brought to you by CORE
provided by Research Papers in Economics

EUROPEAN CENTRAL BANK

EUROSYSTEM

ECB LAMFALUSSY FELLOWSHIP PROGRAMME

WORKING PAPER SERIES
NO 1395 / NOVEMBER 2011

ON THE IMPORTANCE OF PRIOR RELATIONSHIPS IN BANK LOANS TO RETAIL CUSTOMERS

by Manju Puri,
Jörg Rocholl
and Sascha Steffen

EUROPEAN CENTRAL BANK

EUROSYSTEM

WORKING PAPER SERIES

NO 1395 / NOVEMBER 2011

ECB LAMFALUSSY FELLOWSHIP PROGRAMME

ON THE IMPORTANCE OF PRIOR RELATIONSHIPS IN BANK LOANS TO RETAIL CUSTOMERS[1]

by Manju Puri[2], Jörg Rocholl[3], and Sascha Steffen[4]

In 2011 all ECB publications feature a motif taken from the €100 banknote.

NOTE: This Working Paper should not be reported as representing the views of the European Central Bank (ECB). The views expressed are those of the authors and do not necessarily reflect those of the ECB.

This paper can be downloaded without charge from http://www.ecb.europa.eu or from the Social Science Research Network electronic library at http://ssrn.com/abstract_id=1572673.

1 We thank the Deutscher Sparkassen- und Giroverband (DSGV) for providing us with the data. Sascha Steffen's contribution to the paper has been prepared under the Lamfalussy Fellowship Program sponsored by the European Central Bank. We thank Rebel Cole, Hans Degryse, Valeriya Dinger, Radhakrishnan Gopalan, Reint Gropp, David Musto, Lars Norden, Martin Weber, Vijay Yeramilli, participants at the EFA 2010 Frankfurt meeting, the FDIC-JFSR Bank Research Conference, the FMA 2010 meeting, the CAREFIN 2010 Conference at Bocconi, the German Finance Association Meeting (DGF), and seminar participants at Drexel University, Erasmus University Rotterdam, Georgia Tech University, University of Cologne, University of Mannheim, and University of Michigan for comments and suggestions.

2 Duke University, Durham, NC 27708, USA, and NBER; e-mail: mpuri@duke.edu.

3 European School of Management and Technology, Schloßplatz 1, 10178 Berlin, Germany; e-mail: rocholl@esmt.org.

4 University of Mannheim, L 5,2 , 68131 Mannheim, Germany; e-mail: steffen@bank.bwl.uni-mannheim.de.

Lamfalussy Fellowships

This paper has been produced under the ECB Lamfalussy Fellowship programme. This programme was launched in 2003 in the context of the ECB-CFS Research Network on "Capital Markets and Financial Integration in Europe". It aims at stimulating high-quality research on the structure, integration and performance of the European financial system.

The Fellowship programme is named after Baron Alexandre Lamfalussy, the first President of the European Monetary Institute. Mr Lamfalussy is one of the leading central bankers of his time and one of the main supporters of a single capital market within the European Union.

Each year the programme sponsors five young scholars conducting a research project in the priority areas of the Network. The Lamfalussy Fellows and their projects are chosen by a selection committee composed of Eurosystem experts and academic scholars. Further information about the Network can be found at http://www.eu-financial-system.org and about the Fellowship programme under the menu point "fellowships".

© European Central Bank, 2011

Address
Kaiserstrasse 29
60311 Frankfurt am Main, Germany

Postal address
Postfach 16 03 19
60066 Frankfurt am Main, Germany

Telephone
+49 69 1344 0

Internet
http://www.ecb.europa.eu

Fax
+49 69 1344 6000

All rights reserved.

Any reproduction, publication and reprint in the form of a different publication, whether printed or produced electronically, in whole or in part, is permitted only with the explicit written authorisation of the ECB or the author(s).

Information on all of the papers published in the ECB Working Paper Series can be found on the ECB's website, http://www.ecb.europa.eu/pub/scientific/wps/date/html/index.en.html

ISSN 1725-2806 (online)

CONTENTS

Abstract 4

Non-technical summary 5

1 Introduction 6

2 Data and summary statistics 13
- A Loan and borrower characteristics 13
- B Relationship characteristics 17

3 Empirical results on private information 18
- A Univariate results 18
- B Multivariate results 21

4 Private information and borrower incentives to default 39

5 Conclusion 43

References 45

Appendices 48

Tables 51

Abstract

This paper analyzes the importance of retail consumers' banking relationships for loan defaults using a unique, comprehensive dataset of over one million loans by savings banks in Germany. We find that loans of retail customers, who have a relationship with their savings bank *prior* to applying for a loan, default significantly less than customers with no prior relationship. We find relationships matter in different forms, scope, and depth. Importantly, though, even the simplest forms of relationships such as transaction accounts are economically meaningful in reducing defaults, even after controlling for other borrower characteristics as well as internal and external credit scores. Our results suggest that relationships of all kinds have inherent private information and are valuable in screening, in monitoring, and in reducing consumers' incentives to default.

JEL: G20, G21

Keywords: Retail banking, relationships, default rates, monitoring, screening

Non-Technical Summary

This paper analyses the importance of relationships between banks and depositors on borrower default rates. Loan officers incorporate private information in the credit decision process as well as in monitoring. We ask, is this relationships specific information valuable to banks as well as borrowers? Are default rates effectively reduced? Does private information help banks to become better at screening and to what extend does it influence the monitoring process? In addition to that, we analyse borrower incentives to default conditioning on the intensity of the relation with the bank.

We use a unique dataset that has information on consumer loans applied for as well as originated by savings banks in Germany. Savings banks do 40% of retail banking in Germany, so this is a significant source of credit for retail customers. The sample spans the time period between 2004 and 2008 and has information on the performance of more than 1 million loans made by 296 different savings banks. The default rates for these loans are calculated in compliance with the Basel II requirements. In addition to the performance data, the dataset contains detailed information on loan and borrower characteristics and in particular 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. These relationships comprise of the existence of a current or savings account, the usage of credit or debit cards, of credit lines, the amount of funds in these accounts as well as the existence and performance of a prior loan. The available data also include detailed information on each borrower, including age, income, employment status, and the length of the relationship with the bank.

In other words, the data comprise information about the existence, scope and depth of the relationship and, in contrast to prior literature, not only related to repeat loan relationships, but also other (cross-selling) products, for example, checking and savings accounts at the time the customer applies for the loan. Using selection methods, it is possible to address the question whether banks use their private information rather in screening than in monitoring borrowers. Using additional information about transaction account behaviour of our sample borrowers, we are able to separate screening and monitoring from the question as to whether or not borrowers with relationships are less inclined to default.

We find that loans of retail customers, who have a relationship with their savings bank *prior* to applying for a loan, default significantly less than customers with no prior relationship. We find relationships matter in different forms (transaction accounts, savings accounts, prior loans), in scope (credit and debit cards, credit lines), and depth (relationship length, utilization of credit line, money invested in savings account). Importantly, though, even the simplest forms of relationships such as transaction accounts (e.g., savings or checking accounts) are economically meaningful in reducing defaults, even after controlling for other borrower characteristics as well as internal and external credit scores. We are able to access data on loan applications to assess how banks screen. We find that relationships are important in screening but even after taking screening into account relationships have a first order impact in reducing borrower default. Our results suggest that relationships of all kinds have inherent private information and are valuable in screening, in monitoring, and in reducing consumers' incentives to default.

1. Introduction

Understanding how banks make loans and under which conditions borrowers default on these 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 based on hard, verifiable information collected by the bank? Allowing discretion to the bank could allow for the information obtained from relationship specific assets to be incorporated to improve the quality of loans made. Likewise, what is the value of a bank relationship to a customer? Is the bank better able to prevent default because of prior relationships? Is a borrower less inclined to default on a loan if she has an extensive relationship with his bank, because of the inherent value of the relationship? These are open questions that are of interest to academics, banks, consumers, and regulators.

There is a vast theoretical literature on the relationships between banks and their customers.[1] Boot (2000) states, *"The modern literature on financial intermediaries has primarily focused on the role of banks as relationship lenders… (However) existing empirical work is virtually silent on identifying the precise sources of value in relationship banking."* The importance of these relationships has been documented in various contexts and in particular for banks' lending to corporate customers.[2]

[1] See, for example, Campbell and Kracaw (1980), Diamond (1984, 1991), Ramakrishnan and Thakor (1984), Fama (1985), and Haubrich (1989).
[2] See James and Wier (1990), Petersen and Rajan (1994), Berger and Udell (1995), Puri (1996), Billet, Flannery, and Garfinkel (1995), Drucker and Puri (2005), and Bharath, Dahiya, Saunders, and Srinivasan (2006).

Our paper adds to this literature studying bank-depositor relationships. In particular, it focuses on the importance of existing relationships for both the bank, which can collect information, and the customer, who has an incentive to maintain his relationship, by analyzing the loan approval decision and subsequent loan performance. Given the significance of retail lending and deposit-taking for banks, and given that banks are a valuable source of personal and consumer loans, understanding the role of bank and retail depositor relationships is important. We ask both, how and what kind of relationships matter in the granting of loans, as well as whether they affect default rates.

