CRISEI

Centro di Ricerca Interdipartimentale in Sviluppo Economico e Istituzioni

Discussion Paper Series

The how and why of soft information production in bank lending

Salvatore Capasso, Stefano Monferrà, Gabriele Sampagnaro

CRISEI - Università di Napoli - Parthenope

Università degli Studi di Napoli- Parthenope

CRISEI

The how and why of soft information production in bank lending

Salvatore Capasso*, Stefano Monferrà**, Gabriele Sampagnaro**

*Dipartimento di Studi Aziendali ed Economici
** Dipartimento di Studi Aziendali ed Quantitativi

Università degli Studi di Napoli, Parthenope

The how and why of soft information production in bank lending

by Salvatore Capasso, Stefano Monferrà, Gabriele Sampagnaro

University of Naples Parthenope

Abstract

Although soft information plays a central role in the relationship banking literature, the mechanism behind its production is unexplored and remains unclear. Assuming that this type of information is collected primarily through instances of face-to-face contact, this paper analyzes the determinants of lender-borrower interactions. The findings suggest that a bank may prefer to gather qualitative information mainly regarding non-opaque firms with low credit risk and low coverage of financial products provided by the same bank. The results are consistent with this study's proposed classification of active and passive soft information collection and might help suggest new insights regarding the intensity and quality of bank-firm relationships.

A strand of the literature on small business lending has emphasized the role played by soft information and the soft/hard information dichotomy. In particular, this literature ties the collection and use of soft information to relationship lending practices as opposed to transaction-based banking, which primarily employs hard information.

Typically, "soft" information is not directly verifiable by a third party and includes hypothetical and/or intangible information such as economic projections, assessments of a firm's management quality, and employee morale. Soft information is typically thought to be accessible exclusively to a primary incumbent lender and cannot be unambiguously documented in a report that a loan officer can pass on to his superiors. In fact, the use of private and non-verifiable information across a bank's organizational layers can be difficult; whereas "hard" information accrues importance as it ascends the hierarchical ladder, "soft" information loses importance in the same process (Liberti and Mian, 2009).

Previous research in relationship banking generally suggests that soft information can improve contracting efficiency and affect a firm's costs and/or its access to credit (see, among others, Diamond, 1991; Boot and Thakor, 1994; Petersen and Rajan, 1994; Berger and Udell, 1995, Blackweel and Winters, 1997; Boot, 2000 Degryse and Cayseele, 2000; Elsas, 2005; Puri et al., 2011; Gambini and Zazzaro, 2013; Bartoli et al., 2013; Gobbi and Sette, 2014; Becks et al., 2015),

collateral requirements (Chakraborty and Hu, 2006; Brick and Palia, 2007; Bharath et al., 2011; Behr et al., 2011), and other important organizational features, such as the portability of soft information across banksøorganizational layers and delegation of the loan approval process (Liberti and Mian, 2009; Agarwal and Hauswald, 2010, Nemoto, et al. 2013). In addition, the role of soft information is considered by the copious literature on bank specialization in SME lending (Stein, 2002; Berger et al., 2005, Berger and Udell, 2006; Berger et al., 2007 Delgado et al. 2007, La Torre et al., 2012).

Although past studies provide comprehensive knowledge regarding the role played by soft information in lending practices (particularly for small business lending), the reasons why a bank gathers this type of qualitative information is under-explored (primarily due to the absence of data). More precisely, past studies consider soft information as a priori goods or as an effect of close lender-borrower relationships, although the issue of the source of soft information may appear more complex in terms of both modality and the incentives for gathering this type of information.

This paper attempts to shed light on this issue and to answer the following question: õHow and why do banks collect soft information?ö The question is important because collecting information may be costly and some information can be gathered intentionally, whereas other information cannot.

To fill the data gap discussed above, this study assumes that inside information is transmitted by means of repeated contact between a bank and its customers. Moreover, this paper attempts to identify the determinants of bank-firm interactions by examining a rich dataset containing information about historical contact between bank and borrower.

To that end, a classification system is proposed for types of possible bank-firm contact. Thus, the paper distinguishes between active and passive types of contact, in which the former are defined as contact promoted intentionally by the bank, whereas the latter contact is unintentional (from the perspective of the bank). More precisely, we propose separating the active from the passive collection of soft information. Active collection refers to obtaining data directly from entrepreneurs through meetings promoted by the bank, including direct face-to-face encounters at the firmøs offices (to allow the bank to inspect and appreciate (or not) the firmøs organization). Passive collection involves the flow of information that a bank receives unintentionally as a consequence of its regular and normal client monitoring activities. Examples of this type of information might include the regularity of a borrower/clientøs payments, the reliability of balance sheet data, and a comparison with banksøother clients in the same industry.

Hence, establishing this distinction may be useful for obtaining an answer as to why banks collect soft information. If we assume that a bank collects soft information only passively, we might

reasonably conclude that the question is not relevant or appropriate. In fact, in this case, banks may have collected soft information randomly and may employ it based on their particular strategy and specialization. If soft information is collected actively (intentionally) and is costly, it seems puzzling that banks would want to collect it.

Based on this preliminary discussion, the purpose of this paper is to explore ó for the first time ó the determinants of the production of soft information empirically by exploiting a unique and comprehensive database that contains bank-firm level information on a large sample of clients of one of the top five largest commercial banks in Italy. The richness of these data makes it possible to overcome obstacles in empirically analyzing these issues. First, our data allow us to measure the amount of soft information produced because we collected the number of times that meetings was made between the bank and the firm. This information allowed us to construct a measure of a bankøs overall knowledge of its borrowing clients. Regressing our measure of bank information production on a set of bank-firm variables, we find evidence of factors that matter in producing soft information. Our findings show that a long-standing relationship and the number of lenders play an important role in the production of soft information. However, the most important result is that as the production of soft information increases, the risk of the borrower decreases. Moreover, as the number of products sold in the past by the bank to the firm increases, so does the number of bank visits to firm headquarters. These results support the notion that banks should actively collect information (and incur costs) regarding high-quality firms in their portfolio in an attempt to extract more profits in the future by reducing borrower risk.

A novel feature of our analysis is that it allows us to obtain direct measures of soft information and identify a bankøs potential preferences regarding its production by proposing a pioneering classification of soft information sources. Hence, the paper contributes to the previous literature on relationship lending by providing empirical evidence on an under-explored topic. To the best of our knowledge, Uchida et al. (2012) is the only study to address the issue of soft information production and report interesting and robust findings. However, our study differs from Uchida et al. (2012) in many respects. First, the two studies analyze different topics. Whereas we are interested in identifying the possible reasons why a bank collects soft information and thus analyze a set of bank/firm-specific variables that affect the production of soft information, Uchida et al. (2012) focus their attention on the role of loan officers in producing soft information. Second, whereas we use a large sample of a direct measure of bank-firm interaction (distinguishing among different modes of contact), Uchida et al. (2012) use data provided by a questionnaire that may be influenced by borrowersø perceptions of the bank. Furthermore, and more importantly, we employ data that

contain information about the portfolio of bank products purchased by each firm, which may help us understand the role of banksø sales strategy in relationship lending.

The remainder of the paper is structured as follows: Section II describes the estimation design and hypothesis, section III discusses the empirical strategy, and section IV discusses the results. Finally, section V reports conclusions.

II. Estimation Design and Hypothesis

A. Active vs. passive production of soft information

As discussed above, the literature on relationship lending uses the term õsoftö to denote information that cannot be objectively verified by a third party and that is difficult to completely summarize in a numeric score (Petersen, 2004), such as economic projections, the statement of managementøs future plans, assessments of management quality, the product market position, and entrepreneurial honesty. However, this list is not exhaustive due to the evanescent character of soft information, which makes it impossible for the finance literature to explicitly state the distinction between hard and soft information (Petersen, 2004). However, hard information is almost always coded as numbers and includes items such as financial statements, payment histories, credit register data, stock returns, and quantity output numbers. Although the lack of a strict definition of soft information makes it impossible for econometricians to codify and observe it, we assume that soft information originates mainly from the multitude of instances of personal contact that loan officers have with borrowers, which is the basic assumption of this study and can be formally developed as follows:

<<Soft information is gathered primarily by means of contact between the bank (branch) and the borrower>>.

Accordingly, with this basic assumption, the history of bank-firm contact can paint a picture of the level of a bankøs knowledge accumulation and allow us to formulate a first attempt at answering our main research question: *How and why do banks collect soft information?*

Regarding the õhow to collectö segment of that question, although our assumption provides a preliminary answer, we must note that a bank may have many types of contact with its clients and that only certain of these types of contact play a role in generating and/or improving high-value relationship information. In attempting to distinguish those types of contact that yield soft information from others, we develop a three-step classification.

