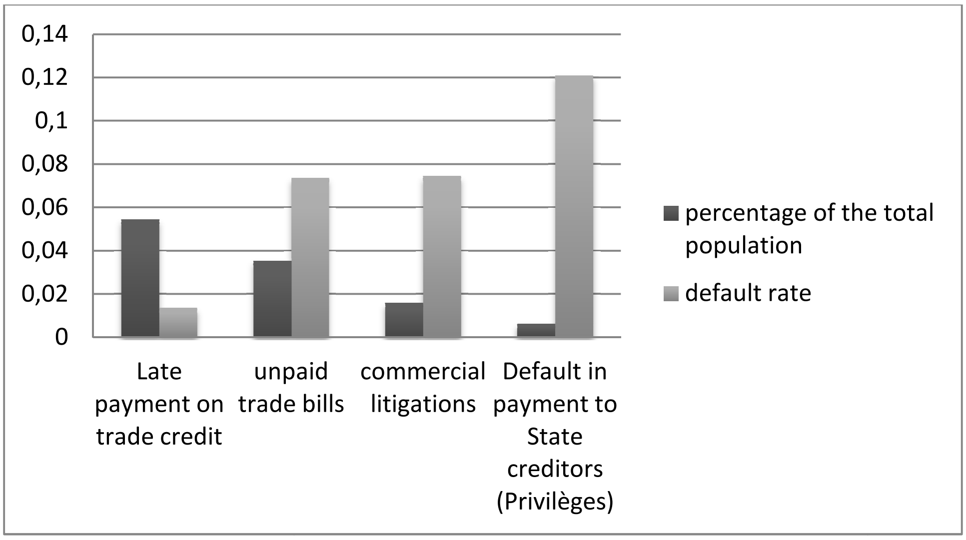

Figure 2: default rate and proportions of firms with incident of payment

Figure 2 presents the 6-months default rate of the pooled sample of firms. This a priori analysis is a suggestion that bankruptcy in a short horizon is affected all sort of delinquency of payment. It is worthy to note than the default rate of the total sample is equal to 0.9%. The comparison between default rates of firms with any pattern of delinquency with those with no delinquency provide 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 incident of payment. Despite the low proportion of firms with "privileges", the latter have the highest impact on default rate (12.09% as a 6 month default rate). Indeed firms can signal a critical financial situation when it starts to not pay government. Unpaid trade bills and commercial litigations have similar default rate. However, our database contains more information about unpaid trade bills than commercial litigations (see table 1).

Table 1

Default rate and incident of payments

delinquencies	With/Without delinquency	number of observation		percentage of the total population	default rate
Late payment on trade credit	with	number of firms	207495	5.45%	1.36%
		number of default firms	2824		
	without	number of firms	3600103	94.55%	0.86
		number of default firms	31086		
unpaid trade bills	with	number of firms	134284	3.53%	7.35%
		number of default firms	9873		
	without	number of firms	3673314	96.47%	0.65
		number of default firms	24037		
commercial litigations	with	number of firms	60177	1.58%	7.45%
		number of default firms	4481		
	without	number of firms	3747421	98.42%	0.79%
		number of default firms	29429		
Default in payment to State creditors (Privilèges)	with	number of firms	23745	0.62%	12.09%
		number of default firms	2870		
	without	number of firms	3783853	99.38%	0.82%
		number of default firms	31040		

NB: The default rate is calculated in an horizon of 6 months.

All sectors are represented in the sample of analysis with a slightly higher proportion of firms in construction, trade and manufacturing industries (respectively 16.06%, 25.43% and 9.4% from the total sample). Results in Table 2 confirm again what is usually observed in the French market: firms operating in industries like manufacturing or construction are more risky than other firms. Their 6-month default rates is respectively equal to 1.06% and 1.49%.