The first key contribution of this paper is to recognize that relationships have multiple dimensions which is essential in understanding both how banks collect private information as well as how borrower and bank incentives are shaped. There are many different ways of thinking about relationships. One could look at the length of relationships, the scope of relationships, or the kind of relationships - whether it is a simple transaction account or a multi-prong relationship. The literature has largely defined relationships in the context of giving repeat loans to corporate firms, but in principle simple transaction relationships, or having multiple products with the bank could matter.[3] A second key contribution of our paper is that we examine the impact of different kinds of relationships that existed *prior* to granting the loan in reducing default rates. Specifically, we show that these relationships matter in various forms, scope, and depth, and even simple transaction or savings accounts make a difference. This is distinct from information obtained from concurrent transaction or checking accounts opened at the time of making the loan. From a practical point of view, our results imply that banks can make better

[3] See e.g. Santikian (2009) who studies banks' profit margins based on the cross-selling of non-loan products to firms.

credit decisions by requiring potential borrowers to open simple savings or checking accounts and observing their transactions before deciding on the loan application. A third key contribution of this paper is that we examine the sources of value of relationships at the loan origination stage and find that relationships play an important role at screening loan applicants, suggesting that the private information inherent in relationships is important. Even after taking screening into account, relationships still have a first order impact in reducing borrower defaults. This suggests a distinct value of existing relationships not just in screening but beyond potentially from better monitoring based on private information as well as reduced incentives to default by the customer. To the best of our knowledge, these results are new to the literature and illustrate the value of relationships to both banks and customers.

A major limitation in studying the importance of retail banking relationships is the availability of data in the context of an appropriate experiment design. This paper accesses a unique, proprietary dataset which comprises the universe of loans made by savings banks in Germany as well as their ex-post performance. These data are recorded on a monthly basis for each individual loan and are provided by the rating subsidiary of the German Savings Banks Association (DSGV). The data span the time period between November 2004 and June 2008 and comprise information on the performance of more than 1 million loans made by 296 different savings banks. The default rates for these loans are calculated in compliance with the Basel II requirements. In addition to the performance data, we have detailed information on loan and borrower characteristics and in particular 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. These relationships comprise the existence of a current or savings account, the usage of credit or debit

cards, the amount of funds in these accounts as well as the existence and performance of a prior loan. The available data also comprise detailed information on each borrower, including age, income, employment status, and the length of the relationship with the bank. All characteristics are taken from an internal scoring system that is used by all our sample banks and available for all loan applications. In addition, for a subset of the loan applications we also have detailed borrower information that is not part of the internal scoring system and only known to the savings banks. Finally, for a substantial number of loan applications we also have information from an external scoring system. The important aspect for our analysis of the bank behavior is that the scoring system provides a credit assessment of each loan applicant and a recommendation for the loan decision, but the final decision remains with the bank and its loan officers. The final loan granting decision is thus made by each individual bank, 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 they approve a loan application. In other words, there are some subjective elements in the screening process that might very well be different for each respective bank and loan officer. These data thus provide an ideal opportunity to investigate the sources of value of relationships from being able to collect more information on a customer.

Our first set of tests examines whether loans with prior relationships have lower default rates after controlling for observable borrower characteristics. We use a number of proxies for the different forms of relationships: First, we examine the impact of relationships through *transaction accounts* on default rates using five measures: (i) the existence of checking accounts, (ii) relationship length, (iii) the usage of debit and credit cards, (iv) the existence of credit lines

and (v) the usage of credit lines. Second, we examine the impact of relationships through *savings accounts* on default rates using two measures: (i) the existence of savings accounts and (ii) the amount of assets held in the savings accounts. Third, we examine the impact of relationships through *repeat lending* on default rates. To summarize our results, we find that relationships that have been built prior to loan origination significantly reduce the probability of default of subsequently issued loans after controlling for borrower risk characteristics as well as internal and external credit scores. This result is consistent with relationships both providing banks with a unique advantage in monitoring their borrowers and creating incentives for customers to default less often. We also examine the relative importance of each of our relationship proxies. While prior literature highlights the importance of repeat lending relationships, this proxy turns out to have a rather small impact on default rates relative to, for example, transaction account related measures.

While these results establish a correlation between having prior relationships and default rates, one can still ask what determines a relationship itself. If relationships are not random but are related to certain (unobservable) borrower characteristics, relationship borrowers might be of higher quality which explains lower default rates. We address this using a simultaneous equation model in which we augment the main probit equation with an additional probit equation that explains what factors determine relationships. To facilitate identification, we include an instrument that proxies for the availability of savings banks to customers in their region. We test the null hypothesis that both probit equations are uncorrelated and cannot reject this hypothesis at conventional levels. These results suggest that there are no unobservable borrower characteristics that bias our estimates of the impact of prior relationships on default rates.

In a second set of tests we examine the sources of value of relationships. Do existing banking relationships with retail consumers help banks to better screen these consumers when they apply for loans and thus to reduce the default rates for these loans? Is there value to relationships beyond screening? If so, does it stem from private information or other sources?

In order to separate screening from other benefits of relationships, we need to explicitly analyze the loan granting process as we cannot observe the loan performance for those customers whose loan application has been rejected. We use a simultaneous equation model augmenting the default model with a second probit model that explains the loan granting decision. We find that borrower characteristics that increase the likelihood of getting credit are negatively correlated with default rates, which is consistent with banks using a screening policy to reduce default rates. We further test the null hypothesis that the error terms of the loan granting and the default model are uncorrelated (i.e. discretion does not matter for screening) and reject this hypothesis at any confidence level. We also find that after controlling for sample selection, our proxies for relationships are still negative and significant. Relationships thus provide value to banks in screening, but they also provide value beyond this.

To investigate further the source of value of relationships, we make use of the detailed information about transaction account behavior for a subset of our sample borrowers, which is only known to the bank, but not included in the internal rating. Our results suggest that private information is important both for screening and subsequent monitoring, but the different relationship proxies still have explanatory power even after controlling for private information.

These results suggest that other factors beyond private information are important for loan performance and borrower defaults. One potential explanation of our results is that there are reduced borrower incentives to default because of the potential value of relationships to the borrower.

Our paper adds to the existing literature in several ways. There is a recent literature that analyzes the benefits of bundling loans and checking accounts (Mester, Nakamura, and Renault (2007) and Norden and Weber (2009)).[4] These papers explore the information banks gain over the duration of the loans from checking account activity. Mester, Nakamura, and Renault (2007) find that transaction accounts provide financial intermediaries with a stream of information for the monitoring of small-business borrowers that gives them an advantage over other lenders.[5] Similarly, Norden and Weber (2009) show that checking account activity provides valuable information for banks as an early warning signal for the default of small firms and their subsequent loan contract terms. Related to these two papers, Agarwal, Chomsisengphet, Liu, and Souleles (2009) document for credit card customers that monitoring and thus the availability of information on the changes in customer behavior result in an advantage to relationship banking. Our paper differs from theirs along several dimensions. While it is common to ask borrowers taking a loan to open an account and important to study how the information in the account helps the bank, i.e. instead of analyzing the benefits of providing jointly a loan and a checking account to the same borrower, we examine the impact of relationships that existed *prior* to granting the loan. Next, we show that relationships matter in various forms, scope and depth. Further, instead

[4] This literature is related but distinct from the literature examining the importance of relationships for small firm credit (Berger and Udell, 1995; Cole, 1998; Petersen and Rajan, 1994) .
[5] For small and medium-sized business borrowers, there is also a growing literature on the collection and use of soft information (Agarwal and Hauswald, 2007) as well as the use of discretion by banks (Cerqueiro, Degryse, and Ongena, 2007).

of analyzing the behavior of one bank we examine the loan making decision of 296 different banks. Finally, we find evidence suggesting screening, monitoring, and borrower incentives as distinct sources of value of relationships. Our paper also adds to the literature on traditional bank specialness (such as James (1987), Lummer and McConnell (1989), Best and Zhang (1993), Billet et al. (1995) and Dahiya et al. (2003)). These papers document a positive impact of loan announcements on a borrower's stock return at time of loan origination and provide evidence that banks perform a special role in the financial system as monitors and information providers. Our results are consistent with the traditional view: relationships are valuable in screening and monitoring borrowers. However, we also find evidence for bank specialness beyond the "traditional role" as a strong relationship with a bank reduces a customer's incentives to default.

The rest of the paper is organized as follows. The next section describes the data that are used for our analyses and provides summary statistics. Section 3 presents the empirical analyses on private information, Section 4 shows the results suggesting borrower incentives to default, Section 5 concludes.

2. Data and Summary Statistics

A. Loan and Borrower Characteristics

We obtain the performance data for the universe of consumer loans by savings banks in Germany.[6] These loans are usually given on an unsecured basis, i.e. without collateral, and it is

[6] The sample thus does not comprise applications for mortgage loans, checking accounts, or credit cards. Credit cards are used differently in Germany than in the United States. They are issued by a bank and are directly linked to the credit card holder's current account in that bank. Payments are automatically deducted from this checking

not possible to sell or securitize these loans unless they default.[7] The data for these loans are recorded on a monthly basis for each individual loan and are provided by the rating subsidiary of the German Savings Banks Association (DSGV). The data span the time period between November 2004 and June 2008 and comprise information on the performance of 1,068,000 loans made by 296 different savings banks. The default rates for these loans are calculated in compliance with the Basel II requirements.[8] According to this definition, 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, or (vii) the banks needs to write-off the loan.[9,] Our data includes flags for each of these default events and the associated date.[10] Defaults are uniquely determined by each given savings bank; there are no cross-default clauses in German retail lending. In addition to performance data, we have detailed information on all the 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.

account at the end of each month. Customers can thus not default on their credit cards, but their payments may exceed the credit line on their current account. In this case, the bank faces the repayment and default risk.

[7] Given some public debate about the lending practices at one given savings bank, savings banks made clear to their retail customers that no loan would be sold.

[8] See "Solvabilitätsverordnung (SolvV) §125", the "Baseler Rahmenvereinbarung Tz. 452-453 and the "EU-Richtlinienvorschlag, Anhang VII, Teil 4".

