Hidden Gems and Borrowers with Dirty Little Secrets: Investment in Soft Information, Borrower Self-selection and Competition

Reint Gropp*
Andre Guettler†

This paper empirically examines the role of soft information in the competitive interaction between relationship and transaction banks. Soft information can be interpreted as a valuable signal about the quality of a firm that is observable to a relationship bank, but not to a transaction bank. We show that borrowers self-select to relationship banks depending on whether their observed soft information is positive or negative. Competition affects the investment in learning the soft information from firms by relationship banks and transaction banks asymmetrically. Relationship banks invest more; transaction banks invest less in soft information, exacerbating the selection effect.

JEL Classification: G21, G28, G32
Key words: soft information, discretionary lending, relationship lending, competition

---

* Prof. Reint Gropp, PhD, Institute for Economic Research Halle (IWH), Kleine Maerkerstr. 8, 06108 Halle (Saale), Germany, reint.gropp@iwh-halle.de.
† Prof. Dr. Andre Guettler, Ulm University, Institute of Strategic Management and Finance, Helmholtzstr. 22, 89081 Ulm, Germany and Institute for Economic Research Halle (IWH), andre.guettler@uni-ulm.de (corresponding author).

We thank Sumit Agarwal, Patrick Behr, Tobias Berg, Hans Degryse, Dominique Demougin, Fred Furlong, Robert Hauswald, Roman Inderst, Simon Kwan, Steven Ongena, Klaus Schaeck, Marcel Tyrell, Vijay Yerramilli, participants at the 10th annual Corporate Finance Conference at Washington University in St. Louis and the annual meeting of the Swiss Society for Financial Market Research, and seminar participants at the Bangor Business School, EBS Business School, Fed San Francisco, Karlsruhe Institute of Technology, KU Leuven, Maastricht University, National Bank of Poland, University of Konstanz, University of Stirling, Tilburg University, and Ulm University for helpful discussions and comments. We further thank the German Savings Banks Association for providing data and Christian Gruendl for excellent research assistance. Part of this paper was written while Reint Gropp was a Duisenberg Fellow at the ECB and the ECB's hospitality is gratefully acknowledged. A previous version of the paper was entitled "Does Discretion in Lending Increase Bank Risk? Borrower Self-selection and Loan Officer Capture Effects".

This paper empirically examines the competitive interaction between relationship banks and transaction banks. "Relationship banks" establish intense and long-term relations with their borrowers and thereby generate soft, and typically proprietary, information about the borrower that is hard to verify by other parties and subjective by nature (e.g., Stein, 2002). "Transaction banks", in contrast, operate at arm's length with centralized decision rights, base their lending decision on credit scoring models, and do not gather soft information. Their loan officers rely on information that is verifiable by third parties and is largely financial, because only financial information can credibly be communicated along hierarchical lines in large banking organizations. We hence interpret soft information as a signal about the quality of a firm that is observable to a relationship bank, but not to a transaction bank (Inderst and Mueller, 2007).

In this paper, we test three main predictions. First, bank size is related to the amount and direction (positive or negative) of the collected soft information. As predicted by theory (Stein, 2002) and consistent with prior empirical evidence (Cole et al., 2004; Berger et al., 2005; Liberti and Mian, 2009), smaller banks should be more likely to deviate from the financial rating and, hence, use more discretion in lending. Hauswald and Marquez (2006) and Inderst and Mueller (2007) furthermore suggest that firms with positive soft information would tend to self-select to relationship banks, because relationship banks can take the positive signal into account in the lending decision. This creates an adverse selection problem, in which transaction banks end up with borrowers that, on average, are associated with negative soft information. In response, transaction banks may apply a negative adjustment to the rating of their borrowers. However, if they do, even borrowers with slightly negative soft information may be better off obtaining a loan from a relationship bank, resulting in an even worse pool of borrowers with respect to the soft information and so forth. Ultimately, in the absence of any offsetting factor, transaction banks would no longer participate in the market for information-sensitive business loans and some positive net present value firms may no longer receive credit. The selection effect may explain why banks ultimately specialize in either relationship or transaction banking.

Second, theory also predict that there are interaction effects with the degree of interbank competition. In the literature, competition has an important effect on the degree to which banks will invest in gathering soft information (Petersen and Rajan, 1995). The literature is ambiguous as to whether competition would increase or decrease banks' investment in gathering soft information. Boot and Thakor (2000), for example, argue that there are two effects. When competition is introduced, banks' marginal rents from relationship lending are smaller and each bank thus reduces its investments in soft information. However, competition affects a bank's profits from relationship and transaction lending asymmetrically. A relationship orientation helps to partially insulate the bank from pure price competition, so that an increase in competition from other banks hurts a bank's profits from transaction lending more than its profits from relationship lending. Thus, increased competition between banks may encourage them to shift from transaction to relationship lending. Empirically, this effect may result in a greater specialization of banks, where relationship banks increase their investment in soft information further and transaction banks reduce their investment further, exacerbating the selection effect described earlier. However, it is also possible that in more competitive markets, transaction banks also start investing more in soft information, reducing the selection effect (Bharath et al., 2007).

Third, to distinguish the idea that relationship banks are better at discovering hidden gems from other potential explanations for our findings (such as private benefits of loan officers), we examine whether, ex post, the probability of default is higher for firms with positive soft information than for other firms. Hence, we link the ex ante use of hard versus soft information by relationship and transaction banks in the lending decision to the ex post default probability of the borrower.

We use a matched bank-borrower dataset of German savings banks to test these predictions. It allows us to construct a proxy for the case when the non-verifiable soft information about the firm was positive, as opposed to when it was negative, that is consistent across the banks in the sample. In our data, all banks use the same rating algorithm to forecast a firm's probability of default (PD). Therefore, the comparability of ratings across banks is ensured. The soft information is consistently available for a *cross section* of banks that can be classified either as largely transactional or largely relationship oriented. These banks compete with pure transaction banks, such as Deutsche Bank and with pure relationship banks, such as the large number of extremely small cooperative banks in Germany. In addition, we are able to construct a precise measure of interbank competition, because, due to legal restrictions on the geographical scope of their operations, savings banks operate on a local basis only.

We first establish that there is sufficient variation in the degree to which banks use soft information in lending decisions. We find that smaller banks are more likely to deviate from the financial rating and, hence, use more discretion in lending. Smaller banks also exhibit some other characteristics associated with relationship banks in the literature. For example, they entertain significantly longer relationships with their borrowers and they have higher costs per euro loan extended. Hence, we use size as a proxy for being a relationship bank in this paper.

The effect, however, is asymmetric. Firms are more likely to receive an improved rating by relationship banks because of positive soft information. At the same time, transaction banks tend to worsen the rating more frequently because of negative soft information, consistent with a broad downward ratings adjustment and adverse selection. We show that the effect is stronger for firms with weak financials. As in Boot and Thakor (2000) the incremental benefit of soft information is decreasing in borrower quality. For financially risky firms, positive soft information is more important than for firms that are strong based on financials alone. Hence, ex ante, the customers of small banks appear riskier based on financial information alone.

Our results regarding the effect of competition on banks' investment in soft information are also interesting and novel. We find evidence that investment in gathering information is not necessarily reduced by all banks as competition increases. There is evidence of increased specialization: smaller banks invest *more* in gathering soft information from risky borrowers, while larger banks reduce their investment. Hence, the selection effect is more pronounced in competitive markets. The evidence is broadly consistent with the theoretical ideas in Boot and Thakor (2000). We argue that the selection effect is at the root of this: as relationship banks invest more in discovering hidden gems, the likelihood that a customer approaching a transaction bank for a loan is a borrower with a dirty secret increases, exacerbating adverse selection and making small business lending less profitable for transaction banks.

Finally, we find that our measure of soft information helps predicting defaults for financially weak firms, but we can reject the hypothesis that loan officers simply use greater discretion to grant loans to worse customers who provide, for example, greater private benefits to the loan officers. We also provide evidence that firms at large banks are more likely to default conditional on their financial rating and further observable characteristics. This result confirms our selection hypothesis that firms with poor soft information are more likely to select to large banks. Finally, we show that the transaction banks' informational disadvantage is compensated for by greater cost-efficiency in lending.

I. Literature

Our paper builds on a large body of literature on the role of relationships in banking. Relationship lending theory is based on the idea that financial intermediaries have a competitive advantage in the production of information about borrowers (Boyd and Prescott, 1986). In particular, Cole et al. (2004) and Berger et al. (2005) show that smaller banks have stronger borrower relationships than larger banks due to a smaller number of managerial layers between the loan officers and the bank management in small banks (Stein, 2002; Williamson, 1967). Liberti and Mian (2009) and Agarwal and Hauswald (2010b) provide evidence that the greater the hierarchical distance, the smaller the importance of soft information in the process of credit approval.[1] Thus, smaller banks are better at producing soft information on the borrower than larger banks due to their organizational structure.

Most of the previous literature on bank-borrower relationships focused on their implications for the borrowers.[2] Berger and Udell (1995) show that stronger relationships lead to lower collateral requirements and lower interest rates charged. Further, stronger bank-borrower

---

[1] Related evidence is provided in Canales and Nanda (2012) who show that decentralized banks extend larger loans to small firms and in those cases where soft information may be important.
[2] One exception is Gruendl (2011) who analyze the effects of relationship lending on bank risk taking.

relationships may increase the availability of credit for the borrower (Petersen and Rajan, 1994; Berger and Udell, 1995) especially in deteriorating credit conditions (Elsas and Krahnen, 1998; Beck et al., 2017; Bolton et al., 2016). Jiménez and Saurina (2004) show that stronger bank-borrower relationships increase the willingness to lend to riskier borrowers.[3] Sutherland (2017) is using the introduction of a US commercial credit bureau to analyze how credit relationships are affected by information sharing. He finds that the introduction shortens maturities in new loan relationships, and reduces lenders' willingness to provide financial support to their delinquent borrowers.