Table 2
Default rate distribution by firm's industry

Industry	Proportion	Default rate
Agriculture, forestry and fishing	1.08	0.42
Manufacturing	9.4	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	0.52
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	0.8
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

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

As we create many variables, the model is more likely to be complex and over-fitted. The more independent variables, the more probable the model had to carry mutually dependent and thus redundant predictors. Variance inflation factor (VIF) is a common way for detecting multicollinearity. Mathematically speaking: VIF = 1/(1-R-square). As advanced in the literature (Janke and Tinslay, 2005**),** if a VIF exceeds 10, the variable entry to the model become problematic. The definition of potential explanatory variables is listed in Appendix 1 with their VIF. Out of 56 variables tested in the model, 39 variables respond have less than 10 for their VIF. In addition to multicoolinearity issues, correlation may affect the results. The general rule usually used in the literature is to keep variables with a Pearson coefficient less than 0.7. As expected, variables of amount of delinquency (i.e. X1, X6 and X9) are highly correlated with variables of number of incidents (respectively X2, X7, X10). We then get rid of either the first or the second variable. The model has 22 potential variables (star variables in Appendix 1) of delinquencies to explain and predict firm's bankruptcy in a short horizon.

4.3 Univariate Analysis and segmentation

The 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 determining with characteristics are worth 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 to determine whether or not the variable is able to separate between bad firms (defaulted) from good firms(non-defaulted firms).
The 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 concerned data is unavailable). We decide to create different subpopulations and conduct different score. This alternative is recognized to provide us with better scores.

For our current study, we form 4 different segments in the following way:

1. Subpopulation with no delinquency features (no litigations, 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 lower default rate for the latter subpopulation comparing to 0.9% the default rate of the whole sample.
2. Subpopulation with no delinquency information's, neither with positive signals or 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/numbers of commercial litigations. In addition to that, late payment on trade credit seems to be statistically insignificant and have a negative impact on default probability. We report some fuzzy patterns of late payment on appendix 2. Table A in the Appendix 2 shows that the 6-months default rate is extremely high (6.3%) when no late payment on trade credit are recorded, whereas it is equal to 1.4% if ones make at least one arrears within the last 6 months. Table B and C from the Appendix 2, reports a non-monotonic evolution of a 6-months default rate when the amount of delays over the total amount of outstanding become more and more high. For instance, having more than 50% of the amount of outstanding in 180 delays is less riskier than having no arrears in terms of default rate within the last 6 months (Appendix 2 table B). Having no arrears within the last 6 months is much riskier than having about 25% of the outstanding in 90 delays (Appendix 2 table C). 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 population into 2 other subpopulations:

3. Firms with late payment on trade credit incidents combined with other payment delinquencies (i.e. commercial litigations or unpaid trade bills or so-called “Privilèges URSSAF”).
4. Firms with only late payment on trade credit payment incidents. (i.e. the firms do not appear on the data base of the other payment of incidents)

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

4.4 Scoring results

For reasons explained in the previous section, we obtain 4 different sub-populations. The Table 3 reports 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 incident of payment such as unpaid trade bills, or commercial litigations 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 to late payment data in our possession.

Table 3

Construction of 4 sub-populations

Sub-population	**Total number of firms**	**Proportion of the total sample**	**numbers of defaulted firms**	**default rate**
Firms with delinquencies, excluding late payment on trade credit (G)	197190	5.20%	13276	6.70%
Firms with only late payment on trade credit (R)	184192	4.90%	1125	0.60%
Firms with positive patterns of payment (P)	867785	22.80%	4429	0.50%
Firms with no information (N)	2558431	67.10%	15080	0.60%
Total	**3807598**	**100%**	**33910**	**0.90%**

For each sub-population, we follow the same process. We run univariate logistic regressions estimate default probability, to check the accuracy ratio and to provide idea about 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 neither good firms that pay in 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 between good and bad firms. For this group no logistic regression is computed. The average 6-month default rate is however equal to 0.6% and this corresponds to 67% of 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 presents 22% of 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 exercises no arrears during last 6 months (X34). The second one is constructed from an internal confidential variable in General Electric. We denote X** a specific variables used in-house by risk managers to identify firms that encounter severe incidents. X** is equal to 0 to specify a firm that honors all its engagements. If, in contrary, several alerts have been reported concerning payment behavior, delays or others incidents, X** take a positive number different to 0. We distinguish 4 different risk classes depending on the observed default rates. (see table 4).

We notice that default rate is driven downward when firms have positive outstanding but respect delays payment (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 on favor of the firms.