[9] 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.

[10] Sales and securitizations of individual loans are uncommon in Germany, and when they occur they are for commercial and industrial loans rather than retail credit.

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 more than 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.[11] The interesting feature for our analysis is that the scoring system does provide 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 they approve a loan application. In other words, there are some subjective elements associated with the banks' screening process which 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.

[11] 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. We use this information in our analysis shown in Table 9.

Table 1 reports the descriptive statistics for loans and borrowers. Over the first twelve month after the loan origination, 0.6% of the approved loans default according to the above default definition. The default rate increases to 1.3% when the loan performance over the full sample period is considered.[12] 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.[13] The loan repayment in percent of the borrower's income amounts to more than 20% only for 6.6% of the borrowers, for 54.5% of our borrowers it is less than 20%. For all other borrowers, this information remains undisclosed. Most borrowers work in the service industry and have been in their current job for more than two years.

The internal rating system does not comprise information on loan amounts, maturities, or interest rates. However, more than 20 million monthly performance observations allow us to make inferences in terms of loan maturities. Note that we can split our sample loans into two categories, (1) loans that have either been repaid in full or defaulted, and (2) loans that have not been repaid and have not yet defaulted or loans in default for which the banks have not closed the account in expectation of future payments. In both categories, we analyze loans that have not defaulted and infer that the average maturity is 14.5 months in both categories The performance data also allow making inferences that pertain to loan amounts. We know the monthly repayment rate (i.e. interest plus principal repayment) and can calculate the loan maturity of the repaid loan.

[12] These relatively low default rates are very typical for consumer loans in Germany. According to 2008 estimates by Creditreform (a German business information service), the average default rates for consumer loans in Germany amount to 2-3% over the lifetime of the loan, while they amount to 5-6% in the UK and more than 6% in the United States.(http://www.creditreform.de/Deutsch/Creditreform/Info-Center/Fachartikel/International_Business/Archiv/Verschuldung.jsp)

[13] The average monthly income of our sample borrowers corresponds to the average German inhabitant. For example, according to the German Census Bureau, in 2006, the median net income in Germany was € 1,800 per person which is very similar to the loan applicants in our sample.

We thus can calculate the total repayment of these borrowers. On average, borrowers repay EUR 237 per month and EUR 3,100 in total.

B. Relationship Characteristics

Table 2 provides detailed information on the loan applicants' relationship status including its length and scope. It reports, in particular, whether loan applicants have an existing relationship with the savings bank at which they apply for a new consumer loan and, if so, which types of products they currently use or have used so far. Only 2.5% of the loan applicants have had no relationship with their savings banks 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.

The data set not only contains information on the checking accounts that loan applicants hold at the savings banks, but also on their assets and prior loans. Table 2 shows that only 23.2% of the borrowers have no savings account with their savings bank. While 19.7% of the loan applicants have assets of less than €50, 36.3% have assets between €50 and €2,000, and 18.5% have assets of more than €2,000. A substantial share of the borrowers already had prior loan lending relationships with their savings bank before the current loan. 19.2% of the loan applicants have had a loan in the past, and 12.1%, 17.4%, and 19.2% of loan applicants have had a loan within the last year, the last two years, and the last three years, respectively.

3. Empirical Results on Private Information

Our objective in this paper is to examine the sources of value of relationships in reducing default rates on consumer loans.

A. Univariate Results

To analyze whether relationships reduce default rates, we first examine the average 12-month default rates in subsamples of relationships versus non-relationship borrowers[14] and find significant differences. While the average default rate is 0.6% for relationship borrowers, it is 1.6% for non-relationship borrowers, respectively. The difference is significant at the 1 percent level. We also analyze differences in ex-ante borrower risk. More precisely, we compare the risk distribution of loans given to relationship versus non-relationship customers using Cramer's V which is a Chi-Square measure taking into account the number of observations in each

[14] We define a relationship borrower as someone who has a transaction account relationship with the savings bank before applying for a loan.

subsample. We cannot reject the null that the risk distribution does not differ between both subsamples (Cramer's V is 0.045). In other words, while we find significant differences in default rates, we cannot find differences in ex-ante borrower risk which suggests that relationships are of first order importance in explaining as to why relationship borrowers exhibit significantly lower default rates.

We next test the performance of consumer loans against a number of variables that capture the existence, length, and scope of the relationship that a customer has with her savings bank. The results are reported in Table 3 and show that customers with relationships, and in particular with more intense relationships, default less often than other customers and that these results are highly significant both from an economic and a statistical perspective.

As the first piece of evidence, model (1) of Table 3 shows that customers with an existing relationship have a 1.0% lower default rate than customers with no existing relationship. This difference in default rates is statistically significant at the 1% level. This is economically large given the average default rate amounts to only 0.6% and corresponds to the difference in default rates of relationship (0.6%) versus non-relationship loans (1.6%). Further, the difference in default rates between new and existing customers is more than 1.5 times higher than the unconditional mean. Model (2) shows that the default rates monotonically decrease with the length of existing relationship. The benchmark case here is customers with a relationship of more than 15 years. The default rates for customers with relationships between 9 and 15 years are 0.2% higher than for the benchmark case, and they increase up to 1.5% for relationships of less than two years.

The results in model (3) of Table 3 suggest that default rates decrease with the scope and thus the intensity of the relationship between customer and bank. We introduce four indicator variables equal to 1 if the borrower has (i) a credit and a debit card, (ii) only a debit card, (iii) only a credit card or (iv) neither a credit nor a debit card. Borrowers without prior relationships are the omitted group. All coefficients on these indicator variables are negative and significant suggesting that relationship customers are less likely to default which is consistent with our previous finding. Nonetheless, the biggest reduction in default rates is associated with borrowers which have both a debit and credit card (only a debit card), which default 1.2% (1.1%) less often relative to non-relationship customers. Model (4) shows that default rates also depend on the existence of prior credit lines. The loans by customers with existing credit lines loans default by 0.6% less. Model (5) considers in more detail the actual usage of these credit lines. Customers with credit lines have a higher default rate than customers without credit line only if their usage is larger than 150% of the credit line. For all other customers with credit lines, the default rates are significantly lower than for the benchmark group rates. In general, the default rates are positively correlated with the usage of the credit line, i.e. customers with a positive account balance exhibit the lowest default rates. Model (6) of Table 3 combines the different measures used so far and looks at them simultaneously. The results are very similar to the previous results, in particular the relationship length and the usage of debit and credit cards are still negatively related to default rates, while the extent of the usage of credit lines is still positively related to default rates.

Starting with model (7), we analyze the effect of savings accounts on default rates. The results show that the existence of a savings account decreases default rates by 0.5%. Model (8) shows that customers with no savings accounts and with savings accounts of less than 50 Euros have a 0.7% and 0.6% higher default rate, respectively, than customers with more than 2.000 Euros on their savings account. Overall, the volume of assets on a savings account is negatively correlated with customer default rates; even customers with savings account assets of more than 50 but less than 2.000 Euros are more likely to default than customers with assets of more than 2.000 Euros.

These results provide initial evidence that customers with existing relationships with the savings bank at which they apply for a loan have lower default rates and that these default rates further decrease with the length and scope of the relationships.

B. Multivariate Results

In this section, we analyze whether existing relationships reduce the default probability of consumer loans controlling for a wide array of borrower characteristics. Our analysis proceeds in two steps. We start by reporting the results separately for customers who have held transaction accounts, savings accounts, and had repeat lending relationships with their savings banks before they receive the current loan. Then we combine these measures in one specification in order to analyze their relative importance.

B.1. Relationships from Transaction Accounts

Table 4 reports the results for customers who have had a transaction account with their savings bank before applying for a loan. This table presents the results of a probit regression. 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 relationships characteristics as a result of relationships via transaction account (relationship length, credit and debit cards, credit lines and usage of credit lines). Models (2) to (6) consider those borrowers that have a checking account with the savings bank (i.e. we drop loans by "new customers"). In model (2), the omitted relationship variable is customers with a relationship longer than 15 years; in model (3) borrowers without a debit and credit card are omitted; in model (5) customers without credit lines are omitted; in model (6) customers with a relationship longer than 15 years, the group of customers with no credit and debit card and without credit line are simultaneously omitted. The coefficients for borrower industries[15] as well as intercept and time fixed effects are not shown. Only the marginal effects are shown. Heteroscedasticity consistent standard errors clustered at the bank level are shown in parentheses (Petersen (2009)). The control variables are the monthly income of the loan applicant, her repayment burden, which is measured by the ratio of the expected monthly loan repayment amount - if the loan application is approved - and the available income, the loan applicant's age as well as her job stability.[16] This is a dummy variable that takes a value of 1 if the borrower has been in her current job for more than two years and 0 otherwise. The analysis also controls for the industry in which the borrower works and includes time fixed

[15] "Industry" has to be understood in a very broad sense and comprises the most important industries borrower work in, for example, the service sector, public sector, construction, whether the borrower is unemployed or retired, but also the following industries: communications and information; energy and water supply, mining; hotel and catering; municipalities; agriculture; banking; insurance; not for profit company. But it also comprises: housewife; apprentice; high school student; student; army; houseman and civil service.

[16] All variables are defined in Appendix I.

effects. The results in Table 4 show that default rates are decreasing in the borrower's income and tend to increase in her repayment burden. The default for this variable is a ratio that exceeds 20% of the loan applicant's monthly income. The borrower's age does not have a significant effect on default rates for borrowers below the age of 30 in some models, in comparison to the default age of larger than 60 years. However, borrowers between the age of 30 and 60 have a higher default probability than borrowers at the age of 60 and above throughout. Job stability also has an important impact on default rates. Customers who have been in their current job for less than two years default 0.3% to 0.5% more often than customers who have been in their current job for more than 2 years. This result is statistically significant at the 1% level.

