Centre de Recherche en Économie et Management
Center for Research in Economics and Management

University of Caen

University of Rennes 1

The Relevance of Soft Information for Predicting Small Business Credit Default: Evidence from a Social Bank

Simon Cornée
CREM, UMR CNRS 6211, University of Rennes 1,
and CERMI, France

June 2012 - WP 2012-26

Revision date : September 2015

Working Paper

The Relevance of Soft Information for Predicting Small Business Credit Default: Evidence from a Social Bank[1]

Simon CORNEE
(Université de Rennes 1, CREM UMR CNRS 6211, and CERMi)[2]

This version July 2015

(Accepted for publication in *Journal of Small Business Management*)

ABSTRACT

Using a unique, hand-collected database of 389 small loans granted by a French social bank dealing with genuinely small, informationally opaque businesses (mainly social enterprises), our study highlights the relevance of including soft information (especially on management quality) to improve credit default prediction. Comparing our findings with those of previous studies also reveals that the more opaque the borrower, the higher the predictive value of soft information in comparison with hard. Finally, a cost-benefit analysis shows that including soft information is economically valuable once collection costs have been accounted for, albeit to a moderate extent.

Keywords: Credit Default Prediction, Credit Rating, Relationship Lending, Social Banking.

JEL codes: G21, M21.

---

[1] The author gratefully thanks Yiorgos Alexopoulos, Frédéric Basso, Carlos Borzaga, Olivier Brandouy, Charles Calomiris, Joeffrey Drouard, Pascal François, Sébastien Galanti, Georges Gallais-Hammono, Christophe Godlewski, Silvio Goglio, Frans van Helden, Panu Kalmi, Marc Labie, Eric Lamarque, Olivier L'Haridon, Frédéric Lobez, Roy Mersland, Franck Moraux, Patrick Navatte, Adrian Pop, Christian Rauch, Catherine Refait-Alexandre, Damien Rousselière, Ariane Szafarz, Lawrence Wall, Laurent Weill, and Jacob Yaron, the participants at the 29th International Symposium on Money, Banking, and Finance (June 2012), the 4th EURICSE International Workshop on Cooperative Finance and sustainable development (June 2013), as well as an anonymous referee for helpful comments and discussions.

[2] Address: Faculté des Sciences Economiques, 7 Place Hoche CS86514, 35065 RENNES CEDEX, FRANCE. Email: simon.cornee@univ-rennes1.fr. Phone: +33 (0)2 23 23 38 87.

1. INTRODUCTION

Within the credit market, the severity of asymmetric information existing between insiders and outsiders renders the provision of external funding to small, informationally opaque businesses and atypical firms (for example, cooperatives) particularly challenging (Brewer, 2007). The problems of adverse selection and moral hazard may typically hamper the supply of external finance (Leland and Pyle, 1977; Myers, 1977; Stiglitz and Weiss, 1981). Financial intermediaries, in particular banks, possess a special ability to solve problems associated with asymmetric information in screening, pricing, and monitoring borrowers more easily (Diamond, 1984; Bhattacharya and Thakor, 1993). Banks benefit from an informational advantage over direct investors, especially in the case of small, informationally obscure borrowers (Binks et al., 1992; Binks and Ennew, 1996; Meyer, 1998). First, they can process large-scale hard information, which is defined as quantitative, explicit knowledge reported through formal instruments, such as audited financial statements, history of repayments, checking of accounts, and other financial usage (Petersen, 2004). Second, banks can amass soft, qualitative information via intense lending relationships with borrowers. Because loan officers are often embedded in their socioeconomic environment, they are able to collect intimate knowledge on the firm owner as well as private, idiosyncratic business facts, such as critical suppliers, customer dependencies, and positioning in the industry (Uzzi and Gillespie, 1999; Uzzi and Lancaster, 2003; Scott, 2006).

The past few decades have borne witness to the primacy conferred on hard information by banks in their screening and pricing operations. In the wake of the seminal works by Beaver (1966) and Altman (1968), credit risk models have been developed and refined exclusively on the basis of quantitative factors, such as financial ratios and accounting

or market-based measures (Altman and Saunders, 1997). Originally designed for large corporations, these models have been tailored to small and medium-sized enterprises (henceforth SMEs) (Ciampi and Gordini, 2013). For example, banks have increasingly resorted to small business credit scoring (henceforth SBCS) to evaluate applicants for small loans, typically under €250,000 (Akhavein et al., 2005; Berger and Frame, 2007). This evolution has been spurred on by the progress in ICT and industry concentration and encouraged by banking regulatory authorities (BCBS, 2000a, 2000b). In spite of the propagation of transactional lending technologies, relationship lending is still considered to be one of the most potent technologies in terms of overcoming information problems, particularly that of adverse selection (Berger and Udell, 2002). By gathering soft information on opaque credit applicants at the selection stage, banks are able to appraise borrower creditworthiness with more accuracy and subsequently make better credit decisions. However, there is surprisingly little academic research assessing the beneficial aspects of soft information in credit default prediction. The aim of the present paper is to bridge this gap by examining the extent to which soft information contributes to predicting small business default and whether it is economically valuable to do so.

Our empirical study is based on a unique, hand-collected dataset, which includes detailed information from 389 individual loans granted by a French social bank (Cornée and Szafarz, 2014). Social banks can be regarded as financial intermediaries specialized in the provision of external debt funding to a specific category of SMEs that prioritize social over financial goals, namely social enterprises (henceforth SEs) (Borzaga and Defourny, 2001). In addition to the fact that a burgeoning and innovative yet understudied financial sector is interesting to examine, the social bank under scrutiny offers a fertile ground for our investigation. It is a relational bank that serves genuinely small and opaque firms by making

intensive use of soft information — if that were not the case, our study would lose its rationale.

Our contribution to the existing literature is twofold. First, we replicate the study of Grunert, Norden, and Weber (henceforth GNW) (2005), applying their methodology to the novel context of social banking. We identify various credit default models to investigate the predictive value of soft information with respect to and as a complement to hard information. Second, we extend the work of GNW (2005) by sketching a cost-benefit analysis to quantify the economic impact of using soft information for a social bank.

The remainder of this paper is organized as follows: Section 2 provides an overview of the literature and states the testable hypotheses; Section 3 describes the data; Section 4 examines whether soft information improves credit default prediction and if so, to what extent; Section 5 proposes a cost-benefit analysis of soft information; Section 6 provides robustness checks; and Section 7 concludes.

2. SOFT INFORMATION, DEFAULT PREDICTION, AND PROFITABILITY

2.1. DOES SOFT INFORMATION IMPROVE DEFAULT PREDICTION?

There is a wide array of evidence attesting the importance of soft information in small business finance. A large number of studies document that banks amplify credit availability for borrowers with whom they have established an intense lending relationship, whether directly through facilitated debt provision or indirectly by reducing collateral requirements (for example, Petersen and Rajan, 1994; Angelini et al., 1998; Degryse and Van Cayseele, 2000; Lehmann and Neuberger, 2001). Recent studies also show that loan approval decisions are not entirely explainable by rule (that is, statistical methods based on hard information) but also by discretion, which for a banking institution means relying on the judgment of its loan officers derived from previously collected soft information (Cerqueiro et al., 2011; Puri et al.,

2011; Gropp et al., 2012). Moreover, in practice, banks' internal credit models are often characterized by a mixture of quantitative and qualitative factors (Elsas and Krahnen, 1998; Machauer and Weber, 1998; Brunner et al., 2000; Treacy and Carey, 2000). Taken altogether, this evidence strongly suggests that soft information plays a critical role in approving loans and determining credit conditions, especially when banks deal with opaque borrowers, however very few academic studies have attempted to assess its role. The most stringent test is provided by GNW (2005).[3] Drawing from a sample of large German SMEs, the authors compare the predictive accuracy of three models: a quantitative model solely fitted with financial data, a qualitative model based on soft-information factors, and a mixed model combining both types of factors.[4] The outcome is that the mixed model systematically outperforms the other two with virtually the same magnitude, thereby suggesting that hard and soft information exhibit an equally important predictive power.

Let us now examine the extent to which GNW's (2005) conclusions may be applicable to the context of social banking. Social banks present obvious similarities with other types of relational banks, such as credit cooperatives and community banks, in terms of organizational architecture and credit-granting procedures (see for example, De Young et al., 2004; Scott, 2004; Cornée et al., 2012; Kalmi, 2012). However, social banks distinguish themselves from their counterparts through their accomplishment of an explicit social mission, which translates into providing external debt financing to SEs (San-José et al., 2011; Becchetti et al., 2011; Cornée and Szafarz, 2014; Barigozzi and Tedeschi, 2015). SEs may be viewed as a specific type of SMEs, aimed at achieving a double-bottom line by putting the emphasis chiefly on

[3] In an unpublished paper, Lehmann (2003) reports similar results from a sample of 1,000 German SMEs. Altman et al. (2008) show that, based on a sample of 5.8 million UK-based SMEs, the introduction of non-financial factors adds substantial power to the overall accuracy performance. However, their non-financial factors (that is, legal actions by creditors, age of the firm, activity sector, late filing days, and the existence of audited accounts) cannot be regarded as *true* soft information, but rather as objective, publicly available data.