Table 4

default rate and risk classes for firms with a positive signal of payment behavior

		X34 = 1	X34 = 0	Total
X** = 0	total number of firms	216882	524327	741209
	number of defaulted firms	387	1720	2107
	default rate	0.2 (P1)	0.3 (P2)	0.30%
X** # 0	total number of firms	24924	101652	126576
	number of defaulted firms	272	2050	2327
	default rate	1.1 (P3)	2.0 (P4)	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%

NB : X34 is equal to 1 when no late payment on trade credit exists within the observation date. The alternative (X34=0) refers to firms with unknows amount of outstanding and no late payment

- **Default rate and risk categories for firms of group R:**

For the third subpopulation (4.6% of population) dealing with the sole late payment on trade credit information we conduct a univariate logistic regression in order to evaluate variables predictive power of default. Between all remaining explanatory variable related to late payment on trade credit practices, those that are rake taken one month before a given observation date (the most recent data available on arrears are used) are the most significant and respond the intuitive hypothesis that default rate are higher with high 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 2 months. The stepwise selection process decides to keep 2 variables. However, it must be pointed out that the latter univariate regression results just give us a pinch of idea of potential powerful variables. Further research must be conducted in this area. When conducting a stepwise model, we obtain several negative sign obliging us to reject almost all variables. The Gini of the model is equal to 45.2%.

Table 5
Logistic regression results for firms with only late payment on trade credit

Explanatory variables	**attributes**	**DF**	**coefficient**	**Standard Error**	**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-5.8743	0.0488	14516.0946***	<.0001
X24	under 25%		0			
X24	between 26% and 75%	1	0.4219	0.09	21.99***	<.0001
X24	more than 75%	1	0.7147	0.0733	95.051***	<.0001
X**	# 0		0			
X**	>0	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

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

It is worthy to notice 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 correspond to firms for which no additional delinquency information. This implies that late payment data that we have would not have a real predictive power as expected; this may be explained by the fact that SME firms are more prompt to exercise late payments and this seems to be a frequent practice. Late payment is frequent in French industry and this does not signal a severe financial distress. We suspect a relatively low predictive power for late payment. To confirm this result, we compare the extend of late payment discriminant power with the other incident of payment for the following sub-population.

Table 6

Default rate for group R of firms with only late payment on trade credit

explanatory variable	proposrtion of the total sub-popoulation	default rate	logistic coefficient
X**			
=0 (no severe incident of payment)	84.0%	0.3%	0.0000
>0 - présence d'incident(s)	16.0%	2.1%	1.8268
X24			
between 0% à 25%	78.5%	0.5%	0.0000
between 26% à 75%	10.0%	0.8%	0.4219
between 76% à 100%	11.6%	1.2%	0.7147
constante			-5.8743

To construct segments of risk categories we first classify score 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 scorecard building process. In final, we obtain 6 risk categories with an increasing default risk from segment R1 to segment R6. We finally plot the ROC curve corresponding to this sub-population (see Figure 4).

Table 7

Default rate by risk classes for Group R

Risk Categories	Number of non-defaulted firms	number of defaulted firms	Number of total firms	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 total population with an average default rate of 6.4%. Firms of this group have several past incidents of payment as commercial litigations, unpaid trade bills and “Privilèges URSSAF” combined with late payment on trade credit.
Table 8 reports the results of the logistic regression conducted on kept variables after a stepwise proceeding. The distribution of default rate by risk classes is displayed in Table 9.

Table 9
Default rate by risk classes for Group G

Risk Classes	Number of non-defaulted firms	number of defaulted firms	Number of total firms	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%
G8	2 209	850	3 059	27.8%
Total	**183 914**	**13 276**	**197 190**	**85.7%**

The Accuracy ratio is equal to 67.5% and the model work well with only 7 variables. Our model is well-fitted with the retained variables as default rates increase with high number/amount/recency of incident of payment.
We notice that other information about delinquencies (i.e. privilèges) have a better predictive power than late payment on trade credit. We succeed 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 lowest discriminant power to predict default at a short run.
We argue that the model is statistically robust and stable. 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 firm’ payment incidents to predict short term bankruptcy. Unpaid trade bills and “Privilèges URSSAF” are the most discriminant variables.