The coefficients of our relationship proxies are in most cases significant at the 1% level and similar in magnitude compared to Table 3. As shown in model (1), the existence of a relationship lowers the default probability by 0.6%. Model (2) shows the results for different relationships length categories. The results suggest that defaults decrease with the length of a relationship and are least likely for the customers with the longest relationship duration. Borrowers with a relationship length less than 2 years have a 1.4% higher probability to default compared to customers with more than 15 years of relationships, ceteris paribus. Apparently, even the existence and the first few months of a relationship have a significant effect on default rates. This finding is consistent with anecdotal evidence we obtain talking to loan officers at a large private bank in India who does lending to SMEs that are also difficult to evaluate. One of their key models is to ask firms to open a checking account and observe them for 6 months before making a loan decision. The loan officers claim they could substantially reduce default rates with this

model. It is noteworthy that the anecdotal evidence from India matches our results on retail lending in Germany.

Model (3) takes into account the intensity of a relationship by analyzing the impact of different combinations of credit and debit cards that transaction account customers had before applying for a loan. Customers that had both credit and debit cards or simply debit cards have the lowest default probability and have 0.3% lower default probability than customers who have held neither a credit nor a debit card. Model (4) tests for the effect of the existence of a credit line in a customer's transaction account. The results suggest that that the existence of a credit line significantly lowers the customer's default probability. Model (5) considers credit lines again more carefully, and the results suggest that the usage of credit lines is positively correlated with default which is consistent with the findings of Mester, Nakamura, and Renault (2007) and Norden and Weber (2009). The coefficients are very similar to those in the previous univariate analysis. Finally, model (6) considers the different relationship variables simultaneously. The results are again very similar to those for the separate analysis of the different characteristics. Taken together, the results for the transaction accounts suggest that the existence of a prior relationship between bank and customer reduces the subsequent loan default rates for the customer, and that these default rates decrease in particular for longer and more intense relationships.

B.2. Relationships from Savings Accounts

Table 5 repeats the previous analysis for customers who have held a savings account before receiving a consumer loan using probit regressions. 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 relationships characteristics as a result of relationships via savings account (the existence of savings accounts and assets held in these accounts). In model (2), the omitted relationship variable is assets > 2,000 Euros. The coefficients for borrower industries (as described in Appendix I) as well as intercept and time fixed effects are not shown. Only the marginal effects are reported. Heteroscedasticity consistent standard errors clustered at the bank level are shown in parentheses. The control factors are the same as before and comprise the borrower's income, her repayment burden, her age as well as her employment status as characterized by her job stability and the industry in which she works. The impact of the control variables is very similar to the earlier results in Table 4. In particular, borrowers tend to default less with an increase in their monthly income and when they are older than 60 years, while they tend to default more with an increase in their repayment burden. Customers also default more often when they have only been in their current job for less than two years.

The relationship variables are again highly significant and carry the expected sign. Model (1) shows that customers who no savings accounts when applying for a consumer loan have a significantly higher default probability than customers with savings accounts. Model (2) analyzes whether or not the amount of assets held in these accounts is important. We split theses amounts in different size categories where the asset class of more than €2,000 is omitted. In comparison to the omitted group, customers with assets between €50 and €2,000 have a slightly higher likelihood of defaulting, and this increase in default likelihood amounts to 0.4% for customers with assets of less than €50 and 0.5% for customers with no assets. Thus the assets that a customers holds with a bank when applying for a loan have significant predictive power for the

likelihood that the loan will finally be repaid, even after controlling for several important borrower characteristics.

B.3. Repeat Lending Relationships

Table 6 considers the impact of repeat lending relationships on subsequent consumer loan defaults in the same way as the previous analyses consider the impact of transaction and savings accounts and their characteristics on these defaults using probit regressions. 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 relationships characteristics based on repeat lending with different look-back windows. *Prior Loan within 2 yr (1yr) look-back* are dummy variables equal to 1 if the borrower was granted a loan within 2 years (1 year) prior to the current loan.[17] *# Prior Loan Defaults* measures the number of loans the borrower defaulted on in the past and which were originated during our sample period. The coefficients for borrower industries as well as intercept and time fixed effects are not shown. Only the marginal effects are reported. Heteroscedasticity consistent standard errors clustered at the bank level are shown in parentheses. The control variables are thus again the same ones as before and comprise several important borrower characteristics. In the same way as before, loan default rates decrease for borrowers with higher income and increase for borrowers with a higher debt repayment burden as measured by the ratio of the monthly repayment amount and the available monthly income. For the age cohorts, all age cohorts default significantly more often than those customers with age 60 and above. Finally, customers with less time on their current job default more often than other customers.

[17] We do not have information on prior loans which were granted to our sample borrowers before our sample period.

The relationships variables of interest are the existence of a prior loan relationship and how long this relationship dates back. Model (1) shows the results for the existence of a prior loan relationship and prior default. The results suggest that the existence of a prior loan relationship significantly reduces the default likelihood by 0.3%. As expected, whether or not a borrower defaulted on a prior loan increases the likelihood of default on the current loan by 2.2%. Models (2) and (3) consider whether the prior loan was granted within the last 2 or 1 years before the current loan, the results, however, do not change compared to model (1).

B.4. Multiple Relationships and Default Rates

The results so far consistently show that customer relationships significantly reduce the likelihood of default. This result holds – in separate analyses - for customers who have had prior transaction accounts, savings accounts, and consumer loans, and the results are particularly strong for longer and more intense relationships in each of these cases. Clearly, customers often have more than one of these relationships with their savings bank, e.g. they have both a transaction account and a savings account. Thus it is important to consider the relative importance of these different relationships. Table 7 reports the results for the simultaneous consideration of the different relationships variables that are tested separately in Tables 4 to 6. This table presents the results of a probit regression. The dependent variable is a binary variable equal to 1 if the borrower defaults within the first 12 months after loan origination. Model (1) repeats the analysis from model (6) in Table 4 and model (2) adds whether or not the borrower also had a savings account. Model (3) considers whether borrowers had simultaneously checking and savings accounts at their bank. Model (4) adds whether or not the borrower had a prior loan during our sample period controlling for previous loan defaults to model specification (2). The

coefficients for borrower industries (as described in Appendix I) as well as intercept, bank and time fixed effects are not shown. Only the marginal effects are shown. Heteroscedasticity consistent standard errors clustered at the bank level are shown in parentheses. The control factors are the same ones as before and comprise again the borrower's income and debt repayment burden as well her age and employment status. The results for these control factors are very similar to those obtained before.

Model (2) adds whether or not borrowers have savings account to model (1). The coefficients hardly change and the magnitude of the coefficients is higher for the variables associated with checking accounts. As there is a probably an overlap in borrowers which have both checking and savings accounts, we model this explicitly in model specification (3). Model (3) shows that if borrowers have both a checking and a savings account before applying for a loan, relationship specific information obtained from checking accounts is important. The coefficients of savings accounts as well as the interaction term are insignificant. Model (4) adds whether or not the borrower had a loan prior to the current loan. Again, the coefficient of this variable is smaller compared to the checking account variables.

Taken together, the multivariate specifications shown in Table 4, 5, 6, and 7 control for several detailed borrower characteristics, and the results show that – even after controlling for these characteristics – relationships are valuable to banks. In particular, our results suggest that relationships that exist prior to applying for the current loan give banks an advantage in monitoring the borrowers and reduce default rates. Furthermore, they suggest that relationship specific information from checking accounts is relatively more valuable compared to savings

accounts or repeat lending relationships. We next extend the previous analysis in two ways to shed more light on the underlying mechanisms for our results and to check their robustness.

B.5. Internal and External Ratings

First, we employ the internal credit score used by the bank instead of controlling explicitly for the different borrower characteristics. This allows us to see whether relationships provide value even above and beyond the information captured in the internal credit score, which represents the key building block of a bank's credit decision. The results are presented in Table 8. The results for the internal rating classes show that the internal rating classes are consistent and capture well the customers' default risk. The default rates decrease monotonically for higher internal rating classes as compared to rating class 12, which is the default and worst rating class employed. This pattern holds for each of the six models presented in Table 8. More importantly for the purpose of our paper, all the relationship variables remain significant and of similar magnitude as in the previous specifications. Model 1 shows that the existence of a relationship lowers the likelihood of a borrower default by 0.3%. Likewise and in the same way as before, the length of a relationship is negatively related to the likelihood of default. While it increases by 1.2% for customers with a relationship of less than 2 years default, it only increases by 0.2% for customers with a relationship between 9 and 15 years, both in comparison to the default of relationships of more than 15 years. Model 3 and Model 4 show the respective value for the existence of credit and debit cards as well as credit lines: The more information is provided by the relationships through existing checking accounts, the more valuable these relationships are. Finally, Model 5 and 6 show the results for relationships through savings accounts and prior loans, respectively. The results suggest that the existence of a savings account reduces the likelihood of default by

0.3%, while the existence of a prior loan reduces this likelihood by 0.4%. After controlling for internal credit scores, the results are thus very similar to those obtained before.