[4] More precisely, GNW's (2005) sample consists of 409 unlisted, conventional SMEs with variable turnovers of between €25 and €250 million, which borrow quite large sums (average: €1.5 million) from major German banks.

social goals (Borzaga and Defourny, 2001). The fact that SEs do not just pursue profit maximization but also the attainment of non-economic objectives renders their business more complex to assess compared with conventional SMEs (Chell, 2007). As a result, evaluation instruments that rely primarily on financial analysis may be insufficient for capturing critical aspects of SE survival, such as relational capital, acquisition of non-market resources, or social value-creation process (Gagliardi, 2009; Di Domenico et al., 2010). Two factors further magnify the informational opacity of SEs (Chess, 2007). On the one hand, SEs are often newly established and innovative businesses, and as such lenders cannot use statistical analysis on past cases to reduce the informational wedge. On the other hand, SEs operate on a very small scale by strongly embedding their activities in their local communities, thereby limiting transparent disclosure of business facts. Because social banks face exacerbated informational problems that are inherent in their borrowers' specificities, we anticipate that soft information is more valuable — in relation to hard information — in our setting than in GNW's (2005). The testable consequences of this conjecture are summarized in H1 and H2. Importantly, we consider these hypotheses (especially H1) to be stepping stones in our reasoning because we need to determine if they are supported before we can convincingly test our other hypotheses (H3 and H4).

H1: A mixed model combining soft and hard information dominates a hard-information model.

H2: A soft-information model is more accurate than a hard-information model. Accordingly, a mixed model is less likely to dominate a soft-information than a hard-information model.

Another instructive finding of GNW (2005) — although it is not explicitly highlighted by the authors — lies in the fact that there is heterogeneity across internal, soft-information factors regarding their ability to predict default, with "management quality" being more powerful than "market position". This result is in line with the small business literature, which

considers the managers' endowment in human capital to be a key internal factor explaining survival probability (see for example, Bosma et al., 2004; Carter and Van Auken, 2006; Unger et al., 2011). We anticipate that this should also be the case for SEs. The human capital of management affects most facets of SE survival. For instance, the continuation of non-market resources, such as volunteering or the preservation of the relational capital of a SE, is conditioned on the ability of its managers to promote a participatory governance structure (Borzaga and Defourny, 2001; Chell, 2007). In the same vein, business skills among SE managers are essential to adequately link operational issues with mission statements and to gain access to scarce financial resources (Sunley and Pinch, 2012). The prominence of management quality is likely to be amplified in our setting, given that SEs are often startup or young businesses and that human capital is more important than other factors in the initial years of business (Unger et al., 2011). This discussion is summarized in hypothesis H3:

H3: Soft-information factors related to management quality are likely to have a better predictive power than other soft-information factors.

2.2. Is the collection of soft information profitable?

Assuming that the inclusion of soft information in credit default models actually improves predictive accuracy, does it make economic sense for a (social) bank to produce and exploit this type of information? The answer to this question essentially depends on the bank's size or on its degree of decentralization. Soft information presents low transferability properties as it cannot be transferred among agents without losing informative power within the banking organization (Grant, 1996). For instance, the qualitative knowledge collected by loan officers at local branch level cannot be reliably communicated across hierarchical layers as it cannot be verified by anyone but themselves (Liberti and Mian, 2009). In other words, qualitative information is best portrayed by the subjective judgmental expertise of its collector (Cornée, 2014). These low transferability properties of soft information entail inevitable internal

contracting costs, and the magnitude of these costs is also likely to increase with the complexity and size of the banking organization (Stein, 2002).

Hence, large banking institutions have increasingly adopted transactional lending technologies to reduce internal contracting problems and the costs associated therewith. Because quantitative data are generally standardized and audited by a third party, interpretation of them is not contingent on the agents in charge of their collection. The advent of SBCS for opaque borrowers whose loan size is under $250,000 is a good example of the development of hard-information-based technologies. Labor costs reduction is undoubtedly the main incentive when banks switch to SBCS (Berger and Frame, 2007).

These conclusions may not apply to small (or decentralized) banks, which have a comparative advantage over large banks in assessing opaque borrowers. First, small banks with few managerial layers may significantly reduce the internal contracting costs associated with the collection of qualitative facts (Stein, 2002; Cole et al., 2004; Berger et al., 2005). Second, the costs of soft information collection for small banks — as compared with those incurred in large institutions that increasingly lend to distant borrowers — is likely to be less onerous because they are often grassroots, community-embedded institutions employing relationship lending technologies (Petersen and Rajan, 2002; Brevoort and Hannan, 2004; De Young et al., 2004).

To the best of our knowledge, with the exceptions of Stein and Jordão (2003) and Stein (2005), there exists little research examining whether using soft information in default models makes economic sense. These two studies report simulations accommodating some real-work complexities (for example, relationship lending), which suggest that more powerful models may be more profitable than weaker ones. More specifically, Stein and Jordão (2003) find that a mid-sized bank might generate additional profits to the order of about $4.8 million a year after adopting a moderately more powerful model.

In the context of social banking (at least in our setting), we conjecture that, overall, the recourse to soft information in credit default models should be beneficial. Internal contracting costs incurred by the use of soft information are likely to remain low as the bank exhibits a flat structure involving only two hierarchical layers. In addition, while loan supervisors located at the headquarters may override initial recommendations and have the final say on contracting elements, loan officers operating at grass-root level are vested with a substantial amount of authority and autonomy. In effect, the headquarters do not generate any new information on applications beyond the relevant information transcribed into a standard file by loan officers after in-depth interviews with credit applicants. Further solid proof of this is that the applications selected by loan officers are rarely rejected by loan supervisors. This discussion is summarized in hypothesis H4.

H4: The inclusion of soft information in a credit default model is likely to generate a benefit for the social bank under observation.

3. DATA AND PRELIMINARY ANALYSIS

The data used in this study come from the portfolio of a French social bank.[5] The bank under scrutiny is a financial cooperative, established in 1988. It operates throughout France under the supervision of the *Banque de France*, the French Central Bank. In 2008, it was composed of 21,467 members, and its total assets amounted to €184 million. The bank puts traditional financial intermediation rules into practice insofar as it bans all forms of speculative financial transactions (San-Jose et al., 2011).

Our sample consists of the complete credit files of 389 loans extended over the 2001–2004 period. Table 1, which captures the general features of the sample, contains two important messages. On the one hand, the data indicate that our sample is exclusively made

[5] The same dataset was used in Cornée and Szafarz (2014).

up of genuine small-sized enterprises. This is implicitly attested in the borrowing firms' legal forms, which are typical of small businesses. More explicit evidence is provided in their modest average turnovers and numbers of employees. This is confirmed by the low loan amounts underwritten by borrowers and the large fraction of loans that would be considered microloans by EU standards. On the other hand, the data also show that the unusually high proportion of start-ups — as compared with empirical studies using disaggregated loan-level data (for example, Petersen and Rajan, 1994; Lehmann and Neuberger, 2001) — mechanically pushes the age of the borrowing firms downward.

Table 1 Sample Description

Legal form of borrowing firms:	
% of proprietorships or other forms of individual unlimited companies	44%
% of private limited liability companies	37%
% of cooperatives or other legal structures of nonprofit organizations	19%
Characteristics of borrowing firms:	
% of start-ups	49%
% of existing firms having a banking relationship prior to loan extension	31%
Turnover amount: mean *[median]*	€546,000 *[€119,000]*
Number of employees: mean *[median]*	7.59 *[5.28]*
Age in years: mean *[median]*	5.28 *[1]*
Loan characteristics:	
Loan size: minimum; maximum	€5,000; €250,000
Loan size: mean *[median]*	€46,900 *[€30,520]*
% of loans < €25,000 (considered microloans by EU standards)	43%
Maturity in years: minimum; maximum	1; 20
Maturity in years: mean *[median]*	7.22 *[6]*

Due to data availability issues, average turnover and employee numbers are computed on 55 firms, average age is computed on 350 firms, and location is computed on 367 firms. All the other statistics are based on the whole sample (389 firms).