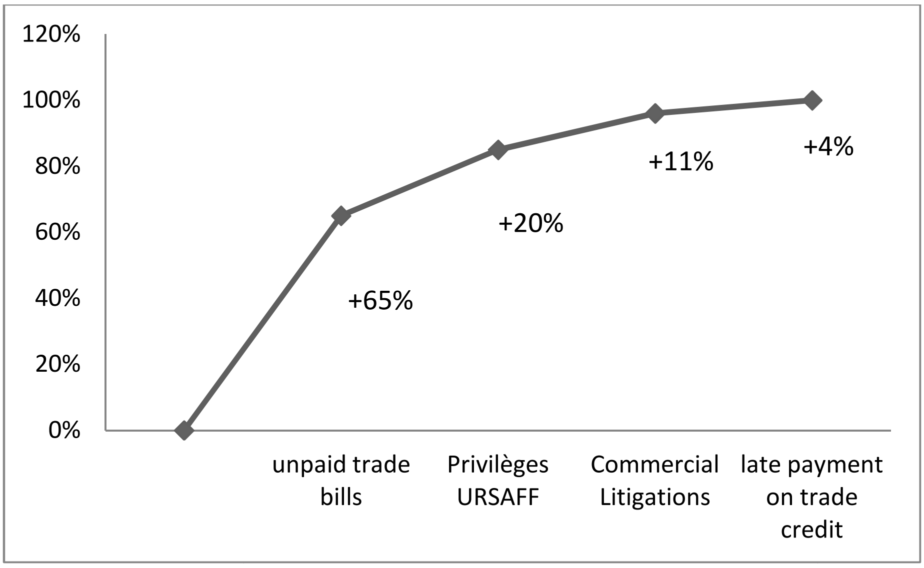

Figure 3: contribution of late payment on the model's overall performance for group G

The Roc curves of both subpopulations R and G show the impact of the inclusion of other late payment variables (see figure 4 below). Again, we confirm that variables of trade credit kept are not discriminant enough in comparison to other late payment variables. We remind that purpose of our paper, is to find alternatives to model SME bankruptcy in a short horizon (i.e. 6 months).

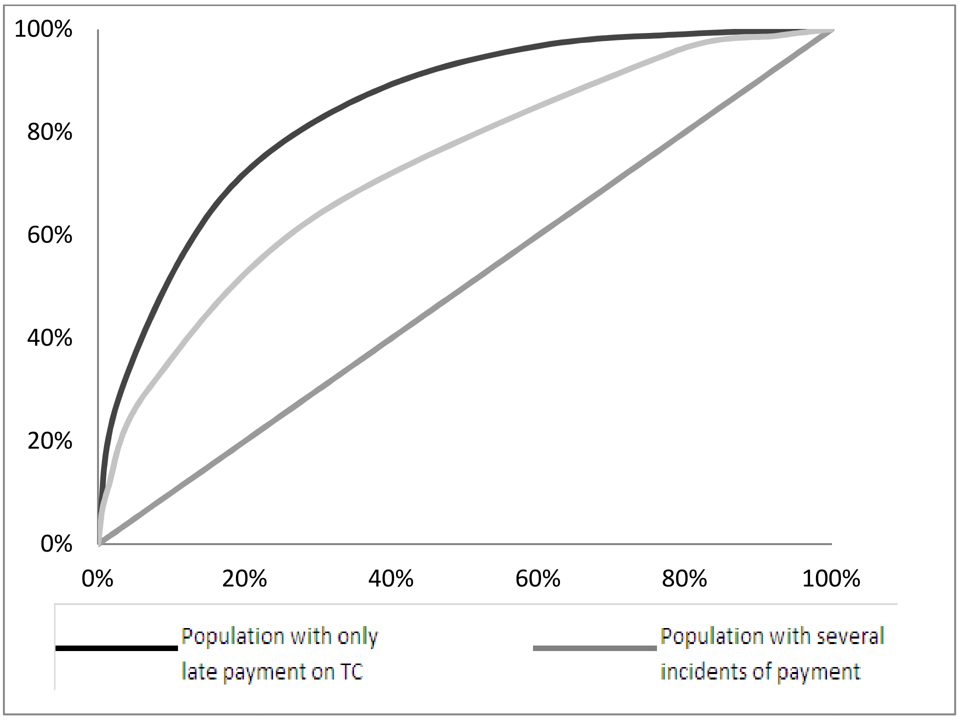

Figure 4 : ROC curves for subpopulations R and G

4.5. Validation of the final model

In this section we merge risk categories identified previously all together to identify the risk categories for the final model. We obtain 18 risk categories with increasing default rates from 0.2% to 27.8%. The corresponding accuracy rate is equal to 70.5% as show in table 10.