In much of the previous empirical literature, soft information is not directly observed and instead indirectly approximated. For instance, Cerqueiro et al. (2011) investigate the importance of discretion in loan rate setting. They use a heteroscedastic regression model to see which factors determine the dispersion in banks' loan rates to small and medium-size enterprises (SME).[4] There are four notable exceptions where the authors have access to a direct measure of soft information like this paper.[5] First, Degryse et al. (2011) use very detailed data from *one bank* and show that soft information explains observed loan officer discretion. In addition, soft information is found to be important in determining loan size. Our present paper differs from Degryse et al. (2011) in that we analyze the selection of borrowers to relationship and transaction banks, respectively, because

---

[3] Closer bank-borrower relationships can also create informational monopolies for the bank, which result in hold-up problems and deteriorating loan terms (e.g., Boot, 2000).

[4] Garcia-Appendini (2011) and Agarwal and Hauswald (2010a, 2010b) are further examples of the use of indirect approximations. Darmouni (2016) provides another empirical approach to identify non-public soft information separately from information observable to all banks.

[5] Herpfer (2017) uses a more general approach to identify soft information. He finds that banker fixed effects have higher explanatory power than bank fixed effects. Furthermore, relationship loans are associated with fewer bankruptcies which suggest that personal relationships lead to better informed lending.

we have consistent data on the use of soft information for a *cross section* of banks. Second, Puri et al. (2011) use retail loan applications and find that loan applications that were rejected based on financial credit scoring are more likely to be approved based on soft information in the case of existing borrowers and those of lower credit quality. In this paper, we rather use data on the role of soft information in commercial lending. It is possible that the production of soft information is more important for this type of borrower given the higher degree of information asymmetry between bank and borrower.

Third, Agarwal and Hauswald (2009) use the difference between publicly available credit scores and internal ratings to infer the degree to which soft information was used in a set of loan applications within one bank. They compare online and in-person loan applications, arguing that in the case of online applications, there is no soft information content due to the absence of personal interaction. In line with our paper, they show that firms with lower quality financials tend to apply in person rather than online in order to benefit from the use of soft information, and that transaction loans enjoy lower interest rates, but tend to be available only to higher quality borrowers. Fourth, Brown et al. (2012) analyze the role of loan officer discretion in credit assessment at nine banks. They show that loan officers use their discretion to smooth a borrower's credit rating. However, they argue that the smoothing is unlikely to be driven by soft information.

Our paper also relates to the literature on the relationship between size and risk. Especially in the wake of the financial crisis of 2007/2008, the debate about divestures of banks into smaller operational units in order to reduce the too big to fail (TBTF) concern was prominently pursued. The main focus so far has been on the effect that larger banks increase risk due to moral hazard because of explicit or implicit public guarantees (e.g., Merton, 1977; Bhattacharya et al., 1998).

According to theory, large banks, which are perceived as TBTF, are more likely to be bailed out and therefore have incentives to increase risk. These predictions have been empirically tested by many studies. For instance, Boyd and Runkle (1993) and Gropp et al. (2011) find evidence for a positive correlation between size and risk. In addition, most papers point towards higher failure probabilities at larger banks (e.g., De Nicoló, 2001).

II. Institutional Background

Germany is a well-suited laboratory to study the questions of this paper. The German banking market is almost evenly split between three types of banks: savings banks (the focus of this paper) and regional state banks[6], credit cooperatives, and commercial banks. It is characterized by a low level of concentration with 452 different savings banks (in our observation period), more than 1,000 credit cooperatives, and around 300 privately owned commercial banks. Savings banks thus compete both with banks that can be characterized as transaction banks, such as the large commercial banks (Deutsche Bank, Commerzbank), and with banks that are pure relationship banks, such as the cooperative banks.

Small savings banks typically have only one or two branches and flat hierarchies and seem very likely to be able to assign a large amount of discretion to loan officers, while large savings banks may operate much like transaction banks with numerous branches, many layers of hierarchy

[6] Each savings bank is affiliated with one regional state bank ("Landesbank") and each state bank is affiliated with one of the German states ("Länder") or a group of states. The state banks facilitate the transfer of liquidity from savings banks with excess liquidity to those with liquidity shortfalls. In addition, the state banks secure market funding through the issuance of bonds. For an in-depth description of the German banking market, see Hackethal (2004).

and centralized decision rights. For instance, the smallest savings bank in our sample had total assets of EUR 102 million, while the largest savings bank had total assets of EUR 30 billion.[7] Hence, we feel we have sufficient cross sectional variation in the use of soft information in lending decisions to study our question. At the same time, all savings banks that are members of the German Savings Banks Association use the same rating system. As we use the rating system to measure the use of soft information in lending decisions (explained in more detail below), we have a measure that is consistent across all banks in the sample.

Taken as a group, savings banks in Germany have more than EUR 1 trillion in total assets and around 22,000 branches. German savings banks focus on traditional banking business with virtually no off-balance sheet operations. Their main financing sources are customer deposits, which they transform into loans to households and firms. They do not compete with each other, as a regional separation applies: each savings bank uniquely serves its local market (similar to the geographic banking restrictions that existed up to the 1990s in the U.S.). This will be important in our empirical analysis, as it allows us to treat bank size as a proxy for relationship lending that is exogenous to the bank. The regional separation also implies that our firms are linked to one savings bank only, i.e., none of the firms in our sample borrow from two (or more) savings banks nor switch between savings banks over time.

The savings banks in our sample are on average relatively profitable in the observation period 2002-2006: the average pre-tax ROE is 8.9% while the average cost to income ratio is 80.6%. Notwithstanding the differences in governance, savings banks appear very similar to

[7] We further document the marked cross sectional bank size variation in Panel B of Table A1 in the appendix. Table 1 provides the data definitions and sources for all variables employed.

private commercial banks of comparable size in continental Europe. The pre-tax ROE of commercial banks is 9.8% in continental Europe and 8.2% in the UK (186 small banks, 2002-2004, data from Bankscope). Similarly, cost to income ratios are 81.6% in continental Europe and 70.6% in the UK. Overall, German savings banks look like a fairly typical set of commercial banks in continental Europe.

III. Data

A. Matching Bank and Borrower Information

Our main dataset consists of matched bank-borrower information. We start with a dataset of commercial borrowers from the savings banks. It provides annual balance sheets and income statements of all commercial loan customers of the German savings banks affiliated with the German Savings Banks Association. The borrowers are largely SMEs.[8] Given their relatively small size, they strongly rely on bank loans. In contrast to some of the previous literature (e.g., Agarwal and Hauswald, 2009; Puri et al., 2011), our unit of observation is the *borrower*, rather than a loan facility.[9] This corresponds to the rating information described below, as banks rate their customers ("issuer rating"), rather than individual loan facilities ("issue rating").

---

[8] These firms are typically not listed given their moderate size and the savings banks' business model. Hence, available information is sparser than for publicly traded firms. See for instance, Schenone (2010) who provide empirical evidence that a firm's IPO reduces the relationship bank's information advantage.

[9] Note that we do not have information on loan applications or booked loans. The latter implies that we do not have loan-level data on interest rates or maturity.

This dataset's unique feature is its hard and soft information for each loan customer. Specifically, the data set contains 77,364 subcomponents of credit ratings for the years 2002-2006 of 60,696 borrowers as of the end of the respective calendar year.[10] The rating information are based on an internal and proprietary rating algorithm and aim to forecast the firms' PD. All savings banks use the same rating algorithm. It produces a score from 1 to 21, where 1 equals AAA, 2 equals AA+, etc. and the last three rating classes indicate different states of default. Thus, the higher the numerical rating, the riskier is the borrower. The rating information that we use in this paper comprise *two* components. The first is a financial rating that incorporates hard financial statement information on the borrower. The second component, so called combined rating, also incorporates qualitative information. While the financial rating component may be subject to manual adjustments ("overrides") to incorporate firm specific information that is otherwise not captured by the rating system, there are no overrides for the qualitative rating assessment. The latter reflects the subjective and non-verifiable nature of the qualitative rating component.[11] Note that we do not use the *final* credit rating, which is based on the hard and soft information but further adds the information of red flags (such as un-tolerated overdrafts) and the influence of guarantees or letters of support by parent companies and the negative influence of holding structures (e.g., because of profit transfer agreements). The final credit rating is also subject to overrides by bank internal control units (four-eyes principle).[12]

[10] Our observation period starts in 2002 because a new, Basel II legitimate, rating system was introduced in that year by all German savings banks. Given our focus on the competitive interaction of banks and in order to avoid adding unnecessary noise to the analysis, we chose not to extend the data to the financial crisis starting in 2007.
[11] That is not to say that loan officers never attempt to manipulate hard information (see Berg et al., 2016).
[12] See the discussion and the mixed empirical evidence in Berg (2015) and Brown et al. (2015).

The difference between the financial rating and the combined rating reveals the non-financial soft information on the borrower that was used in the lending decision, such as management quality, the firm's strategy, and perceived product or service quality. We interpret this difference as a signal that the borrower can send to a relationship bank but not to a transaction bank. Depending on whether the deviation from the financial rating is negative or positive, i.e., the final rating is higher or lower than the combined rating that also includes qualitative information, we use this as a proxy for a firm's information, which is only observable to the bank to the degree it invests in gathering soft information.

We construct five different variables based on the rating information: i) the absolute difference between the financial and the combined rating; ii) the probability of observing a better combined rating compared to the financial rating (*GoodSoft*); iii) the probability of observing a worse combined rating compared to the financial rating (*BadSoft*); iv) the strength of the rating difference in numerical rating notches in the case of observing positive soft information; v) the strength of the rating difference in numerical rating notches in case of observing negative soft information. Hence, in the empirical analysis below we can distinguish between rating adjustments based on positive and negative soft information. This enables us to explicitly test for borrower selection based on soft information. Our dataset also enables us to link the use of soft information with ex post defaults and to check for biases in the use of soft information by loan officers.

