On the Use of Soft Information in Bank Lending Decisions

Manju Puri[*], Jörg Rocholl[†], and Sascha Steffen[‡]

September 6, 2010

Preliminary and Incomplete

Please do not circulate without permission from authors

Abstract

[*] Duke University and NBER. Email: mpuri@duke.edu. Tel: (919) 660-7657.
[†] ESMT European School of Management and Technology. Email: rocholl@esmt.org. Tel: +49 30 21231-1292.
[‡] University of Mannheim. Email: steffen@bank.bwl.uni-mannheim.de. Tel: +49 621 181-1531.

Understanding how banks make loans is important and has been at the forefront of the current financial crisis. An important question is how should the process of loan making by banks be regulated to minimize risks? For example, should the loan making process be entirely codified so that the potential for discretion does not exist, and loans are made on hard, verifiable information collected by the bank? On the one hand, allowing discretion could allow for the information obtained from relationship specific assets and soft information to be incorporated to improve the quality of loans made. On the other hand the downside of using unobservable information is that discretion can be used to promote favoritism or cronyism. How allowing discretion in incorporating soft information to make decisions affects the quality of loan making is an important, open question of interest to academicians, banks, consumers and regulators.

Some important questions that arise in understanding the use of soft information is how is such information incorporated into loans? Is it used to justify loan approvals or to justify loan rejections? How does this information ultimately relate to the quality of the loan? A major limitation in studying such questions is quantifying the impact of soft information, as it is, by its very definition, inherently difficult to measure.

We have been able to access a unique, proprietary database of over 1 million loans which are the universe of retail loans made by savings banks in Germany. This data provide an ideal experimental ground to assess these questions and in particular, to understand how soft information is incorporated into loans. In particular, we have the universe of applications for consumer retail loans to German savings banks between 2004 and 2008. Each application has a scorecard that has to be filled out by the lending officer. The scorecard details and methodology is uniform across all of the savings banks in our sample. The scorecard asks the loan office to record all the data items that include borrower characteristics and all pertinent information such as age of the borrower, income, years on job, repayment capacity (computed as loan amount as a percentage of income). The total score is then translated into an internal rating category, which is associated with a certain default probability. However, the scorecard does not automatically result in an accept/reject decision. The loan office uses it as input but can overrule the recommendation that would be associated with the numerical scoring. E.g., if the score is below the cut-off the loan office can approve the loan, or even if the score is above the cutoff the loan office can reject the loan. However, in each of these cases the loan officer has to give an explicit reason as to why she overrode the internal numeric scoring recommendation. These reasons are collated in the data and are observed. We can use these reasons to code the "soft" information and see when and how it is incorporated into the loan making process. We know which loans would have been accepted based on the internal rating system and which would have been turned down. We also can follow the loan officer's decision to identify those loans which have finally been accepted and have the performance data for all loans. We are thus able to assess the relative performance of those loans accepted based on soft information versus those accepted automatically by the internal rating system.

Our initial set of tests therefore examines how soft information is incorporated in the loan making process. Does the loan officer use soft information to justify overruling the numeric scoring to approve the loan, or is soft information used more often to justify rejecting applications that otherwise have numeric scores over the cutoff? A first look at the data suggests that the use of soft information is widespread and economically significant. While the overall acceptance rate of consumer loans in our sample is 97.8%, we find that less than 82% have been approved by the internal rating system. More than 16% of the acceptances are overrides and the majority of overrides are based on loan officer's soft information. The remaining overrides are justified by "hard" information reasons. Those loans are accepted either because they are collateralized or because the application is associated with a restructuring or refinancing of an existing loan and does not result in an increase of the exposure to the individual borrower. But the use of soft information is also common with regard to rejections as only 61% of the final rejections are based on scoring.

We test this more formally and ask, given that the application has been rejected based on credit scores, when do loan officers use their discretion in the loan granting process and override the initial rejection (what we call "lowside overrides")? We find that loan officers use private (soft and hard) information to reverse initial rejections conditioning on the existence of a relationship with the applicant. We use transaction account related relationship proxies (Puri et al. (2010)) and find that lowside overrides are more likely the stronger the relationship is (based on the existences, length and scope of the relationship). We find that overrides are significantly more likely for applicants that have extensively drawn down their credit limit of their transaction accounts or have even exceeded the limit which suggests that, at least to some degree, overrides are meant to restructure existing debt. The coefficients of the internal ratings that we use as risk control variables provide consistence evidence: applicants in the four highest rating classes are least likely to be accepted based on private information. More interestingly, we find that rejections of medium quality applicants are most likely to be reversed as opposed to both high and low quality customers. While high quality applications are most likely to be rejected based on "Policy Rules" (for example, related to age or employment history), low quality borrowers are more likely to have additional negative external credit bureau information.

Our dataset, however, is censored as we can only observe overrides for those applications that have initially been rejected. Therefore, we use a binomial Probit model with selection to account for sample selection. We augment the selection equation where the dependent variable is one if the application is rejected with a second Probit model where the dependent variable is one if the loan officer reverses the initial rejection and find consistent results.

We then ask, given initial rejections of the loan applications, when are they more likely to be reversed based on soft information as opposed to private hard information (either restructuring of existing debt or new but collateralized loans)? We use a multinomial Logit model to analyze this further and find that those applicants that have extensively used their credit limits are more likely to have their loan restructured (or have to provide sufficient collateral) as opposed to being accepted based on soft information. We also find that young applicants are more likely to be accepted based on soft information. Intuitively, they are also less likely to have a credit history with the bank on which hard information decisions can be based on.

The next question that arises is whether the use of soft information results in higher quality loans being made by the bank. An important theoretical and practical question is whether the use of private soft information improves the approval decision in retail loans and thus the subsequent success of these loans – or in the extreme – whether the loan officer's discretion is simply due to favoritism that occurs at the bank's expense. Our next set of tests examines the default performance of the loans to address this question. Thus we explicitly examine if loans rejected by the internal numeric scoring but approved based on internal soft information and the loan officer's discretion perform better or worse than comparable loans that are approved by the numeric scoring system without additional soft information. We do not find any evidence that loans approved based on private soft information perform any differently than loans approved using the internal numeric scoring recommendation alone.

The rest of the paper is organized as follows. Section 1 describes the data. Section 2 explains the credit scoring process and how private information is incorporated into the loan making process. Section 3 conducts an analysis of loan applications to determine the factors that lead to a loan officer to override the internal scoring. Section 4 examines if the use of soft information leads to differential default rates for loans. Section 5 concludes.

I. Data

A. Sample

Our sample consists of the universe of applications for consumer loans to German savings banks between 2004 and June 2008, and it comprises detailed information on the approval decision as well as the performance of the accepted loans, i.e. the advantage of our dataset over previous studies is that our sample is not drawn, and may be systematically different, from a larger population (see, for example, Greene (1992)). All applications are made in person at one of the branches of the savings banks. We obtain these data from Sparkassen Rating- und Risikosysteme GmbH (s-rating), a fully owned subsidiary of the German Savings Bank Association (DSGV).[1] This subsidiary was founded in 2004 and develops and enhances procedures and instruments in rating and scoring of both business and retail clients. The goal is to support the credit and portfolio decisions of savings banks during the loan application process. For this purpose, s-

[1] "The […] DSGV is the umbrella organisation of the 'Sparkassen-Finanzgruppe' and its 446 savings banks, 11 Landesbanken, 11 Landesbausparkassen, 12 public insurance companies and many more financial service providers. It is funded by the regional savings banks associations together with the Landesbanken. It represents the interests of the Sparkassen-Finanzgruppe on banking policy, regulatory law and other banking industry issues on a national and international level. It also organizes decision-making and stipulates strategic direction within the Group, acting in cooperation with the regional associations and other Group institutions." (http://www.dsgv.de/en/dsgv_portraet/aufgaben_und_ziele/index.html)

rating implemented application scorecards to evaluate customer risks in three major business areas: checking accounts (and associated credit facilities), consumer loans, and mortgage loans. We focus on consumer loans for several reasons. First, this product is highly standardized and, therefore, homogenous not only within the savings bank sector, but also across all other banks. Second, credit-scoring methodologies are used in this standardized process to screen retail customers, and the resulting internal ratings meaningfully reflect borrower risk based on codified hard information. Third, all savings banks in our sample apply a uniform credit-scoring methodology.

The dataset also comprises monthly performance data for each loan. For the calculation of the default rates, we apply the same methodology as the German savings banks use ("s-rating") to calibrate their internal rating models. We compute one-year default rates for each approved loan in our dataset and thus constraint our dataset to those loans for which we have at least 12 monthly performance observations. We also use the same events as s-rating to measure a borrower's default, which comply with the Internal Risk Based Approch (IRBA) in Basel II.[2] A borrower defaults, if one of the following events occurs: (i) the borrower is 90 days late on payment of principal or interest, (ii) the borrower's repayment becomes unlikely, (iii) the bank builds a loan loss provision, (iv) the liabilities of the borrower are restructured with a loss to the bank, (v) the bank calls the loan, (vi) the bank sells the loan with a loss, and (vii) the banks needs to write-off the loan.[3] Our data includes flags for each of these default events and the associated default date. There are no cross-default clauses in German retail lending, i.e. borrowers who default on one loan do not simultaneously default on all loans at the same bank. In addition to performance data, we have detailed information on all loan and borrower characteristics that the bank employs to assess a borrower's creditworthiness. In particular, we have information on the existence and extent of prior relationships that loan applicants have had with the savings banks at which they apply for a new loan.