The bank's social mission is theoretically problematic for the validity of our study, since the social and financial criteria used in the bank's screening process could potentially interfere with one another. Cornée and Szafarz (2014) reject this possibility, however, by showing that for any given project, the assessment of its social dimension is not interlinked with its financial and economic analysis. However, the bank's social mission negatively impacts the field personnel's productivity and translates into additional screening costs. Cornée and Szafarz (2014) compute that the total costs incurred by social screening represent

31.84 percent of the bank's loan officers' workload. Finally, the bank's social mission may impact its portfolio composition. For example, its commitment to the right to credit may explain why the bank served such a large proportion of start-ups.

Our study period stretches from 2001 to 2008. As revealed in Table 2, we first consider a credit-granting period stretching from January 01, 2001 to November 25, 2004. Over this period, the bank extended a total of 476 loans. Hence our 389-loan sample represents 81.72 percent of the whole population.[6] We then look at a standardized four-year observation period for each loan to identify potential default events. For example, if a loan was extended on November 25, 2004, its corresponding observation period stretches from November 25, 2004 to November 25, 2008.[7]

Table 2 Sample Yearly Composition

Credit-granting period	Number of loans (in %)	% of representativeness	Number of defaulting loans (in %)
2001: 01/01–12/31	50 (12.85%)	57.47%	7 (14.00%)
2002: 01/01–12/31	84 (21.59%)	79.25%	24 (28.57%)
2003: 01/01–12/31	129 (33.16%)	90.21%	31 (24.03%)
2004: 01/01–11/25	126 (32.39%)	90.00%	29 (23.02%)
Total	389 (100.00%)	81.72%	91 (23.89%)

Table 2 indicates that 23.89 percent of borrowers experienced a default event within four years following loan extension. Admittedly, the proportion of loans in default appears quite high according to conventional retail banking standards. However, this figure should be kept in perspective, because the significant proportion of start-ups in our sample automatically elevates the default rate. First, the fact that start-ups are shown to be markedly

[6] The fact that the representativeness of the sample improves in the second half of the study period is due to advancements in the bank's information system. Besides, it is most likely that we can reject the possibility that the sample suffers from selection bias as the loans were unintentionally excluded.

[7] An analogous empirical strategy is employed by Cornée and Szafarz (2014). Given the conditions of access to the data, this four-year convention may be viewed as "optimal". It simultaneously maximizes: i) the duration of the observation window to record potential default events for each loan and ii) the number of loans with an equivalent observation period. This four-year observation window implies that default events occurring after this period cannot be detected (for example, a loan that experienced a default in the fifth year following its extension date is viewed as a non-defaulting loan). However, Cornée and Szafarz (2014) indicate that the vast majority of loans that experienced default at some point over their life are observable using this four-year convention (about 87 percent). In Section 6, we show that our results are robust when censorship is accounted for (as not all defaults are observable) and that they are not sensitive to a change in the observation window.

riskier than existing firms (32.80 percent of defaulting loans versus 14.50 percent respectively) is in line with the findings of previous studies (for example, Kalleberg and Leicht, 1991). On a similar note, SEs — the target borrowers of the bank — are also considered to be riskier than conventional SMEs (see for example, Chell, 2007). Second, the "default" denomination we use is extensive and encompasses various types of repayment issues that did not have the same degree of stringency. On the basis of out-of-sample figures from 2007, Cornée and Szafarz (2014) estimate that, on average, only 3.5 percent of the bank's loan portfolio results in liquidation. Third, we cannot exclude the possibility that the bank adopted more prudent behavior in its risk management — while financing riskier market segments — given its ethical orientation. For instance, it might have considered certain loans to be "in default" that would not necessarily have been regarded as such in a conventional bank. Conversely, the high proportion of defaulters in our sample is positive from a strict econometric vantage point. Such a percentage indicates that the sample is well-balanced in terms of default and non-default populations, thereby suggesting that our analysis should not be biased (King and Zeng, 2001).

Table 3 Definition of the Variables

VARIABLES	DEFINITION
Default variable	
DEF	= 1 if a default occurred within four years following loan extension; 0 otherwise.
Hard-information rating	
FIN	In-house financial rating: from 3 (best) to 1 (worst)
Soft-information ratings	
MGT	Assessment of "management quality": from 3 (best) to 1 (worst).
PROJECT	Assessment of "project quality": from 3 (best) to 1 (worst).
Type of borrowing firm	
STARTUP	= 1 if the loan was extended to a start-up firm; 0 otherwise.
Year dummies	
Y2001	= 1 if the loan was extended in 2001; 0 otherwise.
Y2002	= 1 if the loan was extended in 2002; 0 otherwise.
Y2003	= 1 if the loan was extended in 2003; 0 otherwise.
Y2004	= 1 if the loan was extended in 2004; 0 otherwise.

Table 3 describes the variables used in the study. The default variable, DEF, is a dummy variable that equals 1 if one or more of the following events occurred within four years after loan extension: moratorium, allowance of loan provisions, withdrawal of credit, disposition of collateral, or liquidation. This definition of default occurrence, suggested by the Basel Committee on Banking Supervision, was used by GNW (2005). The hard-information rating is represented by the FIN variable. This financial rating was established in-house by the bank after a thorough inspection of a borrower's financial condition. The analysis dealt with a variety of elements including past financial statements, expected future cash flow, collateral, and personal guarantees (Berger and Udell, 2006). The elements taken into consideration for the analysis could change from one borrower to the next according to their situation. Typically, a start-up would probably be evaluated on different criteria than those for an existing firm. It is likely that much more emphasis would have been put on future cash flows because there would have been no past financial statements.[8]

The soft-information ratings correspond to the subjective judgment of loan officers for credit applicants. The MGT variable provides an overall view of the management character (honesty, prudence, ethics) and capacity (experience, training, motivation). The PROJECT variable qualitatively assesses the global relevance of the investment project (Strengths, Weaknesses, Opportunities, Threats matrix). Importantly, FIN, MGT, and PROJECT are raw, non-transformed ratings, drawn directly from individual credit files. These ratings were given at the same time as credit-granting decisions were made and were not revised thereafter. In addition, the ratings were all originally assigned by loan officers on the basis of the information they had previously collected. However, their construction differs substantially in nature. As explained in Section 2.2, MGT and PROJECT could not be appropriately revised

[8] The bank produced an alternative financial rating, given on a one-to-five scale, on the basis of guidelines from the *Banque de France*. This alternative rating does not really suit start-ups' specificities, since it is mainly computed from borrowers' past financial statements. While the FIN variable was used by the bank in its credit risk management, this alternative financial rating aimed instead to comply with regulatory purposes.

by the loan supervisor at the headquarters, as soft information does not easily flow across hierarchical layers. In contrast, the loan supervisor was able to exert their full authority to homogenize FIN ratings across borrowers by using their expertise, economic indicators computed internally by the bank on priority sectors, and standard financial analysis techniques.

Finally, two sets of control variables are included. First, the type of borrowing firm differentiates the loans extended to start-ups (STARTUP) from those of existing businesses (defined as all firms but start-ups). Second, the year dummies aim to capture potential changes in the global economic conjuncture, the bank's lending strategy, and borrowers' creditworthiness. These year dummies also account for two issues revealed in Table 2, namely that excluded files are proportionately more frequent during the first years of observation and the distribution of default occurrences is not perfectly smooth.

Table 4 Descriptive Statistics

Panel A: Ratings by default status

	Full sample	**DEF=0**	**DEF=1**	***t***
FIN	1.97[a] (0.03)[b]	2.01 (0.03)	1.84 (0.04)	2.92**
MGT	2.76 (0.02)	2.81 (0.04)	2.58 (0.05)	4.48***
PROJECT	2.27 (0.02)	2.31(0.05)	2.14 (0.04)	2.91**

[a] reports means. [b] reports standard deviations. *** Equality between DEF=0 and DEF=1 rejected at $p < 0.10\%$; ** equality between DEF=0 and DEF=1 rejected at $p < 1\%$.

Panel B: Correlation matrix

	DEF	FIN	MGT	PROJECT
DEF	1.00			
FIN	−0.15**	1.00		
MGT	−0.22***	0.19**	1.00	
PROJECT	−0.15**	0.18**	0.15**	1.00
STARTUP	0.22***	−0.05	−0.09[†]	−0.16**

All are Spearman rank correlations. *** Zero correlation rejected at $p < 0.10\%$; ** zero correlation rejected at $p < 1\%$; * zero correlation rejected at $p < 5\%$; [†] zero correlation rejected at $p < 10\%$.

Panel A in Table 4 displays the rating categories averaged within the full sample as well as within the DEF=0 and DEF=1 subsamples. As expected, the means of the three rating categories are lower for defaulters than for non-defaulters. This is the first indication that a

strong link can be established between credit ratings and default status. Panel B of Table 4 confirms this robust association by revealing that the DEF variable exhibits significant negative correlations for all rating categories. The moderately positive correlation between the FIN, MGT, and PROJECT variables shows that these three rating variables only partially overlap, thereby revealing potentially complementary contributions in their ability to predict default occurrences. Finally, the (logical) negative correlations between STARTUP and the rating variables can be explained by the riskiness of start-ups as well as by other problems, such as informational asymmetry. Altogether, this shows the necessity of explicitly accounting for the STARTUP variable in the econometric analysis.