In principle, the difference between the financial rating and the combined rating that also incorporates qualitative information may reflect three different items (Degryse et al., 2011): (i) private hard information from the transaction accounts of the firm and its owner. This information is not publicly observable, but verifiable by senior management. (ii) Soft information that is not verifiable by senior management. (iii) Loan officer discretion. In the following we assume that relationship banks and transaction banks do not differ in the ability to take (i) into account and use the terms “soft information” and “discretion” interchangeably. This approach is supported by the findings in Degryse et al. (2011), who show for very detailed borrower information from one bank in Argentina that only non-verifiable soft information but not verifiable hard information guide loan officer discretion. Even though all savings banks use the same rating system, our empirical approach cannot rule out that different banks use the rating system differently. Any such systematic deviation between small and large banks that is not captured by our control variables may thus affect our results.

Merging borrower level with the bank level dataset comes at a cost: in order to ensure some degree of anonymity of customers, the matching of borrowers to savings banks is possible only aggregated in groups of 5-12 savings banks. In total, there are 62 savings bank groups with rating data available. The aggregation was done by the Savings Banks Association and savings banks of the same region were lumped together, except that larger savings banks were put into “large bank” groups. This approach helps in preserving enough heterogeneity with respect to bank size.[13] Hence, while we have precise information on the individual bank and on the individual customer, we only know that the customer banked with one of the banks belonging to a particular group, but not exactly which bank.

In the previous literature, bank size is found to be a good indicator for tighter bank-borrower relationships (Cole et al., 2004; Berger et al., 2005). Berger et al. (2005) show that large

[13] Panel C of Table A1 in the appendix shows within bank group variation in bank size. We document that savings banks with total assets of beyond EUR 7 billion are all put into the fourth bank group size quartile.

banks tend to approve or reject loan applications primarily via credit scores that are purely based on financial information. Potential soft information on the borrower is not taken into consideration. Liberti and Mian (2009) argue that if the number of hierarchy levels between the loan officer and the management is larger, decisions based on soft information are overruled more often in management decisions, and loan officers thus have fewer incentives to gather soft information. We show below that small and large banks also differ in a number of other dimensions that are consistent with our labeling of small banks as relationship banks and large banks as transaction banks, including the length of the relationship with customers and number of branches per euro of loans. Hence, we use bank size as a proxy for relationship bank. We construct a *Large bank* dummy that equals 1 if the average bank assets per group of savings banks is in the largest quartile. Assets are very common in the literature and well-suited as they are relatively stable and not as much affected by the business cycle as a bank's revenues or profits. In the robustness section, we also report results for two further bank size proxies: the average number of bank branches and the average number of bank employees.

A further variable of particular interest in this paper is competition. Because savings banks are regionally restricted in their operations, we can control for the regional level of bank competition (Boyd and De Nicoló, 2005) by using the ratio of branches of direct competitors (commercial banks and cooperative banks) to savings bank branches per group of savings banks and year in their region.

We use regional variables to control for cross sectional differences in loan demand that may be related to cross sectional differences in the use of soft information. The number of mergers per savings bank group per year is used to control for weakening bank-borrower relationships in the wake of a merger (Di Patti and Gobbi, 2007).[14] We also employ the average debt per capita of the community where the savings bank is located to control for heterogeneity in local public finances. We further control for the regional GDP per capita. In the specifications below we capture further time-variant regional differences by including state-year fixed effects. We also use industry-year fixed effects to control for time-varying regional differences in industry structure that may affect the regional prevalence of the use of soft information.

An important advantage of the dataset is the possibility to relate the *ex ante financial risk* and *ex post defaults* of the banks' commercial loan customers. Besides the financial rating described above we use an Altman-type (1968) Z-Score for a borrower's ex ante financial risk. We use the Z-Score in addition to the financial rating, in order to avoid potential biases in the financial rating. For all commercial loan customers in the data we also have an ex post default measure, which equals 1 if the firm repaid principal or interest more than 90 days late in the calendar year after the year-end credit rating was recorded and 0 otherwise. We also control for borrower size, as Stanton (2002) shows that managers are more efficient in monitoring a few large loans than many small ones. We calculate the length of the bank-firm relationship, as relationship banks gather more soft information over time. Furthermore, we use the borrowers' legal form to distinguish between closely held firms and incorporated firms, as closely held firms have much lower accounting and transparency standards. We use a dummy variable, *Opaque borrower*, that equals 1 for the former and 0 for the latter type of firms. Finally, we use commission payments

[14] However, Berger et al. (1998) provide evidence that reduced small business lending is offset by the reactions of other banks.

divided by total wage expenditures to proxy for the variable bonus payments to loan officers in order to control for potential heterogeneity with respect to bank size.

B. Descriptive Statistics

Panel A of Table 2 provides descriptive statistics for the main variables. We first discuss variables which are at the borrower level. The average absolute change in rating, based on soft information on the borrower, is around 2 notches, which indicates a significant influence of soft information on the final rating decision. Good soft information (*GoodSoft*), i.e., cases where the final rating indicates a lower risk due to soft information than the combined rating, are observed with a frequency of 24.5% and have an average magnitude of 2.48 numerical rating notches. Bad soft information is more frequently observed with 60% and on average slightly less strong with 2.37 notches. The rating remains unchanged for 15% of the borrowers.

On first sight, it may seem surprising that bad soft information is more common than good soft information given that we only observe firms that had an active lending relationship with savings banks. Sufficiently bad soft information could result in rejected loan applications which we cannot observe in the data. We cannot rule out such kind of sample selection but given the descriptive statistics, it does not seem to play a major role. There are also (at least) two reasonable explanations for the high fraction of negative soft information cases: i) the financial rating component is not well calibrated (i.e., it is tilted towards too positive ratings); ii) the loan officers are too cautious. The latter argument seems valid if reputation is reduced by a larger extent for a wrong soft information assessment (i.e., a firm with positive soft information defaults ex post) than reputational gain due to a correct soft information assessment (i.e., a firm with positive soft information does not default ex post). The former situation outweighs the latter by a large degree and may thus yield rather cautious usage of soft information. We further elaborate on this issue in Section IV.D.

The median assets of the borrowers are EUR 515 thousand. More than 50% of the borrowers are classified as being opaque. Both demonstrate that the savings banks mostly engage in SME lending. The average length of the bank-borrower relationship is 3.7 years. The average Z-Score for the borrower is 3.40 while the average financial rating is 12.40 (corresponding to a long-term credit rating of BB). Both measures approximate financial risk from an ex ante perspective. On average, 4.7% of the borrowers in our sample default in the calendar year following the rating record.

Next we show bank group descriptive statistics. The size of the average bank group is EUR 3.1 billion. The dispersion of bank size is large. The 95th percentile of the bank assets is more than 14 times the 5th percentile (see also the discussion below and Table A1 in the appendix). This pattern suggests that we may have sufficient within-savings bank variation in the use of soft information to assume that bank-borrower relations are of different strength, but below we formally test for whether large savings banks are less likely than small ones to adjust the ratings of their customers based on soft information. The number of direct competitors is less than one on average, indicating about one branch of a competitor for each savings bank branch. Merger activity was extensive during our sample period. On average, the savings bank groups were involved in a merger every third year.

IV. Results

A. Borrower Self-selection

As a first step, we check whether large savings banks deviate less from financial ratings than small savings banks. Such a finding would suggest that we indeed observe cross sectional variation in the use of soft information in our sample. We present univariate results in Panel B of Table 2. We split the borrowers according to their bank groups' average assets at the 75th percentile: small banks consist of the first three quartiles and large banks of the top quartile. The last column shows the significance levels of univariate regressions that test for differences of small versus large savings banks. We find that the average absolute difference between the financial rating and the combined rating, |Δ Rating|, is significantly smaller for large than for small savings banks. Larger banks thus seem to deviate less from the financial rating than smaller banks. This is consistent with the previous literature that larger banks produce less soft information (Berger et al., 2005; Uchida et al., 2012). As expected, other proxies for relationship intensity are highly correlated with bank size. Relationship length with borrowers is about half a year shorter for large banks, large banks spend less resources per euro loan extended, and large banks are less likely to lend to opaque and smaller firms.

Most importantly for the purposes of this paper, the use of soft information is not symmetric for cases of good and bad soft information. Observing good soft information is 4.4% less likely for large than for small savings banks, which accounts for around 18% of the unconditional upgrade likelihood (see Panel A of Table 2). In addition, when small savings banks improve the financial rating, it is done by significantly more rating notches. Observing bad soft information is more likely for large than for small savings banks. Hence, we obtain first evidence for the hypothesis that borrowers with positive soft information self-select to smaller and more relationship-oriented banks that are more likely to take this information into account.

Encouraged by the univariate results we test variants of the following specification:

$$Y_{it} = \alpha_{rt} + \alpha_{jt} + \beta * Large\ bank_{bt} + \gamma X + e_{it} \quad (1)$$

where Y_{it} denotes one of the five measures for soft information on borrower i at year t,[15] α_{rt} and α_{jt} denote state-year and industry-year fixed effects. Industry-year fixed effects are important, because they address the concern that small banks might operate in regions where the industrial structure is tilted towards firms where soft information is more important. State-year fixed effects control for time-variant shocks on the state level. X represents a vector of borrower, bank group, and regional characteristics. Specifically, we control for the size of the borrower, whether the borrower is opaque or not, the borrowers' Z-Score and financial rating, and the relationship length between the firm and the savings bank in years. Standard errors are clustered at the bank group level since we expect the discretion in lending and the quality of loans to be correlated with geographic location.