First, we separate close (face-to-face) contact from remote contact (by phone, email, etc.) and consider the former to be the type most valuable for collecting inside information.

Second, we distinguish close contact that is intentional from contact that is unintentional. The bank solicits the first type of contact, whereas the firm initiates contact with the bank in the second type. These two types of interactions are classified as active contact and passive contact, respectively.

Both classes of contact may then be submitted to a third classification if we consider the location of the contact. Because close contact (both active and passive) coincides with face-to-face meetings, interactions can be held at a firm's headquarters or at a bank's branch. We believe that among active types of contact, interaction at a firm's location may signal the interest of the lender to build (or reinforce) its relationship with the borrower. Thus, considering the two dimensions of analysis (intentionality and the location of contact) leads to the classification reported in Figure 1.

Specifically, we identify four types of close contact (and four types of soft information sources):

Quadrant I) Direct face-to-face meetings held at a bank branch at the initiative of the bank;

Quadrant II) Direct face-to-face meetings held at a firm location at the initiative of the bank (i.e., visit by the bank);

Quadrant III) Direct face-to-face meetings held at a bank branch at the spontaneous initiative of the borrower; and

Quadrant IV) Direct face-to-face meetings held at a firm location at the spontaneous initiative of the borrower.

We include the last class only to make the classification exhaustive. However, we consider it an extremely rare case because direct face-to-face encounters at a firm location solicited by the firm are not common in small business lending (due to the low bargaining power of borrowers).

Figure 1. Type of Soft Information Sources

Location of the meeting

	BANK	FIRM
BANK	Active I	Active⁺ II
FIRM	III Passive	IV Passive

Who solicit the meeting

B. Hypothesis

Whereas in the previous section, we attempt to answer õhowö banks produce soft information, in this section, we formulate a hypothesis about õwhyö banks collect soft information. In this sense, we believe that there are at least three reasons why banks might push to have close contact with customers, which leads us to formulate three hypotheses.

B1. Opaqueness Hypothesis (Hp. 1).

Opaqueness plays a crucial role in lending and can represent a major obstacle in small firm financing, particularly with respect to lending from large banks that engage in impersonal or arms-length interactions with their customers, whereas relationship lending can mitigate opacity problems through the assessment of reliable soft information collected by loan officers through repeated contact and interactions (among others, Berger and Udell, 1998; Cole et al., 2004).

However, although the role between opacity and relationship lending is well established and conventional, the relationship between opacity and the gathering of soft information is unexplored. Specifically, although we know that soft information can help the interaction between opaque firms and lenders, it is unclear whether banks prefer to intentionally collect information from opaque firms. In other words, although opaque firms may benefit from disclosing soft information to a lender, the role of opacity as a õcauseö of soft information production is ambiguous because it is arguable that it may be collected for other reasons. In fact, the expectation that a firmøs opaqueness works as an incentive for the lender to collect inside information would lead to incorrect conclusions. For example, suppose that banks decide to query customers with intentional and close contact (active contact) to clarify documents or numbers in their possession (hard information) that appears to be opaque (e.g., financial statements). This line of reasoning ó in which the borrowerøs informational opaqueness creates friction in the bank-firm relationship and repeated contact between the parties is used to alleviate this friction ó implies that banks have poor screening processes in which they agree to lend to unqualified borrowers (highly opaque firms) and then later gather qualitative information about these firms to better assess their portfolio. The results do not change if we consider passive contact rather than active contact. In fact, in this case (i.e., opaque firms intentionally contact the bank, which then collects inside information passively), the idea that banks lend to questionable borrowers first and become convinced of their creditworthiness only later (after collecting soft information) is not credible. Thus, we believe that the production of soft information is negatively related to the opacity of borrowers and formulate the following *opaqueness hypothesis*:

Hp1. "Lack of transparency does not induce banks to collect inside information about opaque borrowers"[1].

B3. Risk Hypothesis (Hp. 2).

The second expectation is related to borrower risk. Lenders will finance firms only when they have high expectations of being repaid; thus, they will favor borrowers with characteristics that increase the likelihood of repayment before or at the maturity date. In addition, institutional changes, such as Basel II and the effects of the financial crisis, may enhance the tendency of banks to increase contact with low-risk borrowers in attempting to improve their allocative efficiency and preserve capital. Accordingly, we develop the following *risk hypothesis.*

Hp2. "Banks prefer to establish contact with low-risk borrowers to improve their capital requirements".

B3. Sales Strategy Hypothesis (Hp. 3).

In addition to risk, banks may have and receive incentives to collect inside information through the product channel.

In fact, the scope of the bank-borrower relationship is another factor that may also affect the amount and type of credit-relevant information desired, as argued by Petersen and Rajan, according to whom "in addition to interaction over time relationships can be built through interaction over multiple products" (Petersen and Rajan, 1994 p. 6). The provision of multiple products may affect lending behavior in multiple ways, such as by determining the cost of loans (Degryse and Cayseele, 2000) and collateral requirements (Chakraborty and Hu, 2006). However, before considering the effect, it is useful to consider the determinants of selling financial products. Although it is obvious to consider that providing financial products is the *raison d'être* of a bank, it is notable that banks provide a larger overall package of products beyond lending. Sales of combinations of loan and non-loan products makes the SME segment particularly profitable for banks and may cause the bank's business strategy to focus on cross-selling activities (De La Torre et al., 2010). Because non-lending products (fee-based services) are not subject to capital adequacy requirements and may be attractive in terms of profitability, cross-selling may be part of a strategy for banks to maximize

---

[1] It is important to understand that this hypothesis is not contrary to the conventional view that relationship lending is the main way for banks to cope with opaqueness. Instead, it just states that opaqueness is not the cause for gathering soft information (although it can be resolved through it).

their scarce resource (i.e., capital).

In this framework, if relationship contact is driven primarily by the risk profile of borrowers (Hp 2), we believe that banks prefer to have direct contact with borrowers characterized by a low level of banksøproduct saturation. In other words, we test the following *sales strategy hypothesis*:

Hp3. Banks prefer to have direct contact with borrowers with low risk and also low levels of product saturation to increase the provision of financial products.

III. Empirical Strategy
A. Data and Sample Description

One of the primary limitations of one study that aimed to examine empirically the role of soft information in lending practices was in the observation of real measurements of the amount of information accumulated by the lender. To circumvent this limitation, we collect data from the commercial loan portfolio of one of the five largest commercial banks in Italy (henceforth, õthe bankö), which contains the amount of contact between the bank and a large sample of its clients from 2009-2012. We also assemble comprehensive information on the lending and non-lending dimension of each borrowerøs relationship, including balance sheet data.

The bank operates in all Italian provinces throughout a branch network of more than 3,000 offices and offers a full suite of commercial and personal loan products, residential mortgages, cash management, deposit services, securities trading, portfolio management, bank insurance, financial derivatives, and advisory services. The bankøs clients include small and medium enterprises (SMEs), which is consistent with the SME definition of the European Commission[2].

To be eligible for selection, firms must meet three selection criteria:

1. They must have maintain a credit relationship with the bank from 2009-2012,
2. They must have operated in the same industry over the period of analysis;
3. Each borrower must not have had a distressed loan during the same period so that contact with the bank was not due to insolvency conditions.

From the population obtained by these filters, firms with missing values for any of the main variables of interest for the four-year period under analysis were dropped (casewise deletion). All firms operated in all Italian provinces. Furthermore, the distribution by geographical area is symmetric and aptly reflects the distribution of the Italian population of firms: firms based in the northern regions represent 56% of the population, and the remainder is based in the center (24%),

---

[2] The European Commission defined SMEs as firms with an annual turnover not exceeding 50 million euros and/or an annual balance sheet total not exceeding 43 million euros.

the southern regions (14%) and the islands (6%). Comparing this distribution with data on the national firm population provided by the domestic bureau of statistics, northern (southern) regions are slightly overrepresented (underrepresented). The industrial distribution also reflects national data, except for manufacturing and services industries, which are slightly overrepresented.

These characteristics make the sample well suited to analyze relationship lending because Italian firms typically apply for loans at local bank branches instead of accessing the bond or equity markets for their external financing. Further, none of the sample firms have issued public debt, and none are publicly listed. Furthermore, the size and characteristics of the legal form of the sample and the domestic banking law ensure that banks do not have equity holdings.

For each borrower, we have information on the number and types of contact per year with the bank, credit rating, contract terms, share of debt provided by the bank, number of lenders, date of origination of the relationship with the bank, and industries. Further data can be matched with firm balance-sheet data that are used in this study primarily to measure the level of borrower opacity.