4. SOFT INFORMATION AND CREDIT DEFAULT PREDICTION

The purpose of a credit rating — regardless of its qualitative or quantitative nature — is to classify loan applicants according to their quality, that is to say their probability of default over a given time horizon (Krahnen and Weber, 2001). The default probability of a credit applicant is most appropriately estimated by means of binary logistic or probabilistic regression models (Greene, 1992).[9] Models that yield binary outcomes are supportive for the bank's decision-makers as they mirror the dichotomous decisions they are confronted with on a daily basis, namely whether they should grant credit or not. Like GNW (2005), we use probit regressions to estimate the probability of default.[10]

Table 5 displays four different specifications with DEF as a dependent variable. In columns (1) and (2), the HI and the SI models are fitted with hard and soft information respectively. The mixed model (HISI) that combines the two types of information is in

[9] Default events may also be viewed as survival events (Glennon and Nigro, 2005). This occurs especially when banks consider the time path of default as being as important as the event itself in a complex modeling approach, such as the net cash flow modeling framework. It is highly likely the bank under observation conferred much more importance on the default event than the time path of default given its simple credit management practices, thereby privileging models that yielded binary outcomes.

[10] While the coefficients in logit and probit models often differ markedly, the two models almost always yield identical predictions (Greene, 1992). In our case, logit estimations (not reported) bring similar results.

column (3). In column (4), the HISI2 model is an abridged version of the HISI model with the removal of the PROJECT variable. Overall, there is clear evidence to suggest a significant negative impact for all ratings with regard to default probability. This supports the idea that both hard *and* soft information are relevant when it comes to predicting default occurrences. Interestingly, the financial rating does not get crowded out by the other ratings and control variables, thereby indicating that the bank's rating system is robust enough to enable us to draw general conclusions from the results. Lastly, the highly significant and negative coefficient on STARTUP confirms that the explicit inclusion of this control variable in the analysis is relevant.

Table 5 Probability of Default: Probit Estimations

Panel A: Regression results

Models	(1) HI	(2) SI	(3) HISI	(4) HISI2
FIN	−0.47** (0.161)		−0.33* (0.167)	−0.36* (0.165)
MGT		−0.60*** (0.156)	−0.54** (0.160)	−0.56*** (0.158)
PROJECT		−0.28† (0.163)	−0.24 (0.166)	
STARTUP	0.61*** (0.148)	0.55*** (0.150)	0.56*** (0.152)	0.59*** (0.150)
Constant	−0.46 (0.382)	0.88 (0.576)	1.25* (0.612)	0.81 (0.526)
Year Dummies	Yes	Yes	Yes	Yes
Observations	389	389	389	389
Log Likelihood	−196.03	−191.05	−189.07	−190.11
McFadden's R^2	0.0723	0.0971	0.1065	0.1016
Likelihood-ratio $\chi 2$	30.61	41.11	45.07	42.99
ROC area	0.6804	0.7228	0.7330	0.7215
Brier Score	0.1666	0.1603	0.1593	0.1608
Naïve Brier Score	*0.1792*	*0.1792*	*0.1792*	*0.1792*

*** $p < 0.10\%$; ** $p < 1\%$; ** $p < 5\%$; † $p < 10\%$.

Panel B: Model comparisons

Criterion	Δ_{SI-HI}	$\Delta_{HISI-HI}$	$\Delta_{HISI-SI}$	$\Delta_{HISI-HISI2}$
Likelihood-ratio $\chi 2$	+[a] **[b]	+**	+*	+
ROC area	+†	+**	+	+†
Brier Score	+	+*	+*	+

[a] reports the sign of the difference between two competing models. [b] indicates whether the test reports significant difference. *** $p < 0.10\%$; ** $p < 1\%$; * $p < 5\%$; † $p < 10\%$.

We now turn our attention to comparing the performance of the HI, SI, HISI, and HISI2 models. As there is no single test to comprehensively evaluate a classification model,

two criteria are measured in addition to the Likelihood-ratio, which assesses the fit of the models, namely the area under the ROC curve and the Brier Score.[11] In Panel A of Table 5, we report comparative evaluations of the models' forecast quality. Panel B reports the results of the tests conducted between models for each of the evaluation criteria. We first examine how hard and soft information complement each other in predicting credit defaults. The Likelihood-ratio test shows that the HISI model has a better fit than the HI and SI models at the one and five percent levels respectively. A similar test rejects the null hypothesis that the HI and SI models have an equal goodness of fit at the one percent level. A formal test of equality of the ROC areas cannot reject the null hypothesis that the ROC areas are equal for the SI and HISI models. This test also indicates that both the SI and HISI models dominate the HI model at the ten and one percent levels respectively. Examination of the Brier Scores yielded by the three models partially corroborates the ROC inspection. In the first place, it is noteworthy that the three models are useful since they are always more accurate than naive forecasting (that is, 17.92 percent). The null hypothesis that the Brier scores are equal for the HI and HISI models is rejected at the five percent threshold. A similar conclusion can be drawn for the comparison between the SI and HISI models. However, the null hypothesis that the HI and SI models are equal cannot be rejected. Taken altogether, these findings strongly support hypothesis H1. The HISI model clearly exhibits a better fit and stronger predictive ability than the HI model. Hypothesis H2 is also supported but perhaps to a lesser extent. The SI model tends to dominate the HI model for two out of the three criteria. At the same time, the difference between the HISI and SI models is substantially weaker than that between the HISI and HI models, both in terms of economic magnitude and statistical significance.

[11] Previous studies have already used the ROC inspection and Brier Score to evaluate ratings systems and credit default models (GNW, 2005; Behr and Güttler, 2007; Krämer and Güttler, 2008). To examine the differences between areas under ROC and the Brier Scores, we perform the test provided by Cleves (2002) and a two-tailed Williams-Kloot test (Vinterbo and Ohno-Machado, 1999) respectively.

The non-significance of the PROJECT variable in column (3) of Panel A in Table 5 gives a first indication that the soft-information factors do not have the same predictive power. Panel B in Table 5 provides more formal proof in the results from the HISI and HISI2 models comparison. The two models prove to be almost identical since the significance evaluation test results are either weak or nonexistent, thereby suggesting that the collection of soft information on the managing team should be privileged by the bank. This provides strong support for hypothesis H3.

5. COST-BENEFIT ANALYSIS OF SOFT INFORMATION

We now outline a cost-benefit evaluation of using soft information. We focus our comparative cost-efficiency analysis on the HI and HISI models. In our view, this comparison offers more promising managerial and policy implications than any that could be carried out with the SI model. The implementation of pure soft-information models indeed seems "unrealistic" in the light of contemporary banks' lending practices, bank regulation, and the conventional academic view on credit risk evaluation.

5.1. GENERAL OVERVIEW

The net benefit of using soft information is the difference between the benefits provided by soft information and its costs. Its benefits stem from a more powerful prediction model, which allows the bank to better select its borrowers, thereby avoiding costs of classification errors. The costs of using soft information are the overhead expenses incurred by its collection, production, and analysis by loans officers. For a year *t*, ΔCCE_t is the year-*t* reduction in costs due to classification errors, which is permitted by the inclusion of soft information in the prediction model. $CCSI_t$ represents the costs inherent in soft information collection. Thus, the aggregated net benefit of soft information (*NBSI*) for the 2001–2004 period is computed as follows: $NBSI = \sum_{t=2001}^{2004}(\Delta CCE_t - CCSI_t)$.

5.2. BENEFITS OF SOFT INFORMATION COLLECTION

To compute ΔCCE_t, we need to know the potential profits and costs that the bank will generate by making a loan. The potential profits consist of interest revenues and underwriting fees. As indicated in Table 6, the applied interest rate is the spread between the actual interest charged by the bank and the three-month PIBOR, which is the bank's reference refinancing rate. The interest revenues are equal to the discounted cashed-in interests. To simplify calculations, we work out annual installments (constant annuities), even though not all borrowers repay according to such a scheme. Like Stein (2005), we assume that underwriting fees represent 0.50 percent of the amount lent by the bank. These fees are not discounted since they are paid upfront by borrowers when loans are disbursed.