Second, we employ a loan applicant's external rating as an additional control variable. The external credit score is provided to the savings banks by a German credit bureau, and it is available for a subsample of 86,628 loan applications. We construct eight different rating classes based on the external credit score with 1 being the lowest risk. The average rating is 4.3. Controlling for external credit bureau information allows us to make sure that our results are in fact due to the information about a specific customer that is generated from the relationship with the savings bank and not to any other information that is obtained from external parties which is available to outside (i.e. non-relationship) lenders. The results are presented in Table 9. We find that high quality customers based on the external credit score are less likely to default. For example, customers with the highest external rating class are 0.3% less likely to default compared to customers in rating class 8 (the omitted group). The coefficients of our relationship proxies are very similar to those before. For example, the coefficient for the existence of a relationship in Model 1 is identical to that in Table 4. The coefficients for the length and intensity of a relationship in Model 2 and Model 3 are again similar, but slightly smaller than those in Table 4, implying that there is indeed valuable information captured in the external ratings. Finally, the existence of a credit line (Model 4), a savings account (Model 5), and a prior loan (Model 6) are shown to reduce customer default rates. Taken together, the key results remain robust even after explicitly controlling for internal and external ratings; relationships provide information above and beyond the existing information from internal and external sources.

B.6. Endogeneity of bank-depositor relationships

The previous sections established that relationships reduce the likelihood of borrower default. We argued that relationship specific information improves banks' monitoring ability which results in lower default rates. However, the relationship between banks and borrowers is unlikely to be exogenous and banks might establish and continue relationships only with high quality customers. We use a wide array of borrower characteristics such as income, age, and employment among others to control for observed borrower heterogeneity, but relationship and non-relationship borrowers might still be different on an unobserved dimension that we are not able to control for in our models. If this was indeed the case, it would be less clear to what extent our results are driven by relationships rather than unobserved higher quality of relationship customers.

Before we proceed with formal tests to address this, we note that there are at least two arguments to support the notion that relationships are unlikely to be endogenous. First, as mentioned earlier, we use extensive borrower controls to net out any differences between relationship and non-relationship borrowers. Further, the risk distribution of both types of borrowers is not significantly different, i.e. they are not different based on ex-ante risk. The second argument is based on the institutional setting in German banking. Savings banks are mandated to serve local customers and provide financial services (and transaction accounts in particular) to all customers in their region. Savings banks are therefore unlikely to establish relationships only with high quality customers taking this political mandate at face value.

We address endogeneity of relationships more formally using a simultaneous equation model in which we augment the main probit equation (default model) with an additional probit equation that explains which factors influence relationships (relationship model). We use a bivariate probit model as both default and the existence of a relationship are binary variables and test the null hypothesis that the contemporaneous error terms of both equations are uncorrelated instrumenting for relationships and in particular the existence of a checking account, which is usually the first relationship that a customer builds with a bank. Identification requires an exogenous variation along the relationship / non-relationship margin that is uncorrelated with borrower default and, therefore, we propose an instrument that measures the availability of savings banks to customers in their region.[18] More precisely, we use the natural logarithm of the number of branches over population as our main instrument. This variable is constructed using the number of all branches of each savings bank and the number of inhabitants of the particular region the bank is operating in. The underlying intuition is that a customer is more likely to have a checking account with a savings bank if the bank has more branches in that region relative to the population. Our instrument thus proxies for the average distance between depositors and savings banks. The smaller this distance the more likely the customer has a relationship with the bank. The regional principle, i.e. savings banks can only engage in business with people living in their region, facilitates the use of this instrument in our setting. We collect data for each savings bank on a very detailed basis. We know for each bank the number of branches operating in each of the 439 regions or districts ("Kreisebene") in Germany. Appendix 3 provides more information about all German banks, the total number of branches in Germany and the average number of branches in each district. Our key identifying assumption is that the availability of

[18] See, for example, Berger et al. (2005) or Hellmann et al. (2008) who use a similar line of arguments to identify relationship building of banks with firms.

savings banks in a particular region influences the initiation and existence of a bank-depositor relationship but does not explain the default behavior of subsequently issued loans.

We include a second instrument in some specifications that additionally captures the availability of savings banks relative to all other banks that have branches in the same region. Using the branch level information about all German banks detailed in Appendix 3, we construct a Herfindahl-Hirschmann Index (HHI) for each region. Evidently, savings banks have the largest branch-network throughout Germany followed by Deutsche Postbank AG (now owned by Deutsche Bank AG) and the cooperative banks (Volks- und Raiffeisenbanken).[19] The mean HHI is 0.22, the minimum HHI is 0.12 and the maximum HHI 0.45, respectively.

Technically, the relationship model and the default model constitute a bivariate qualitative dependent variable model where the error terms are uncorrelated with our instrument, are distributed as bivariate normal with mean zero and each has a unit variance (Greene (2003) and Pindyck and Rubinfeld (1998)). ρ is the correlation between the error terms. If the correlation is zero, we get consistent coefficients with the probit estimation of the default model, i.e. there are no unobservable characteristics that make relationship customers less risky than non-relationship customers. The model is estimated using the Maximum Likelihood Estimation (MLE) approach.[20]

[19] Hackethal (2004) provides more information about the German banking system.
[20] Application of this approach with two binary dependent variables can be found, for example, in Evans and Schwab (1995) who study the causal effect of attending high school on the probability of attending college and Hellmann et al. (2008) who study the relation between a bank's venture capital investments and future lending.

The results of the bivariate probit model are presented in Table 10. We report both the 1st stage (relationship equation) and the 2nd stage (default equation). The relationship models include all control variables as shown in the previous analyses along with the instruments. The first column reports the results from the probit model for comparison. Model (1) includes our main instrument (*Log(Branches/Population)*), model (2) adds *Log(HHI)* as additional instrument. Panel A of Table 10 reports the results from the 1st stage relationship equation. The coefficient of the instrument model (1) confirms our expectation that an increase in the number of branches relative to the population also significantly increases the likelihood that loan applicants have a checking account relationship with their savings bank. Staiger and Stock (1997) propose a test for the strength of the instrument under the null hypothesis that the instrument is not significantly different from zero. We can reject this hypothesis at any confidence level and our instrument clearly passes the threshold for this F-Test (the F-statistic is 61.45). In model (2), we add *Log(HHI)* as a second instrument. While the coefficient of *Log(Branches/Population)* does hardly change, the coefficient of *Log(HHI)* is also positive and significant suggesting that the more savings bank branches relative to other bank branches exist in a particular region the more likely does the applicant has a relationship with the savings bank.[21] Panel B of Table 10 reports the results of the 2nd stage default equation. The coefficient and marginal effects are shown. Model (1) shows the results using Log(Branches/Population) as instrument, model (2) adds Log(HHI) as second instrument consistent with the order of the 1st stage tests in Panel A. Most importantly, the result for the existence of a checking account in the 2nd stage does not differ from the results before: Customers with an existing checking account still have a significantly

[21] The F-statistic of the first stage regression is also significant rejecting the null hypothesis that both instruments are equal to zero. An overidentification (Hansen-J)-test cannot reject the null hypothesis that the instruments are uncorrelated with the error term of the outcome equation. In an earlier version of this paper, we use *Log(HHI)* as sole instrument and also reject the null that the instrument is weak at any confidence level. The results for both relationship and default equation are qualitatively similar to using the combination of both instruments.

lower default probability than other customers, and the coefficient is significant at the 1% level. The diagnostic section reports the Wald test under the null hypothesis that the correlation between the error terms is zero. We cannot reject this hypothesis at conventional levels suggesting that there are no unobservable factors that would simultaneously affect the existence of a checking account and default probability. These results suggest that our main results remain unchanged even after controlling for the possibility that relationships may proxy for unobserved higher quality of relationship borrowers.

B.7. Default probabilities and sample selection: Screening and monitoring

In the previous specifications, we test whether relationships in various forms, scope and depth affect the likelihood that a borrower defaults on a new loan. The results – both for the separate and the joint analysis of different relationship variables – suggest that relationships that existed at the time of loan origination reduce loan default rates. However, our sample is censored because we can observe the performance of the loans only if the applicant received credit. As shown by Heckman (1979), censored samples can lead to biased estimates if the errors in the default equation are correlated with the way as to how our sample was selected, or, in other words, with the banks' screening process. If this screening process is based on quantitative credit scores alone (i.e. which can be controlled for in our selection equation) or a deterministic function thereof, screening does not lead to biased estimates in the default equation if we do not control for the selection process (Boyes et al., 1989). If the banks' screening process is not deterministic but includes elements of subjective assessment which are also correlated with the errors in the default equation, the estimators in the default equation might be biased.

A similar argument provided for using the bivariate probit model earlier applies here: being approved for a loan and default are both qualitative variables which has to be accounted for in modeling the selection problem. Technically, the loan approval model and the default model constitute a bivariate qualitative dependent variable model in a similar way as the relationship and the default model discussed above but with partial observability (Poirier (1980)) as the applicants who were denied credit are not included in the default equation, i.e. the dependent variable is not always observed. Indexing individual customer applications by i and the savings bank to which the application is submitted by j, the selection equation is

$$z_{ij}^{*} = \gamma' w_{ij} + \mu_{ij} .$$

The regression model is

$$y_{ij} = \beta' x_{ij} + \varepsilon_{ij} ,$$

where $(\mu_{ij}, \varepsilon_{ij})$ are assumed to be bivariate normal $[0,0,1,\sigma_{\varepsilon},\rho]$.

z_{ij}^{*} is not observed; the variable is observed as $z_{ij} = 1$ if $z_{ij}^{*} > 0$ and 0 otherwise with probabilities $\Pr(z_{ij} = 1) = \phi(\gamma' w_{ij})$ and $\Pr(z_{ij} = 0) = 1 - \phi(\gamma' w_{ij})$. $z_{ij} = 1$ indicates that the savings bank j accepts the loan application i (selection model); ϕ is the standardized normal cumulative distribution function. y_{it} is the default model. This model corresponds to the probit model with sample selection and maximum likelihood estimation provides consistent, asymptotically efficient estimates of the parameters in both equations (Van den Ven and Van Pragg (1981)).