The variable of interest is the dummy $Large\ bank_{bt}$ that takes the value 1 for the largest bank size quartile. This dummy approximates the likelihood that a savings bank is a transaction bank. Table A1 in the appendix shows that bank size largely remains unchanged over time (Panel A), while we observe substantial cross sectional variation (Panel B). Following Zhou (2001) we

[15] Note that our state and industry information are time-invariant, i.e., each firm remains in its industry and its state during our sample period.

do not use bank fixed effects because in the absence of significant time series variation we run the risk of estimating spurious effects of bank size on soft information.[16]

We prefer size as our proxy for a relationship bank relative to other proxies used in the literature (e.g., distance between bank branch and borrower), because in the institutional setting of savings banks in Germany size is largely exogenous. By law savings banks are limited to lending within their region. Hence, their size is predominantly determined by the size of the community they reside in. Moreover, they are not permitted to take over other non-savings banks and cannot be taken over by non-savings banks. Within the savings bank sector, there are mergers, but in all cases they are due to the financial difficulties of the savings banks involved, rather than strategic growth considerations.[17]

Panel A of Table 3 shows regression results with the five different measures for discretion in lending as dependent variables and the *Large bank* dummy as the main independent variable of interest.[18] The first column of Panel A shows that |Δ Rating| tends to be slightly higher for large banks though it is not statistically significant. As in the case of the univariate results, the total effect is not symmetric for cases of good and bad soft information. Column 2 shows that larger banks seem to be 1.6% less likely to improve the financial rating based on positive soft information.[19] The point estimate is statistically significant and also economically substantial,

---

[16] The argument is that the size effects would be identified from very small time series variations that may in part be random or due to mismeasurement. In a robustness check, we follow Zhou (2001) and estimate a specification that uses averages across time for each bank group. We find coefficients of interest that are robust in economic magnitude and sign, but are statistically insignificant, which we attribute to the small sample size (N = 62). Results are available from the authors upon request.
[17] For further institutional detail on this particular point, see Bian et al. (2017).
[18] We use OLS models throughout, since differences compared with using Probit models for the binary dependent variables in columns two and three are negligible.
[19] It is possible that loan officers at small banks use their discretion inefficiently to upgrade borrowers that provide loan officers with greater private benefits. This is investigated below.

since it accounts for around 6.7% of the unconditional upgrade likelihood (see Panel A of Table 2). If large banks improve the financial rating based on positive soft information they do so by fewer rating notches than small banks (column 3). In addition, larger banks are more likely to use soft information to downgrade a borrower's financial rating based on negative soft information (column 3) and they do so by more notches compared to small banks (column 5). We thus find evidence in favor of the selection hypothesis: borrowers with positive soft information are more likely to obtain a loan from small relationship lenders, while borrowers with negative soft information are less likely to receive credit from smaller banks. We further interpret the finding that larger banks tend to downgrade borrowers more often as evidence that large banks attempt to take the selection effect into account by downgrading borrowers across the board.

B. Financially Riskier Borrowers

If firms with better soft information self-select towards smaller banks that are more likely to take soft information into account, is this effect stronger for firms with weak financials? For firms with weak financials it should be particularly valuable if positive soft information is taken into account in the lending decision. We measure the extent of positive soft information by the probability to observe an improved combined rating compared to the financial rating (*GoodSoft*). As a measure of the financial risk of a borrower we use the borrowers' financial rating which is strictly limited to financial characteristics and does not include soft information.[20]

[20] We furthermore use the borrowers' Z-Score as an additional control variable for financial risk. Although not explicitly shown in the table, the Z-Score enters the regression significantly at the 1% level.

Table 4 shows the regression results.[21] We regress *GoodSoft* on borrower risk, bank size, local competition, and the controls. We form interaction terms to capture the bank size-borrower risk relationship that we discovered in the univariate analysis. Specifically, the dummy variable *Risky borrower* equals 1 for borrowers in the riskiest financial rating quartile.

Column I shows the individual effects for firm financial risk without interaction terms. We find that riskier borrowers are more likely to be upgraded due to positive soft information. The specification of column II also includes the interaction term *Large bank * Risky borrower*. We find that larger banks are 8.9% less likely to improve the financial rating of ex ante financially risky borrowers. Note again that the unconditional probability of receiving a rating upgrade is 24.5% (see Panel A of Table 2), thus, our point estimate of 8.9% also seems to be economically significant. This result is in line with the idea that riskier borrowers (based on financial characteristics) who have substantial positive soft information have much higher chance of obtaining a loan from a bank that takes the soft information into account.

Selection would also predict that large banks should be more likely to adjust the rating of borrowers downward, because they are concerned that the borrowers with negative soft information are particularly likely to apply for a loan at a large bank. We expect this effect to be particularly strong for financially risky borrowers. Unreported results that are available from the authors upon request show that this is indeed the case: riskier borrowers are more likely to be downgraded by a large bank than by a small bank.

[21] We again use OLS models since differences compared with using Probit models for the binary dependent variables are negligible.

Overall, we find a striking asymmetry: Small, relationship banks are more likely to adjust the credit rating of borrowers upward. We observe the exact opposite for large, transaction banks, who are more likely to adjust the credit rating of borrowers downward. Both of these effects are especially operative for firms that are financially risky. Hence, the investment in soft information by relationship banks is important for financially risky firms that have positive soft information only observable to the relationship bank, as they, in the absence of relationship banks, may be unable to obtain credit at all or at the same terms.

C. Borrower Self-selection and Competition

We next check whether the selection effect and the differential investment in soft information are related to local competition. Hauswald and Marquez (2006) argue that banks will invest less in the acquisition of (soft) information in more competitive markets, because they have to share rents with the borrowers. In contrast, the predictions in Boot and Thakor (2000) are more differentiated: while they acknowledge the existence of the effect of increased rent sharing with borrowers, they also argue that banks may invest more in gathering soft information in competitive markets in order to avoid direct price competition. We investigate these arguments in our sample. We use data from the Bundesbank on the number of branches of competitor banks per savings bank branch in the savings bank's local market and define a dummy variable, *High competition*, that takes on the value of 1 if the bank operates in a market that is above the median of 0.823 branches per savings bank branch. We then interpret a negative relationship between higher competition and *GoodSoft* as a reduction in the banks' investment in the generation of soft information.

In the first column of Table 4 we document an overall tendency to reduce investment in soft information in more competitive markets, which is consistent with Hauswald and Marquez (2006).[22] We obtain a negative coefficient on the individual *High competition* dummy variable. If banks invest less in information acquisition in more competitive markets, this may suggest that financially risky firms with positive soft information may no longer be able to obtain credit in these markets. We investigate this issue in more detail by including two-way interaction terms between the competition level and bank size, and, in separate regressions, with borrower risk. First, we concentrate on the differential effect with respect to bank size. In column III we find that larger banks invest less in information acquisition in more competitive markets.

Second, we focus on the differential effect with respect to the borrowers' ex ante risk level. In column IV of Table 4*,* we show results in which we analyze whether banks maintain their investment in soft information for financially riskier borrowers. Using the financial rating as risk measure, we find that banks are slightly more likely to upgrade riskier borrowers in competitive markets even though the interaction term does not enter the regression significantly. In the last column of Table 4 we analyze a bank's investments in soft information in more competitive markets by using three-way interaction terms between competition, the financial risk of the borrower, and the size of the bank. We find that large banks are 15.8% less likely to improve the financial rating based on positive soft information for ex ante financially riskier firms in competitive markets. The point estimate is not only statistically significant at the 1% level but it

[22] Agarwal and Hauswald (2010b) show in a different context that banks strategically use soft information in response to local competition

also accounts for around two thirds of the unconditional upgrade likelihood (see Panel A of Table 2).

These results support the idea that, overall, the generation of soft information is reduced in more competitive markets. However, we also find evidence in favor of specialization in more competitive markets: larger banks reduce their investment in information, while small banks do not. Hence, the selection effect of ex ante financially riskier borrowers selecting towards relationship banks is even more pronounced in more competitive markets.

D. Ex Post Credit Outcomes

Relationship banks lend to borrowers that exhibit ex ante weaker financial characteristics. However, these borrowers tend to be upgraded based on positive soft information that transaction banks are unable to use. We now examine whether this use of soft information results in overall riskier outcomes ex post. Clearly, if banks use the soft information in an unbiased way, the customers with ex ante weaker financial information may not necessarily exhibit higher probabilities of default ex post. On the other hand, if loan officers use their discretion to provide loans to borrowers that entail a private benefit to themselves or are otherwise in some way "captured" by their customers, these borrowers will show a higher ex post default frequency compared to other borrowers. In order to differentiate between the two possibilities, we regress the default outcome of the borrower, which is either 1 in the case of a default in the following calendar year after the year-end rating was recorded or 0 otherwise, on our proxies for the use of soft information.

Table 5 shows results for this exercise. Glancing at the results in the table as a whole, most soft information proxies tend to obtain significant coefficients, which suggests that soft information seems to matter for predicting borrower default, even conditioning on financial information. This finding is consistent with the previous literature (Degryse et al., 2011; Grunert et al., 2005).

In column I we see that |Δ Rating| obtains a coefficient close to zero, suggesting that if loan officers deviate from judging based on financial information alone, i.e., use soft information in their decision, it is not clear whether a firm is more or less likely to default ex post. In column II, we distinguish between upgrades and downgrades. This permits a distinction between higher defaults because loan officers improved a firm's financial rating too much based on positive soft information and higher defaults because loan officers deteriorated a firm's financial rating too little based on negative soft information. It turns out that if positive soft information was observed at a firm it is as likely to default as a borrower whose rating was not changed due to soft information (the coefficient is -0.31% and insignificant). In contrast, if negative soft information was detected at a firms it is 0.66% *more* likely to default (significant at the 1% level) relative to firms that received a loan purely based on financial information. If we compare both type of firms, we find that the latter are 0.97% more likely to default relative to the former, controlling, as before, for the financial rating and the host of other covariates. These results are also economically significant given that the average unconditional default frequency is 4.7% (see Panel A of Table 2).