Because we have information on their physical location, we are able to measure the distance between the borrower and the lender to take account of possible friction that may arise as a result of long distance and the frequency of bank-firm meetings. Moreover, the physical location permits the measurement of the banking concentration level at the province level through the calculation of the Herfindahl-Hirschman index. Finally, because our data regarding the commercial dimension of the bank are highly detailed, we collected data related to the portfolio of loan and non-loan products sold by the bank to its clients during and before the period under examination.

B. Variables Definition

In the analysis that follows, we work with the set of variables briefly described in Table 1.

About here Table 1

The most important variable of our set of analysis is $MEETING_{it}$, which consists of the number of meetings between the banks and firm i during year t. Meetings can be distinguished by three types according to the place in which a meeting occurred and the party that requested it. Specifically, we distinguish among the following types: *I)* meeting at a firm's headquarters (requested by the bank), *II)* meeting at a bank branch (requested by the bank), and *III)* meeting at a bank branch (requested by the firm). Unreported descriptive statistics demonstrate that the first two types of meetings

comprise, respectively, approximately 71% and 4% of the total number of meetings observed during the period of analysis, whereas the third type represents 25%.

As shown in Table 1, the mean value of $MEETING_{it}$ is 0.49, whereas the median, the 1st percentile and the 99th percentile are 0, 0 and 6, respectively. Because this variable is key to this study, further analysis was necessary to reject some concern about its representativeness: specifically, some unreported analysis noted no difference between firms with and firms without meetings at the industry and province level.

Regarding variables on the right-hand side, we have four variables that can be considered proxies for the nature of the relationship between the firm and its bank: 1) LENDERS is the number of banks with which the firm has transactions; (2) REL_LENGTH is the number of years that the bank has been providing services to the firm since first contact with the latter; (3) SHARE is the proxy for the relative magnitude of the bank with respect to all the other creditors of a firm and is measured by the bank's share in total debt financing of the firm; and (4) SCOPE[*] is the number of financial products the firm purchases from the lender (at the year level).

Next, there are four variables that capture other bank-firm characteristics: (1) OPACITY is the ratio of intangibles to total assets and represents a proxy for the lack of informational transparency from an accounting perspective; (2) RISK contains the borrower creditworthiness, as calculated by the banks'internal borrower rating system, which consists of 5 risk classes for solvent borrowers (i.e., 1 = the least risky class; 5 = the riskiest class)[3]; (3) DIST is a proxy of the distance between the firm and the bank and is a time-invariant dummy that assumes a value of one if the province in which the firm is located and the bank's local branch are different and zero otherwise; (4) HHI (Herfindahl-Hirschman index) measures the market concentration of the local credit market (computed in terms of bank branches) at the province level of firms.

We also build a specific set of dummy variables to control for borrowers'industry (specifically, 14 dummies), borrowers'province (110 dummies) and year (4 dummies).

C. Methodology

Our empirical methodology is designed to test the empirical predictions regarding the determinants

---

[3] The real number of internal rating classes is different. However, we decided to normalize to 5 the number of classes to preserve the privacy disclaimer of our data provider. All empirical evidence reported in this study is obtained using the real (and hidden) number of rating classes.

of soft information production after considering face-to-face meetings between bank and firms as the main source of this production. We employ a firm-level panel data analysis because our dataset includes 60,000 observations over the 2009-2012 four-year period. The dataset is unbalanced because some firms appear twice, whereas others appear for all four years. The general static panel data regression model can be represented as follows:

$$Production\ of\ Soft\ Informatio\,n_{it} = \beta_0 + \sum_{k=1}^{K} \beta_{1k} Bank\_Firm\ Var_{it} + \sum_{n=1}^{N} \beta_{2k} Control\ Var_{it} + \alpha_i + \varepsilon_{it}$$

where we index all variables with an i for borrowing firms and a t for the time period (t = 2009,í, 2012), and $_i$ and $_{it}$ are error terms. Specifically, $_i$ varies across firms but not across years, whereas $_{it}$ is different for each firm at every time period.

We regress our dependent variable, MEETING$_{it}$, against the set of variables extracted by the firm-bank credit file: duration of the relationships (REL_LENGTH), credit concentration (LENDERS), percentage of credit granted (SHARE), borrower's credit risk (RISK), bank-firm interaction across products (SCOPE), lender-borrower distance (DIST) and the index of market concentration (HHI). The regression also includes dummies (not presented in the table) for the year of each observation and the firm's industry and firm's location, which is important in Italy (considering the high level of economic differences among regions).

Although the dataset employed is characterized by a large number of observations, the period of analysis is limited (four years); thus, the panel regression with random effects might be suggested as the more suitable approach. Furthermore, the availability of some time-invariant variables that may have significant effects on the dependent variable and cannot be omitted (i.e., bank-firm distance, industry and province dummies) calls into question the use of a fixed effect approach. However, to address the choice of the model formally, a Hausman test was conducted. The results show that it is possible not reject the null hypothesis that the differences in coefficients are not systematic. Because there are reasons both to include and exclude fixed effects, we opt to estimate our main results without fixed effects and employ fixed effects as a robustness check (see Section III.E).

D. Main Results

D1. Opacity and Risk.

Table 2 shows the regression results. We begin with column (1), which shows a negative and statistically significant (at the 1% level) relationship between the measure of the opacity of the borrower and the frequency of contact with the bank. This result indicates that the bank prefers (does not prefer) to visit clients that are informatively transparent (opaque) from an accounting

perspective. Therefore, the idea according to which soft information can be considered (among other things) as a type of supplement that a bank officer collects to compensate for the lack of transparency of the balance sheet can be rejected.

About here Table 2

However, we believe that the most interesting finding of this first estimation is the result for RISK, which has a negative and statistically significant coefficient at the 1% level; i.e., as the borrower's credit risk increases, the number of face-to-face meetings between the bank and the firm decreases. In other words, the production of soft information is more intense for high-quality firms. One possible explanation for this finding is that banks are privileged to conduct physical face-to-face encounters to improve their knowledge about low-risk clients in an attempt to reinforce the relationship and/or detect the possibility of developing more business in the future with this type of client. It seems plausible that banks are trying to strengthen their relationship with high-quality firms to extract additional (not risky) future profits. We believe that this is the most important finding of the study because it might shed light on the real mechanism of soft information production that has not been previously explored. However, we are concerned that the origination of the face-to-face contact is not random because each contact might be solicited by the bank or the firm. Particularly when the contact is generated at the firm's initiative, the bank-firm interaction can extend beyond the bank's strategy because it can reflect the firm's needs of a different nature (e.g., credit deterioration, claims, assistance, payments, monitoring). Thus, we address these concerns in Section III.F.

D2. Number of Lenders

The sign of the coefficient of LENDERS is positive and statistically significant at the 1% level, which suggests that an increase in the number of lenders positively affects the one-to-one interaction between the bank and the firm. In particular, the result shown in column (1) suggests that the relationship with a new lender increases the number of occasions of annual contact by 0.0652. If the meeting is solicited by the bank, this increase might be interpreted as the bank's attempt to improve the borrower's knowledge with multiple banking relationships. Moreover, the result of LENDERS may also be interpreted as further evidence of the tendency of the bank to incentivize the collection of soft information for non-opaque firms because firms that borrow from multiple banks are less informationally opaque than firms with one lending relationship and are

more likely to borrow from a large bank (Berger et al. 2001).

However, because we are interpreting these initial findings by holding all other variables constant, it might be interesting to detect whether the bank prefers to meet clients who have multiple lenders without considering their credit risk or if it prefers to relate with low-risk borrowers. In attempting to answer to this question, we add an interaction term between LENDERS and RISK in column (3). The sign of the interaction term is negative and statistically significant at the 1% level, indicating that the frequency of face-to-face contact with multiple banking is decreasing in risk. For example, the coefficient suggests that the amount of change in MEETING with a one-unit change in LENDERS for firms included in the best class of risk is +0.0834 (0.0964-0.013$\times$1), whereas the change for firms included in the worst class of risk is +0.0314 (0.0964-0.013$\times$5).

This result indicates that the frequency of meetings is higher for low-risk borrowers who have shown a willingness to borrow from multiple lenders. Furthermore, if we assume that the number of lenders is more likely to increase in markets with high competition, this finding is consistent with the notion that banks have a tendency to cultivate their extensive ties with customers, particularly in banking markets with low market concentration (Presbitero and Zazzaro, 2011).

D3. Share of debt

Consistent with previous findings, it is reasonable that one bank intensifies the collection of soft information for borrowers with a low share of loan and non-loan products because it appears to be the ideal state in which the bank engages in profitable commercial activities. To this end, we add to the set of variables SHARE, which exhibits a negative coefficient that is statistically significant at the 1% level. In particular, the coefficient shown in column (1) indicates that a one-percentage point decrease in the share of the debt held by the bank predicts an increase in contact of approximately +0.08. Notably, the sign and the statistical significance of SHARE hold for all specifications (except for column 3, where the coefficient is statistically significant at the 5% rather than the 1% level). Although this finding is consistent with the notion that more informed banks (with a high fraction of debt) are less willing to collect new information, it is consistent with the share of wallet maximization processes. In fact, if the objective of contact is mainly to provide new services to the borrower, these services are progressively less effective after a certain limit until they become useless when the borrower reaches a saturation level with respect to its own financial needs.