---

[2] Definitions of these default events can be found in the "Solvabilitätsverordnung §125", the "Baseler Rahmenvereinbarung Tz. 452-453 and the "EU-Richtlinienvorschlag, Anhang VII, Teil 4".

[3] The second event is used if the default cannot be categorized into one of the other default events. For example, if the repayment of the borrower is 'unlikely', but the bank does not build a loan loss provision because the loan is fully collateralized, this category is chosen as default event.

There are a number of unique characteristics of these data that make them particularly suitable for the purpose of our study: First, they contain detailed information on individual loan applicants, including information on their credit risk and their relationship status. Second, they comprise detailed monthly information on the performance of each individual loan and in particular its default. Third, the data on both the loan applicants and loan performance are highly reliable, as they comply with the Basel II requirements. Fourth, the data are very comprehensive as they cover the bulk of the universe of savings banks in Germany, which hold a market share in retail lending of about 40 percent in Germany. Also, the "regional principle" is an important institutional setting associated with German savings banks. This implies that borrowers can only do business with savings banks within the region they are domiciled in. Consequently, we do not have to worry about endogenous matching of borrowers and banks in our sample. Finally, all borrower and relationship characteristics are taken from an internal scoring system that is used by all our sample banks.[4] The interesting feature for our analysis is that the scoring system provides a credit assessment of the applicant, but it serves as a guideline rather than a mandatory prescription. The final loan granting decision is made by each individual bank also using its own discretion and taking into account its respective ability and willingness to take on risks. Furthermore, loan officers have some discretion themselves as to whether or not to approve a loan application. In other words, there are some subjective elements associated with the banks' screening process that might very well be different for each respective bank. Overall, the large and comprehensive sample of loans by savings banks and the detailed information on loan applicants' relationship status and credit risk as well as on the performance of the approved loans provides a unique opportunity to analyze the sources of value of relationships.

B. Descriptive Statistics

Table I reports summary statistics for consumer loans. On average, 97.8% of the loan applications have been accepted. Over the first twelve month after the loan origination, 0.6% of the approved loans default according to the above default definition.[5] We define an existing relationship as an applicant who has a checking account with the savings bank prior to applying for the loan. Only 2.5% of the loan applicants have had no relationship with their savings banks

[4] In principle, savings banks can also use information from external rating agencies, but they have to pay for this information. It is thus available only for 86,628 loan applications.

[5] The default rate increases to 1.3% when the loan performance over the full sample period is considered.

prior to the loan application. At the same time, many of the existing customers have been customers of the savings banks for a substantial period of time. For example, 47.6% of the loan applicants have been customers of the savings banks for more than 15 years, and more than 80% of them have been customers for at least 5 years.

The majority of customers have checking accounts with the savings banks prior to the loan application. Checking accounts can be combined with debit and credit cards. The combination of debit and credit cards is the most common type among customers; 46.5% of them have both types of cards. 3.8% of the customers only have a debit card, while 18.3% of the customers only have a credit card. 28.9% of the customers have no cards. Furthermore, 94.5% of the loan applicants have an existing credit line at the time when they ask for a loan. These credit lines are not used in 30.1% of the cases. If they are used, the usage ranges mostly between 20 % and 80% of the limit of the credit line.

Table II presents summary statistics for borrowers and applicants for our sample loans. Loan applicants have an average monthly income of €1,769, and most of them are in the age cohort between 30 and 45 years, followed by the age cohorts between 50 and 60 years. 7.4% of the applicants have debt-to-income ratios (DTI) of less than 5%. More than 54% of the applicants have a DTI of less than 20%. Most borrowers work in the service industry and have been in their current job for more than two years.

In the following section, we explain in detail how credit scoring is used in practice at our sample banks and how the different kinds of information, i.e. public information as well as private hard and soft information, are utilized in the credit decision.

II. The Use of Private Information in Credit Decisions

A. An Overview of the Credit Scoring Process

The credit rating for retail customers is based on a quantitative score that uses an application scorecard at the loan origination stage to facilitate and standardize the decision process across all savings banks. The idea of the scorecard is to record all data items that have predictive power related for an applicant's default risk. These data comprise characteristics about the borrower and information about previous relationships with the bank. This information is translated into an individual score for each item and then summed up to a total score. Finally, the total score is mapped into an internal rating category, which is associated with a default probability of the borrower. The loan officer records all data items to calculate the total score and the internal rating in the scorecard process. A score cannot be calculated, and the process is left unfinished if one of the data items is missing. Every process receives a status whether the loan application in that process is approved or rejected by the bank or rejected by the applicant. From the beginning of the scoring process until the classification into the status, the scoring process is pending. The maximum time-span the process can be pending is 16 weeks. After 16 weeks, the scorecard process is automatically classified as 'rejected by applicant'.[6] If the loan application is approved, performance data about the loan (delinquencies, default, etc.) are recorded on the scorecard on a month by month basis. The scorecard thus contains both loan application and performance data for consumer loans from all German savings banks that use the scorecard. Three data centers implement and administrate the scorecards and the export of the data into the central data-pool of all savings banks. The information from these data centers are used to calibrate the internal rating system in accordance with Basel II and MaK ("Mindestanforderungen an das Kreditgeschäft"), which accentuates the high quality of the data used in this study.

B. Codified Information in Credit Scores and Credit Decisions

Credit risk scoring is used to evaluate the level of risk associated with a loan applicant and assigns probabilities that an applicant with a given credit score will be good or bad. These probabilities or odds are a result of a history of default frequencies of borrowers with the same scores or underlying characteristics, i.e. the characteristics used in scorecards are statistically determined to have predictive power in separating good and bad customers. Credit scores together with other business considerations or "Policy Rules" are then used as a basis for credit decision. The characteristics used in the scorecards are to some degree standardized across the

[6] We exclude all loans that are flagged as "rejected by applicants" from our analysis.

different banking groups (including the group of savings banks) as consumer lending is a mass customization business, albeit the relative weight associated with the characteristics varies between institutions. We can broadly distinguish between three types of characteristics[7]: (1) *Demographic and socioeconomic factors*: Age (of applicant and co-signers), job and industry (including unemployed, student, retired, etc.), how long the applicant has been in his current job, education, single versus married, gender etc.[8] Information about the living conditions of the applicants are used to predict their default probability. (2) *Financial situation* of the applicant: characteristics include income, debt-to-income ratio, wealth as well as information from external credit bureaus. (3) *Existing relationships* with the bank: account age, credit limits, usage of credit limits, number of accounts, credit and debit cards etc. In other words, only hard verifiable information is codified in the scorecard which has a statistically proven influence on default rates and leads to the same credit score or expected probability of default regardless of who collects and handles the data (which is the definition of hard information following, for example, Petersen (2004)). The factors are used depending on their respective availability. For example, institutions that do not have access to credit bureau data (or where credit bureau data is unavailable) rely on other factor such as socioeconomic characteristics.

Each attribute (if applicant age is the characteristic, than "18 – 23 years" is the attribute) is assigned a score based on statistical analyses (for example, logit regressions). The information value or predictive strength of the attribute as well as its correlation with other attributes determine its respective score. The scores of each attribute are summed to the total score of the applicant which itself is associated with a specific probability of default. This process is identical for each of our 296 sample banks that use the same scorecard. Based on these scorecards, each bank individually decides how to set cutoff points, for example, decline all loan applications with a credit score below 500. We know, for all of our sample banks, the credit score and internal rating together with the suggested "accept / reject" decision that would automatically result as the credit decision based on that score.

[7] These characteristics are taken from a report on behalf of the Federal Ministry of Food, Agriculture and Consumer Protection about chances and risks for consumers associated with credit scoring (xxx). They are more general and do not reflect the characteristics used by the group of savings banks.
[8] Employing socioeconomic factors in credit scoring is a sensitive topic both in the US as well as in Germany. Consequently, banks generally do not use information about nationality or addresses in scorecards in accordance with the Equal Opportunities Act in the US and antidiscrimination laws in Germany.

However, the scorecard is only a tool to facilitate the credit decision, but the scoring outcome does not result in an automatic credit decision.[9] The credit decision based on internal credit scores can either be accepted or it can be overridden by the loan officer in both ways: he can reject the initially accepted or accept the initially rejected applicants. The next section describes the loan officer's discretion and how this is coded in the data in more detail.

B. Overriding Credit Decisions Based on Credit Scores

Before we explain the discretion loan officers have in the loan granting decision, we highlight another reason as to why credit decisions based on credit scores might be reversed which are policy rules of the respective bank. Usually, policy rules are not part of the scorecard in order not to interfere with the scoring process but are part of prudent risk management (Siddiqi (2006)). Examples of possible policy rules are legal age requirements (decline the application if the applicant is younger than 18 years old or has just recently turned 18), employment (decline the application if the applicant has been employed for only a few months), or credit bureau information (decline if there are too many delinquencies at the credit bureau). In other words, even if the applicant is above the cutoff score, policy rules might result in a rejection of the loan application. Therefore, rejections are common even for the highest rating classes. Furthermore, as cutoffs and policy rules are business decisions of each individual bank, two applicants with the same credit score might be accepted at one bank but rejected at another bank. As with the discretionary overrides, overrides based on policy rules have to be reported using an "override reason code". The internal rating system of the savings banks require that reason codes have to be reported in the following circumstances: (1) whenever a loan application is rejected, and (2) whenever a loan officer overrides the decision based on credit scoring. Applicants with credit scores below the cutoff who are approved are called "lowside overrides", applicants with credit scores above the cutoff who are rejected are called "highside overrides".