The potential costs of default the bank can suffer by making a loan are twofold. The first component, which is the more substantial, is the net loss given default excluding workout fees. Like Cornée and Szafarz (2014), we approximate the value of this component by using the loan loss provisions, which reflect the bank's expectations of future losses (principal and interests) on defaulted loans. As suggested by Eales and Bosworth (1998), to obtain a net loss given default, loan loss provisions are netted from the underwriting fees and the interests collected by the bank during the pre-default period. The second component is made up of internal and external workout fees (for example, legal assistance). Like Stein (2005), we assume that these fees represent two percent of the original loan size. The negative cash flows generated by these two default cost components are then discounted using the bank's weighted average cost of capital according to the average period between loan extension and default occurrence (that is, 1.99 years).

Table 6 Soft Information Benefits: Variables and Assumptions

Variable	Value and definition	Source
Interest spread (per annum)	(Actual interest rate) – (3M-PIBOR)	Own calculation
Underwriting fees (upfront)	0.50% of original loan amount	Stein (2005)
Loss given default (excluding workout fees)	27.54% (average)	Cornée and Szafarz (2014)

Workout fees (on default)	2.00% of original loan amount	Stein (2005)
Pre-default period (in years)	1.99 (average)	Own calculation
Bank's weighted average cost of capital (discount rate)	6.00% (assumed baseline value)	Cornée and Szafarz (2014)

We now compare the profits and costs associated with the use of the HI and HISI models. There are four possible scenarios for both models, each of which entail costs or profits for the bank. The bank benefits from an opportunity gain by avoiding default costs when the model correctly predicts defaults (true positives). In the case of true negatives, the bank benefits from interest revenues and underwriting fees since the model correctly predicts non-defaults. Conversely, the bank suffers from costs when the model does not accurately predict defaults and non-defaults. Errors I (false positives) cause actual default costs to the bank since the model has inappropriately accepted "bad projects". Errors II (false negatives) incur an opportunity cost to the bank, which is equivalent to the interest revenues and underwriting fees of the "good projects" rejected by the model.

Like Paleologo et al. (2010) and others, we aim to minimize both errors I and II as they incur costs of virtually the same order of magnitude.[12] Consequently, we choose the optimal cut-off as the one that maximizes both sensitivity (true positive rates) and specificity (true negative rates). We compute the costs and profits generated by the HI and HISI models by using the optimal cut-off point of the HI model (0.32). As argued by GNW (2005), this procedure is conservative since our results would be even stronger if the optimal cut-off value were also used for the HISI model. Under the assumptions made, we find that $\sum_{t=2001}^{2004}(\Delta CCE_t)$ amounts to €352, 622.

5.3. Costs and net benefit of soft information collection

The benefits computed above should be compared with $\sum_{t=2001}^{2004}(CCSI_t)$, the burden associated by the collection of soft information. As described in Section 2.2, all information on credit applications is collected by loan officers. While we obtain a reasonable estimate of

[12] The discounted default costs and discounted opportunity costs average €8,952 and €6,303, respectively.

the overall operational costs associated with loan officers' activity (€871,156, see Cornée and Szafarz, 2014), two issues remain. First, it is difficult to determine accurately the proportion of overall operation costs (incurred by loan officers) devoted to information collection. We suggest varying this proportion from 25 to 75 percent (as depicted in Table 7) to avoid significant measurement errors.[13] Second, using the conservative assumption that all loan origination overheads are fully captured by information collection, we seek to establish which part of these overheads is dedicated to *soft* information collection.[14] Once again, to circumvent a potential measurement issue, we carry out a sensitivity analysis by varying the proportion of overheads incurred by soft information collection from 50 to 75 percent, as displayed in Table 7. We reject the possibility that this proportion is 25 percent or 100 percent. The first case is implausible because under a relationship lending technology, loan officers spend more time collecting soft information than hard. The second case, which implies that loan officers only collect soft information, is far-fetched given banks' current lending practices and regulatory reporting constraints.

Table 7 Net Benefits in Euros of Soft Information Collection

		% of total information collection overheads dedicated to soft information collection	
		75%	**50%**
% of loan officers' overheads devoted to information collection	**75%**	−137,403	25,938
	50%	25,938	134,833
	25%	189,280	243,727

Each cell of Table 7 presents the variable *NBSI*, in other words the outcome of the cost-benefit analysis for a given scenario. The scenarios, which can all be regarded as

[13] It is unlikely that this proportion could reach 100 percent. Information collection takes place at the stage of loan origination, and it is only one of the missions assigned to loan officers (Pollinger et al., 2007). Other missions typically involve marketing (prior to loan origination) and loan monitoring (after loan disbursement). Costs of information collection as a percentage of the amount lent — which varies from 1.20 to 3.59 percent depending on the scenario — are reconcilable with Pollinger et al. (2007).

[14] Loan origination also encompasses information collection and analysis as well as underwriting (Pollinger, 2007).

plausible, are characterized by a sizeable variability. *NBSI* varies from −€137,403 to €243,727. These figures represent −19.77 and 35.07 percent of the bank's net operating incomes aggregated over 2001–2004, and −0.75 and 1.34 percent of the amount lent by the bank over the same period.[15] Nonetheless, the dominant picture to emerge is that the use of soft information is likely to incur a profit for the bank. The message is most similar when the analysis is conducted *ceteris paribus* solely on the existing firms, with figures varying from −0.83 to 1.26 percent of the amount lent (over 2001–2004). Interestingly, the outcome is markedly more positive (oscillating from 1.08 to 2.65 percent of the amount lent) when the analysis is limited to firms in a credit relationship with the bank. Overall, this provides support for hypothesis H4.

6. ROBUSTNESS CHECKS

In this section, we report on four series of robustness checks, conducted to test the strength of the results displayed in Table 5 (see also Appendix A). First, we use a Cox proportional hazard model to check for two potential issues: i) the possibility that the bank envisages the time path of default as being as important as the event itself[16] and ii) censorship, because all defaults are not observable within the study period. The results presented in columns (1) to (4) of Table 8 and the corresponding model comparisons in Table 9 corroborate our previous findings. We use the Cox model's standard criteria to evaluate predictive accuracy, that is the Harell's C and the Akaike Information criterion (AIC) (whose value should be minimized). To compare models on the basis of the AIC, we compute an evidence ratio as suggested by Burnham et al. (2011).[17] According to this ratio, the SI and HISI models are, respectively,

[15] Additional computations (not reported here) indicate that the results of the cost-benefit analysis are insensitive to a variation in the discount rate and to the suppression of the underwriting or workout fees.
[16] For example, this might be the case when loans are guaranteed by public collateralization schemes that cannot be triggered immediately after loan extension in the case of a default.
[17] The key assumptions for the results to be valid are satisfied. First, censoring is minimal (about 13 percent of non-observed defaults) and non-informative, as the exclusion of non-observed defaults is not intentional.

97.03 and 139.77 times more likely than the HI model to be closer to the best model. These ratios are considered to be "strong" to "very strong" (Burnham et al., 2011). In sharp contrast, these ratios are extremely weak when the HISI model is tested against the SI and HISI2 models.

Second, we provide evidence that our findings are not driven by our sample composition, namely the high proportion of start-ups. In column (5) of Table 8, we enrich the HISI model of Table 5 with three interaction terms to check whether the effect of soft information is stronger than that of hard information for start-ups.[18] None of these interaction terms is significant. In addition, the model comparisons carried out in Table 9 on both the start-up and existing firm subsamples show that the tests are markedly more powerful for existing firms.

Table 8 Robustness Checks: Regression Results

	Cox regressions				Probit regressions	
	(1)	(2)	(3)	(4)	(5)	(6)
Models	HI	SI	HISI	HISI2	HISI	HISI
FIN	−0.61** (0.221)		−0.39† (0.231)	−0.43† (0.225)	−0.70** (0.243)	−0.44† (0.238)
MGT		−0.78*** (0.198)	−0.70** (0.205)	−0.73*** (0.203)	−0.03 (0.230)	−0.30 (0.247)
PROJECT		−0.41† (0.242)	−0.35 (0.244)		−0.48* (0.227)	−0.62* (0.261)
STARTUP	0.94*** (0.225)	0.85*** (0.227)	0.87*** (0.227)	0.90*** (0.225)	0.05 (1.149)	0.53** (0.155)
FINAN*STARTUP					0.33 (0.316)	
MGT*STARTUP					0.30 (0.261)	
PROJECT*STARTUP					−0.39 (0.286)	
EXPERIENCE						−0.20 (0.219)
FINAN*EXPE.						0.12 (0.083)
MGT*EXPE.						−0.05 (0.068)
PROJECT*EXPE.						0.01 (0.070)
CONSTANT					1.93* (0.804)	2.37** (0.856)
Year dummies	Yes	Yes	Yes	Yes	Yes	Yes
Observations	389	389	389	389	389	389
Log-likelihood	−515.85	−510.30	−508.91	−509.97	−187.53	−184.53
Harell's C	0.6668	0.6962	0.7066	0.6973	.	.