D4. Banking Products

Now, we turn to a discussion about the role of sales activities in the production of soft information.

Because we collected detailed data on the set of bank products sold to each client, we are able to add $SCOPE_{it}^{*}$ to our set of variables, and we consider it an expression of the breadth of the relationship because the provision of services may act as a good measure of the closeness of the lending relationship (Petersen and Rajan, 1994; Degryse and Van Cayseel, 2000). This variable measures the number of loan and non-loan services purchased from the bank by client i during year t. Considering that the number of financial products purchased from a bank might depend (among other things) on the amount of credit that the same bank grants the borrower, we chose to scale $SCOPE_{it}^{*}$ for the amount of credit granted from the bank during year t (outstanding loans and loan commitment) and multiply the result by 1,000,000 for readability.

In this way, we find SCOPE, which represents the number of banking products per 1 million euros of loans. The standardization of the credit amount of the number of banking products (with respect to a non-standardized measure) is preferred for at least two reasons. First, and most importantly, this standardization may help neutralize the effect of a firm's demand for credit and thus enables an interpretation of the variable from a supply-side perspective. Second ó and as consequence of the previous effect ó the standardization for credit granted may help obtain a measure of intensity of cross-selling strategies. In fact, given two firms with the same amount of credit, it is arguable that firms with a higher level of SCOPE might have purchased more non-loan services as a consequence of selling products or services in addition to the main product (loans).

Thus, SCOPE can be used as a measurement of the probability of the future commercial expansion of the bank and might indicate the bank's commercial appetite for firms: As the value of SCOPE decreases, the frequency of contact with the lender increases, which, in a such framework, can be interpreted as the expected efforts of the bank to satisfy the firm's financial needs.

Consistent with this prediction, the coefficient is negative and statistically significant in all specifications, suggesting that one-on-one encounters are more frequent for firms that have smaller shares of products.

In column (4), we include the interaction between SCOPE and RISK to detect how the change in the frequency of meetings caused by a one-point increase in SCOPE is related to the borrower's risk. The coefficient of the interaction terms is positive and statistically significant at 1% (and significantly different from zero), indicating that the effect of SCOPE increases with risk. Specifically, in the table of predictive margins (unreported for purposes of brevity), we observe that when we consider firms included in the 1st and 5th classes of risk, an increase of SCOPE causes changes in MEETING of -0.0003 (-0.0004+0.0001*1) and +0.0001 (-0.0004+0.0001*5), respectively. Analogously, in the case of a decrease in SCOPE, the signs of the margins are the

opposite, indicating that banks prefer contact with borrowers with low SCOPE and low risk conjointly, which is consistent with the õ*sales strategy hypothesis*ö (Hp3). Moreover, the very low value of coefficients should not mislead because the value depends on the presence of an adjusting factor of 1,000,000 þ[4] in the formula of SCOPE.

D5. Length of relationship

The production of soft information may be correlated to the length of the bank-firm relationship, although not unambiguously. In particular, it is possible to formulate two opposite mechanisms of interaction. On one hand, it seems reasonable to expect a negative correlation between REL_LENGHT and MEETING to the extent that it might be expected that the bank may seek more contact with newer customers in an attempt to retain them. On the other hand, long standing relationships lead to cross-selling activities that multiply bank-firm interactions over time. The results for REL_LENGHT seem to consider the latter case the most probable behavior because the coefficient of the variable is positive and statistically significant at the 1% level.

However, we believe that saturation and diversification of borrowersø financial needs over time reduce demand for and the supply of bank products. In other words, after a certain level of banking products are purchased, firms may appear less willing to buy new products; as a consequence, the bank might focus on and switch its commercial (and contact) efforts to different clients.

To test this mechanism, we add a new explanatory variable called CUM_SCOPE_{it} t, which indicates the number of banking products that the bank sold to firm *i* from the beginning of the relationship to time *t*. The variable considers all types of banking products provided by the bank, including loan and non-loan services[5], and captures the intensity of the bank-firm relationship in terms of the coverage of financial needs; the average value is 73, and the median is 48. Contrary to the results for SCOPE, we are not able to scale the variable for credit granted by the bank because CUM_SCOPE is a cumulative value that summarizes the historical interactions and cannot be related to a unique year, as a result.

The results for the new variable are reported in column (5), which shows that the coefficient of CUM_SCOPE is positive and statistically significant at the 1% level, indicating that an increase in one product sold in the past increases the number of occasions of direct contact during the current

---

[4] For example, if we adjust the calculation of SCOPE to 1,000 (euros) rather than 1,000,000, the coefficient of the interaction term would be 0.08, and the marginal effects for the 1st and 5th classes of risk would be -0.3 and +0.1, respectively.

[5] The number of products is 981.

year. More importantly, we add to the same column an interaction between the new variable and REL_LENGHT to capture how the change in the duration of the relationship affects the dependent variable at different levels of products sold in the past. The coefficient of the interaction term is negative and statistically significant at the 1% level (and different from zero at the 1% level), which suggests that the frequency of face-to-face contact is decreasing in CUM_SCOPE. We divide the distribution of CUM_SCOPE into deciles and report the marginal effects of REL_LENGHT on the dependent variable for each decile of CUM_SCOPE in Table (A1) of the Appendix. The results are completely consistent with our expectations because the effect of the interaction is positive when the number of products sold in the past is low (deciles 1-4), whereas the effect is negative (i.e., the amount of MEETING decreases when REL_LENGHT increases by one year) for a higher number of products (deciles 5-10).

This finding reveals the different behavior of banks with young and not-young clients in terms of the amount of contact when the level of products is considered. Moreover, it represents further evidence in support of the *sales strategy hypothesis* to the extent that it indicates how the products channel ó and the distance from the saturation point in the clientøs spending patterns ó works as an incentive to meet clients more frequently and consequently to create the conditions to collect soft information.

D6. *Residual variables*

We explore the effect of bank-firm distance (DISTANCE) and market power on the dependent variable as factors that might affect the incentives of both banks and firms to have direct face-to-face meetings or on-site visits.

As opposed to previous results, we find that both variables show a negative coefficient. At first glance, these findings suggest that *ceteris paribus*, close contact between banks and firms is more likely when the borrower is nearby and in markets with higher competition. However, these results are not reliable because all coefficients are not ó or are only weakly ó statistically significant.

Furthermore, to assess whether the intensity of the relationship is influenced by the profitability of the company, in column (2), we add the return on investment ratio (ROI) to the set of explanatory variables. The result shows a positive coefficient but is not significant, which suggests the irrelevance of profitability with respect to our purposes.

E. Robustness Check.

This section provides additional analyses to check the robustness of the results obtained thus far. First, the results might be influenced by typical problems of endogeneity and simultaneous causality. The most relevant problem in our case is represented by the number of banking products sold during year t (SCOPE $_{it}$) because this number might be endogenous to the frequency of contact. The difficulty in selecting an appropriate instrumental variable that is able to satisfy the typical requirement of a 2SLS approach (relevance condition and exclusivity restriction) lead us to opt for a different strategy. In particular, our strategy to reduce endogeneity problems is to use banking products lagged by one period as a variable on the right-hand side. Thus, to try to õexogenizeö banking products when estimating the causal effect on the dependent variable, we use SCOPE$_{i, t-1}$.

The results are reported in column (1) of Table 3 and show that SCOPE$_{t-1}$ has the same coefficient of the baseline regression (Table 2, column 1). Moreover, whereas the coefficient was statistically significant at the 5% level in the baseline regression, the results reported in Table 3 indicate a coefficient that is statistically significant at the 1% level.

The second check reported in this section is related to fixed effects regression. As discussed above (Section III.C), we prefer a random effects model because we must observe the effect of some time-invariant variables (i.e., distance) on the dependent variable. Furthermore, the large number of observations within the period of analysis is limited (four years), and province and industry dummies are included; these factors call into question the use of a fixed effects approach. However, to address the choice of the model formally, a Hausman test was conducted, and the results show that it is possible not to reject the null hypothesis that the differences in coefficients are not systematic. Thus, we include fixed effects at the industry and province levels and report the results in column (2). A comparison of columns (1) and (2) reveals that the results remain substantially unchanged when we opt for a fixed effects rather than a random effects model, although the latter shows an R-squared more than twice as high.