Panel A of Table III presents information related to the accept / reject decisions based on internal scoring and loan officer discretion. The high override rates for both rejected and accepted

[9] This process is consistent with German consumer protection laws that outcomes such as credit decisions that affect consumers both positively and negatively are not supposed to be solely based on an automatic decision process.

applications are striking. Despite the low average rejection rate of 2.2%, only 62% of the applications are rejected based on internal credit scores. 29% of the rejected applications were coded as "miscellaneous" what we refer to as soft information because it is a proxy for judgmental overrides. 9.4% of the rejections were coded as hard information which includes violation of the regional principle, negative external credit bureau information or other internal warning signs (for example, from a transaction account relationship).

Furthermore, 16.2% of all loan applications are lowside overrides. In other words, taking these results at face value, the average acceptance rate of the internal rating system is less than 82%. This is an important aspect for our analysis as we know the override reason code and can thus monitor the performance of these loans to assess whether or not they perform worse relative to those loans that have automatically been accepted by the internal rating system. We find that 176,622 applications have been approved because of discretionary overrides by loan officers. 78% of these overrides are classified as soft information overrides and include reason codes such as promising customer, VIP customer or, the judgmental code, miscellaneous. 15.2% of the overrides are coded as hard information reasons and relate to restructuring or refinancing of loans. In other words, these loans do not increase the exposure to the respective borrower but restructure the existing debt ("Hard Information: Restructuring"). This is, for example, particularly relevant for customers that have already drawn down or even exceeded their credit limit of their transaction accounts. 13.5% of the overrides are due to collateralization of new loans("Hard Information: Collateral").

Panel B of Table III further segregates the lowside overrides by internal rating classes. Column (5) presents the percentage of lowside overrides within each rating class. While only 0.29% of the accepted applications in the best rating class are overrides, 86.11% of the acceptances in the lowest rating class are overrides. Columns (6) to (8) report the percentage of overrides based on soft (column (7)) and hard information (columns (7) and (8)). Soft information is used primarily for medium quality borrowers as only 11% of loans in rating class 1 are soft information overrides. Further, we find a marginal decrease in the use of soft information in rating classes 11 and 12. There is a similar distribution of overrides because of collateralization of these loans. More interestingly, however, is the restructuring of existing loans. 86.46% of the highest quality overrides are related to restructurings of existing debt. The percentage is sharply decreasing for medium quality applicants and, again, increasing for rating classes 11 and 12. Note that restructuring can simply be a refinancing of an existing loan as opposed to a restructuring of, for example, bad debt. We investigate this further below for borrowers that already have extensively drawn down their credit limits of their transaction accounts.

To summarize our discussion, credit decisions use different forms of information: (1) public hard information and (2) private hard information codified in the internal rating and (3) soft and hard private information that is not codified in credit scores but is used to override and reverse credit decisions based on scorecards. Hard private information is differentially used for new and the restructuring of existing loans. In the following sections, we assess the relative importance of uncodified versus codified information in the loan granting process and the impact of overrides on the performance of these loans.

III. Loan Application Analysis

The previous section shows that loan applications by retail customers may still be accepted by savings banks even if the internal scoring for the loan applicant warrants a loan rejection. This happens when the loan officer overrides the internal credit scoring for a loan applicant and grants the loan to this applicant by indicating a reason. We will now analyze in detail which borrower characteristics may lead a loan officer to override a negative outcome of the internal scoring and to still grant a loan.

A. Internal Scoring versus Private Information

We first estimate a probit model using only those loans that are rejected based on the internal credit scoring. Our dependent variable takes a value of one if a loan that is rejected by the internal credit scoring is granted by a loan officer who uses her discretion and overrides the scoring, and a value of zero if the loan remains rejected. The results are reported in Table IV.

The main explanatory variables comprise different relationship characteristics that measure the existence, length, and intensity of a bank-customer relationship as well as the internal rating for each loan applicant. Model 1 shows that the loan officer is more likely to override the internal scoring if the loan applicant is an existing customer of the bank. Existing customers are 4.2% more likely to receive a loan than new customers if their loan is rejected by the internal scoring, even after controlling for borrower risk as captured by the internal rating. The internal rating exhibits an inverse U-shape in determining a customer's chance to receive a loan even after the negative outcome of the internal scoring. The default group comprises loan applicants in credit rating group 12, which represents those customers with the worst credit rating. In comparison to these customers, customers with credit ratings of 5 to 11 fare better as they are more likely to receive a loan due to the loan officer's discretion. However, there is much a lower likelihood for customers in credit rating groups 1 to 4, which are the customers with the best credit rating. As an example, customers with the best credit rating in group 1 have an almost 40% lower probability of receiving a loan due to the loan officer's discretion than customers with the worst credit rating in group 12. In sum, this means that the likelihood of receiving through the overriding decision first increases with a customer's credit rating and then significantly decreases. Put differently, customers with medium credit rating are most likely to receive a loan due to the loan officer's discretion.

The inverse U-shape that we observe in Model 1 for the credit rating groups can also be found in Models 2 to 5 in which we employ a very similar empirical specification and test the importance of other relationship variables. Model 2 estimates whether the length of a relationship has an impact on the likelihood of receiving a loan after receiving a negative outcome from the internal credit scoring. The results are in line with those from Model 1 in that longer rather than shorter relationships will induce a loan officer to override a negative loan scoring outcome. The default group here comprises those customers with relationships of longer than 15 years. In comparison to this group, all other groups of customers are less likely to receive a loan, and the loan officer is the less likely to override the internal scoring the shorter is the relationship between a savings bank and a loan applicant. As an example, customers with relationships of less than two years are 9.2% less likely than customers with relationships of more than 15 years to receive a loan due to the loan officer's use of discretion. Model 3 considers the intensity of bank-borrower relationships, and the results are again in line with the previous ones. While the loan officer uses her discretion even for customers that have only a debit card, she uses the discretion even more when the customer has a credit card or a combination of credit and debit card. The default group here comprises those customers without any card. Taken together, Model 1 to 3 suggest that a loan officer is more likely to use her discretion if she knows the customer either for a longer time or can collect more information from the borrower through the type of business the borrower has had with savings bank.

Model 4 analyzes whether the existence of a credit line in a loan applicant's checking account has an impact on the loan officer's decision to override the internal credit scoring. The evidence suggests that an existing credit line increases the loan applicant's likelihood of receiving a loan by 10%. Model 5 investigates this relation further by considering the extent to which a credit line is used. The default group comprises customers whose credit line is not used. The results suggest that customers with only moderate use of the credit line, i.e. between 0 and 80%, are not more likely to receive a loan due to the loan officer's use of discretion than customers who do not use their credit line at all. In contrast, customers with a credit line use of more than 80% are increasingly more likely to receive a loan. In particular, customers with a credit line use of more than 150% are almost 7% more likely to receive a loan than customers with no credit line use. Taken together, the results suggest that customers with existing credit lines and in particular those customers with a substantial use of their credit lines are much more likely to receive a loan due to the loan officer's discretion than other customers. It is important to note that interest rates for loans on current accounts that exceed the given credit limit become substantially more expensive than those for loans within the give credit limit, often leading to the need for debt restructuring in the form of consolidating a loan on a current account into a regular consumer loan. The evidence here thus suggests that debt restructuring is another reason for loan officers to override the internal credit scoring and to grant a loan to a customer even if the internal credit scoring provides a negative outcome.

Finally, Model 6 summarizes the earlier findings by sorting loan applicants into three groups: 1) Applicants with high quality, covering credit rating groups 9 to 12, 2) applicants with medium quality in credit rating groups 5 to 8, and 3) applicants with low quality, comprising the remaining credit groups. The results are very similar to those in Models 1 to 5 for the internal credit rating and reinforce the inverse U-shape for the likelihood of receiving a loan due to a loan officer's discretion. While medium-quality applicants are 4.7% more likely to receive a loan than low-quality applicants, high-quality applicants are almost 29% less likely to receive a loan than low-quality applicants, which serve as the default group in this empirical specification.

The empirical specification in Table IV does not take into account the fact that the loans for which the loan officer can use her discretion, i.e. the loans that are initially rejected by the internal scoring, are pre-selected and in particular do not comprise those loans that are accepted by the internal scoring. We now explicitly take the loan selection process into account by running a Heckman procedure in which we first estimate which loans are rejected by the internal credit scoring (selection) and then analyze which factors determine whether these initially rejected loans are finally accepted or not (outcome).

The results are reported in Table V. For the selection model, we add an instrument, which captures the level of competition in a region in which a savings bank operates. While the selection model can be identified without such an instrument, we would in that case rely deterministically on the non-linearity of the selection equation. Specifically, for the instrument we calculate the Herfindahl-Hirschmann Index (HHI) based on the number of branches that operate in each of the 439 regions or districts in Germany. The evidence in, for example, Jayaratne and Strahan (1996), Black and Strahan (2002), and Zarutskie (2006) suggest that more competition in a banking market leads to an increase in credit supply in that market. A savings bank is thus expected to be less likely to approve a loan application if there are fewer competitors in the region in which it operates.