Second, the condition of proportional moral hazards is met as Schoenfeld's test is rejected in all specifications. The evidence ratio, notated as ER, is computed as follows: $ER = e^{(-\left(\frac{1}{2}\right)\Delta_{AIC})}$, where Δ_{AIC} is the difference in AIC between two models.

[18] The fact that the STARTUP coefficient loses its significance (as compared with Table 5) with the introduction of these interaction variables in specification (5) may reveal potential autocorrelation issues. Interaction variables remain non-significant when rerunning specification (5) without STARTUP.

AIC	1041.75	1032.60	1031.87	1031.85	.	.

*** $p < 0.10\%$; ** $p < 1\%$; * $p < 5\%$; † $p < 10\%$.

Third, we test the validity of the FIN, MGT, and PROJECT ratings. These ratings, especially those derived from soft information, may suffer from heterogeneous quality or manipulation as they are subjectively attributed by loan officers. We discard this possibility in specifications (6) of Table 8. None of the three interaction terms involving EXPERIENCE (that is, the number of years spent by an employee as a loan officer in the bank) is significant. In a similar vein, clustering the standard errors at the "loan officer" unit to account for potential differences across loan officers does not alter the results (regressions not reported here). Moreover, we replace the FIN variable with an alternative rating produced by the bank on the basis of the *Banque de France* guidelines (see footnote 6). As shown in Table 9, using FIN is a conservative approach since our results prove to be more robust with this alternative rating in lieu of FIN.

Table 9 Robustness Checks: Comparative Analysis

Robustness checks	Criterion	Δ_{SI-HI}	$\Delta_{HISI-HI}$	$\Delta_{HISI-SI}$	$\Delta_{HISI-HISI2}$
Survival analysis (n=389)	LR $\chi2$	+[a] **[b]	+***	+†	+
	Evidence ratio	97.03	139.77	1.44	1.05
Sample restricted to start-ups (n=189)	LR $\chi2$	+ **	+*	+	+†
	ROC area	+	+	–	+
	Brier Score	+	+	+	+
Sample restricted to existing firms (n=200)	LR $\chi2$	+*	+**	+*	–
	ROC area	+	+†	+†	–
	Brier Score	+	+*	+*	+
Replacing FIN by the *Banque de France*'s financial rating (n=389)	LR $\chi2$	+***	+***	+	+†
	ROC area	+**	+**	+	+†
	Brier Score	+**	+**	+	+
Two-year observation period to record defaults (n=389)	LR $\chi2$	+*	+*	+	+
	ROC area	+	+*	+	+
	Brier Score	+	+†	+	+
Six-year observation period to record defaults (n=134)	LR $\chi2$	+**	+**	+*	+†
	ROC area	+	+	+	+
	Brier Score	+	+*	+	+

[a] reports the sign of the difference between two competing models. [b] indicates whether the test reports significant difference. *** $p < 0.10\%$; ** $p < 1\%$; * $p < 5\%$; † $p < 10\%$.

Fourth, we show that our findings are not sensitive to any increase or reduction in the observation period for recording potential defaults. For example, Table 9 indicates that taking a period of two or six years (instead of four years) does not affect our previous results.

7. CONCLUSION

7.1. DISCUSSION OF THE MAIN FINDINGS

Our paper examines the relevance of using soft information to predict small business credit default. Our findings are fourfold. First, our paper corroborates GNW's (2005) result in a new context, namely that of social banking. In other words, we show unequivocally that including soft — in addition to hard — information increases the predictive accuracy of credit default models. Interestingly, however, this result diverges from those of GNW (2005) insofar as we show that soft information tends to be more valuable than hard. The fact that our sample includes small-sized, opaque SEs while theirs is composed of larger, more transparent SMEs may explain this discrepancy and suggests that the more informationally opaque the borrower, the higher the predictive value of soft information in comparison with hard, thereby validating our second finding. Our third finding relates to the fact that soft ratings are heterogeneous in terms of their predictive ability, with information on "management" being markedly more powerful than information on "project". Our fourth and main finding stems from quantifying the economic impact of soft information in credit default models. Our cost-benefit analysis allows us to disentangle two counteracting effects, which are that while soft information leads to cost reduction thanks to improved predictive accuracy, its collection generates a substantial increase in labor costs. In our case, the former effect dominates the latter, thus indicating that soft information is likely to be economically valuable.

Importantly, there are two factors that serve to increase confidence in the outcome of the cost-benefit analysis. First, the bank under scrutiny provides a stringent benchmark

because of the unusually high proportion of start-ups among its borrowers. As suggested by our calculations at the end of Section 5.3, the outcome would have been even more positive if the bank had dealt with proportionally more existing firms with which it had maintained long-term relationships. This is in line with previous research, which shows that soft information is more economical and conveys higher predictive value as a function of the intensity of the relationship (Petersen and Rajan, 1995; Presbitero and Zazarro, 2011). This factor is probably further strengthened by the specific type of borrowers funded by the bank and the strong informational opacity thereof. Second, our cost-benefit analysis delivers a lower bound of the outcome as it excludes any potential indirect benefits of soft information. For example, improved predictive accuracy may help the bank fulfill its claims regarding ethical, responsible credit-granting practices. Above and beyond the bank's perspective, increasing forecasting abilities may generate positive spillover effects for the whole economy as the bank channels savings into safer and more efficient projects.

Our findings can also be discussed in the light of conflicting evidence provided by the literature on the economic sense of including soft information in credit default models. Studies carried out before the outbreak of the 2007 financial crisis suggest that getting rid of soft information was economically well founded. For example, De Young et al. (2008) show that lenders switching from a relational to a transactional lending technology, such as SBCS, suffer from classification costs inherent in a loss of predictive accuracy in their default model. Nonetheless, these lenders are ready to accommodate these classification costs in exchange for the ancillary benefits associated with reductions in labor costs and the increase in securitized loan volume (scale economies, fee generation). Moreover, the profitability of this approach was found to motivate banks to amplify credit availability to opaque, risky borrowers and low-income communities — albeit raising interest rates and increasing credit risk (Berger et al., 2002; Berger and Frame, 2007). Yet, the sub-prime mortgage crisis

dramatically highlighted the limits of not considering soft information. Rajan et al. (2015) document that the statistical default models that were used in the active securitization period prior to the crisis break down in a systematic manner because they severely underestimate default risk for opaque borrowers, in relation to whom soft information is more precious (for example, borrowers with low documentation and high loan-to-value ratios). Furthermore, the failure of default models is deemed to be one of the key factors explaining the recent credit crisis (Diamond and Rajan, 2009). This directly echoes our own finding that the undesirable consequences of ignoring soft information are likely to increase as a function of borrower informational opacity.

7.2. LIMITATIONS AND RESEARCH PERSPECTIVES

We must acknowledge that our empirical analysis has been constrained by the nature of our data, which may in turn diminish the strength of some of our results. The main limitation lies in the fact that we did not gain access to the loan applications turned down by the bank. Access to such data would have allowed us to account for potential selection bias through the specification of a Heckman-like two-stage correction model. Regarding the construction of the ratings, finer-grained data would have been welcomed as their current measurement may be regarded as rather dull instruments, especially in relation to start-ups. On a similar note, the deficient data in the cost-benefit analysis forced us to rely on various scenarios to circumvent significant measurement issues, thereby leading to significant variability in the outcome.

Further investigation is also needed to generalize our results, because drawing global conclusions from a specific institution is always hazardous. Other potential generalization issues may typically stem from the particular type of relational banks represented by the social bank in our study as well as from the national context. While the French context is likely to be representative of the European Union context and, to a lesser extent, other developed countries, we do not think this is the case for developing countries. Therefore,

comparative analyses with different types of institutional settings and countries represent a promising research avenue.

Besides, relying on soft information automatically empowers loan officers because they are delegated more authority in the information production process, which may in turn induce detrimental effects. For example, certain population segments may be excluded from borrowing because credit officers exhibit unjustified preferences and/or stereotypes (Agier and Szafarz, 2013). These agency problems incur multifaceted costs, which have not yet been accurately measured.

Finally, the dynamic aspect of relationship lending, which is critical for lenders relying on soft information, should be examined more accurately. Special attention should also be paid to the borrower's perspective. In fact, not all small businesses are prone to engaging in long-term relationships because of putative hold-up effects (Boot, 2000). Thus, understanding the conditions under which both contracting parties can fruitfully cooperate is essential (Binks and Ennew, 1997).