The model is estimated using MLE. The explanatory variables in the loan granting and default equation are identical. In different model specifications, we add *Log(Branches/Population)* and *Log(HHI)* as instruments to the selection equation for identification.[22] The intuition for using *Log(Branches/Population)* as an instrument is similar to our endogeneity tests. The more savings bank branches are available to customers the more likely will these customers apply for loans at one of these branches. However, while savings banks are expected to provide their services to all customers in their region, this political mandate does not extend to loan market relationships. In other words, a different way to phrase the question we are analyzing in this section is: Do savings banks establish loan market relationships only with (in an unobservable way) high quality customers? At the same time, we treat bank-depositor relationship as completely exogenous based on our previous results. *Log(HHI)* captures the level of competition for each savings bank as measured by the number of competitor branches that operate in the same region in which a savings bank operates. The choice of this variable is motivated by the evidence in papers such as Jayaratne and Strahan (1996), Black and Strahan (2002), that more competition in banking markets has a positive effect on credit supply. This means that a savings bank is expected to be less likely to approve a loan application if there are fewer competitors. The empirical results suggest that this is indeed the case and thus confirm the evidence for U.S. banks. The higher the HHI in a given savings bank region, i.e. the fewer competitors operate in that region, the lower is the acceptance of consumer loans within these savings banks.

[22] The selection model can be identified without using an instrument but would then rely deterministically on the non-linearity of the selection equation.

The results are reported in Table 11. Panel A of Table 11 shows the results from the selection equation, Panel B the results from the default equation, respectively. Model (1) includes our main instrument (*Log(Branches/Population)*), model (2) adds *Log(HHI)* as additional instrument. The diagnostic section reports the correlation between the two models (ρ) and its p-value. Again, if ρ =0, both models can be separately estimated without selection bias. If ρ≠0, then unobservable borrower quality is clearly important in the loan granting process that potentially biases our estimates in the default equation if we do not carefully control for the selection.

Panel A of Table 11 shows coefficient estimates and the respective standard errors. We find that loan applicants are more likely to be approved if they have a previous transaction account relationship with the savings bank. If the variables in the loan granting and default equation have opposite signs, then this variable affects the loan approval decision and default probability differently, for example, an applicant is more likely to be rejected based on one variable and this characteristic is also positively associated with default probability. This is consistent with the interpretation that a bank's screening policy is designed to minimize default rates. We find that all variables that are significant in both the approval and default model carry the opposite signs, which is consistent with this interpretation and in line with what we would expect.

The coefficient of *Log(Branches/Population)* is positive and significant indicating that a greater presence of savings banks increases the likelihood of a loan market relationship keeping the population size constant. However, the coefficient of *Log(HHI)* is negative and significant. I.e. more competition increases the likelihood that loan applications are approved which echoes the

results found in Jayaratne and Strahan (1996), Black and Strahan (2002). The estimate of the correlation that maximizes the likelihood is insignificant. The coefficient of the relationship variable is still negative and significant. Moreover, correcting for sample selection does not change the magnitude of the coefficient. Having a relationship with the bank reduces the probability of default by 0.4%, ceteris paribus, suggesting that prior relationships allow for better monitoring by banks reducing default rates. Table 12 reports the results using alternative proxies for bank-borrower relationships. Our previous results, however, remain unchanged.

Taken together, observing both loan applications and the performance of the originated loans allows us to contribute an important aspect to the literature on the value of relationships, namely, the value of ex-ante relationship specific information (that existed before the start of the application process) in both screening and ex-post monitoring of the borrower.

4. Private Information and Borrower Incentives to Default

In the previous discussion, we highlight the benefit of relationships beyond screening of loan applicants and link this to an enhanced monitoring ability of relationship lenders. Another interpretation of our results could be that lower default rates of relationship borrowers mirrors lower incentives of these borrowers to default.[23] This could be because of a number of reasons. In particular, the borrower may have concerns about asset setoffs or may have concerns about

[23] Another reason why default is not automatic is that loan officers may restructure a loan in order to avoid recognizing a default. Such restructurings are more likely to favor customers that have a variety of accounts and long relationship with the bank as well as customers with high account balances. The interpretation would therefore be more consistent with favoritism of particular customers rather than private information. As we can only observe loan performance of loans that have been originated during our sample period, we exclude all loans to customers who have more than one loan in our sample. This approach, of course, is much more conservative and does not only exclude restructuring loans but also new loans to the same customer and our sample drops by 32 percent following this method. However, our results continue to hold.

future credit rationing.[24] Even though the borrower could divert their assets from the bank prior to default, if they maintain multiple accounts or access many services there can be costs to doing this. Further, borrowers who maintain an extensive relationship with their bank (particularly, as it is frequently the case in Germany, if this is the sole bank relationship they have) worry about jeopardizing this relationship if they default on their loans, for fear of reducing their future credit availability. For example, 95 percent of our sample borrowers have a credit line in combination with their transaction account. These loan commitments facilitate short term borrowing if households are credit constrained. Or, to say it differently, it is indicative for households to be financially constrained if they do not have access to credit lines. This argument is similar to the results in Sufi (2009), who shows that U.S. firms with low levels of cash flow and high covenant violations are less likely to obtain credit lines. Given even sparser outside financing options, credit lines (and relationships in general) might be even more important for households. Put together this suggest that borrowers may have reduced incentives to default on loans taken from a bank with which they have relationships.

In this section, we investigate this further. In order to do this we are able to access detailed information about transaction account behavior for a subset of our sample borrowers. This information is private and thus only known to the bank and is available immediately before the loan is applied for. However, this information is not part of the internal rating even though it is available in a quantitative score. Our empirical approach includes this behavioral score as control variable in our regressions in addition to all other relationship proxies. The behavioral score is the most direct measure of a bank's private information from the transaction account relationship

[24] In Germany a bank has the legal right to access other customer assets held with the bank if the customer defaults on a loan.

with the borrower taking even positive and negative information into consideration. The coefficients of our relationship proxies should therefore show any additional effect of bank-borrower relationships on default rates net of a banks' private information. Our behavioral score is a comprehensive measure of the transaction account activity of a borrower and comprises, among other factors, the number of transaction accounts, the maximum balance during the last six months, the percentage of months with a negative balance during the last 6 months, the usage of the credit limit, the percentage of months in excess of the credit limit, sum of account credits of the last months relative to the average account credits of the last six months, the number of return debit notes during the last six months and the longest period of a declining maximum account balance. The factors are weighted with respect to their power in predicting borrower defaults (based on out of sample data) and summed to a single behavioral score. The average score is 576. A higher score indicates positive information from the borrower's transaction account behavior.

We report the results in Table 13. The dependent variable is the 12-month-default rate. Models (1) to (5) successively include the transaction account based relationship proxies. Model (6) includes all transaction and savings account related proxies as well as dummy for a prior loan lending relationship. We also include the behavioral score as private information proxy. We show only marginal effects and multiply the marginal effect and standard error of the behavioral score with 1,000 for illustrative purposes here and in all following tables. In all models, the behavioral score is negative and significant suggesting that private information that exists prior to the loan origination date reduces borrower defaults, which is consistent with our earlier results. The economic significance is similar to what we observe with regard to the relationship

proxies. A one standard deviation increase in the behavioral score decreases the default rate by 0.2%. In other words, we find that private information is a significant determinant of loan default rates. Furthermore, the coefficients of our relationship proxies remain significant even after controlling for private information. These results suggest that other factors are important at explaining borrower defaults and our results point to borrower incentives. Having a credit line has the largest effect on the 12-month default rate supporting our earlier conjecture that losing a credit line and becoming credit constraint is a serious concern for borrowers. Model (6) includes several dummies for savings account balances and, obviously, borrowers are less likely to default if they have larger balances in their accounts that can be seized in the event of default.

In a next step, we use selection models similar to above to separate the effect of private information into screening and monitoring. We use a Heckman selection model with two probit equations and *Log(Branches/Population)* as instrument. Table 14 reports the results using the existence of a relationship as relationship proxy and the behavioral score as measure of the bank's private information. The coefficients of the control variables remain unreported. Consistent with our results in Tables 11 and 12 we find that the availability of savings banks in the regions increases the likelihood that borrowers obtain loans. We also find evidence for the importance of screening. In the selection equation, the coefficient of the behavioral score is positive and significant, i.e. positive information from transaction accounts (or being a higher quality customer) increases the chance of being approved by the loan officer as does having a relationship. This supports our interpretation: netting out the private information component from the relationship measure, our results suggest that loan officers anticipate the reduced incentives of borrowers to default and incorporate this information in the lending decision. The

default equation shows the corresponding results with both private information and relationships being negatively associated with defaults. The contemporaneous correlation between the error terms of selection and default equation is insignificant. Table 15 shows the results from a Heckman models using the alternative relationship proxies with qualitatively similar results.

5. Conclusion

Using a unique database of the universe of consumer loans by savings banks in Germany, we investigate if relationship loans default less, and the sources of value of relationships. We find that relationships between banks and retail customers *prior* to a loan application significantly reduce the default rates of loans given to these customers. We find relationships matter in different forms (transaction accounts, savings accounts, prior loans) and scope (credit and debit cards, credit lines) and depth (relationship length, utilization of credit line, money invested in savings account). Our results suggest that relationships even in the form of simple transaction and savings account are economically important, even after controlling for detailed borrower characteristics and their internal and external credit scores. Hence, from a practical viewpoint, our results suggest that having people open simple savings or checking accounts can enable banks to make better credits.