The results in the first-stage estimation of Table V confirm this expectation and thus the evidence for U.S. banks. The dependent variable here takes a value of one if a loan is rejected and a value of zero otherwise. In regions with a higher HHI, a customer loan application is more likely to be rejected than in regions with a lower HHI. The other variables in the selection model show the expected results. Model 1 and Model 2 show that customers with existing relationships and in particular with longer relationships are less likely to be rejected in the loan application process than new customers, i.e. they are more likely to receive a loan than new customers even after controlling for their internal credit scoring. The estimation in Model 3 extends this analysis and suggests that customers are also more likely to receive a loan if their relationship is more intense as measured by the use of different combinations of credit and debit cards. Customers with credit or debit cards as well as customers with both types of cards are much less likely to be rejected than customers without any card. Model 4 and 5 consider the extent and use of credit limits and find that the existence of a credit line reduces the likelihood of being rejected, while the impact of the credit line depends on its use. A moderate use of a credit line in fact increases the chances of receiving a loan, while a substantial use of a credit line reduces these chances significantly. These results suggest that the benefits from the information, which savings banks can collect from customers who have a credit line and who moderately use it, may first dominate, while for a more extensive use of credit line the default risk may dominate. Finally, as shown in Model 6, customers with medium and in particular high quality are much less likely to be rejected than customers with low quality, reflecting the consistency with which the internal ratings are constructed.

More importantly for the purpose of our study, the results in the second stage estimation of Table V are very similar to those in Table IV. Both the relationship and internal rating variables show the same patterns as before. In particular, the inverse U shape still holds for the internal rating variables. With the rising credit quality of a loan applicant, loan officers are first more likely to override the negative outcome of the internal scoring, but then they are significantly less likely to override this outcome for the group of customers with the best credit ratings. This pattern can be consistently found in Models 1 to 6 and suggests again that medium-quality customers benefit most from the loan officer's discretion.

For the relationship variables, the results are also very similar to those in Table IV. The results in Model 1 and Model 2 suggest again that the loan officer uses her discretion more for customers with existing and longer relationships, respectively, and Model 3 provides evidence that the same holds for customers with a more intense relationship, as measured by the use of different combinations of credit and debit cards, i.e. a customer about whom the loan officer knows more is also to benefit from the loan officer's discretion. Finally, Model 4 and Model 5 confirm the earlier evidence that customers who overextend their credit line are more likely to receive a loan as well, suggesting again that debt restructuring is an important other determinant for loan officer to override the internal credit scoring and grant a loan despite a substantial credit in the current account.

In sum, the results in this section provide evidence that there are two main reasons why loan officers use their discretion in the loan application process and do not rely exclusively on the internal credit scoring. First, they override the internal credit scoring for loan applicants with medium quality and with longer and more intense relationships for which they can collect more information. Second, loan officers override the internal scoring when a loan applicant has overextended the credit limit in his checking account and is thus likely to be in need of a debt restructuring.

B. Soft versus Hard Private Information

Next we shed more light on the characteristics of the different types of customers for whom loan officers use their discretion. For this purpose, we run a multinomial logit model in which we only use the 176,621 observations for those loans that are initially rejected by the internal scoring mechanism and which are subsequently accepted by the loan officer for one of three reasons: first, the loans that are rejected by the internal scoring mechanism and subsequently accepted by the loan officer due to the private soft information that she has about the loan applicant; second, the loans that are rejected by the internal scoring mechanism and subsequently accepted by the loan officer with the intention to restructure the loan applicant's existing debt at the savings banks, i.e. based on hard information on the exposure the loan applicant already has with the savings bank; and third, the loans that are again rejected by the internal scoring mechanism and subsequently accepted by the loan officer if the loan applicant has sufficient collateral, i.e. again hard information, but this time based on the other assets owned by the loan applicant. The base case in the multinomial logit estimation is the first case, i.e. the case in which the loan application is accepted based on private soft information. The results are reported in Table VI.

The model estimates the case of a loan acceptance based on hard information on restructuring and the case of a loan acceptance based on hard information on collateral versus the base case. The results show that private soft information is particularly valuable for younger loan applicants for whom there is not a sufficiently long credit history. Both in the right and the left column, the coefficients for the different age groups become increasingly and statistically significantly negative for younger loan applicants. The default group here comprises those customers who are more than 60 years old. Soft information is of particular importance here as the loan officer has to compensate for the lack of other sources of information by judging the loan applicant's personality. For the other variable of interest, the existence and use of a credit line on a customer's current account, customers are more likely to be accepted based on soft information the lower is the use of their credit limit. Put differently, loan officers use their hard information on restructuring in particular for loan applicants with an already existing high exposure to the savings bank. This holds again for the results in both the left and the right column of Model 2.

IV. Loan Performance Analysis

The results in the previous section raise the important question of whether the loan officer discretion is beneficial or detrimental to the bank, i.e. whether the bank would be better off by not allowing the loan officer to override the outcome of the internal scoring. In particular, does the loan officer's ability to use soft information improve the assessment of a loan applicant over and beyond that provided by hard information from an internal credit scoring? Or do cognitive biases or the closeness to the loan applicant induce the loan officer to grant loans that should rather be rejected? We will focus next on this question by analyzing the performance of loans that are accepted in the various ways mentioned above.

Table VII reports the results. The dependent variable is a dummy variable that takes a value of one if the borrower defaults within the first 12 months after the loan origination and a value of zero otherwise. We distinguish again between four ways in which loans can be accepted. The first and default group comprises loans that are accepted in the regular way, i.e. by passing the internal scoring mechanism. While the first group thus does not contain an intervention or use of discretion by the loan officer, the next three groups are characterized by active loan officer involvement. The second group includes the initially rejected loans that are accepted by the loan officer due to private soft information. In the third group, there are the initially rejected loans that are accepted by the loan officer based on hard information on restructuring. The fourth and final group includes the initially rejected loans that are accepted by the loan officer based on hard information on collateral. We control again for various characteristics that capture the length and intensity of bank-borrower relationships and the loan applicant's internal credit rating.

Model 1 shows that the default rates differ significantly for the various ways in which the loans are accepted. While the default rates for loans that are accepted based on private soft information and for loans that are accepted based on private hard information on the collateral that is provided by the loan applicant do not differ from the default rates for the group of regularly accepted loans, the default rates for loans that are accepted based on private hard information with the intention to restructure the loans exhibit significantly higher default rates. The debt restructuring process thus still results in a higher default rate. More importantly for the purpose of this paper, the evidence suggests that soft information and the use of discretion is valuable as it allows the loan officer to accept loan applications that are rejected by the internal scoring mechanism, but which have the same performance as those loans that are accepted by the internal scoring mechanism. Put differently, without the loan officer discretion, many loan applications would be rejected that would deserve an acceptance, i.e. the access to financing for creditworthy borrowers is increased by the loan officer discretion, while, at the same time, it does not hurt the bank as default rates remain the same.

This result can also be found for Models 2 to 5, independently of the specific relationship characteristics that are used as control variables. Throughout these models, loans that are accepted based on private soft information and private hard information based on collateral perform as well as loans that are accepted by the internal credit scoring mechanism. Likewise, loans accepted based on private hard information with the intention to restructure existing debt with the savings bank perform worse than the loans that are accepted by the internal scoring. The results for the different relationship control variables are in line with the existing empirical evidence. In particular, customers with an existing relationship are less likely to default than new customers, as shown in Model 2. Furthermore, Model 3 shows that the default probability is continuously decreasing in relationship length, i.e. customers are the less likely to default the longer they have a relationship with their savings bank. Model 4 provides evidence that the default probability also decreases in the intensity in which customer and bank interact.

Customers with debit or credit cards as well as customers with a combination of these cards default significantly less often than customers without any card. Finally, Model 5 shows that the existence of a credit line significantly reduces default rates.

In sum, the evidence in Table VII suggests that there is indeed value in private soft information. Loan officers are right in using their discretion to override the internal credit scoring, and savings banks are right in giving this discretion to their loan officers ex ante.

The empirical specification in Table VII considers only loan applications that are finally accepted, as performance observations are obviously not available for rejected loan applications. To take into account the initial loan approval decision, we follow the same methodology as before and use a Heckman model that considers in its first stage the determinants for the loans to be selected (selection) and in its second stage the determinants for loans to default (outcome). The instrument in the first-stage specification is again the level of bank concentration in a region in which a savings bank operates, as measured by the HHI. The dependent variable in the first stage of Table VIII is a dummy variable that takes a value of one if a loan application is accepted and a value of zero otherwise. The results suggest that the different relationship variables show the same pattern as before. Customers with longer and more intense relationships with their savings banks are more likely to receive a loan than other customers. The results thus mirror those found in the earlier selection model for the loan acceptance analysis.

More importantly for the purpose of this paper, the results in the second stage of Table VIII are very similar to those in Table VII. Loans that are accepted based on hard information on restructuring are significantly more likely to default than loans that are accepted by the internal scoring mechanism. In contrast, loans that are accepted based on private soft information and loans that are accepted based on private hard information based on collateral do perform as well as those loans that are accepted by the internal scoring mechanism. In fact, loans that are accepted based on soft private information even perform significantly better in Model 3 and Model 5 than the regularly accepted loans; their default rates in these models are significantly lower. Likewise, loans that are accepted based on private hard information on collateral perform better than the default group of loans in Model 5. Taken together, the results provide further evidence that loan officers are right in using their discretion, to the mutual benefit of savings banks and their customers. Savings banks are thus right in allowing their loan officers to use their discretion to grant a loan even if the internal scoring mechanism provides a negative outcome on a loan applicant.