A similar picture emerges from the regression where we consider the strength of the rating improvement and the strength of the rating deterioration, given a firm's financial rating was improved or worsened based on soft information, respectively (columns III and IV of Table 5). Firms that received a higher improvement (by more notches in the rating system) were significantly *less* likely to default (by -0.3%) and firms that received a stronger downgrade were significantly *more* likely to default (by 0.3%). These results indicate that banks are too cautious in using soft information to adjust their view on a borrower's credit risk based on financial characteristics, but we do not find evidence of loan officers using discretion to improve the firms' financial rating in order to obtain a private benefit.

In columns V to VIII of Table 5 we analyze whether the relation between soft information and default is stronger for borrowers with riskier financials. Specifically, we include interaction terms between the soft information proxy used and a dummy variable that equals 1 if the borrowers' financial rating is in the riskiest quartile. The evidence is consistent with banks investing more in soft information where the pay-off may be greatest: financially risky borrowers. In column VI we compare the downgrade probability of financially risky borrowers that received an improved combined rating based on positive soft information (-1.16%) with financially risky borrowers that received a worsened rating based on negative soft information (-0.35%); the difference is -0.81% and highly significant. Banks thus appear to invest in generating soft information about borrowers where the pay-off is largest, namely borrowers that have ex ante weak financial characteristics.

In further unreported regressions, we check the robustness of our results with subsamples for which two-year and three-year risk outcome measures (the maximum we can go with our data) are available. We are interested in whether banks grant credit to their financially weakest borrowers in order to delay their default and the realization of losses ("evergreening", e.g., Peek and Rosengren, 2005). In the text we rely on the idea that a biased use of soft information may at least in part become visible over longer horizons. We do not observe this in the data. Over a two-year and a three-year horizon we continue to find that upgraded borrowers are significantly less likely to default than downgraded borrowers, i.e., no evidence of evergreening. We also test for potential reverse causality in the relationship between discretion and the probability of default (Degryse et al., 2011). If banks use soft information to predict defaults, defaults may become less likely if upgrades based on soft information improve loan terms such as interest rates and maturity. We take the following approach to the problem: We restrict the sample to firms with multiple lenders. In this case the savings bank is only one of several lenders and reverse causality should be less of an issue. We find qualitatively unchanged results for this subsample and, hence, reverse causality seems to play no role in explaining our findings of Table 5.

E. Positive-correlation Test for Self-selection

So far, we have shown that small banks appear to use soft information more extensively (see Table 3) and that soft information seem to predict default rates (see Table 5). We next complement these results with a more direct test for self-selection which uses a positive-correlation test that is frequently applied in the insurance literature (Finkelstein and McGarry, 2006).

We regress the default dummy on the financial rating that only contains hard information and the large bank dummy. We further control for other observable characteristics available to the bank. We expect to observe higher default rates for borrowers at large banks conditional on the financial rating and further observable characteristics. The results are shown in Table 6.

Column I provides the univariate results without controlling for observable characteristics. We find no evidence for a positive correlation between bank size and default rates. In columns II to V, we further add the observable characteristics used in the previous tests. In column II, we find that the large bank dummy is positive and significant at the 10% level. Conditional on the financial rating and other observables, borrowers at large banks are 0.78 percentage points more likely to default. This result is consistent with the notion that firms with negative soft information self-select to large banks because these banks invest less in soft information collection. These results further confirm our previous findings that small banks use soft information more extensively and that soft information predict default rates. Our results are economically significant given the unconditional default frequency of 4.7% (see Panel A of Table 2).

In column III, we interact the financial rating with the previously used *Risky borrower* dummy variable. Albeit this interaction is not entering the regression significantly, we find a stronger bank size effect for financially sound firms (significant at the 5% level). In column IV, we replace the financial rating with dummy variables for five rating buckets to capture the non-linear relationship between the rating and the default frequency. Results are robust to this adjustment. In column V, we interact the financial rating dummies with the *Risky borrower* dummy and find the most pronounced *Large bank* coefficient (0.93 percentage points).

Overall, we provide evidence that firms at large banks are more likely to default conditional on their financial rating and further observable characteristics. This result confirms our selection hypothesis that firms with poor soft information are more likely to select to large banks

V. Robustness and Extensions

We perform a number of robustness checks and extensions. First, we use different measures of bank size. Second, we check the relationship between savings bank (group) size and the level of local bank competition in order to address the concern that large savings banks operate in more competitive environments than small savings banks. We also address concerns that areas with high and low competition may be different in many aspects, and that these differences drive our results.

Third, we check whether larger banks indeed have lower costs per loan. Fourth, we analyze the impact of soft information on the price of credit. Fifth, we check whether soft information is more important for opaque firms. Finally, we analyze whether large banks were granting larger loans to firms.

We first verify whether the selection result is robust to using different measures of bank size. For space limitations, we do not report these regressions, but they are available from the authors upon request. We find that using the number of bank branches or the number of bank employees yields qualitatively similar results to those described in Section IV. Our results also remain qualitatively unchanged if we use the natural logarithm of bank assets instead of the *Large bank* dummy variable.

Further, we analyze the relationship between savings bank (group) size and the level of local bank competition to address concerns that our results may be driven by omitted regional characteristics. We apply a double sort on bank group size (quartiles) and the level of local bank competition (below and above the median). Table A2 in the appendix shows that there is no evidence that larger savings banks operate in more competitive markets. In particular, the largest banks are located in banking markets that exhibit below-average competition.

We also address concerns that areas with high and low competition are different in many aspects. We hence add, one-by-one, the available covariates used in Table 4 with the additional two-way interaction terms (with the *Large bank* and *Risky borrower* dummy variables) and the additional three-way interaction term. We find in three out of the eight cases significant effects for the further triple interaction term. However, our triple interaction term of interest ("*Large bank* *

*Risky borrower * High competition*") keeps being economically substantial and always remains statistically significant at the 1% level. Results are available from the authors upon request.

Next, we examine whether there are differences in costs between small and large banks in granting loans. Having an informational advantage by gathering soft information may go hand in hand with higher screening/monitoring costs at relationship banks (Boot and Thakor, 2000; Hauswald and Marquez, 2006). Compared to transaction banks, margins and charter values may be lower at relationship banks, which may result in a greater willingness to accept riskier borrowers (e.g., Keeley, 1990; Hellmann et al., 2000).

We rely on three bank (group) level measures for costs per loan: i) sum of staff costs over average assets per bank group and year (in percent); ii) number of bank branches (in hundreds) over the average assets per bank group (in EUR billion) and year; iii) number of bank FTEs (in thousands) over the average assets per bank group (in EUR billion) and year. Table 7 shows the results for which we regress the three proxies on the *Large bank* dummy variable. This variable of interest enters significantly in the regressions for all three proxies. We find that larger banks have lower staff costs, use fewer branches, and have fewer employees (per unit of assets). This result is consistent with a cost advantage for large banks in screening/monitoring that they may use to offset the informational disadvantage and the associated selection problem.

Additionally, the savings banks make use of broadly the same compensation system for loan officers, which largely rely on fixed-salary contracts.[23] In our dataset, the median commission

[23] Agarwal and Ben-David (2017) show that loan origination-based incentive compensation increases loan origination and the bank's credit risk.

payments over regular staff expenses, which approximate the share of loan officer bonus payments in total salaries, is 2.8% (see Panel A of Table 2).[24] It thus seems very unlikely that our main results are driven by differences in loan officer incentives due to differences in contracts.

Agarwal and Hauswald (2009) show that the trade-off between relationship loans and transaction loans is between loan availability and pricing. Relationship loans are more readily available, but carry a higher interest rate. So far we have not examined pricing, in part because we do not have loan level data, but borrower level data. Hence, we do not have access to loan level interest rate data. However, we are able to calculate a proxy for the average interest rate a firm pays (interest rate expenses / bank credit). Unreported results, available from the authors upon request, show that soft information per se does not impact interest rates charged by savings banks significantly, but that banks, given they adjust the rating based on soft information, adjust interest rates downward in response to positive and upward in response to negative soft information. Next we constructed a matched sample of firms with the same financial rating, representing on the one hand cases where positive soft information was used, and on the other, cases where no soft information was involved in the credit decision. We then compared interest rates across the two groups. Consistent with Agarwal and Hauswald (2009) we find that in the group in which positive soft information was used borrowers paid 8 basis points higher interest. The difference is

[24] The level of variable payments is comparable with lenders in the US. According to survey results from ChaseCompGroup, between 20 and 30 percent of US mortgage lenders with assets below USD 1bn do not pay their loan officers any bonus (Survey results were downloaded on August 17, 2014, from www.chasecompgroup.com/downloads/CCG_MortgageLenderSurveyFindings_Jan2011.pdf). Salary.com shows that the average share of bonuses in total compensation of loan officers in the U.S. is 3.7% (Accessed on August 17, 2014 at http://swz.salary.com/salarywizard/Commercial-Loan-Officer-I-Salary-Details.aspx)

statistically significant using the Mahalanobis nearest-neighbor matching. While statistically significant, it is economically small.

To tackle the incentives to generate soft information from a different angle, we use the firms' legal form to distinguish between more and less opaque borrowers (Berger et al., 2005; Cole et al., 2004). We find that opaque borrowers are more likely to receive a rating upgrade based on positive soft information. However, this effect does not vary by bank size and thus does not constitute an additional selection effect.

Finally, we check whether large banks were granting larger loans. Although we do not know whether large savings banks grant larger loans to firms with low financial ratings, we do observe the firms' total loan amount granted by savings banks. We thus regress the natural logarithm of the firms' total loan amount granted by savings banks on the *Large bank* dummy (and available control variables and the set of fixed effects of Table 3). Unreported results show that the *Large bank* dummy does not enter any of the different specifications significantly (no fixed effects and control variables, with fixed effects, with both). These results hold for the full sample of firms and firms with low financial ratings.