Furthermore, we model RISK as a set of dummy variables to obtain the effect of different levels of risk on banksøcontact strategy. The results are reported in column (3) and show how the frequency of contact is perfectly decreasing in risk because the marginal effects are 0.3374 and -0.2292 (both statistically significant at the 1% level) when firms belong to the first (the best) or the fifth (the worst) classes of risk, respectively.

Finally, to control for an alternative measure of opacity, we change OPAC to *ln(TotAsset)*, which is the natural logarithm of one plus total assets, to address the skewness of total assets and to accommodate borrowers that report zero assets: The results reported in column (4) show a positive coefficient (statistically significant at the1% level), suggesting that the frequency of contact

decreases as the size decreases (and thus so does the production of inside information). Considering that the size of firms is a measure of opacity (Cole, 2004), this finding confirms our previous results.

About here Table 3

F. Type of soft information production: splitting active and passive contact

The results shown in the previous sections indicate that (*ceteris paribus*) the collection of soft information is higher for firms with low risk, a low level of credit concentration and that have a low number of products sold by the bank. These findings might be interpreted as a consequence of a bank strategy to contact the firm to improve its market share and extract future profits with a low level of risk. However, because our dependent variable is an aggregate value of direct face-to-face encounters, to measure correctly those cases in which the bankøs privileging of a firm results in direct face-to-face encounters and those other cases in which the bank has (passive) encounters with its clients, the distinction between the active and passive collection of soft information should be focused upon, as suggested at the beginning of this study (Section II.A). In fact, to examine the existence of a sales strategy beyond the activity of bank-firm contact, it is necessary to disaggregate the number of direct encounters in four mutually exclusive and jointly exhaustive categories, as shown in the matrix in Figure 1.

Thus, the types of soft information considered by quadrants I and II are denoted by ACTIVE and ACTIVE$^+$, respectively, whereas the contact related to quadrants III and IV are grouped with the term PASSIVE.

Following this classification, we can run our baseline regression for each of the previous two variables. Specifically, we prefer to extend the baseline regression only to ACTIVE$^+$ and PASSIVE due to the limited representativeness of ACTIVE. (As discussed above, this type of contact accounts for only 4% of the total number of instances of contact.) The results are shown in Table 4, where columns (1) and (2) report the outcome when the dependent variables are ACTIVE$^+$ and PASSIVE, respectively.

The results are different for columns (1) and (2). More precisely, the results obtained in previous sections remain unchanged when the dependent variable is ACTIVE$^+$ (column 1). However, when we analyze the results for passive contact (column 2), we note that there is not strong statistical evidence for OPACITY or the lagged value of SCOPE, whereas the borrowerøs risk (RISK) has a negative coefficient and is statistically significant but is 10 times smaller. Furthermore, comparing columns (1) and (2) shows that the regression with ACTIVE+ as a dependent variable has an R-

squared approximately 3.5 times larger than with PASSIVE as a dependent variable.

This finding indicates that the production of soft information described in this study, and particularly the opacity and sales strategy hypothesis, are confirmed when the bank takes the initiative to contact the borrower at the borrower's headquarters.

About here Table 4

IV. DISCUSSION AND EVIDENCE ON LOANS RATE
A. Soft information and quality selection process

The aim of this study is to analyze the determinants of the production of soft information on bank clients. We began our empirical investigation assuming that soft information is collected primarily through the bank's close contact (i.e., not-remote contact) with borrowers. The findings show that, *ceteris paribus*, the bank has close contact with firms characterized by high transparency, low risk, and low use of their own banking products (loan and non-loan services). Notably, when we split the instances of contact by who solicits the meeting (the bank or the borrower), these findings are confirmed when contact originates by virtue of the bank's own initiative.

Taken together, our results indicate that the production of soft information follows a quality selection process according to which the bank prefers to tighten its relationship with clients who are less risky and more profitable in terms of providing future services.

The finding of this mechanism of soft information production might lead to an important implication for studies of relationship lending.

A significant portion of the abundant literature on relationship lending is devoted to analyzing the beneficial effect of a tight relationship both for banks and (particularly) for firms. A large part of these studies examine the effects of relationship lending on the cost of loans (among others, Degryse and Cayseel, 2000; Brick and Palia, 2007; Bharath et al., 2011) and credit availability (among others, Petersen and Rajan, 1994; Berger and Udell, 2002; Jiangli et al., 2008; Puri, et al., 2011) and provide mixed results. On one hand, this strand of the literature depicts relationship lending primarily as a practice that is able to alleviate information asymmetry; on the other hand, a close relationship may determine hold-up problems (Sharpe, 1990; Rajan 1992) and soft budget-constraints (Dewatripont and Maskin, 1995).

However, if we consider the portion of studies that highlight the beneficial effects of relationship lending, we see that the possibility that a strong relationship that stems from a predetermined quality selection process is not clearly considered. In other words, banks may select firms with the best qualities and then focus on developing tight relationships with them by accumulating soft

information. In this mechanism, the effects of tight relationships may be biased by a preliminary scheme related to bank preferences. Under this reasoning, it is possible that borrowers obtain more credit and/or a lower loan rate because they should be considered by the banks as clients in conformance with their commercial strategies in terms of risk and future profits and then receive repeated points of contact with the bank. In particular, firms that match the characteristics required by a bankøs strategy might have received past contact with high relationship value, such as direct contact (face to face) and, in particular, intentional contact (active soft information collection). Consequently, the strength of a relationship positively affects the credit availability and/or loan rates because it can depend on a prior bank-specific preference that may appear to be latent, in cases with a lack of data.

B. Loan rate and high-value relationship contact

The issue of the quality selection process discussed above led us to a further analysis. In this section, we estimate the effect of soft information production on loan rates. The effect of relationship lending on the cost of loans is well examined in the literature. Because relationship lending lowers information asymmetries and monitoring costs, in a competitive loan market, banks transfer monitoring cost savings to their clients in terms of lower interest rates (Diamond, 1989, 1991). Boot and Thakor (1994) also argue that the rates charged for loans should decrease as a borrower-lender relationship matures. Empirically, some proxy of the strength of relationship, such as borrowing concentration (Blackwell and Winters, 1997) and the duration of bank-firm interaction (Berger and Udell, 1995; Brick and Palia, 2007), can alleviate loan rate premiums. Furthermore, the age of the firms play a role to the extent that older firms are more likely to pay lower loan rates (Petersen and Rajan, 1994; Harhoff and Körting, 1998; Degryse and Van Cayseele, 2000).

To examine the effect of the accumulation of soft information through repeated contact on interest rates, we collect SPREAD, which represents the spread paid over the cost of funding for a sample of loan levels in our data (section II.A) and regress it on the following set of explanatory variables: the amount of the loan granted (AMOUNT), the duration of the loan (DUR), the internal rating of the borrower (RATING), the length of the borrower-lender relationship (REL_LENGHT), and the number of face-to-face borrower-lender meetings (MEETING), both lagged and not lagged.

About here Table 5

Consistent with our previous discussion, we expect that repeated contact lowers the lenderøs cost of

lending to small firms. In particular, we expect that reduced information asymmetries through inside information collected during face-to-face meetings are passed on to the borrower by means of a lower interest rate.

The results are reported in Table 5. We begin with the duration of the loan (DUR), which is a variable that has not been employed thus far, and has a median and average value of 5 and 6.08 years, respectively, whereas the 5th percentile is 1,5 years, and the 10th percentile is 3 years. These descriptive statistics suggest that the loans analyzed are primarily medium/long-term mortgages because most of the loansø duration distribution exceeds 18 months, which in the Italian market represents the (informal) cut-off period to distinguish short-term loans (such as unsecured lines of credit) and medium and mortgage loans (secured loans).

As reported in all specifications of Table 5, the coefficient of DUR is negative (-0.0129) and highly statistically significant. We explain this finding with the effect of the collateralization of loans, which is more intense for long-lasting loans, although we are not able to collect data on collateral.

The correlation between the duration and the amount of loans is high, as expected (ρ=0.50). However, when both variables are included in the regression, we find a lack of importance for AMOUNT, with a coefficient that is not significantly different from zero.

To test the relations between risk and loan rates, we include 4 dummies (n-1), indicating the internal rating. As clearly expected, the loan rate is increasing in risk to the extent that firms included in the first class of risk (rating 1) experience a decrease of approximately 24 basis points compared with other firms, whereas firms co-located in the worst class of risk (rating 5) pay an excess interest rate of approximately 11 basis points.

More importantly, we add MEETING to the regression and find a negative coefficient that is statistically significant at the 1% level. In particular, we observe that an increase of one meeting reduces the loan rate by 2 basis points. The coefficient value passes to approximately 1.5 basis points in column (2) when we include the lagged value of MEETING at one year to control for possible simultaneity effects.