V. Conclusion

This paper analyzes whether there is value in soft information that is used in the application process for retail consumer loans.

---To be added---

References

---To be added---

Figure 1

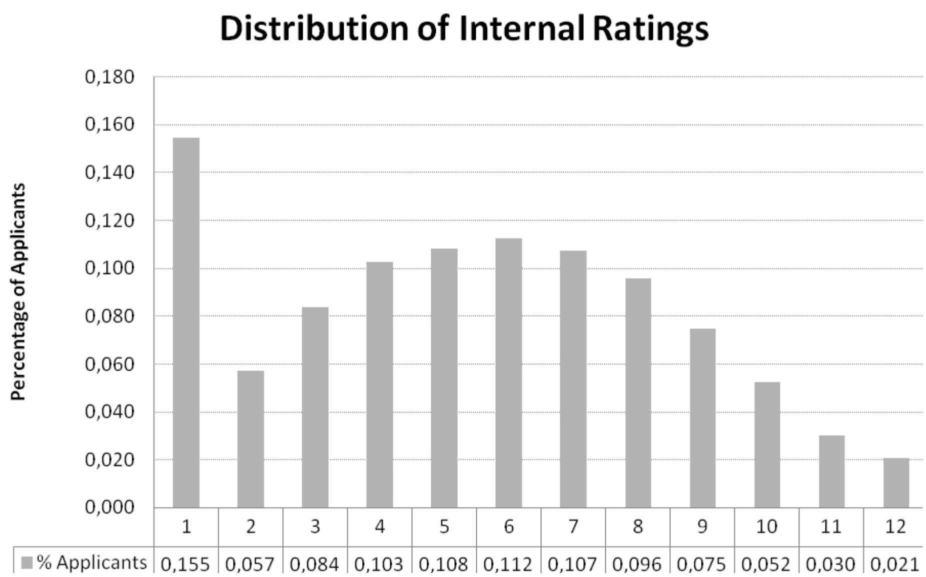
Distribution of Internal Ratings
Percentage of Applicants
0,180
0,160
0,140
0,120
0,100
0,080
0,060
0,040
0,020
0,000
1 2 3 4 5 6 7 8 9 10 11 12
% Applicants 0,155 0,057 0,084 0,103 0,108 0,112 0,107 0,096 0,075 0,052 0,030 0,021

Figure 2

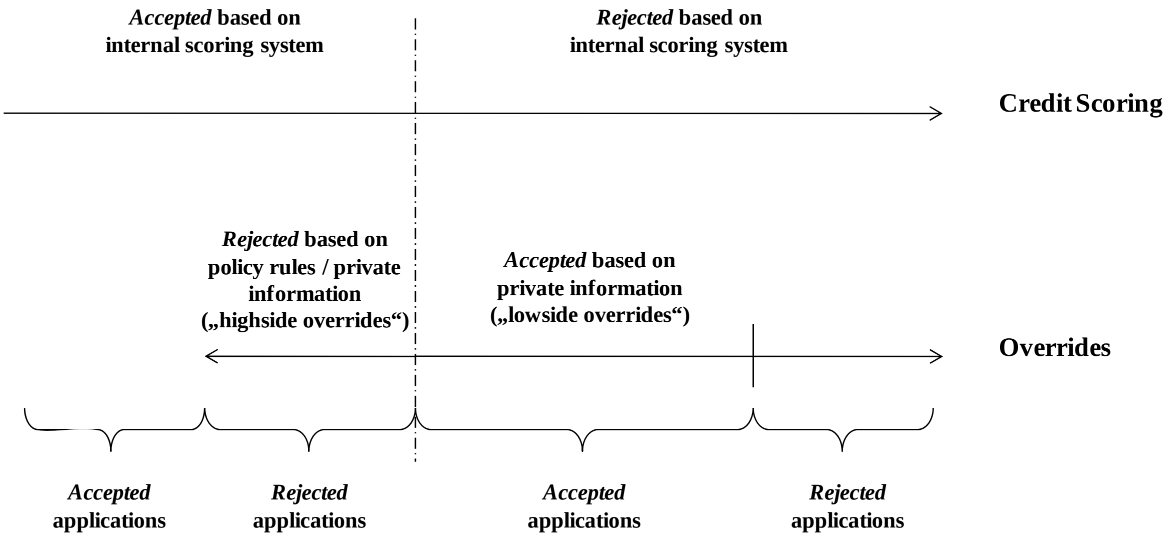
Accepted based on internal scoring system
Rejected based on internal scoring system
Credit Scoring
Rejected based on policy rules / private information („highside overrides“)
Accepted based on private information („lowside overrides“)
Overrides
Accepted applications
Rejected applications
Accepted applications
Rejected applications

Table I. Panel A
Definition of Consumer Loan Characteristics

Loan Characteristics	
Approved	Dummy variable equal to 1 if the applicant is approved.
Default Rate	Dummy variable equal to 1 if the borrower defaults within 12 months after loan origination.
Relationship Characteristics	
Relationship (Yes/No)	Dummy variable equal to 1 if the loan applicant had a checking account with the same bank before the application. The regional principle excludes the possibility that a borrower has relationships with multiple sample banks.
Relationship Length	
Relationship <2years	Dummy variable equal to1 if the relationship length is shorter than 2years.
Relationship >=2, <5years	Dummy variable equal to1 if the relationship length is between 2 and 5 years.
Relationship >=5, <9years	Dummy variable equal to1 if the relationship length is between 5 and 9 years.
Relationship >=9, <15years	Dummy variable equal to1 if the relationship length is between 9 and 15 years.
Relationship >=15	Dummy variable equal to1 if the relationship length is longer than 15 years.
Scope of Relationships: Cards	
Debit and Credit Card	Dummy variable equal to 1 if the borrower has both credit and debit card from the savings bank.
Debit Card	Dummy variable equal to 1 if the borrower has a debit card but not a credit card from the savings bank.
Credit Card	Dummy variable equal to 1 if the borrower has a credit card but not a debit card from the savings bank.
No Cards	Dummy variable equal to 1 if the borrower has neither a credit card nor a debit card from the savings bank.
Scope of Relationships: Credit Line	
Credit Line (Yes/No)	Dummy variable equal to 1 if the borrower has a credit line (which is an overdraft facility associated with the checking account).
Used > 150%	Dummy variable equal to 1 if the borrower has used more than 150% of the credit line.
Used >120%, <= 150%	Dummy variable equal to 1 if the borrower has used more than 120% but less of equal to 150% of the credit line.
Used > 100%, <=120%	Dummy variable equal to 1 if the borrower has used more than 100% but less of equal to 120% of the credit line.
Used > 80%, <=100%	Dummy variable equal to 1 if the borrower has used more than 80% but less of equal to 100% of the credit line.
Used > 20%, <=100%	Dummy variable equal to 1 if the borrower has used more than 20% but less of equal to 100% of the credit line.
Used > 0%, <=20%	Dummy variable equal to 1 if the borrower has used more than 0% but less of equal to 20% of the credit line.
Positive account balance	Dummy variable equal to 1 if the borrower has a positive checking account balance.
Market Data	
Log(HHI)	Natural logarithm of the Hirschman-Herfindahl-Index (HHI) which measures the competition among banks. The number of branches of each particular bank are used to construct the HHI.

Table I. Panel B
Definition of Borrower and Applicant Characteristics

Borrower Characteristics	
Income	
Log(Income)	Log(Income) is the natural logarithm of the monthly net income of the applicant measured in Euro.
Debt-to-Income Ratio (DTI)	This variable is defined as the sum of monthly repayment (principal plus interest) over monthly net income. We use 5 different categories: less than 5%, 5% to 11%, 11% - 13%, 13% - 20% and more than 20%. The higher the ratio, the higher the likelihood that the borrower might have troubles to repay the loan.
< 5%	Dummy variable equal to 1 if DTI is below 5%.
>= 5%, < 11%	Dummy variable equal to 1 if DTI is between 5% and 11%.
>=11%, < 13%	Dummy variable equal to 1 if DTI is between 11% and 13%.
>=13%, < 20%	Dummy variable equal to 1 if DTI is between 13% and 20%.
>20%	Dummy variable equal to 1 if DTI is above 20%.
Age	
18 to 23 years	Dummy variable equal to 1 if the borrower is between 18 and 23 years old.
23 to 25 years	Dummy variable equal to 1 if the borrower is between 23 and 25 years old.
25 to 30 years	Dummy variable equal to 1 if the borrower is between 25 and 30 years old.
30 to 45 years	Dummy variable equal to 1 if the borrower is between 30 and 45 years old.
45 to 50 years	Dummy variable equal to 1 if the borrower is between 45 and 50 years old.
50 to 60 years	Dummy variable equal to 1 if the borrower is between 50 and 60 years old.
> 60 yers	Dummy variable equal to 1 if the borrower is more than 60 years old.
Job / Industry	
Unemployed	Dummy variable equal to 1 if the borrower is unemployed.
Service Sector	Dummy variable equal to 1 if the borrower works in the service sector.
Public Sector	Dummy variable equal to 1 if the borrower works in the public sector.
Retired	Dummy variable equal to 1 if the borrower is retired.
Construction Work	Dummy variable equal to 1 if the borrower is construction worker.
Other	"Other" comprises the following industries: Communications and information; Energy and water supply, mining; Hotel and catering; Municipalities; Agriculture; Banking; Insurance; Not for profit company. It also comprises: Unemployment; Housewife; Apprentice; High School Student; Student; Army; Houseman; Civil Service
Years in Job	
<=2 years	This variable is a measure of job stability. The variable takes the value 1 if the borrower was 2 years or less in her durrent job.
> 2 years	The variable takes the value 1 if the borrower was more than 2 years in her current job.
Ratings	
Internal Rating	
Rating 1 – Rating 12	We segregate the internal rating score in 12 different rating categories based on the default probability of the borrower. Category 1 is the lowest, category 12 the highest default probability, respectively.