VI. Conclusion

We start from the idea that soft information can be viewed as information about a borrower that is observable to a relationship bank but not to a transaction bank (Inderst and Mueller, 2007), resulting in an adverse selection effect: Firms with positive soft information optimally self-select to relationship banks, firms with negative soft information to transaction banks. Transaction banks therefore face disproportionately many borrowers with negative soft information and adjust their lending behavior accordingly. The interaction between relationship banks and transaction banks is also affected by the competitive environment, which from a theoretical perspective may result in more or less investment in gathering soft information by both types of banks.

We use a matched bank-borrower dataset of German savings banks that has three distinct advantages: First, we observe whether the lender used positive or negative soft information in the lending decision. Second, due to restrictions on the geographic operations of the banks in our sample, we can accurately measure their competitive environment. Third, we have information on ex post borrower defaults and can therefore check whether soft information was used efficiently.

Using these unique data, we are able to uncover two empirical results that to our knowledge have so far not been documented in the literature. First, we find that borrowers with riskier financial characteristics are more likely to obtain credit from relationship banks than from transaction banks if they have positive soft information. In contrast, financially riskier firms with negative soft information are more likely to turn to a transaction bank. This evidence supports adverse selection in the market for small business loans that is different from the Stiglitz and Weiss (1981) type adverse selection the literature has largely focused on (see for instance, Bae and Goyal, 2009). Second, we show that competition affects relationship banks' and transaction banks' investment in gathering soft information asymmetrically. Relationship banks tend to increase their investment in gathering soft information, while transaction banks reduce it. An increase in competition results in a further specialization of banks, but does not necessarily reduce the overall investment in gathering soft information. Hence, interbank competition may not reduce the availability of credit to firms with weak financials but strong soft information, as some previous theory has suggested (e.g., Hauswald and Marquez, 2006). All of these results are more pronounced for firms where soft information can, ex ante, be expected to play a greater role in the lending decision: firms with weak financials.

Further research should maintain the strengths of this paper, i.e., directly observing the soft information component for a large set of banks using the same rating algorithm, and being able to overcome restrictions of our dataset, such as obtaining loan level and loan term information including loan applications, for all banks in a major financial market using individual bank level data. Having access to loan application data would also alleviate concerns about sample selection issues. Ideally, the econometrician should also know the distance between firms and banks, the quantity and quality of firm-bank interactions as well as detailed information about the loan officer handling a credit relationship. These measures would yield further and potentially more robust and valid proxies of the firm-bank relationship.

References

Agarwal, S. and R. Hauswald (2009): The Choice between Arm's-Length and Inside Debt, Unpublished manuscript.

Agarwal, S. and R. Hauswald (2010a): Distance and Private Information in Lending, Review of Financial Studies, 23, 2758-2788.

Agarwal, S. and R. Hauswald (2010b): Authority and Information, Unpublished manuscript.

Agarwal, S. and I. Ben-David (2017): Loan Prospecting and the Loss of Soft Information, Journal of Financial Economics, forthcoming.

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

Bae, K.-H. and V. K. Goyal (2009): Creditor Rights, Enforcement, and Bank Loans, Journal of Finance, 64, 823-860.

Beck, T., H. Degryse, R. De Haas, and N. van Horen (2017): When Arm's Length is Too Far: Relationship Lending over the Credit Cycle, Journal of Financial Economics, forthcoming.

Bian, B., R. Haselmann, T. Kick, and V. Vig (2017): The Political Economy of Bank Bailouts, Unpublished manuscript.

Berg, T., M. Puri, and J. Rocholl (2016): Loan Officer Incentives, Internal Rating Models and Default rates, Unpublished manuscript.

Berg, T. (2015): Playing the Devil's Advocate: The Causal Effect of Risk Management on Loan Quality, Review of Financial Studies, 28, 3367-3406.

Berger, A. N. and G. F. Udell (1995): Relationship Lending and Lines of Credit in Small Firm Finance, Journal of Business, 68, 351-381.

Berger, A. N., A. Saunders, J. M. Scalise, and G. F. Udell (1998): The Effects of Bank Mergers and Acquisitions on Small Business Lending, Journal of Financial Economics, 50, 187-229.

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

Bharath, S., S. Dahiya, A. Saunders and A. Srinivasan (2007): So what do I get? The Bank's View of Lending Relationships, Journal of Financial Economics, 85, 368-419.

Bhattacharya, S., A. W. Boot, and A. V. Thakor (1998): The Economics of Bank Regulation, Journal of Money, Credit and Banking, 30, 745-770.

Bolton, P., X. Freixas, L. Gambacorta, and P. E. Mistrulli (2016): Relationship and Transaction Lending in a Crisis, Review of Financial Studies, 29, 2643-2676.

Boot, A. W. A. and A. V. Thakor (2000): Can Relationship Banking Survive Competition? Journal of Finance, 55, 679-713.

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

Boyd, J. H. and G. De Nicoló (2005): The Theory of Bank Risk Taking and Competition Revisited, Journal of Finance, 60, 1329-1343.

Boyd, J. H. and E. C. Prescott (1986): Financial Intermediary-coalitions, Journal of Economic Theory, 38, 211-232.

Boyd, J. H. and D. E. Runkle (1993): Size and Performance of Banking Firms: Testing the Predictions of Theory, Journal of Monetary Economics, 31, 47-67.

Brown, M., M. Schaller, S. Westerfeld, and M. Heusler (2012): Information or Insurance? On the Role of Loan Officer Discretion in Credit Assessment, Unpublished manuscript.

Brown, M., M. Schaller, S. Westerfeld, and M. Heusler (2015): Internal Control and Strategic Communication within Firms. Evidence from Bank Lending, Unpublished manuscript.

Canales, R. and R. Nanda (2012): A Darker Side to Decentralized Banks: Market Power and Credit Rationing in SME Lending, Journal of Financial Economics, 105, 353-366.

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

Cole, R. A., L. G. Goldberg, and L. J. 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, 227-251.

Darmouni, O. (2016): Estimating Informational Frictions in Sticky Relationships, Unpublished manuscript.

De Nicoló, G. (2001): Size, Charter Value and Risk in Banking: An International Perspective, in the Financial Safety Net: Costs, Benefits and Implications for Regulation, Proceedings of the 37th Annual Conference on Bank Structure and Competition, Federal Reserve of Chicago, 197-215.

Degryse, H., J. Liberti, T. Mosk, and S. Ongena (2011): Is Loan Officer Discretion Advised When Viewing Soft Information? Unpublished manuscript.

Di Patti, E. B. and G. Gobbi (2007): Winners or Losers? The Effects of Banking Consolidation on Corporate Borrowers, Journal of Finance, 62, 669-695.

Elsas, R. and J. P. Krahnen (1998): Is Relationship Lending Special? Evidence from Credit-file Data in Germany, Journal of Banking and Finance, 22, 1283-1316.

Engelmann, B., E. Hayden, and D. Tasche (2003): Testing Rating Accuracy, Risk, January, 82-86.

Finkelstein, A. and K. McGarry (2006): Multiple Dimensions of Private Information: Evidence from the Long-Term Care Insurance Market, American Economic Review, 96, 938-958.

Garcia-Appendini, E. (2011): Lending to Small Businesses: The Value of Soft Information, Unpublished manuscript.

Gruendl, C. (2011): The Effects of Relationship Lending on Bank Risk Taking: Evidence from the German Savings Banks, Unpublished manuscript.

Gropp, R., H. Hakenes, and I. Schnabel (2011): Competition, Risk-Shifting, and Public Bail-out Policies, Review of Financial Studies, 24, 2084-2120.

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

Hackethal, A. (2004): German Banks and Banking Structure, in The German Financial System, ed. by J. P. Krahnen and R. H. Schmidt, Oxford University Press, 71-105.

Hauswald, R. and R. Marquez (2006): Competition and Strategic Information Acquisition in Credit Markets, Review of Financial Studies, 19, 967-1000.

Hellmann, T. F., K. C. Murdock, and J. E. Stiglitz (2000): Liberalization, Moral Hazard in Banking, and Prudential Regulation: Are Capital Requirements Enough? American Economic Review, 90, 147-165.

Herpfer, C. (2017): The Role of Bankers in the U.S. Syndicated Loan Market, Unpublished manuscript.

Inderst, R. and H. M. Mueller (2007): A Lender-based Theory of Collateral, Journal of Financial Economics, 84, 826-859.

Jiménez, G. and J. Saurina (2004): Collateral, Type of Lender and Relationship Banking as Determinants of Credit Risk, Journal of Banking and Finance, 28, 2191-2212.

Keeley, M. C. (1990): Deposit Insurance, Risk, and Market Power in Banking, American Economic Review, 80, 1183-1200.

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

Merton, R. C. (1977): An Analytic Derivation of the Cost of Deposit Insurance and Loan Guarantees: An Application of Modern Option Pricing Theory, Journal of Banking and Finance, 1, 3-11.

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

Petersen, M. A. and R. G. Rajan (1995): The Effect of Credit Market Competition on Lending Relations, Quarterly Journal of Economics, 110, 407-443.

Peek, J. and E. Rosengren (2005): Unnatural Selection: Perverse Incentives and the Misallocation of Credit in Japan, American Economic Review, 95, 1144-66.

Puri, M., J. Rocholl, and S. Steffen (2011): Rules Versus Discretion in Bank Lending Decisions, Unpublished manuscript.

Schenone, C. (2010): Lending Relationships and Information Rents: Do Banks Exploit Their Information Advantages? Review of Financial Studies 23, 1149-1199.

Stanton, K. R. (2002): Trends in Relationship Lending and Factors Affecting Relationship Lending Efficiency, Journal of Banking and Finance, 26, 127-152.

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

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

Sutherland, A. (2017): The Economic Consequences of Borrower Information Sharing: Relationship Dynamics and Investment, Unpublished manuscript.