We next investigate the reasons behind lower default rates or put differently where the sources of value of prior relationships lie. We are able to access data that has information not only on loan performance, but also on the determinants for the loan approval decision, including the quantifiable credit information on the customer. Our results suggest that relationship customers

have a much higher likelihood of receiving a loan than new customers. But, these customers also have a significantly lower default probability than new customers after controlling for the bank's loan acceptance decision. Importantly, prior relationships allow the bank to produce information that goes beyond publicly available information and allows it to better assess loan applicants' creditworthiness. These results suggest that relationships provide value to banks in the screening process of loan applications by retail customers. At the same time, relationships also provide value beyond the improvement in the initial screening process.

The results in this paper highlight that relationships matter in multiple dimensions. We find that the private information banks accumulate over the course of a relationship is an important factor in consumer lending. It matters in screening loan applicants. However, even after a loan has been originated, relationships help reduce default. We have evidence that private information from relationships is important here in monitoring but relationships also potentially add value in shaping borrowers' incentives to default. They thus fill an important gap left by the existing literature on the benefits of bank relationships.

Our results suggest that relationships are an important source of private information used by banks. There are a number of important open questions for future research. Are there other sources of unobservable private information that loan officers incorporate into their decisions? Does use of such discretion improve the lending decision or lead to favoritism or cronyism? We hope to address these questions in future research.

References

[1] Agarwal, S., Chomsisengphet, S., Liu, C., and Souleles, N., 2009. Benefits of Relationship Banking: Evidence from Consumer Credit Markets, Working Paper.

[2] Agarwal, S., Hauswald, R., 2007. Distance and Private Information in Lending. *Review of Financial Studies* 23, 2757-2788.

[3] Berger, A. N., Miller, N. H., Petersen, M. A., Rajan, R. G., and Stein, J. C., 2005. Does function follow organizational form? Evidence from the lending practices of large and small banks. *Journal of Financial Economics* 76, 237-269.

[4] Berger, A., Udell, G., 1995. Relationship lending and lines of credit in small firm finance. *Journal of Business* 68, 351-381.

[5] Bharath, S., Dahiya, S., Saunders, A., Srinivasan, A., 2007. So what do I get? The bank's view of lending, *Journal of Financial Economics*, 85, 368-419.

[6] Billett, M., Flannery, M., Garfinkel, J., 1995. The effect of lender identity on a borrowing firm's equity return. *Journal of Finance* 50, 699-718.

[7] Black, S., Strahan, P., 2002. Entrepreneurship and Bank Credit Availability, *Journal of Finance* 57, 2807-2833.

[8] Boot, A., 2000. Relationship banking: What do we know? *Journal of Financial Intermediation* 9, 7-25.

[9] Boyes, W. J., Hoffman, D. L., Low, S., 1989, An Econometric Analysis of the Bank Credit Scoring Problem, *Journal of Econometrics*, 40, 3-14.

[10] Campbell, T., Kracaw, W., 1980. Information Production, Market Signaling, and the Theory of Intermediation, *Journal of Finance* 35, 863 – 882.

[11] Cerqueiro, G., Degryse, H., Ongena, S., 2007. Rules, discretion and loan rates, in Evanoff D. (ed.), *Federal Reserve Bank of Chicago Conference on Bank Structure and Competition Proceedings*, 100-109.

[12] Cole, R. A, 1998. The importance of relationships to the availability of credit, *Journal of Banking and Finance* 22, 959 – 977.

[13] Dahiya, S., Puri, M., Saunders, A., 2003. Bank borrowers and loan sales: New evidence on the uniqueness of bank loans. *Journal of Business* 76(4), 563-582.

[14] Diamond, D.W. 1984. Financial Intermediation and Delegated Monitoring. *Review of Economic Studies* 51,393-414.

[15] Diamond, D.W. 1991. Monitoring and Reputation: The Choice Between Bank Loans and Directly Placed Debt. *Journal of Political Economy* 99,689-721.

[16] Drucker, S., Puri, M., 2005. On the benefits of concurrent lending and underwriting. *Journal of Finance* 60, 2763-2799.

[17] Evans, W. N., Schwab, R. M., 1995. Finishing High School and Starting College: Do Catholic Schools Make a Difference? *Quarterly Journal of Economics*, 110, 941-974.

[18] Fama, E., 1985, What's Different about Banks? *Journal of Monetary Economics* 15, 29-36.

[19] Greene, W. H., 1992, A Statistical Model for Credit Scoring, NYU Working Paper.

[20] Greene, W. H., 2003, Econometric Analysis, 5th ed. Upper Saddle River, NJ: Prentice Hall.

[21] Haubrich, J., 1989. Financial Intermediation, Delegated Monitoring, and Long-Term Relationships, *Journal of Banking and Finance* 13, 9-20.

[22] Heckman, J. H., 1979, Sample Selection Bias as a Specification Error, *Econometrica* 47, 153-162.

[23] Hellmann, T., Lindsey, L. Puri, M., 2008, Building Relationships Early: Banks in Venture Capital. *Review of Financial Studies* 21, 2, 513-541.

[24] James, C. M., 1987. Some evidence on the uniqueness of bank loans. *Journal of Financial Economics* 19(2), 217-235.

[25] James, C., Wier, P., 1990. Borrowing relationships, intermediation, and the cost of issuing public securities. *Journal of Financial Economics* 28, 149-171.

[26] Jayaratne, J., Strahan, P., 1996. The finance-growth nexus: Evidence from bank branch deregulation. *Quarterly Journal of Economics* 111, 639-670.

[27] Lummer, S. L., McConnell, J. J., 1989. Further evidence on the bank lending process and the capital-market response to bank loan agreements. *Journal of Financial Economics* 25(1), 99-122.

[28] Mester, L., Nakamura, L., Renault, M., 2007. Transaction accounts and loan monitoring. *Review of Financial Studies* 20, 529-556.

[29] Norden, L., Weber, M., 2009. Credit Line Usage, Checking Account Activity, and Default Risk of Bank Borrowers. *Review of Financial Studies,* forthcoming.

[30] Petersen, M., 2009. Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches, *Review of Financial Studies*, 22, 435-480.

[31] Petersen, M., Rajan, R., 1994. The benefits of lending relationships: Evidence from small business data, *Journal of Finance* 49, 1367–1400.

[32] Pindyck, R., Rubinfeld, D., 1998. Econometric Models and Economic Forecasts. 4th ed. New York: McGraw-Hill.

[33] Poirier, D. J., 1980, Partial Observability in Bivariate Probit Models, Journal of Econometrics, 12, 210-217.

[34] Puri, M., 1996. Commercial banks in investment banking: Conflict of interest or certification role? *Journal of Financial Economics* 40, 373-401.

[35] Rajan, R. 1992. Insiders and Outsiders: The Choice Between Informed and Arm's-Length Debt. *Journal of Finance* 47,1367-1406.

[36] Ramakrishnan, R.T., and Thakor, A.V., 1984. Information Reliability and a Theory of Financial Information. *Review of Economic Studies* 51, 415-432.

[37] Santikian, L., 2009. The Ties That Bind: Bank Relationships and Small Business Lending. Working Paper.

[38] Stein, J. C., 2002. Information Production and Capital Allocation: Decentralized versus Hierarchical Firms. *Journal of Finance* 57, 1891-1921.

[39] Van de Ven, W., Van Praag, B., 1981, The Demand for Deductibles in Private Health Insurance – A Probit Model with Sample Selection. *Journal of Econometrics*, 17, 229-252.

Appendix 1

Definition of Relationship Variables

Relationship Dummy	
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 & Checking Account Information	
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.
Scope of Relationships: Assets held in the Bank (Yes/No)	
Assets (Yes/No)	Dummy variable equal to 1 if the borrower has accounts with the savings bank other than checking accounts (e.g. savings account, brokerage account).
Scope of Relationships: Assets held in the Bank (Amount)	
< 50 EUR	Dummy variable equal to 1 if the borrower has less than 50 Euro in accounts other than checking accounts.
>50 EUR , < 2000 EUR	Dummy variable equal to 1 if the borrower has between 50 and 2000 Euro in accounts other than checking accounts.
>=2000 EUR	Dummy variable equal to 1 if the borrower has more than 2000 Euro in accounts other than checking accounts.
Transaction Account Behavior	
Behavioral Score	A quantitative score of a borrower's transaction account activity prior to loan origination. It comprises information such as the number of transaction accounts, the maximum balance during the last six months, the percentage of months with a negative balance during the last 6 months, the usage of the credit limit, the percentage of months in excess of the credit limit, sum of account credits of the last months relative to the average account credits of the last six months, the number of return debit notes during the last six months and the longest period of a declining maximum account balance. The factors are weighted with respect to their power in predicting borrower defaults (based on out of sample data) and summed to a single behavioral score.
Instruments	
Log(Branches/Population)	Natural logarithm of the number of savings bank branches over the population in the area the respective savings bank is operating in.
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.

Appendix 2

Definition of Control Variables

Income	
Log(Income)	Log(Income) is the natural logarithm of the monthly net income of the applicant measured in Euro.
Repayment Burden (% of Income)	This variable measures the applicant's burden to repay the loan and 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 the repayment burden is below 5%.
>= 5%, < 11%	Dummy variable equal to 1 if the repayment burden is between 5% and 11%.
>=11%, < 13%	Dummy variable equal to 1 if the repayment burden is between 11% and 13%.
>=13%, < 20%	Dummy variable equal to 1 if the repayment burden is between 13% and 20%.
>20%	Dummy variable equal to 1 if the repayment burden is above 20%.
Undisclosed	Dummy variable indicating that repayment burden is undisclosed.
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.
Borrower Industry	
Borrower Industry	Borrower Industry' comprises the different industries the applicants are working in (we use dummy variables for each industry): Service, Metalworking, public service, construction, communication. Energy and water supply, etc. Further included are dummies for unemployed applicants, retirees, etc.
Job Stability	
<=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.
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.
External Rating	
Rating 1 – Rating 8	We segregate the internal rating score in 8 different rating categories based on the default probability of the borrower. Category 1 is the lowest, category 8 the highest default probability, respectively.