Table II. Panel A
Summary Statistics for Consumer Loans

This table presents summary statistics for the sample of 1,091,999 consumer loan applications to 296 savings banks in Germany from 2004 through June 2008. Summary statistics of loan characteristics are calculated at the loan application level.

	No. of observations	Mean	SD	Distribution				
				Min	p25	p50	p75	Max
Loan Characteristics								
Approved[1)]	1,091,999	0.978	0.147	0	1	1	1	1
Default Rate	1,068,000	0.006	0.077	0	0	0	0	1
Relationship Characteristics								
Relationships (Yes)	1,068,000	0.975	0.156	0	1	1	1	1
Relationship Length								
Relationship < 2 years	1,041,291	0.050	0.218	0	0	0	0	1
Relationship >= 2, <5 years	1,041,291	0.093	0.291	0	0	0	0	1
Relationship >= 5, <9 years	1,041,291	0.160	0.366	0	0	0	0	1
Relationship >= 9, <15 years	1,041,291	0.209	0.407	0	0	0	0	1
Relationship >=15 years	1,041,291	0.488	0.500	0	0	0	1	1
Scope of Relationships: Transaction Accounts								
Debit and credit card	1,041,291	0.465	0.499	0	0	0	1	1
Debit card	1,041,291	0.038	0.190	0	0	0	0	1
Credit card	1,041,291	0.183	0.387	0	0	0	0	1
No cards	1,041,291	0.289	0.453	0	0	0	1	1
Credit line	1,041,291	0.945	0.228	0	1	1	1	1
Used > 150%	1,041,291	0.013	0.114	0	0	0	0	1
Used > 120%, <=150%	1,041,291	0.016	0.127	0	0	0	0	1
Used > 100%, <=120%	1,041,291	0.049	0.216	0	0	0	0	1
Used > 80%, <=100%	1,041,291	0.187	0.390	0	0	0	0	1
Used > 20%, <=80%	1,041,291	0.303	0.460	0	0	0	1	1
Used > 0%, <=20%	1,041,291	0.063	0.243	0	0	0	0	1
Positive account balance	1,041,291	0.301	0.459	0	0	0	1	1

Table II. Panel B
Summary Statistics for Borrowers and Applicants

This table presents summary statistics for the sample of 1,091,999 consumer loan applications to 296 savings banks in Germany from 2004 through June 2008. Summary statistics of borrower characteristics are calculated at the loan application level.

	No. of observations	Mean	SD	Distribution Min	p25	p50	p75	Max
Borrower / Applicant Characteristics								
Borrower Income								
Income (monthly)	1,068,000	1,769	1,378	0	1,256	1,600	2,030	758,087
Age								
18 to 23 years	1,068,000	0.048	0.215	0	0	0	0	1
23 to 25 years	1,068,000	0.047	0.211	0	0	0	0	1
25 to 30 years	1,068,000	0.123	0.329	0	0	0	0	1
30 to 45 years	1,068,000	0.365	0.481	0	0	0	1	1
45 to 50 years	1,068,000	0.123	0.328	0	0	0	0	1
50 to 60 years	1,068,000	0.168	0.374	0	0	0	0	1
> 60 years	1,068,000	0.126	0.332	0	0	0	0	1
Debt-to-Income Ratio (DTI)								
< 5%	1,068,000	0.074	0.262	0	0	0	0	1
>= 5%, < 11%	1,068,000	0.277	0.447	0	0	0	1	1
>=11%, < 13%	1,068,000	0.066	0.248	0	0	0	0	1
>=13%, < 20%	1,068,000	0.128	0.334	0	0	0	0	1
>= 20%	1,068,000	0.066	0.248	0	0	0	0	1
Job / Industry								
Unemployed	1,068,000	0.007	0.086	0	0	0	0	1
Service Sector	1,068,000	0.238	0.426	0	0	0	0	1
Public Sector	1,068,000	0.134	0.341	0	0	0	0	1
Retired	1,068,000	0.116	0.321	0	0	0	0	1
Construction Work	1,068,000	0.057	0.232	0	0	0	0	1
Other[2)]	1,068,000	0.239	0.426	0	0	0	0	1
Years in Job								
<=2 years	1,068,000	0.170	0.376	0	0	0	0	1
> 2 years	1,068,000	0.830	0.376	0	1	1	1	1

Table III. Panel A
Summary Statistics for Accepted/Rejected Applications based on Private Information

This table presents summary statistics for the sample of 1,091,999 consumer loan applications to 296 savings banks in Germany from 2004 through June 2008. 176,622 applications have been rejected at by the internal credit scoring but have subsequently been approved by the loan officer. 23,999 applications have finally been rejected.

	No. of			Distribution				
	observations	Mean	SD	Min	p25	p50	p75	Max
Loan Characteristics								
Approved	1,091,999	0.978	0.147	0	1	1	1	1
Approved (with Reason)	1,091,999	0.162	0.364	0	0	0	0	1
Hard Information: Restructuring	176,622	0.133	0.340	0	0	0	0	1
Hard Information: Collateral	176,622	0.087	0.282	0	0	0	0	1
Soft Information	176,622	0.780	0.479	0	0	0	1	1
Rejected (with Reason)								
Internal Ratings	23,999	0.616	0.111	0	0	0	0	1
Hard Information	23,999	0.094	0.232	0	0	0	0	1
Soft Information	23,999	0.290	0.208	0	0	0	0	1

Table III. Panel B
Summary Statistics for Accepted Applications based on Soft and Hard Private Information

This table presents summary statistics for the sample of 1,091,999 consumer loan applications to 296 savings banks in Germany from 2004 through June 2008. The total number of consumer loan applications is segregated by internal rating scores. Column (3) shows the number of total approvals in each rating category. Column (4) shows the number of applications that have been rejected based on internal rating scores but have subsequently been approved by the loan officer; column (5) shows the percentage of (4) relative to the total number of applications. Columns (6) to (8) report the percentage of loans approved under soft information, under hard information (debt restructuring) and hard information (collateral),

(1) Internal Rating	(2) Total No. Applications	(3) No. of Approvals	(4) No. Approvals With Reason	(5) % of approvals	(6) % soft information	(7) % hard information: restructuring	(8) % hard information: collateral
1	165,951	165,095	480	0.29%	11.04%	86.46%	2.50%
2	61,331	60,978	448	0.73%	35.27%	56.25%	8.48%
3	89,935	89,323	1,129	1.26%	58.19%	33.30%	8.50%
4	110,534	109,694	2,503	2.28%	67.32%	25.21%	7.47%
5	116,477	115,466	5,269	4.56%	71.55%	18.01%	10.44%
6	121,238	119,995	13,113	10.93%	77.98%	11.76%	10.26%
7	116,175	114,687	20,040	17.47%	77.67%	11.97%	10.35%
8	103,962	102,339	26,525	25.92%	78.25%	11.75%	10.00%
9	81,981	79,995	30,049	37.56%	79.77%	11.23%	9.00%
10	59,539	55,981	32,421	57.91%	79.95%	11.53%	8.52%
11	36,905	32,203	25,490	79.15%	79.32%	12.85%	7.83%
12	27,971	22,244	19,155	86.11%	77.10%	16.05%	6.85%
Total	1,091,999	1,068,000	176,622				

Table IV
Probit Regression of Loan Acceptance Rates and Private Information

This table presents the results of a Probit regression relating loan approval decisions based on private information to relationship characteristics and borrower risk. The dependent variable is a binary variable equal to 1 if the loan application has been accepted based on the loan officer's private information, 0 if it was rejected. Our main inference variables are relationships characteristics as a result of relationships via transaction account (relationship length, credit and debit cards, credit lines and usage of credit lines). All variables are defined in Table I. Models (2) to (5) consider those borrowers that have a checking account with the savings bank. In (2), the omitted relationship variable are relationships > 15 years; in (3) borrowers without a debit and credit card are omitted; in (5) customers without credit line are omitted. In (6), low quality applicants are omitted. The coefficients for the applicant's job / industry (as described in Table I) as well as intercept and time fixed effects are not shown. Only marginal effects are shown. Heteroscedasticity consistent standard errors are shown in parentheses. ***,**,* denote significance levels at the 1, 5 and 10 percent level, respectively.