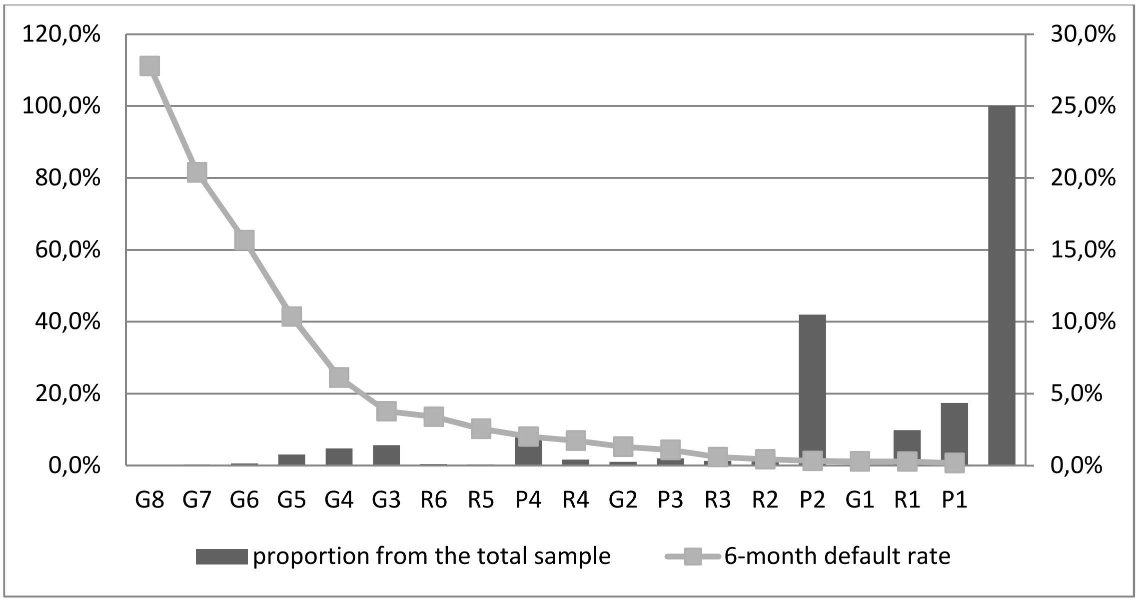

Figure 5: Default rate per risk catagory for the final model

Table 10
Accuracy ratio for the final model

classe "raw PD"	**6-month default rate**	**proportion frmo the total sample**	**cumulative percentage of defaulted firms**	**Accuracy Ratio**
G8	27.8%	0.2%	4.5%	0.0%
G7	20.4%	0.3%	8.5%	0.0%
G6	15.7%	0.6%	14.7%	0.1%
G5	10.4%	3.1%	36.0%	0.8%
G4	6.1%	4.8%	55.4%	2.2%
G3	3.8%	5.7%	69.6%	3.5%
R6	3.4%	0.4%	70.5%	0.3%
R5	2.6%	0.3%	71.0%	0.2%
P4	2.0%	8.1%	81.9%	6.2%
R4	1.7%	1.7%	83.8%	1.4%
G2	1.3%	1.0%	84.7%	0.8%
P3	1.1%	2.0%	86.1%	1.7%
R3	0.6%	1.3%	86.6%	1.1%
R2	0.4%	1.2%	87.0%	1.0%
P2	0.3%	42.0%	96.1%	38.4%
G1	0.3%	0.1%	96.1%	0.1%
R1	0.3%	9.9%	97.9%	9.6%
P1	0.2%	17.4%	100.0%	17.2%
Total	**1.5%**	**100.0%**	**100.0%**	**70.5%**

5. Conclusion

Financial institutions and banks have built many statistical models to measure the risk when lending to firms. However, no single type of model is suitable across all firms. Few attempts have been devoted to small commercial credit risk to deal with informational opacity that characterizes SME. Our paper is an attempt to identify a proper model for SME that could be used us substitute to accounting-based models. We use different explanatory variables derived from the only information about SME's payment behavior. First, we find that so far incident payment such us unpaid trade bills or unpaid debt to State are signals to a severe financial distress. Second, we show that the latter variables have a significant impact on firms bankruptcy in an horizon of 6 months. Finally, late payment on trade credit variables has less predictive power than other incidents of payments. These results are not conclusive as further research should be done to find other significant variables reflecting late credit patterns of SMEs.