Finally, to test how a high-value relationship value affects the dependent variable, we include in columns (3) and (4) the number of instances of contact at a firm location at year t ($ACTIVE^+$) and at year t-1 ($ACTIVE^+_{t-1}$). Consistent with the discussion and results in Section III, the results remain substantially unchanged in columns 1-2 in terms of the sign and statistical significance, whereas, more importantly, we note a stronger value of the coefficient for $ACTIVE^+$ than MEETING. We interpret this finding as further confirmation for how active soft information collection plays an important role in understanding the beneficial effect of relationship lending practice.

V. Conclusion

Although there has been a considerable amount of theoretical and empirical work on the topic of soft information and its role in relationship lending, there has been far less discussion about the originations of this special type of information. A particularly under-explored issue is the examination of how banks collect information and why they do it. The aim of this paper is to take the first steps toward addressing these issues. First, the study is based on the basic assumption that inside information is produced through repeated contact between a lender and its customers. Then, the paper focuses its analysis on close contact (face-to-face contact), and remote contact (by phone, email, web etc.) is excluded. In an effort to analyze the determinants of these instances of contact, the study presents a classification of types of contact based on two dimensions: the intentionality of the contact and the location of the contact. This classification identifies two macro classes of contact: active and passive, in which the former types of contact are promoted by the bank, and the latter represents spontaneous close contact between the firm and the bank. In particular, among õactiveö instances of contact, those related to meeting at the firm location are considered instances of contact with high relationship value and are ideal to identify the non-random paths of the origination of this type of highly informative bank-firm interaction. The basic notion is that because active contact is costly and intentional, such contact likely follows a predetermined combination of strategies set by the bank, aiming to strengthen the relationship with borrowers with specific characteristics.

By using a large sample of bank-firm meetings extracted from the credit file of a large Italian bank between 2009 and 2012, we observe that the banks in the sample prefer to concentrate their active contact primarily with transparent and low-risk firms. This finding may be a consequence of current institutional changes (Basel II) and of the resultant willingness of banks to preserve capital. More importantly, we also test the possibility that contact is focused particularly on commercial purposes, and we find strong evidence that active contact is more frequent when selling and cross-selling of products is more likely.

Finally, our results imply that the production of soft information may follow a quality selection process according to which the bank prefers to have tight relationships with clients that are less risky and more profitable in terms of provision of future services in the future. We believe that these findings have non-negligible implications to the extent that the beneficial effects of relationship lending shown by abundant and rigorous past studies might be affected by a selection effect of firms

not well explored until now, primarily because of a lack of data.

However, these results must be interpreted with caution, and at least three caveats must be mentioned. First, although the bank is highly representative of the domestic banking system in terms of strategy, credit/financial products and geographical coverage, because this study is based on one bank, our results cannot be generalized to all domestic banks. Second, we cannot ignore that the findings of this study are affected by the large size of the bank to the extent that such intensive commercial and risk-allocation strategies may be consistent with large banks, primarily.

Third, soft information is also distinguished from hard information in that the data are recorded. Typically, the person who collects soft (hard) information is the same person ó although it can be a different person ó who evaluates the information and makes a decision. In this study, we do not know whether the visits were made by the same person who makes lending decisions about the firm visited. Therefore, we suggest caution in the interpretation of data. However, it is nevertheless true that in cases in which there is no coincidence between the collector and the user of information, their roles are determined by the bank.

References

Agarwal S, Hauswald R (2010) Distance and private information in lending. Review of Financial Studies 23:2757ó2788.

Aghion, P., Tirole, J., 1997, Formal and real authority in organizations. Journal of Political Economy 105, 1ó29.

Bartoli F., Ferri G., Murro P., Rotondi A., 2013, SME financing and the choice of lending technology in Italy: Complementarity or substitutability?, Journal of Banking and Finance 37, 5476ó5485.

Behr, P., Güttler A. and Entzian A., 2011, How do Lending Relationships affect Access to Credit and Loan Conditions in Microlending?, Journal of Banking and Finance 35, 2169ó2178

Berger, A., Klapper, L., Udell, G., 2001, The ability of banks to lend to informationally opaque small businesses. Journal of Banking and Finance 25, 2127ó2167.

Beck T., Degryse H., De Haas R., and van Horen N., 2015, When Armø Length Is Too Far. Relationship Banking over the Credit Cycle, EBRD working paper n. 169.

Berger, A., Udell G., (2002), Small business credit availability and relationship lending : the importance of bank organisational structure, The Economic Journal, 112, 32ó53

Berger, A.N., Miller, N.H., Petersen, M.A., Rajan, R.G., Stein, J.C., (2005), õDoes function follow organizational form? Evidence from the lending practices of large and small banksö, Journal of Financial Economics 76, 237ó269.

Berger, A.N., Udell, G., 1995. Relationship lending and lines of credit in small firm finance. Journal of Business, 68, 351ó381.

Berger, A., Udell, G., 1998. The economics of small business finance: The roles of private equity and debt markets in the financial growth cycle. Journal of Banking and Finance 22, 613ó673

Berger, A.N., Udell, G.F., 2006, A more complete conceptual framework for SME finance. Journal of Banking and Finance 30, 2945ó2966.

Berger, A., R. Rosen, and G. Udell, 2007, Does market size structure affect competition? The case of small business lending, Journal of Banking and Finance, 31, 11ó33.

Bharath, S. T., Dahiya, S., Saunders, A. and Srinivasan, A., 2011, Lending Relationships and Loan Contract Terms, Review of Financial Studies, 24, 1141ó1203.

Blackwell, D., Winters D., 1997, Banking relationships and the effect of monitoring on loan pricing, in Journal of Financial Research 20, 275ó289.

Boot, A.W.A., 2000. Relationship banking: what do we know? Journal of Financial Intermediation 9, 7ó25.

Boot, A.W.A., Thakor, A.V., 1994. Moral hazard and secured lending in an infinitely repeated credit market game, International Economic Review 35, 899ó920.

Brick I., Palia, D., 2007, Evidence of jointness in the terms of relationshipLending, Journal of Financial Intermediation, 16, 452ó476

Cetorelli, N., Gambera, M., 2001, Banking market structure, financial dependence and growth: International evidence from industry data, Journal of Finance 56, 617ó648

Cetorelli, N., Strahan, P., 2006, Finance as a barrier to entry: Bank competition and industry structure in local US markets, Journal of Finance 61, 437ó461

Chakraborty, A. and Hu, C. (2006), Lending relationships in line-of-credit and nonline-of-credit loans: evidence from collateral use in small business, Journal of Financial Intermediation, 15, 86-107

Cole, R., Goldberg, L., White, L., 2004, Cookie-cutter versus character: The microstructure of small business lending by large and small banks, Journal of Financial and Quantitative Analysis 39, 227ó251.

De la Torre A., Martínez Pería M.S. and Schmukler S.L., 2012, Bank involvement with SMEs: Beyond relationship lending, Journal of Banking and Finance 34, 2280ó2293.

Delgado, J., V. Salas, and J. Saurina, 2007, Joint Size and Ownership Specialization in Bank Lending, Journal of Banking and Finance 31, 3563ó3583.

Degryse, H., and Cayseele, P. (2000). Relationship lending within a bank-based system: Evidence from European small business data, Journal of Financial Intermediation, 9, 90ó109.

Dewatripont M, Maskin E (1995) Credit and efficiency in centralized and decentralized economies. Rev Econ Stud 62:541ó555

Dewatripont M., Maskin E., (1995), õCredit and efficiency in centralized and decentralized economiesö, The Review of Economic Studies 62, 541ó555.

Diamond, D., 1989, Reputation Acquisition in Debt Markets, Journal of Political Economy 97, 828ó862.

Diamond, D., 1991. Monitoring and reputation: the choice between bank loans and directly placed debt. Journal of Political Economy 99, 689ó721.

Gambini, A. and A. Zazzaro, 2013, Long-lasting bank relationships and growth of firms, Small Business Economics 40, 977ó1007.

Gobbi G. and Sette E. (2014) Do firms benefit from concentrating their borrowing? Evidence from the great recession, Review of Finance 18, 527ó560.

Harhoff, D., Körting, T., 1998, Lending relationships in Germany: empirical evidence from survey data, Journal of Banking & Finance 22, 1317ó1353

Jiangli W, Unal H, Yom C (2008) Relationship lending, accounting disclosure, and credit availability during the Asian financial crisis. Journal of Money, Credit, Banking 40, 25ó55

Liberti, J.M., Mian, A.R., 2009, õEstimating the effect of hierarchies on information useö, Review of Financial Studies 22, 4057ó4090.