7.3. MANAGERIAL AND POLICY IMPLICATIONS

Our analysis entails managerial implications for the social bank under scrutiny. Our findings suggest that the bank should balance their risk assessment instruments toward the inclusion of soft information. Our results also offer various ways of further increasing the economic advantage of using soft information. First, loan officers should focus on obtaining precise knowledge about a firm's management quality and expend less effort analyzing the firm's project characteristics. Second, the bank should rebalance its loan portfolio toward a higher proportion of existing firms with which it maintains long-term relationships. Third, given its not-for-profit status and ethical orientation, the bank may seek to rely on volunteer work to review modest loans for which the investment in soft information is both more critical and more difficult to recoup.

The present paper has potential policy implications. The current regulatory framework is not adequately tailored to relational banks, and even less so to social banks, thereby negatively affecting their specificities and hindering their activities (Milano, 2005). In showing the relevance of soft information for predicting credit default, our work modestly contributes to paving the way for credit risk systems that are both robust and representative of the diversity of lending technologies.

REFERENCES

Agier, I., and A. Szafarz (2013), "Subjectivity in Credit Allocation to Micro-Entrepreneurs: Evidence from Brazil," *Small Business Economics* 41(1), 263-275.

Akhavein, J., W.S. Frame, and L.J. White (2005), "The Diffusion of Financial Innovation: An Examination of the Adoption of Small Business Credit Scoring by Large Banking Organizations," *Journal of Business* 78(2), 577-596.

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

Altman, E.I., and A. Saunders (1997), "Credit Risk Measurement: Developments over the Last 20 Years," *Journal of Banking and Finance* 21(11-12), 1721-1742.

Altman, E.I., Sabado, G., and N. Wilson (2008), "The Value of Qualitative Information in SME Risk Management," *Journal of Financial Services Research* 40(2), 15-55.

Angelini, P., R. Di Salvo, and G. Ferri (1998), "Availability and Cost of Credit for Small Businesses: Customer Relationships and Credit Cooperatives," *Journal of Banking and Finance* 22(6-8), 925-954.

Barigozzi, F., and P. Tedeschi (2015), "Credit Markets with Ethical Banks and Motivated Borrowers," *Review of Finance* 19(3), 1281-1313.

BCBS (Basel Committee on Banking Supervision) (2000a), "Range of Practice in Banks' Internal Rating Systems," Discussion Paper (January), Bank for International Settlements.

BCBS (Basel Committee on Banking Supervision) (2000b), "Credit Ratings and Complementary Sources of Credit Quality Information," Working Paper No3 (August), Bank for International Settlements.

Beaver, W.H. (1966), "Financial Ratios as Predictors of Failure," *Journal of Accounting Research* 4, 7-11.

Becchetti, L., M. Garcia, and G. Trovato (2011), "Credit Rationing and Credit View: Empirical Evidence from Loan Data," *Journal of Money, Credit and Banking* 43(6), 1217-1245.

Behr, P., and A. Güttler (2007), "Credit Risk Assessment and Relationship Lending: An Empirical Analysis of German Small and Medium-Size Enterprises," *Journal of Small Business Management* 45(2), 194-213.

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, N.H. Miller, M.A. Petersen, R.G. Rajan, and J.C. Stein (2005), "Does Function Allow Organizational Form? Evidence from the Lending Practices of Large and Small Banks," *Journal of Financial Economics* 76(2), 237-269.

Berger, A.N., and G.F. Udell (2002), "Small Business Credit Availability and Relationship Lending: the Importance of Bank Organizational Structure," *Economic Journal* 112 (477), 34-53.

Berger, A.N., and G.F. Udell (2006), "A More Complete Conceptual Framework for SME Finance," *Journal of Banking and Finance* 30(11), 2945-2966.

Bhattacharya, S., and A.V. Thakor (1993), "Contemporary Banking Theory," *Journal of Financial Intermediation* 3(1), 2-50.

Binks, M.R., and C.T. Ennew (1996), "Growing Firms and the Credit Constraint," *Small Business Economics* 8(1), 17-25.

Binks, M.R., and C.T. Ennew (1997), "The Relationship between UK Banks and their Small Business Customers," *Small Business Economics*, 9(2), 167-178.

Binks, M.R., C.T. Ennew, and G.V. Reed (1992), "Information Asymmetries and the Provision of Finance to Small Firms," *International Small Business Journal* 11(1), 35-46.

Boot, A.W. (2000), "Relationship Banking: What Do We Know?" *Journal of Financial Intermediation*, 9(1), 7-25.

Borzaga, C. and J. Defourny (2001), *The Emergence of Social Enterprise*, London - New York: Routledge, 312-332.

Bosma, N., M. Van Praag, R. Thurik, and G. De Wit (2004), "The Value of Human and Social Capital Investments for the Business Performance of Startups," *Small Business Economics* 23(3), 227-236.

Brevoort, K. P., and T. H. Hannan (2004), "Commercial Lending and Distance: Evidence from Community Reinvestment Act Data," SSRN Working Paper.

Brewer, E. (2007), "On Lending to Small firms," *Journal of Small Business Management*, 45(1), 42-46.

Brunner, A., J. Krahnen, and M. Weber (2000), "Information Production in Credit Relationships: On the Role of Internal Ratings in Commercial Banking," Centre For Financial Studies Working Paper No 2000/10.

Burnham, K.P., D.R. Anderson, and K.P. Huyvaert (2011), "AIC Model Selection and Multimodel Inference in Behavioral Ecology: Some Background, Observations, and Comparisons," *Behavioral Ecology and Sociobiology*, 65(1), 23-35.

Carter, R., and H. Van Auken (2006), “Small Firm Bankruptcy,” *Journal of Small Business Management* 44(4), 493–512.

Cerqueiro, G., H. Degryse, and S. Ongena (2011), “Rules versus Discretion in Loan Rate Setting,” *Journal of Financial Intermediation* 20(4), 503-529.

Chell, E. (2007), “Social Enterprise and Entrepreneurship towards a Convergent Theory of the Entrepreneurial Process,” *International Small Business Journal* 25(1), 5-26.

Ciampi, F., and N. Gordini (2013), “Small Enterprise Default Prediction Modeling through Artificial Neural Networks: An Empirical Analysis of Italian Small Enterprises,” *Journal of Small Business Management* 51(1), 23-45.

Cleves, M. (2002), “From the Help Desk: Comparing Areas under Receiver Operating Characteristic Curves from Two or More Probit or Logit Models,” *Stata Journal* 3, 301-313.

Cole, R.A., L.G. Goldberg, and J.W. White (2004), “Cookie Cutter vs. Character: The Micro Structure of Small Business Lending by Large and Small Banks,” *Journal of Financial and Quantitative Analysis* 39(2), 227-250.

Cornée, S. (2014), “Soft Information and Default Prediction in Cooperative and Social Banks,” *Journal of Organizational and Entrepreneurial Diversity* 3(1), 89-109.

Cornée, S., D. Masclet, and G. Thenet (2012), “Credit Relationships: Evidence from Experiments with Real Bankers,” *Journal of Money, Credit and Banking* 44(5), 957-980.

Cornée S., and A. Szafarz (2014), “Vive La Différence: Social banks and Reciprocity in the Credit Market,” *Journal of Business Ethics* 125(3), 361-380.

Degryse, H., and P. Van Cayseele (2000), “Relationship Lending within a Bank-Based System: Evidence from European Small Business Data,” *Journal of financial Intermediation* 9(1), 90-109.

De Young, R., D. Glennon, D., and P. Nigro (2008), “Borrower–Lender Distance, Credit Scoring, and Loan Performance: Evidence from Informational-Opaque Small Business Borrowers,” *Journal of Financial Intermediation*, 17(1), 113-143.

De Young, R., W.C. Hunter, and G.F. Udell (2004), “The Past, Present and Probable Future of Community Banks,” *Journal of Financial Service Research* 25 (2-3), 85-133.

Diamond, D.W. (1984), “Financial Intermediation and Delegated Monitoring,” *Review of Economic Studies* 51(3), 393-414.

Diamond, D.W., and R. Rajan (2009), “The Credit Crisis: Conjectures about Causes and Remedies”, NBER Working Paper No 14739.

Di Domenico, M., H. Haugh, and P. Tracey, (2010), “Social Bricolage: Theorizing Social Value Creation in Social Enterprises,” *Entrepreneurship theory and practice* 34(4), 681-703.

Eales, R., and E. Bosworth (1998), “Severity of Loss in the Event of Default in Small Business and Larger Consumer Loans,” *Journal of Lending and Credit Risk Management* 80, 58-65.

Elsas, R., and J.P. Krahnen (1998), “Is Relationship Lending Special? Evidence from Credit-File Data in Germany?” *Journal of Banking & Finance* 22(10-11), 1283-1316.

Gagliardi, F. (2009), “Financial Development and the Growth of Cooperative Firms,” *Small Business Economics* 32(4), 439-464.