Appendix 1
List of initial explanatory variables

variable	Description	VIF	variable	Description	VIF
X1*	cumulative amount of unpaid trade bills within the observation window	1.12	X15	the maximum amount of LP of more than 120 days within the observation window	35.56
X2	cumulative number of unpaid trade bills within the observation window	22.19	X16	the maximum amount of LP of more than 150 days within the observation window	1.14
X3*	date of last unpaid trade bills	3.23	X17	the maximum amount of LP of more than 180 days within the last 6 months	3.79
X4	average ratio of the cumulative amount of unpaid trade bill within the observation window by the turnover of the same year	1.04	X18	the average amount of arrears of more than 30 days within the observation window	2.52
X5	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	X19	the average amount of LP of more than 60 days within the observation window	1.66
X6*	cumulative amount of so-called 'Privilèges URSAFF' within the observation window	2.82	X20	the average amount of LP of more than 90 days within the observation window	2.54
X7*	cumulative number of 'Privilèges URSAFF' within the observation window	2.7	X21	the average amount of LP of more than 120 days within the observation window	9.94
X8*	date of last 'Privilèges URSAFF'	1.03	X22	the average amount of LP of more than 150 days within the observation window	1.03
X9*	cumulative amount of commercial litigations within the observation window	1.22	X23	the average amount of LP of more than 180 days within the observation window	2.30
X10	cumulative number of commercial litigations within the observation window	19.76	X24*	the percentage of LP of more than 30 days over the amount of outstanding taken1 month before a given observation date	7.89
X11*	date of last commercial litigations	1.42	X25*	the percentage of LP of more than 60 days over the amount of outstanding taken1 month before a given observation date	1.17
X12	the maximum amount of LP on trade credit of mroe than 30 days within the observation window	1.32	X26*	the percentage of LP of more than 90 days over the amount of outstanding taken1 month before a given observation date	3.23
X13	the maximum amount of arrears of more than 60 days within the observation window	9.28	X27*	the percentage of LP of more than 120 days over the amount of outstanding taken1 month before a given observation date	1.4
X14	the maximum amount of LP of more than 90 days within the observation window	1.22	X28*	the percentage of LP of more than 150 days over the amount of outstanding taken1 month before a given observation date	1.0

Appendix 1
List of initial explanatory variables (Cont)

variable	Description	VIF	variable	Description	VIF
X29*	the percentage of LP of more than 180 days over the amount of outstanding taken1 month before a given observation date	9.9	X42	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
X30	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	X43	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
X31*	the maximum amount of outstanding within the observation window	2.56	X44	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
X32	the minimum amount of outstanding within the observation window	16.61	X45	the percentage of outstanding over LP of more than 30 days 3 month before a given observation date	25.9
X33*	turnover in euros	4.02	X46*	the percentage of outstanding over LP of more than 60 days 3 month before a given observation date	9.0
X34*	Dummy variable that takes 1if the outstanding is positive but no LP recorded during the observation window	4.41	X47*	the percentage of outstanding over LP of more than 90 days 3 month before a given observation date	3.79
X35	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	X48	the percentage of outstanding over LP of more than 120 days 3 month before a given observation date	25.0
X36*	the number of incident of payment in the observation window	9.92	X49*	the percentage of outstanding over LP of more than 150 days 3 month before a given observation date	1.08
X37*	the date of the last incident of payment	3.25	X50*	the percentage of outstanding i LP of more than 180 days 3 month before a given observation date	7.40
X38	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	X51	the percentage of outstanding over LP of more than 30 days 6 month before a given observation date	14.87
X39	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	X52	the percentage of outstanding over LP of more than 60 days 6 month before a given observation date	1.81
X40	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	X51	the percentage of outstanding over LP of more than 30 days 6 month before a given observation date	14.87
X41	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	X52	the percentage of outstanding over LP of more than 60 days 6 month before a given observation date	1.81

Appendix 1

List of initial explanatory variables (Cont)

X53	the percentage of outstanding over LP of more than 90 days 6 month before a given observation date	1.09
X54	the percentage of outstanding over LP of more than 120 days 6 month before a given observation date	12.28
X55	the percentage of outstanding over LP of more than 150 days 6 month before a given observation date	20.17
X56	the percentage of outstanding over LP of more than 180 days 6 month before a given observation date	37.4

NB: LP stands for late payment on trade credit

* variables kept after controlling for correlation and multicollinearity.