Nemoto T., Ogura Y., and Watanabe, W., 2013, The Decision-Making Mechanism of Regional Financial Institutions and the Utilization of Soft Information, Public Policy Review, Vol.9, 87 ó 116

Petersen, M. 2004. "Information: Hard and soft," working paper.

Petersen, M.A., Rajan, R.G., 1994. The benefits of lending relationships: evidence from small business data. Journal of Finance 49, 3-37.

Presbitero, A.F., and Zazzaro, A., (2011), Competition and relationship lending: friends or foes? Journal of Financial Intermediation, 20, 387-413.

Puri, M.; J. Rocholl; and S. Steffen. "Global Retail Lending in the Aftermath of the US Financial Crisis: Distinguishing between Demand and Supply Effects." Journal of Financial Economics, 100 (2011), 556-578.

Rajan R (1992) Insiders and outsiders: the choice between informed and arm's length debt, Journal of Finance 47:1367-1400

Sharpe, S. (1990). Asymmetric information, bank lending and implicit contracts: A stylized model of customer relationships, Journal of Finance 45, 1069-1087.

Stein, J.C., 2002. Information production and capital allocation: decentralized vs. hierarchical firms, Journal of Finance 57, 1891-1921.

Uchida, H., G. F. Udell, and W. Watanabe, 2008, "Bank Size and Lending Relationships in Japan," Journal of the Japanese and International Economies, 22, 242-267.

Uchida, H., G. F. Udell, and Yamori N., 2012, Loan officers and relationship lending to SMEs, Journal of Financial Intermediation, 21, 97-122.

Table 1
Summary of variables

Name	Description	Mean	Median	1%	99%	# Obs
$MEETING_{it}$	Number of instances of face-to-face contact between the bank and firm i at year level (t)	0.93	0	0	8	59,820
$ACTIVE^{+}_{it}$	Number of instances of face-to-face contact promoted by the bank at the firm's headquarters (at year t)	0.67	0	0	7	59,820
$ACTIVE_{it}$	Number of instances of face-to-face contact with firm I promoted by the bank and held at a bank's branch (at the year level)	0.03	0	0	1	59,820
$PASSIVE_{it}$	Number of instances of face-to-face contact promoted by firm i at a bank's branch at the year level (t)	0	0.23	0	4	59,820

Table 2

OPACITY$_{it}$	The ratio of intangibles to total assets of firm i at year t	0.28	0.20	0	1	73,841
RISK$_{it}$	Internal rating of firm calculated by the bank. It has ten categories, in which a rating of 1 is the best quality, and a rating of 5 is the worst.	4.23	5.23	1	10	144,181
LENDERS$_{it}$	The number of lending banks of firm i at year t	3.85	3	1	21	60,778
SHARE$_{it}$	Bank's share in total debt financing of firm i at year t	0.46	0.39	0	1	60,778
REL_LENGTH$_{it}$	The length (in years) of the credit relationship between the bank and firm i at the end of year t	8.43	5.75	0.2	44	60,778
SCOPE$_{it}$	The number of banking products (loan and non-loan services) sold to borrower i during year t, normalized by the amount credit granted by the bank at year t. The value obtained in this way is then multiplied by 1,000,000.	75.38	30	0	667	59,820
CUM_PRODUCTS$_{it}$	The number of banking products (loan and non-loan services) sold to borrower i from the beginning of the relationship with the bank to the end of year t	73	48	0	420	59,820
DIST$_{it}$	Dummy with value of 1 if the firm is located in a province different from the bank	0.085	0	0	1	60,778
HHI$_t$	Market power: Herfindahl index of bank branches at the province level (at year t)	0.10	0.09	0.03	0.61	60,778
ROI$_{it}$	Return on equity of firm i at time t	4.83	3.83	-42.47	55.56	60,778

	The production of soft information				
	(1)	(2)	(3)	(4)	(5)
OPACITY	-0.1850***	-0.1821***	-0.1872***	-0.1847***	-0.1868***
	(0.0287)	(0.0289)	(0.0287)	(0.0287)	(0.0281)
RISK	-0.1806***	-0.1803***	-0.1267***	-0.1865***	-0.1796***
	(0.0061)	(0.0062)	(0.0101)	(0.0065)	(0.0060)
LENDERS	0.0653***	0.0653***	0.0964***	0.0652***	0.0557***
	(0.0038)	(0.0038)	(0.0070)	(0.0038)	(0.0034)
SHARE	-0.0824***	-0.0817***	-0.0668**	-0.0800***	-0.1328***
	(0.0292)	(0.0294)	(0.0295)	(0.0293)	(0.0262)
REL_LENGTH	0.0087***	0.0087***	0.0089***	0.0085***	0.0091***
	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0014)
SCOPE	-0.0001**	-0.0001**	-0.0001***	-0.0004***	
	(0.0000)	(0.0000)	(0.0000)	(0.0001)	
DIST	-0.0376	-0.0378	-0.0379	-0.0377	-0.0457*
	(0.0258)	(0.0259)	(0.0258)	(0.0258)	(0.0253)
HHI	-0.0637	-0.0587	-0.0484	-0.0628	-0.0330
	(0.1856)	(0.1858)	(0.1848)	(0.1856)	(0.1825)
ROI		0.0001			
		(0.0002)			
LENDERS × RISK			-0.0130***		
			(0.0021)		
SCOPE × RISK				0.0001***	
				(0.0000)	
CUM_SCOPE					0.0026***
					(0.0002)
REL_LENGTH × CUM_SCOPE					-0.0001***
					(0.0000)
Constant	1.1146***	1.1115***	0.9561***	1.1245***	1.0034***
	(0.2542)	(0.2541)	(0.2619)	(0.2534)	(0.2376)
Dummy included (n)					
Industry (14)	YES	YES	YES	YES	YES
Province (110)	YES	YES	YES	YES	YES
Year (4)	YES	YES	YES	YES	YES
Obs	59,820	59,669	59,820	59,820	60,778
R-squared	0.1000	0.0998	0.1009	0.1002	0.1049

This table reports a panel data estimation for the production of soft information. Data are gathered from a large sample of customers of an Italian bank, and the period of analysis is 2009-2012. The dependent variable is MEETING, which indicates the number of instances of bank-firm face-to-face contact *per* year. Variables extracted from the balance sheet (OPACITY and ROI) are referred to as year *t-1* (balance-sheet data for 2008 are available). Numbers in parentheses are the standard error corrected for heteroskedasticity. The symbols ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 3
Robustness Check

	(1)	(2)	(3)	(4)
OPACITY	-0.1762***	-0.1811***	-0.1797***	0.1570***
	(0.0332)	(0.0360)	(0.0363)	(0.0098)
RISK	-0.1526***	-0.1576***		-0.1477***
	(0.0071)	(0.0123)		(0.0070)
LENDERS	0.0409***	0.0406***	0.0411***	0.0147***
	(0.0040)	(0.0055)	(0.0055)	(0.0043)
SHARE	-0.1697***	-0.1715***	-0.1576***	-0.0608*
	(0.0334)	(0.0373)	(0.0376)	(0.0338)
REL_LENGTH	0.0036***	0.0035***	0.0032***	-0.0002
	(0.0013)	(0.0012)	(0.0011)	(0.0013)
SCOPE$_{t-1}$	-0.0001***	-0.0001***	-0.0001***	-0.0000**
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
DIST	-0.0454		-0.0414	-0.0594**
	(0.0281)		(0.0418)	(0.0279)
HHI	-0.0154	-0.0144	0.0366	-0.0160
	(0.2169)	(0.2510)	(0.2539)	(0.2151)
Rating#1			0.3374***	
			(0.0312)	
Rating#2			0.1557***	
			(0.0279)	
Rating#4			-0.1518***	
			(0.0259)	
Rating#5			-0.2292***	
			(0.0484)	
Constant	1.5279**	0.6260***	1.1273	0.2014
	(0.7698)	(0.0829)	(0.6920)	(0.7704)
Dummy included (n-1)				
Industry (13)	YES	NO	YES	YES
Province (109)	YES	NO	YES	YES
Year (3)	YES	NO	YES	YES
Industry FE	NO	YES	NO	NO
Province FE	NO	YES	NO	NO
Obs	27,575	27,575	27,954	27,576
R-squared	0.1004	0.04	0.1000	0.1076

This table reports some robustness check for result reported in Table 2. The dependent variable is MEETING. Data are gathered for a large sample of customer of a Italian bank and the period of analysis is 2009-2012. Variable SCOPE is referred to the year *t-1* to mitigate endogeneity issues. Numbers in parentheses are standard error corrected for heteroskedasticity. The symbols ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 4
Active vs. Passive Collection of Soft Information