Uchida, H., G. F. Udell, and N. Yamori (2012): Loan Officers and Relationship Lending to SMEs, Journal of Financial Intermediation, 21, 97-122.

Williamson, O. E. (1967): Hierarchical Control and Optimum Firm Size, Journal of Political Economy, 75, 123-179.

Zhou, X. (2001), Understanding the Determinants of Managerial Ownership and the Link between Ownership and Performance: Comment. Journal of Financial Economics 62, 559-571.

Table 1: Definition of variables

The table gives the definitions of all variables used in the empirical analysis. The savings bank data was provided by the Savings Banks Association. There are seven savings banks in Germany that are not full members in the Savings Banks Association. These banks are not covered in the dataset. Destatis is the Federal Statistical Office of Germany and Bundesbank is the German central bank.

Variable name	Description	Data source
Dependent variables		
|Δ Rating|	Absolute difference in notches between financial rating and final rating. Both ratings range from 1 (AAA) to 21 (C)	Savings banks
GoodSoft	Equals 1 for an improvement of the financial rating because of good soft information, 0 otherwise	Savings banks
BadSoft	Equals 1 for a deterioration of the financial rating because of bad soft information, 0 otherwise	Savings banks
Strength(GoodSoft)	Strength of an improvement of the financial rating based on good soft information in notches	Savings banks
Strength(BadSoft)	Strength of a deterioration of the financial rating based on bad soft information in notches	Savings banks
Default borrower	Equals 1 if the borrower defaults in the calendar year after the year-end rating was recorded, 0 otherwise	Savings banks
Staff costs / Bank assets	Sum of staff costs over average assets per bank and year	Savings banks
Number of bank branches / Bank assets	Number of bank branches (in hundreds) over the average assets per bank (in EUR billions) and year	Savings banks
Number of bank FTEs / Bank assets	Number of bank FTEs (in thousands) over the average assets per bank (in EUR billions) and year	Savings banks
Independent variables		
Large bank	Equals 1 if the savings bank's assets (or the average assets in the savings bank group) are in the top size quartile, 0 otherwise	Savings banks
ln(Bank assets)	Natural logarithm of total assets (in EUR billions) of the savings bank (or the average assets in the savings bank group)	Savings banks
Commission ratio	Commission payments over total salaries per group of savings banks	Savings banks
Direct competition	Branches of direct competitors (commercial banks and cooperative banks) to savings banks branches per group of savings banks. The raw data is provided per region which refers to an area approximately equivalent to a U.S. county ("Kreis") or a municipal area ("Stadt")	Bundesbank
Z-Score borrower	Altman's Z-Score calibrated to the German banking market (Engelmann et al., 2003), defined by Z-Score = 0.717*Working capital/Assets + 0.847*Retained earnings/Assets + 3.107*Net profits/Assets + 0.420*Net worth/Liabilities + 0.998*Sales/Assets A higher Z-Score indicates a lower risk associated with the borrower.	Savings banks
Financial rating borrower	A borrower's financial rating, numerical notches from 1 (AAA) to 21 (C)	Savings banks
Number of mergers	Number of mergers within a group of savings banks per year	Savings banks
Regional debt per capita	Debt per capita of the community where the savings bank (or savings bank group) is located	Destatis
Regional GDP per capita	GDP per capita of the community where the savings bank (or savings bank group) is located	Destatis
ln(Borrower assets)	Natural logarithm of total assets per borrower (in EUR millions)	Savings banks
Opaque borrower	Equals 1 for closely held borrowers that are more opaque, 0 otherwise	Savings banks
Relationship length	Observed bank-borrower relationship time, in years	Savings banks

Table 2: Descriptive statistics and univariate analysis

Panel A shows descriptive statistics of the main variables. The definitions of variables are given in Table 1. Panel B shows the results of the univariate analysis on the impact of discretion in relationship lending. We split the borrowers into four groups depending on the bank groups' average assets, which approximates relationship strength. Column 5 provides the average differences between the largest and the smallest bank size quartiles and the significance level. We use univariate regressions with standard errors clustered at the savings bank group level. *, **, *** indicate significance at the 10%, 5% and 1% level, respectively.

Panel A: Descriptive statistics

Variable	Observations	Mean	Std. dev.	5p	25p	Median	75p	95p
Dependent variables								
|Δ Rating|	77,364	2.023	1.549	0.000	1.000	2.000	3.000	5.000
GoodSoft (Dummy variable)	77,364	0.245	0.430	0.000	0.000	0.000	0.000	1.000
BadSoft (Dummy variable)	77,364	0.598	0.490	0.000	0.000	1.000	1.000	1.000
Strength(GoodSoft)	18,986	2.476	1.626	1.000	1.000	2.000	3.000	6.000
Strength(BadSoft)	46,240	2.367	1.286	1.000	1.000	2.000	3.000	5.000
Default borrower	77,364	0.047	0.212	0.000	0.000	0.000	0.000	0.000
Staff costs / Bank assets (in percent)	2,140	1.355	0.187	1.007	1.246	1.368	1.482	1.637
Number of bank branches / Bank assets	2,140	20.291	9.291	8.148	13.606	19.494	24.769	36.171
Number of bank FTEs / Bank assets	2,140	2.404	0.401	1.719	2.161	2.404	2.675	3.046
Independent variables								
Bank assets (in EUR billion)	77,364	3.094	2.986	0.878	1.434	1.975	2.861	12.527
Commission ratio	77,364	0.028	0.006	0.020	0.023	0.027	0.033	0.039
Direct competition	77,364	0.841	0.252	0.461	0.667	0.823	0.945	1.361
Z-Score borrower	77,364	3.399	3.008	0.523	1.654	2.786	4.353	8.093
Financial rating borrower	77,364	12.395	3.403	8.000	10.000	12.000	14.000	20.000
Number of mergers	77,364	0.364	0.696	0.000	0.000	0.000	1.000	2.000
Regional debt per capita	77,364	1.064	0.403	0.624	0.809	0.960	1.217	1.836
Regional GDP per capita	77,364	26.174	7.129	18.393	20.667	25.892	27.587	43.579
Borrower assets (in EUR million)	77,364	3.336	19.340	0.071	0.223	0.515	1.408	10.260
Opaque borrower (Dummy variable)	77,364	0.515	0.500	0.000	0.000	1.000	1.000	1.000
Relationship length (in years)	77,364	3.662	2.570	1.000	2.000	3.000	5.000	9.000

Panel B: Univariate comparison: small and large banks

Variable	Bank size, measured by average assets		Large - Small
	Small	Large	
Dependent variables			
|Δ Rating|	2.046	1.951	-0.095**
GoodSoft	0.256	0.212	-0.044**
BadSoft	0.588	0.626	0.038*
Strength(GoodSoft)	2.543	2.230	-0.312***
Strength(BadSoft)	2.370	2.361	-0.008
Default borrower	0.050	0.039	-0.011**
Staff costs / Bank assets (in percent)	1.395	1.247	-0.148***
Number of bank branches / Bank assets	22.130	15.460	-6.670***
Number of bank FTEs / Bank assets	2.507	2.134	-0.373***
Independent variables			
Bank assets (in EUR billion)	1.739	7.251	5.512***
Commission ratio	0.027	0.031	0.004*
Direct competition	0.903	0.653	-0.250***
Z-Score borrower	3.308	3.676	0.368***
Financial rating borrower	12.535	11.964	-0.572***
Number of mergers	0.343	0.426	0.083
Regional debt per capita	0.996	1.275	0.279
Regional GDP per capita	23.879	33.213	9.334**
Borrower assets (in EUR million)	3.170	3.845	0.675**
Opaque borrower	0.540	0.438	-0.102***
Relationship length (in years)	3.787	3.280	-0.508

Table 3: Discretionary lending and bank size

The table contain the results of OLS models regressing discretion in lending on bank size. We use the matched bank-borrower dataset including five measures for discretion in lending. The large bank dummy variable approximates relationship strength. See Table 1 for the definitions of all variables. Industry-year and state-year fixed effects are included in all regressions. The standard errors in parentheses are clustered at the savings bank group level. *, **, *** indicate significance at the 10%, 5% and 1% level, respectively.

	|Δ Rating|	GoodSoft	BadSoft	Strength(GoodSoft)	Strength(BadSoft)
Large bank dummy	0.0362	-0.0164*	0.0226**	-0.0858*	0.0746***
	(0.0255)	(0.0086)	(0.0105)	(0.0476)	(0.0257)
Commission ratio	-1.7525	0.1811	-0.4877	-4.1229	-1.1385
	(1.7426)	(0.6639)	(0.7774)	(4.0240)	(1.9730)
Direct competition	-0.0339	-0.0053	-0.0020	0.0292	-0.0383
	(0.0533)	(0.0136)	(0.0182)	(0.1355)	(0.0597)
Number of mergers	-0.0080	-0.0044	0.0001	-0.0196	0.0040
	(0.0081)	(0.0032)	(0.0044)	(0.0203)	(0.0079)
Regional debt per capita	0.0469	0.0145	0.0031	-0.0044	-0.0117
	(0.0367)	(0.0095)	(0.0141)	(0.0483)	(0.0327)
Regional GDP per capita	0.0007	-0.0013*	0.0007	0.0008	0.0022
	(0.0018)	(0.0007)	(0.0009)	(0.0028)	(0.0016)
ln(Borrower assets)	-0.1491***	-0.0115***	-0.0104***	-0.1323***	-0.0938***
	(0.0063)	(0.0015)	(0.0021)	(0.0076)	(0.0070)
Opaque borrower dummy	0.0427**	-0.0020	0.0071	0.0284	0.0258
	(0.0170)	(0.0043)	(0.0056)	(0.0255)	(0.0214)
Financial rating borrower	0.0943***	0.0771***	-0.0712***	0.2667***	-0.0026
	(0.0066)	(0.0013)	(0.0014)	(0.0052)	(0.0047)
Z-Score borrower	-0.0929***	0.0330***	-0.0482***	0.0904***	-0.1422***
	(0.0069)	(0.0014)	(0.0020)	(0.0055)	(0.0080)
Relationship length	0.0167***	-0.0034***	0.0062***	-0.0296***	0.0200***
	(0.0036)	(0.0010)	(0.0013)	(0.0057)	(0.0036)
Industry-year fixed effects	Yes	Yes	Yes	Yes	Yes
State-year fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	77,364	77,364	77,364	18,986	46,240
Adj. R squared	0.1237	0.3470	0.2377	0.3930	0.0943

Table 4: Borrower selection, financial risk and competition

The table shows OLS regression results. We regress the upgrade probability on borrower financial risk, bank competition and bank size. *Risky borrower* is a dummy variable that equals 1 for borrowers of the riskiest quartile according to the financial rating. The dummy variable *Large bank* equals 1 for the largest size quartile. *High competition* is a dummy variable that equals 1 if the competition level is above the median. We omit the other covariates for space considerations. See Table 1 for the definitions of all variables. Industry-year and state-year fixed effects are included in all regressions. The standard errors in parentheses are clustered at the savings bank group level. *, **, *** indicate significance at the 10%, 5% and 1% level, respectively.