	(1)		(2)		(3)		(4)		(5)		(6)	
Left hand side variable	Accepted (Yes=1)		Accepted (Yes=1)		Accepted (Yes=1)		Accepted (Yes=1)		Accepted (Yes=1)		Accepted (Yes=1)	
	Probit		Probit		Probit		Probit		Probit		Probit	
	Marginal Effect	Std. Error	Marginal Effect	Std. Error	Marginal Effect	Std. Error	Marginal Effect	Std. Error	Marginal Effect	Std. Error	Marginal Effect	Std. Error
Relationship Characteristics												
Relationship (Yes)	0.042***	(.003)										
Relationship Length												
Relationship <2years			-0.092***	(.004)								
Relationship >=2, <5years			-0.066***	(.003)								
Relationship >=5, <9years			-0.050***	(.003)								
Relationship >=9, <15years			-0.042***	(.003)								
Scope of Relationships: Cards & Checking Account Information												
Debit and Credit Card					0.023***	(.003)						
Debit Card					0.006***	(.002)						
Credit Card					0.027***	(.004)						
Scope of Relationships: Credit Line												
Credit Line (Yes)							0.101***	(.003)				
Used > 150%									0.068***	(.003)		
Used > 120%, <=150%									0.059***	(.003)		
Used > 100%, <=120%									0.042***	(.003)		
Used > 80%, <=100%									0.022***	(.003)		
Used > 20%, <=80%									0.028***	(.003)		
Used > 0%, <=20%									0.005	(.004)		
Positive Account Balance									0.001	(.003)		
High Quality Applicant											-0.288***	(.005)
Medium Quality Applicant											0.047***	(.001)

Table IV (cont'd)
Probit Regression of Loan Acceptance Rates and Private Information

	(1)		(2)		(3)		(4)		(5)		(6)	
Left hand side variable	Accepted (Yes=1) Probit		Accepted (Yes=1) Probit		Accepted (Yes=1) Probit		Accepted (Yes=1) Probit		Accepted (Yes=1) Probit		Accepted (Yes=1) Probit	
	Marginal Effect	Std. Error	Marginal Effect	Std. Error	Marginal Effect	Std. Error	Marginal Effect	Std. Error	Marginal Effect	Std. Error	Marginal Effect	Std. Error
Borrower Characteristics												
Internal Rating												
1	-0.397***	(.013)	-0.486***	(.013)	-0.417***	(.013)	-0.372***	(.013)	-0.368***	(.013)		
2	-0.236***	(.015)	-0.313***	(.016)	-0.254***	(.015)	-0.218***	(.014)	-0.213***	(.014)		
3	-0.136***	(.009)	-0.209***	(.011)	-0.153***	(.01)	-0.129***	(.009)	-0.115***	(.009)		
4	-0.047***	(.006)	-0.106***	(.007)	-0.059***	(.006)	-0.046***	(.006)	-0.034***	(.005)		
5	0.023***	(.003)	-0.011***	(.004)	0.017***	(.004)	0.023***	(.003)	0.032***	(.003)		
6	0.082***	(.002)	0.062***	(.002)	0.077***	(.002)	0.079***	(.002)	0.084***	(.002)		
7	0.096***	(.001)	0.081***	(.002)	0.092***	(.001)	0.093***	(.001)	0.098***	(.001)		
8	0.106***	(.001)	0.094***	(.001)	0.102***	(.001)	0.103***	(.001)	0.107***	(.001)		
9	0.108***	(.001)	0.098***	(.001)	0.104***	(.001)	0.104***	(.001)	0.109***	(.001)		
10	0.083***	(.002)	0.075***	(.002)	0.080***	(.002)	0.079***	(.002)	0.086***	(.002)		
11	0.046***	(.002)	0.040***	(.002)	0.044***	(.002)	0.043***	(.002)	0.049***	(.002)		
Additional control variables	Time Fixed Effects		Time Fixed Effects		Time Fixed Effects		Time Fixed Effects		Time Fixed Effects		Time Fixed Effects	
N	221,980		208,203		208,203		208,203		208,203		208,203	
Pseudo R2	0.073		0.080		0.075		0.082		0.081		0.041	

Table V

1st Stage (Selection Equation): Loan Rejections and Internal Scoring

This table presents the results from a binomial Probit model with selection. The dependent variable in the outcome equation is a binary variable equal to 1 if the application has been approved based on the loan officer's private information. The dependent variable in the selection equation is a binary variable equal to 1 if the applicant was rejected based on the internal credit score. The control variables used are the same as in the previous models, as well as an additional instrument. Log (HHI) is the natural logarithm of the Hirshman-Herfindahl Index which measures the competition among banks. The number of branches of each particular bank is used to construct the HHI. Only marginal effects are reported. All variables are defined in Table I. Heteroscedasticity consistent standard errors are shown in parentheses. ***,**,* denote significance levels at the 1, 5 and 10 percent level, respectively.

	(1)		(2)		(3)		(4)		(5)		(6)	
Left hand side variable	Rejected (Yes=1)		Rejected (Yes=1)		Rejected (Yes=1)		Rejected (Yes=1)		Rejected (Yes=1)		Rejected (Yes=1)	
Instrument												
Log(HHI)	0.085***	(.001)	0.084***	(.001)	0.082***	(.001)	0.086***	(.001)	0.086***	(.001)	0.089***	(.001)
Relationship Characteristics												
Relationship Yes/No												
Relationship (Yes)	-0.066***	(.002)										
Relationship Length												
Relationship <2years			0.137***	(.002)								
Relationship >=2, <5years			0.084***	(.001)								
Relationship >=5, <9years			0.056***	(.001)								
Relationship >=9, <15years			0.042***	(.001)								
Scope of Relationships: Cards & Checking Account Information												
Debit and Credit Card					-0.105***	(.001)						
Debit Card					-0.109***	(.001)						
Credit Card					-0.082***	(.001)						
Scope of Relationships: Credit Line												
Credit Line (Yes)							-0.032***	(.001)				
Used > 150%									0.191***	(.004)		
Used > 120%, <=150%									0.132***	(.003)		
Used > 100%, <=120%									0.051***	(.002)		
Used > 80%, <=100%									0.004***	(.001)		
Used > 20%, <=80%									-0.054***	(.001)		
Used > 0%, <=20%									-0.065***	(.001)		
Positive Account Balance									-0.069***	(.001)		
High Quality Applicant											-0.523***	(.002)
Medium Quality Applicant											-0.283***	(.002)
Additional control variables	Internal Ratings, Time Fixed Effects		Internal Ratings, Time Fixed Effects		Internal Ratings, Time Fixed Effects		Internal Ratings, Time Fixed Effects		Internal Ratings, Time Fixed Effects		Internal Ratings, Time Fixed Effects	
N (censored)	984,307		966,826		966,826		966,826		966,826		984,307	
N (uncensored)	215,623		202,239		202,239		202,239		202,239		215,623	
Wald Test	ρ=-.0742 (p=0.036)		ρ=-.17554 (p<0.001)		ρ=-.65093(p<0.001)		ρ=.04□56 (□=0.301□		ρ=-.4469 (p<0.001)		ρ=-.1567 (p<0.001)	
	X^2(13)=9512.68		X^2(16)=9922.51		X^2(□5)=13352.45		X^2(13)=10081.17		X^2(19)=11472.45		X^2(4)□5532.43	

	Prob > X²=0.000	Prob > X²=0.000	Prob > X²=0.000	Prob > X²=0.000	Prob > X²=0.000	Prob > X²=0.000

Table V (cont'd)
2nd Stage (Outcome Equation): Loan Acceptance Rates and Private Information

	(1)		(2)		(3)		(4)		(5)		(6)	
Left hand side variable	Accepted (Yes=1)		Accepted (Yes=1)		Accepted (Yes=1)		Accepted (Yes=1)		Accepted (Yes=1)		Accepted (Yes=1)	
Relationship Characteristics												
Relationship Yes/No												
Relationship (Yes)	0.065***	(.005)									0.055***	(.004)
Relationship Length												
Relationship <2years			-0.117***	(.005)								
Relationship >=2, <5years			-0.068***	(.004)								
Relationship >=5, <9years			-0.039***	(.003)								
Relationship >=9, <15years			-0.025***	(.003)								
Scope of Relationships: Cards												
Debit and Credit Card					0.023***	(.001)						
Debit Card					0.024***	(.001)						
Credit Card					0.021***	(.001)						
Scope of Relationships: Credit Line												
Credit Line (Yes)							0.157***	(.008)				
Used > 150%									0.018***	(.004)		
Used > 120%, <=150%									0.014***	(.003)		
Used > 100%, <=120%									0.013***	(.003)		
Used > 80%, <=100%									0.011***	(.002)		
Used > 20%, <=80%									0.028***	(.003)		
Used > 0%, <=20%									0.019***	(.002)		
Positive Account Balance									0.020***	(.002)		
Borrower Characteristics												
Internal Rating												
1	-0.354***	(.047)	-0.274***	(.045)	0.024***	(.001)	-0.366***	(.052)	0.018***	(.008)		
2	-0.276***	(.046)	-0.204***	(.042)	0.022***	(.001)	-0.249***	(.049)	0.034***	(.003)		
3	-0.148***	(.037)	-0.096***	(.029)	0.030***	(.001)	-0.092**	(.037)	0.051***	(.002)		
4	-0.001	(.02)	0.016	(.013)	0.035***	(.002)	0.008	(.025)	0.060***	(.004)		
5	0.091***	(.007)	0.081***	(.003)	0.037***	(.002)	0.091***	(.011)	0.066***	(.005)		
6	0.140***	(.004)	0.113***	(.005)	0.039***	(.002)	0.145***	(.003)	0.070***	(.006)		
7	0.151***	(.006)	0.119***	(.007)	0.038***	(.002)	0.163***	(.005)	0.068***	(.006)		
8	0.150***	(.007)	0.117***	(.008)	0.035***	(.002)	0.168***	(.008)	0.065***	(.006)		
9	0.120***	(.006)	0.095***	(.006)	0.029***	(.002)	0.156***	(.008)	0.058***	(.005)		
10	0.081***	(.005)	0.063***	(.004)	0.022***	(.001)	0.130***	(.007)	0.049***	(.004)		
11	0.038***	(.004)	0.029***	(.004)	0.012***	(.001)	0.079***	(.005)	0.031***	(.003)		
High Quality Applicant											-0.131***	(.022)
Medium Quality Applicant											0.082***	(.001)

Table VI
Loan Acceptances and Soft versus Hard Information

This table presents the results from a multinomial Logit regression where the base case is if the loan application has been accepted based on soft information. The coefficients reported in this table are to be interpreted as the multinomial logits compared to the base case given all other predictor variables are held constant. All variables are defined in Table I. Heteroscedasticity consistent standard errors are shown in parentheses. ***,**,* denote significance levels at the 1, 5 and 10 percent level, respectively.