Appendix 2

Table A

Default rate by firms with late payment on trade credit

Subpopulation		**Total number of firms**	**numbers of defaulted firms**	**default rate**
firms with delinquency	Having late payment on trade credit within the observation window	207495	2824	1.36%
	Having no late payment on trade credit within the observation window	185544	11687	6.29%
firms with no delinquency	no late payment on trade credit	3414559	19399	5.90%
Total		**3807598**	**33910**	**0.90%**

NB: Late payment on trade credit count for arrears of at least 30 days

Table B

a snapshot of the evolution of the default rate at the beginning of the observation window by the ratio of the total amount of LP on trade credit over the total amount of outstanding

	0	**<=5%**	**<=10%**	**<=25%**	**<=50%**	**<=75%**	**>75%**	**total**
LP of more than 30 days	1.20%	0.60%	0.70%	1.10%	1.60%	1.90%	2.40%	1.36%
LP of more than 90 days	1.30%	0.70%	1.30%	1.60%	2.20%	1.90%	2.80%	1.36%
LP of more than 180 days	1.30%	0.70%	1.50%	1.90%	1.70%	1.00%	2.90%	1.36%

NB: defaut rate are cvalculated for firms with at least one delinquency within the observation window (10.3% of total sample)

Table C

a snapshot of the evolution of the default rate at the end of the observation window by the ratio of the total amount of LP on trade credit over the total amount of outstanding

	0	**<=5%**	**<=10%**	**<=25%**	**<=50%**	**<=75%**	**>75%**	**total**
LP of more than 30 days	1.20%	0.60%	0.70%	1.00%	1.80%	2.20%	2.70%	1.36%
LP of more than 90 days	1.30%	0.70%	1.00%	1.00%	2.20%	2.70%	3.60%	1.36%
LP of more than 180 days	1.30%	0.70%	1.10%	1.40%	2.70%	2.90%	4.00%	1.36%

NB: defaut rate are calculated for firms with at least one delinquency within the observation window (10.3% of the total sample)

Table 8
Logistic regression results for firms with several incident of payment

Explanatory Variables	Attributes	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-5.3654	0.1963	746.843	<.0001
X1	**2 unpaid trade bills ou <=2000€**	1	0.4188	0.0276	230.1658	<.0001
	de 3 à 5 impayés	1	0.5879	0.0288	417.4179	<.0001
	<=15000€	1	0.6811	0.0343	393.4447	<.0001
	<=45000€	1	0.8423	0.0355	561.9234	<.0001
	<=99999€	1	1.0565	0.0396	712.0532	<.0001
X7	**1 "Privilège URSAFF"**	1	0.4169	0.0494	71.2711	<.0001
	2 Privilège URSAFF"	1	0.7459	0.0551	183.4736	<.0001
	between 3 à 9 "Privilège URSAFF"	1	0.9582	0.0367	683.2917	<.0001
	10 or more "Privilèges URSAFF"	1	1.0975	0.0383	822.0039	<.0001
X9	**<= 3000€**	1	0.4729	0.0504	87.9979	<.0001
	<= 5000€	1	0.551	0.0493	124.7833	<.0001
	<= 9000€	1	0.7469	0.0503	220.4526	<.0001
	<=18000€	1	1.0558	0.0513	423.6483	<.0001
	>18000€	1	1.1627	0.0548	450.9221	<.0001
X33	**<= 5 000 000€**	1	1.9164	0.195	96.6216	<.0001
	<=15 000 000€	1	1.5462	0.1962	62.0814	<.0001
	<=30 000 000€	1	1.2937	0.2204	34.4432	<.0001
X37	**in the two first months 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
X**	**=0**	1	0.8059	0.1768	294.456	<.0001
	>0	1	0.9569	0.087	435.867	<.0001
X24	**between 50% and 75%**	1	0.2826	0.1049	7.2566	0.0071
	more than 75%	1	0.3828	0.0557	47.1891	<.0001

Somers' D	0.35	Accuracy Ratio	0.675
R-Square	0.0266	Max-rescaled R-Square	0.0682

References