Glennon, D., and P. Nigro (2005), “Measuring the Default Risk of Small Business Loans: A Survival Analysis Approach,” *Journal of Money, Credit and Banking* 37(5), 923-947.

Grant, R.M. (1996), “Toward a Knowledge-Based Theory of the Firm,” *Strategic Management Journal* 17, 109-122.

Greene, W.H. (1992), “A Statistical Model for Credit Scoring,” SSRN Working Paper.

Gropp, R., C. Gruendl, and A. Guettler (2012), “Does Discretion in Lending Increase Bank Risk? Borrower Self-Selection and Loan Officer Capture Effects,” European Banking Center Discussion Paper No 2012-010.

Grunert, J., L. Norden, and M. Weber (2005), “The Role of Non-Financial Factors in Internal Credit Ratings,” *Journal of Banking and Finance* 29(2), 509-531.

Kalleberg, A., and K. Leicht (1991), “Gender and Organizational Performance: Determinants of Small Business Survival and Success,” *Academy of Management Journal* 34(1), 136-161.

Kalmi, P. (2012), “Co-operative Banking,” in: Michell, J. and J. Toporowski (eds.), *Handbook of Critical Issues in Finance*, Elgar, 56-66.

King, G., and L. Zeng (2001), “Logistic Regression in Rare Events Data,” *Political Analysis* 9(2), 137-163.

Krahnen, J.P., and M. Weber (2001), “Generally Accepted Rating Principles: A Primer,” *Journal of Banking and Finance* 25(1), 3-23.

Krämer, W, and A. Güttler (2008), “On Comparing the Accuracy of Default Predictions in the Rating Industry,” *Empirical Economics* 34(2), 343-356.

Lehmann, E., and D. Neuberger (2001), “Do Lending Relationships Matter? Evidence from Bank Survey Data in Germany,” *Journal of Economic Behavior & Organization* 45(4), 339-359.

Lehmann, B. (2003), “Is it Worth the While? The Relevance of Qualitative Information in Credit Rating,” SSRN Working Paper.

Leland, H.E., and D.H. Pyle (1977), “Informational Asymmetries, Financial Structure, and Financial Intermediation,” *Journal of Finance* 32(2), 371-387.

Liberti, J.M., and A.R. Mian (2009), “Estimating the Effect of Hierarchies on Information Use,” *Review of Financial Studies* 22(10), 4057-4090.

Machauer A., and M. Weber (1998), “Bank Behavior Based on Internal Credit Ratings of Borrowers,” *Journal of Banking and Finance* 22 (10-11), 1355-1383.

Meyer, L.H. (1998), “The Present and Future Roles of Banks in Small Business Finance,” *Journal of Banking and Finance* 22, 1109-1116.

Milano, R. (2005), “Ethical Finance in Europe Following the Basel II Agreements,” *Finance & Common Good* 25(1), 15-17.

Myers, S. C. (1977), “Determinants of Corporate Borrowing. *Journal of Financial Economics* 5(2), 147-175.

Paleologo, G., A. Elisseeff, and G. Antonini (2010), “Subagging for Credit Scoring Models,” *European Journal of Operational Research* 201(2), 490-499.

Petersen, M.A. (2004), “Information: Hard and Soft,” Working Paper, NBER.

Petersen, M.A., and R.G. Rajan (1994), “The Benefits of Lending Relationships: Evidence from Small Business Data,” *Journal of Finance* 49(1), 3-37.

Petersen, M.A., and R.G. Rajan (1995), “The Effect of Credit Market Competition on Lending Relationships,” *Quarterly Journal of Economics* 110(2), 407-443.

Petersen, M.A., and R.G. Rajan (2002), “Does Distance Still Matter? The Information Revolution in Small Business Lending,” *Journal of Finance* 57(6), 2533-2570.

Pollinger, J. J., J. Outhwaite, and H. Cordero-Guzmán (2007), “The Question of Sustainability for Microfinance Institutions,” *Journal of Small Business Management* 45(1), 23-41.

Presbitero, A.F., and A. Zazzaro (2011), “Competition and Relationship Lending: Friends or Foes?” *Journal of Financial Intermediation*, 20(3), 387-413.

Puri, M., J. Rocholl, and S. Steffen (2011), “Rules Versus Discretion in Bank Lending Decisions,” Unpublished working paper.

Rajan, U., A. Seru, and V. Vig (2015), "The Failure of Models that Predict Failure: Distance, Incentives and Defaults," *Journal of Financial Economics* 115(2), 237-260.

San-Jose, L., J.L. Retolaza, and J. Gutierrez (2011), "Are Ethical Banks Different? A Comparative Analysis Using the Radical Affinity Index," *Journal of Business Ethics* 100(1), 151-173.

Scott, J.A. (2004), "Small Businesses and the Value of Community Banks, *Journal of Financial Service Research*," 25(2-3), 207-230.

Scott, J.A. (2006), "Loan Officer Turnover and Credit Availability for Small Firms," *Journal of Small Business Management* 44(4), 544-562.

Stein, J.C. (2002), "Information Production and Capital Allocation: Decentralized versus Hierarchical Firms," *Journal of Finance* 57(5), 1891-1921.

Stein, R.M. (2005), "The Relationship between Default Prediction and Lending Profits: Integrating ROC Analysis and Loan pricing," *Journal of Banking & Finance*, 29(5), 1213-1236.

Stein, R.M., and F. Jordão (2003), "What is a More Powerful Model Worth?" Moody's KMV Paper.

Stiglitz, J.E., and A. Weiss (1981), "Credit Rationing in Markets with Imperfect Information," *American Economic Review* 71(3), 393-410.

Sunley, P., and S. Pinch (2012), "Financing Social Enterprise: Social Bricolage or Evolutionary Entrepreneurialism?" *Social Enterprise Journal* 8(2), 108-122.

Treacy, W.F., and M. Carey (2000), "Credit Rating Systems at Large US Banks," *Journal of Banking and Finance*, 24(1-2), 167-201.

Unger, J.M., A. Rauch, M. Frese, and N. Rosenbusch (2011), "Human Capital and Entrepreneurial Success: A Meta-Analytical Review," *Journal of Business Venturing* 26(3), 341-358.

Uzzi, B., and J.J. Gillespie (1999), "What Small firms Get Capital and at What Cost: Notes on the Role of Social Capital and Banking Networks," *Federal Reserve Bank of Chicago Proceedings* (No. 778).

Uzzi, B., and R. Lancaster (2003), "Relational Embeddedness and Learning: The case of Bank Managers and their Clients," *Management Science*, 49(4), 383-399.

Vinterbo, S., and L. Ohno-Machado (1999), "A Recalibration Method for Predictive Models with Dichotomous Outcomes," in *Predictive Models in Medicine: Some Methods for*

Construction and Adaptation, PhD Dissertation, Norwegian University of Science and Technology.

APPENDIX A: ADDITIONAL ROBUSTNESS CHECK

Previous research on relationship lending by banks to SMEs suggests that firms with a relationship may benefit from better conditions. In our case, this may in turn have reduced the default probability of firms with a relationship, thereby "contaminating" the sample and distorting our results. We consider this possibility in Table A1. The dependent variables are the three classical contractual variables: the nominal rate charged to borrowers in 100 basis points (RATE), the loan size in €10,000 (AMOUNT), and the fraction of the loan uncollateralized in % (NONCOLLATERAL). The independent variables are those comprised in the HISI model plus a dummy variable that equals 1 if the firm was in a relationship (RELATIONSHIP) and the refinancing rate (PIBOR3M). We specify a reduced-form estimation (that is, SURE model) to simultaneously assess the impact of loan characteristics while avoiding endogeneity biases.

Table A1 Robustness Checks: Multivariate Conditions for the Credit Conditions

VARIABLES	(1) RATE	(2) AMOUNT	(3) NONCOLLATERAL
FIN	−0.14** (0.047)	−0.00 (0.496)	−0.02 (0.017)
MGT	−0.05 (0.051)	0.01 (0.535)	0.02 (0.018)
PROJECT	−0.10** (0.047)	0.23 (0.494)	0.03† (0.017)
PIBOR3M	0.44*** (0.091)	−0.59 (0.964)	−0.01 (0.032)
STARTUP	0.05 (0.047)	−2.40** (0.497)	0.03† (0.017)
RELATIONSHIP	−0.01 (0.064)	−0.72 (0.676)	0.03 (0.023)
Constant	Yes	Yes	Yes
Year dummies	Yes	Yes	Yes
Observations	389	389	389
R^2	0.60	0.08	0.07

*** $p < 0.10\%$; ** $p < 1\%$; * $p < 5\%$; $^{\dagger}p < 10\%$.

In sum, the results show the variable RELATIONSHIP is non-significant throughout Table A1, suggesting that our sample is not "contaminated" and confirming our findings.