	(1)			
Left hand side variable	Restructuring vs. Soft		Collateral vs. Soft	
	Marginal Effect	Std. Error	Marginal Effect	Std. Error
Scope of Relationships: Credit Line				
Used > 150%	1.488***	(.037)	0.268***	(.045)
Used > 120%, <=150%	1.301***	(.037)	0.240***	(.044)
Used > 100%, <=120%	1.149***	(.029)	0.050	(.034)
Used > 80%, <=100%	0.638***	(.025)	-0.186***	(.027)
Used > 20%, <=80%	-0.041	(.027)	-0.166***	(.026)
Used > 0%, <=20%	-0.875***	(.063)	-0.129***	(.044)
Positive Account Balance	-1.432***	(.036)	-0.038	(.025)
Age				
18 to 23 years	-0.885***	(.058)	-0.503***	(.06)
23 to 25 years	-0.701***	(.058)	-0.556***	(.061)
25 to 30 years	-0.576***	(.055)	-0.541***	(.059)
30 to 45 years	-0.242***	(.053)	-0.340***	(.057)
45 to 50 years	-0.117**	(.056)	-0.196***	(.061)
50 to 60 years	-0.146***	(.053)	-0.059	(.058)
Additional control variables	Debt-to-income ratio, Log (Income), Job / Industry, In job since, Time dummy			
N	176,621			
Pseudo R2	0.104			

Table VII
Default Rates and Soft and Hard Information

This table presents the results of a Probit regression relating borrower defaults to whether or not the application has previously been approved based on deterministic credit scores or loan officer's soft or hard private information. The dependent variable is a binary variable equal to 1 if the borrower defaults within the first 12 months after loan origination. Our main inference variables are whether the applicant has been accepted based on internal ratings (omitted), soft information, hard information (restructuring) or hard information (collateral). Relationship variables come from relationships via transaction account (relationship length, credit and debit cards, credit lines and usage of credit lines). All variables are defined in Table I. Models (2) to (5) consider those borrowers that have a checking account with the savings bank. In (3), the omitted relationship variable are relationships > 15 years; in (4) borrowers without a debit and credit card are omitted; in (5) customers without credit line are omitted. The coefficients of the intercept and time fixed effects are not shown. Only marginal effects are shown. Heteroscedasticity consistent standard errors are shown in parentheses. ***,**,* denote significance levels at the 1, 5 and 10 percent level, respectively.

	(1)		(2)		(3)		(4)		(5)	
Left hand side variable	Default (Yes=1)		Default (Yes=1)		Default (Yes=1)		Default (Yes=1)		Default (Yes=1)	
	Probit		Probit		Probit		Probit		Probit	
	Marginal Effect	Std. Error	Marginal Effect	Std. Error	Marginal Effect	Std. Error	Marginal Effect	Std. Error	Marginal Effect	Std. Error
Accepted (Private *Soft* Info)	-0.001	(.001)	-0.109	(.103)	-0.101	(.094)	-0.001	(.001)	-0.109	(.095)
Accepted (Private *Hard* Info - Restructuring)	0.011**	(.004)	0.011***	(.004)	0.011***	(.004)	0.011**	(.004)	0.011**	(.004)
Accepted (Private *Hard* Info - Collateral)	-0.002	(.001)	-0.002	(.001)	-0.001	(.001)	-0.001	(.001)	-0.001	(.001)
Relationship Characteristics										
Relationship (Yes)			-0.002***	(.001)						
Relationship Length										
Relationship <2years					0.007***	(.001)				
Relationship >=2, <5years					0.004***	(.001)				
Relationship >=5, <9years					0.002***	(.001)				
Relationship >=9, <15years					0.001*	<0.001				
Scope of Relationships: Cards										
Debit and Credit Card							-0.002***	<0.001		
Debit Card							-0.002***	<0.001		
Credit Card							-0.001***	<0.001		
Scope of Relationships: Credit Line										
Credit Line (Yes)									-0.007***	(.001)
Additional control variables	Internal Ratings, Time Fixed Effects		Internal Ratings, Time Fixed Effects		Internal Ratings, Time Fixed Effects		Internal Ratings, Time Fixed Effects		Internal Ratings, Time Fixed Effects	
N	1,068,000		1,068,000		1,041,291		1,041,291		1,041,291	
Pseudo R2	0.067		0.067		0.073		0.067		0.071	

Table VIII
1st Stage (Selection Equation): Screening Loan Applications

This table presents the results from a binomial Probit model with selection. The dependent variable in the outcome equation is a binary variable equal to 1 if the borrower defaults within the first 12 months after loan origination. The dependent variable in the selection equation is a binary variable equal to 1 if the applicant was approved. The control variables used are the same as in the previous models, as well as an additional instrument. Log (HHI) is the natural logarithm of the Hirshman-Herfindahl Index which measures the competition among banks. The number of branches of each particular bank is used to construct the HHI. Only marginal effects are reported. All variables are defined in Table I. Heteroscedasticity consistent standard errors are shown in parentheses. ***,**,* denote significance levels at the 1, 5 and 10 percent level, respectively.

	(1)		(2)		(3)		(4)		(5)	
Left hand side variable	Accepted (Yes=1)		Accepted (Yes=1)		Accepted (Yes=1)		Accepted (Yes=1)		Accepted (Yes=1)	
Instrument										
Log(HHI)	-0.146***	(.009)	-0.151***	(.009)	-0.182***	(.01)	-0.146***	(.01)	-0.199***	(.01)
Relationship Characteristics										
Relationship Yes/No										
Relationship (Yes)			0.152***	(.012)						
Relationship Length										
Relationship <2years					-0.250***	(.012)				
Relationship >=2, <5years					-0.168***	(.01)				
Relationship >=5, <9years					-0.121***	(.009)				
Relationship >=9, <15years					-0.103***	(.009)				
Scope of Relationships: Cards										
Debit and Credit Card							0.115***	(.011)		
Debit Card							0.107***	(.007)		
Credit Card							0.054***	(.017)		
Scope of Relationships: Credit Line										
Credit Line (Yes)									0.299***	(.009)
Additional control variables	Internal Ratings, Time Fixed Effects		Internal Ratings, Time Fixed Effects		Internal Ratings, Time Fixed Effects		Internal Ratings, Time Fixed Effects		Internal Ratings, Time Fixed Effects	
N (censored)	27,084		27,084		24,773		24,773		24,773	
N (uncensored)	1,172,846		1,172,846		1,144,292		1,144,292		1,144,292	
Wald Test	□= -.4366 ($p<0.001$)		$\lambda=-.3452$ ($p<0.001$)		$\lambda=-.23175$ ($p<0.001$)		$\lambda= -.37544$ ($p<0.001$)		$\lambda=-.0395549$ ($p<0.001$)	
	$X^2(15)=183.85$		$X^2(16)=293.05$		$X^2(19)=641.83$		$X^2(18)=241.30$		$X^2(16)=4036.1$□	
	Prob > $X^2=0.000$		Prob > $X^2=0.000$		Prob > $X^2=0.000$		Prob > $X^2=0.000$		Prob > $X^2=0.00$□	

Table VIII
2nd Stage (Outcome Equation): Default Rates and Soft and Hard Information

	(1)		(2)		(3)		(4)		(5)	
Left hand side variable	Default (Yes=1)		Default (Y□s=1)		Default (Yes=1)		Default (Yes=1)		Default (Yes=1)	
Accepted (Private *Soft* Info)	-0.0□2	(.001)	-0.002	(.001)	-0.002**	(.001)	-0.002	(.001)	-0.002***	<0.0001
Accepted (Private *Hard* Info - Restructuring)	0.023***	(.002)	0.023***	(.002)	0.024***	(.001)	0.023***	(.002)	0.024***	<0.0001
Accepted (Private *Hard* Info - Collateral)	-0.002	(.003)	-0.002	(.002)	-0.002	(.001)	-0.002	(.002)	-0.003***	(.001)
Relationship Characteristics										
Relationship Yes/No										
Relationship (Yes)			-0.013***	(.002)						
Relationship Length										
Relationship <2years					0.015***	(.001)				
Relationship >=2, <5years					0.007***	(.001)				
Relationship >=5, <9years					0.003***	(.001)				
Relationship >=9, <15years					0.002***	(.001)				
Scope of Relationships: Cards										
Debit and Credit Card							-0.005***	(.001)		
Debit Card							-0.005***	(.001)		
Credit Card							-0.001	(.002)		
Scope of Relationships: Credit Line										
Credit Line (Yes)									-0.007***	<0.0001
Additional control variables	Internal Ratings, Time Fixed Effects		Internal Ratings, Time Fixed Effects		Internal Ratings, Time Fixed Effects		Internal Ratings, Time Fixed Effects		Internal Ratings, Time Fixed Effects	