Dep. Variable	$ACTIVE^+$	PASSIVE
	(1)	(2)
OPACITY	-0.1448***	-0.0340*
	(0.0275)	(0.0184)
RISK	-0.1268***	-0.0131***
	(0.0058)	(0.0039)
LENDERS	0.0367***	0.0036**
	(0.0035)	(0.0016)
SHARE	-0.1988***	0.0483***
	(0.0275)	(0.0180)
REL_LENGTH	0.0029***	0.0005
	(0.0011)	(0.0006)
SCOPE	-0.0001***	0.0000
	(0.0000)	(0.0000)
DIST	-0.0420*	-0.0029
	(0.0242)	(0.0145)
HHI	0.1367	-0.0938
	(0.1860)	(0.1054)
Constant	1.3639**	0.8991*
	(0.6676)	(0.4592)
Dummy included (n-1)		
Industry (13)	YES	YES
Province (109)	YES	YES
Year (3)	YES	YES
Obs	27,575	27,575
R-squared	0.0334	0.1140

This table reports the panel data estimation for the production of soft information. Data are gathered from a large sample of customers of an Italian bank, and the period of analysis is 2009-2012. The dependent variables are ACTIVE+ and PASSIVE. The former indicates the number of instances of close contact per year promoted by the bank and held at the firmøs headquarters, whereas the latter indicates the number of spontaneous visits at a bank branch of the firm during year t, which indicates the number of instances of bank-firm face-to-face contact *per* year. Numbers in parentheses are standard error corrected for heteroskedasticity. The symbols ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Table 5
Borrowing costs and the role of bank-firm interactions

	(1)	(2)	(3)	(4)
DUR	-0.0129***	-0.0129***	-0.0129***	-0.0129***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
AMOUNT	-0.0000	-0.0000	-0.0000	-0.0000

	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Rating#1	-0.2402***	-0.2405***	-0.2399***	-0.2402***
	(0.0225)	(0.0225)	(0.0225)	(0.0225)
Rating#2	-0.1011***	-0.1012***	-0.1011***	-0.1011***
	(0.0202)	(0.0203)	(0.0202)	(0.0203)
Rating#4	0.0895***	0.0926***	0.0883***	0.0924***
	(0.0274)	(0.0274)	(0.0274)	(0.0274)
Rating#5	0.1085*	0.1146*	0.1068*	0.1158*
	(0.0592)	(0.0592)	(0.0592)	(0.0592)
MEETING	-0.0200***			
	(0.0042)			
MEETING$_{t-1}$		-0.0152***		
		(0.0038)		
ACTIVE_INF			-0.0303***	
			(0.0048)	
ACTIVE_INF$_{t-1}$				-0.0202***
				(0.0043)
Constant	3.5578***	3.5556***	3.5589***	3.5556***
	(0.0280)	(0.0281)	(0.0280)	(0.0280)
Obs	31,064	31,064	31,064	31,064
R-squared	0.1039	0.1039	0.1048	0.1044

This table reports the panel data estimation for loan rates. The dependent variable is SPREAD, which indicates the spread paid over the cost of funding. The numbers in parentheses are the standard error corrected for heteroskedasticity. The symbols ***, **, and * indicate significance at the 1%, 5%, and 10% levels, respectively.

Appendix

Table A1
The effect of duration of relationship and interactions
Average Marginal Effects

Deciles of CUM_SCOPE	Marginal effects	Standard Errors	z-values	P>|z|
1	0.0090	0.0013	6.49	0.000
2	0.0061	0.0011	5.21	0.000
3	0.0033	0.0011	2.83	0.005
4	0.0004	0.0013	0.35	0.728
5	-0.0023	0.0016	-1.43	0.154
6	-0.0052	0.0020	-2.54	0.011
7	-0.0081	0.0024	-3.25	0.001
8	-0.0109	0.0029	-3.73	0.000
9	-0.0138	0.0034	-4.06	0.000
10	-0.0166	0.0038	-4.30	0.000

This table reports the average marginal effects of the REL_LENGHT for different values of CUM_SCOPE. The former indicates the duration of the bank relationship (measured in years), whereas the latter indicates the number of banking products (loan and non-loan services) sold to the borrower since the beginning of the relationship with the bank. The results are obtained by the interaction term included in column 5 of Table 2.
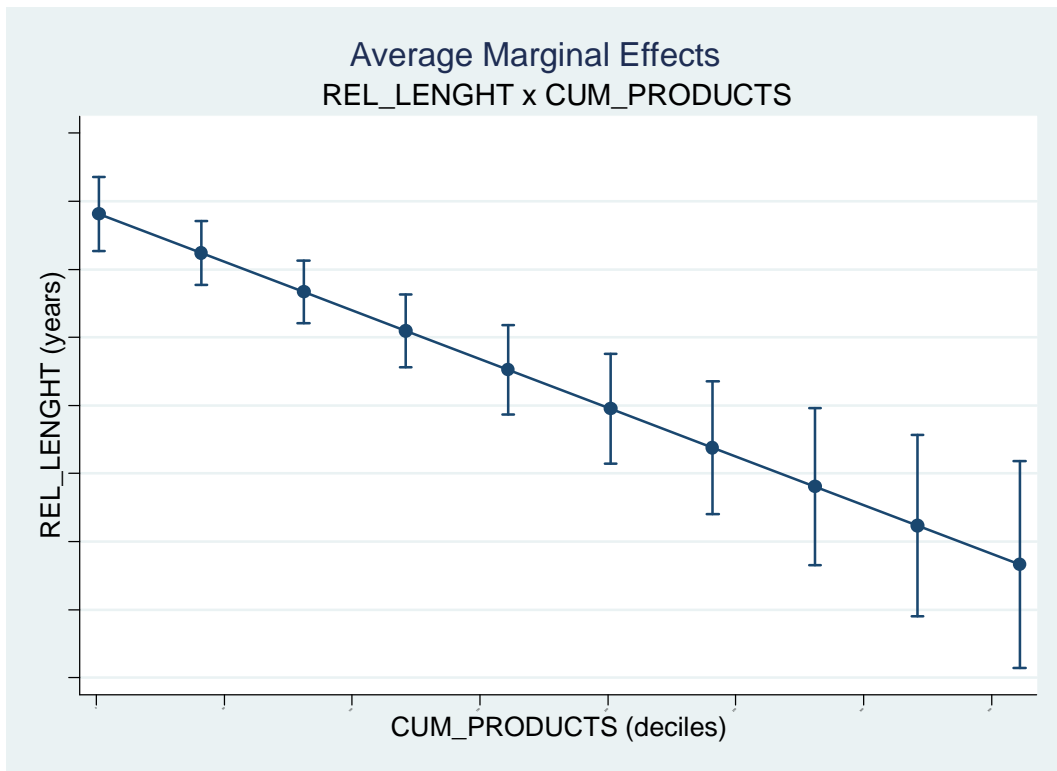
Figure A1.

Average Marginal Effects
REL_LENGHT x CUM_PRODUCTS

REL_LENGHT (years)

CUM_PRODUCTS (deciles)

Table A2
Correlation Matrix

	A	B	C	D	E	F	G	H	I	J	K	L	M
MEETING (A)	1												
ACTIVE[+] (B)	0.8859***	1											
ACTIVE (C)	0.1613***	0.017***	1										
PASSIVE (D)	0.4668***	0.0265***	0.0231***	1									
OPACITY (E)	-0.0174***	-0.0136***	-0.0022	-0.0108***	1								
RISK (F)	-0.1134***	-0.1141***	-0.0423***	-0.0173***	0.0268***	1							
LENDERS (G)	0.0938***	0.1064***	0.0309***	-0.0064**	0.0248***	0.0557***	1						
SHARE (H)	-0.0399***	-0.0601***	-0.0078***	0.0315***	-0.0455***	-0.1481***	-0.5481***	1					
REL_LENGHT (I)	0.0277***	0.0291***	0.0108***	0.0018	0.0757***	-0.2523***	0.0279***	0.0381***	1				
SCOPE (J)	-0.0065***	-0.0126***	-0.0014	0.0099***	-0.0049	0.05***	-0.0712***	0.0481***	-0.0662***	1			
CUM_SCOPE (K)	0.0646***	0.058***	0.0283***	0.0232***	0.034***	-0.0644***	0.1821***	0.0197***	0.3232***	0.0643***	1		
DIST (L)	0.0073***	0.0136***	-0.0007	-0.01***	0.0028	0.0305***	0.0582***	-0.0441***	-0.066***	0.0095***	0.0059***	1	
HHI (M)	-0.014***	-0.0128***	-0.0036*	-0.0053**	0.0007	0.0192***	-0.0021	-0.0068***	-0.0555***	0.0041	-0.0321***	0.02***	1

The table shows pairwise correlation coefficients.

* Correlation coefficient statistically significant at the 10% level

** Statistically significant at the 5% level

*** Statistically significant at the 1% level