	I	II	III	IV	V
Large bank dummy	-0.0146*	0.0024	0.0042	-0.0144*	0.0139
	(0.0086)	(0.0062)	(0.0104)	(0.0086)	(0.0093)
Risky borrower dummy	0.5348***	0.5536***	0.5347***	0.5279***	0.5458***
	(0.0123)	(0.0112)	(0.0123)	(0.0154)	(0.0166)
High competition dummy	-0.0190**	-0.0197**	-0.0155*	-0.0224**	-0.0191*
	(0.0076)	(0.0077)	(0.0079)	(0.0096)	(0.0096)
Large bank * Risky borrower		-0.0887**			-0.0512*
		(0.0347)			(0.0289)
Large bank * High competition			-0.0330***		-0.0039
			(0.0102)		(0.0096)
High competition * Risky borrower				0.0135	0.0128
				(0.0220)	(0.0203)
Large bank * Risky borrower					-0.1579***
* High competition					(0.0317)
Full set of covariates	Yes	Yes	Yes	Yes	Yes
Industry-year fixed effects	Yes	Yes	Yes	Yes	Yes
State-year fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	77,364	77,364	77,364	77,364	77,364
Adj. R squared	0.2914	0.2927	0.2915	0.2914	0.2935

Table 5: Discretionary lending and borrowers' ex post default risk

The table contains marginal effects from Probit estimates with the borrowers' default dummy variable (1 equals default, 0 otherwise) as the dependent variable and the five discretionary lending proxies as the main independent variables for the matched bank-borrower dataset. *Risky borrower* is a dummy variable that equals 1 for borrowers of the riskiest quartile according to the financial rating. We conduct Wald tests in columns II and VI for *Upgrade* = *Downgrade* and in column VI also for the interaction effects *Upgrade* * *Risky borrower* = *Downgrade* * *Risky borrower*. See Table 1 for the definitions of the list of covariates that are omitted from being displayed in the table. Year, industry, and state fixed effects are included in all regressions. The standard errors in parentheses are clustered at the savings bank group level. *, **, *** indicate significance at the 10%, 5% and 1% level, respectively.

	I	II	III	IV	V	VI	VII	VIII
|Δ Rating|	0.0005				0.0027***			
	(0.0004)				(0.0004)			
GoodSoft		-0.0031				0.0056*		
		(0.0020)				(0.0033)		
BadSoft		0.0066***				0.0084***		
		(0.0016)				(0.0017)		
Strength(GoodSoft)			-0.0030*				0.0002	
			(0.0016)				(0.0029)	
Strength(BadSoft)				0.0030***				0.0030***
				(0.0005)				(0.0005)
|Δ Rating| * Risky borrower					-0.0048***			
					(0.0006)			
GoodSoft * Risky borrower						-0.0116***		
						(0.0021)		
BadSoft * Risky borrower						-0.0035*		
						(0.0019)		
Strength(GoodSoft) * Risky borrower							-0.0033	
							(0.0027)	
Strength(BadSoft) * Risky borrower								-0.0004
								(0.0009)
Wald tests								
GoodSoft = BadSoft		-0.0097***				-0.0028		
GoodSoft * Risky borrower = BadSoft * Risky borrower						-0.0081***		
Full set of covariates	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	77,364	77,364	18,986	46,240	77,364	77,364	18,986	46,240

Table 6: Positive correlation test

The table contains marginal effects from Probit estimates with the borrowers' default dummy variable (1 equals default) as the dependent variable and the large bank dummy and the financial rating as the main independent variables for the matched bank-borrower dataset. We run a positive correlation test for self-selection based on Finkelstein and McGarry (2006). The dummy variable *Large bank* equals 1 for the largest size quartile. *Risky borrower* is a dummy variable that equals 1 for borrowers of the riskiest quartile according to the financial rating. In columns IV and V of the table, we replace the *Financial rating borrower* variable with five dummy variables that capture the non-linear relationship between the financial rating and default. The first financial rating bucket (up to a numerical rating of ten) serves as the omitted category. See Table 1 for the definitions of the list of covariates that are omitted from being displayed in the table. The standard errors in parentheses are clustered at the savings bank group level. *, **, *** indicate significance at the 10%, 5% and 1% level, respectively.

	I	II	III	IV	V
Large bank dummy	-0.0011	0.0078*	0.0085**	0.0071*	0.0093*
	(0.0041)	(0.0046)	(0.0046)	(0.0043)	(0.0061)
Financial rating borrower	0.0096***	0.0090***	0.0090***		
	(0.0005)	(0.0005)	(0.0005)		
Large bank * Risky borrower			-0.0014		
			(0.0034)		
Financial rating bucket 11-13				0.0415***	0.0418***
				(0.0026)	(0.0036)
Financial rating bucket 14-16				0.1227***	0.1261***
				(0.0064)	(0.0091)
Financial rating bucket 17/18				0.2090***	0.2103***
				(0.0130)	(0.0169)
Financial rating bucket 19-21				0.3119***	0.3149***
				(0.0135)	(0.0176)
Large bank * Financial rating bucket 11-13					-0.0008
					(0.0036)
Large bank * Financial rating bucket 14-16					-0.0041
					(0.0035)
Large bank * Financial rating bucket 17/18					-0.0007
					(0.0052)
Large bank * Financial rating bucket 19-21					-0.0021
					(0.0056)
Full set of covariates	No	Yes	Yes	Yes	Yes
Observations	77,364	77,364	77,364	77,364	77,364
Pseudo R squared	0.1356	0.1514	0.1514	0.1500	0.1500

Table 7: Costs per loan

The table contains OLS regressions results. The data is on the individual savings bank level. We regress three proxies of screening/monitoring intensity on the *Large bank* dummy variable. See Table 1 for the definitions of all variables. State-year fixed effects are included in all regressions. Standard errors are clustered at the individual savings bank level. *, **, *** indicate significance at the 10%, 5% and 1% level, respectively.

	Screening / monitoring intensity		
	Staff costs / Assets (in percent)	Number of bank branches / Assets	Number of bank FTEs / Assets
Large bank dummy	-0.1299***	-5.9697***	-0.2475***
	(0.0166)	(0.7207)	(0.0314)
Commission ratio	-0.0193**	-0.8620**	-0.0342**
	(0.0076)	(0.3603)	(0.0135)
Direct competition	0.0509*	3.9373**	0.0803
	(0.0283)	(1.6130)	(0.0509)
Number of mergers	0.0472**	2.1849**	0.0780**
	(0.0188)	(0.9789)	(0.0347)
Regional debt per capita	0.0000	0.0015	0.0000
	(0.0000)	(0.0013)	(0.0000)
Regional GDP per capita	0.0000	0.0166	-0.0001
	(0.0010)	(0.0528)	(0.0019)
State-year fixed effects	Yes	Yes	Yes
Observations	2,140	2,140	2,140
Adj. R squared	0.3583	0.2997	0.4966

INTERNET APPENDIX

Table A1: Variation in bank size

The table shows variation in bank size according to the time and cross section dimension as well as within bank group variation. Bank size is measured in total bank group assets. Panel A provides the change in bank group assets to show the time series variation from 2002 to 2006. Panel B shows the cross sectional variation of bank group size by bank size quartiles. Panel C provides the within group variation in bank size.

Panel A: Time series variation (change in bank group assets)

Year	Mean	Median
2002	0.027	0.028
2003	0.007	0.007
2004	0.006	0.004
2005	0.007	0.008
2006	0.010	0.009
Total	0.012	0.010

Panel B: Cross sectional variation of bank group assets

Bank size	Mean	Median
1 - Small	6.99	6.52
2	11.81	11.24
3	14.27	13.85
4 - Large	31.82	32.86
Total	12.78	10.44

Panel C: Within group variation in bank size

Bank group quartile	Mean	Std. dev.	Min.	25p	Median	75p	Max.
1 - Smallest bank group	0.83	0.60	0.11	0.42	0.62	1.05	3.34
2	1.25	0.86	0.13	0.68	1.04	1.56	4.59
3	1.81	1.29	0.21	0.93	1.36	2.36	6.83
4 - Largest bank group	3.64	4.45	0.20	1.15	2.39	4.24	29.86

Table A2: Additional sample characteristics

The table shows the number of observations sorted by bank size and bank competition. See Table 1 for the general data definitions.

	Bank competition		
Bank size quartile	Below median	Above median	Total
1 (Smallest)	14,138	5,507	19,645
2	6,648	12,506	19,154
3	3,463	16,082	19,545
4 (Largest)	14,925	4,095	19,020
Total	39,174	38,190	77,364