Appendix 3

Banks and branches in Germany

This table presents an overview of the different banks and the number of their respective branches which are used to calculate the Herfindahl-Hirschmann-Index (HHI) which is used as an instrument in the paper.

Bank Name	Total # Branches	Average # Branches
Aareal Bank AG	10	0.02
Baden-Württembergische Bank AG	53	0.12
Bankgesellschaft Berlin AG	58	0.13
Bayerische Hypo- und Vereinsbank AG	639	1.46
CC-Bank AG	78	0.18
Citibank	287	0.65
Commerzbank AG	812	1.85
CreditPlus Bank AG	13	0.03
DaimlerChrysler Bank AG	9	0.02
Delbrück Bethmann Maffei AG	11	0.03
Deutsche Bank AG	765	1.74
Dresdner Bank AG	725	1.65
DVB Bank AG	1	0.00
GE Money Bank	85	0.19
Oldenburgische Landesbank AG	206	0.47
ReiseBank AG	70	0.16
SEB AG	180	0.41
Bankhaus Max Flessa KG	28	0.06
Fürstlich Castell´sche Bank	19	0.04
Hanseatic Bank GmbH & Co KG	30	0.07
Privatbankiers insgesamt	178	0.41
Reuschel & Co KG	9	0.02
Landesbank Baden-Württemberg	221	0.50
Landesbank Berlin - Girozentrale	162	0.37
Landesbank Hessen-Thüringen Girozentrale	3	0.01
Norddeutsche Landesbank Girozentrale	115	0.26
Savings Banks	**13,850**	**31.55**
Badische Beamtenbank eG	87	0.20
Deutsche Apotheker- und Ärztebank eG	47	0.11
DZ-Bank AG	11	0.03
norisbank AG	100	0.23
PSD Bank	27	0.06
Sparda-Banken eG	389	0.89
Volks- und Raiffeisenbanken	12,372	28.18
Deutsche Postbank AG	13,772	31.37
Other Banks	1,987	4.53

Table 1

Descriptive Statistics

This table presents summary statistics for the sample of 1,068,000 consumer loans originated by German savings banks from November 2004 through June 2008. The number of observations corresponds to the number of approved loan applications.

	N	Mean	SD	Min	p25	p50	p75	Max
Loan Characteristics								
Approved[1)]	1,091,999	0.978	0.147	0	1	1	1	1
Loan Performance Characteristics								
Default Rate (12 months)	1,068,000	0.006	0.077	0	0	0	0	1
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
Repayment Burden (% of Income)								
< 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
Undisclosed[2)]	1,068,000	0.388	0.487	0	0	0	1	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
Metal Working Industry	1,068,000	0.204	0.403	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[3)]	1,068,000	0.239	0.426	0	0	0	0	1
Job Stability								
<=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
External Rating								
External Rating Score	86,628	4.256	2.359	1	2	4	6	8

1) Number of observations is number of loan applications.

2) In the regressions we combine the >=20% and the undisclosed categories; using undisclosed as separate category does not change our results.

3) "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

Table 2

Relationship Characteristics

This table presents summary statistics for the sample of 1,068,000 consumer loans originated by German savings banks from November 2004 through June 2008. The number of observations corresponds to the number of approved loan applications. All variables are proxies for the existence and scope of bank-borrower relationships: (1) existence of a relationship and relationship length, (2) transaction accounts, (3) savings accounts and (4) prior consumer loans.

	N	Mean	SD	Min	p25	p50	p75	Max
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
Scope of Relationships: Savings Accounts								
No savings account	1,068,000	0.232	0.422	0	0	0	0	1
< 50 Euro	819,913	0.197	0.398	0	0	0	0	1
>= 50, < 2000 Euro	819,913	0.363	0.481	0	0	0	1	1
>= 2000 Euro	819,913	0.185	0.388	0	0	0	0	1
Prior Consumer Loans								
Prior loan (yes)	1,068,000	0.192	0.394	0	0	0	0	1
Prior loan with 1 year look-back	1,068,000	0.121	0.326	0	0	0	0	1
Prior loan with 2 year look-back	1,068,000	0.174	0.379	0	0	0	0	1
Prior loan with 3 year look-back	1,068,000	0.192	0.394	0	0	0	0	1

Table 3

Univariate results

This table reports univariate results relating borrower default rates to various relationship measures. The dependent variable is the 12 month default rate of the borrower. In model (1), non relationship customers are omitted. Models (2) to (6) 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) relationships > 15 years, the group of customers with no credit and debit card and without credit line are simultaneously omitted. In models (7) and (8), we omit the group of customers without savings accounts and assets above 2000 Euro, respectively. All models are estimated using OLS. ***, **, * denote significance levels at the 1, 5 and 10 percent level, respectively.

	(1)		(2)		(3)		(4)		(5)		(6)		(7)		(8)	
Relationship Characteristics																
Relationships (Yes)	-0.010***	<0.001														
Relationship Length																
Relationship <2years			0.015***	<0.001							0.014***	<0.001				
Relationship >= 2, <5 years			0.009***	<0.001							0.008***	<0.001				
Relationship >= 5, <9 years			0.005***	<0.001							0.005***	<0.001				
Relationship >= 9, <15 years			0.002***	<0.001							0.002***	<0.001				
Relationship >= 15 years																
Scope of Relationships: Transaction Accounts																
Debit and credit card					-0.012***	<0.001					-0.004***	<0.001				
Debit card					-0.011***	<0.001					-0.004***	<0.001				
Credit card					-0.008***	<0.001					-0.003***	<0.001				
No cards					-0.005***	<0.001					-0.001***	(.002)				
Credit line							-0.011***	<0.001								
Used > 150%									0.010***	(.001)	0.010***	(.001)				
Used > 120%, <=150%									-0.003**	(.001)	-0.002**	(.001)				
Used > 100%, <=120%									-0.006***	<0.001	-0.005***	<0.001				
Used > 80%, <=100%									-0.012***	<0.001	-0.010***	<0.001				
Used > 20%, <=80%									-0.016***	<0.001	-0.014***	<0.001				
Used > 0%, <=20%									-0.017***	<0.001	-0.015***	<0.001				
Positive account balance									-0.016***	<0.001	-0.015***	<0.001				
Scope of Relationships: Savings Accounts																
Savings account													-0.005***	<0.001		
No savings account															0.007***	<0.001
< 50 Euro															0.006***	<0.001
>= 50, < 2000 Euro															0.002***	<0.001
# of observations	1,068,000		1,041,291		1,068,000		1,041,291		1,041,291		1,041,291		1,068,000		1,068,000	

Table 4

Private information from transaction accounts and borrower defaults

This table presents the results of a probit regression. 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 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 Appendix I. Models (2) to (6) 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) relationships > 15 years, the group of customers with no credit and debit card and without credit line are simultaneously omitted. The coefficients for borrower industries (as described in Appendix I) as well as intercept, bank and time fixed effects are not shown. Only the marginal effects are shown. Heteroscedasticity consistent standard errors clustered at the bank level are shown in parentheses. ***, **, * denote significance levels at the 1, 5 and 10 percent level, respectively.

	(1)		(2)		(3)		(4)		(5)		(6)	
Relationship Characteristics												
Relationship Yes/No												
Relationship (Yes)	-0.006***	(.001)										
Relationship Length												
Relationship <2years			0.014***	(.001)							0.011***	(.001)
Relationship >=2, <5years			0.008***	<0.001							0.006***	(.001)
Relationship >=5, <9years			0.004***	<0.001							0.003***	<0.001
Relationship >=9, <15years			0.002***	<0.001							0.001***	<0.001
Scope of Relationships: Cards & Checking Account Information												
Debit and Credit Card					-0.003***	<0.001					-0.002***	<0.001
Debit Card					-0.003***	<0.001					-0.002***	<0.001
Credit Card					-0.001**	<0.001					-0.001**	<0.001
Scope of Relationships: Credit Line												
Credit Line (Yes)							-0.008***	(.001)				
Used > 150%									0.003***	(.001)	0.003***	(.001)
Used > 120%, <=150%									-0.001	<0.001	-0.00002	<0.001
Used > 100%, <=120%									-0.001***	<0.001	-0.001*	<0.001
Used > 80%, <=100%									-0.003***	<0.001	-0.002***	<0.001
Used > 20%, <=80%									-0.005***	<0.001	-0.004***	<0.001
Used > 0%, <=20%									-0.004***	<0.001	-0.003***	<0.001
Positive account balance									-0.006***	<0.001	-0.005***	<0.001

Table 4 (cont'd)

Borrower Characteristics	**(1)**		**(2)**		**(3)**		**(4)**		**(5)**		**(6)**	
Log (Income)	-0.002***	<0.001	-0.002***	<0.001	-0.001***	<0.001	-0.002***	0.00023	-0.001***	<0.001	-0.001***	<0.001
Repayment Burden (% of Income)												
< 5%	-0.002***	(.001)	-0.002***	(.001)	-0.002***	(.001)	-0.002***	(.001)	-0.002***	(.001)	-0.002***	<0.001
>= 5%, < 11%	-0.003***	(.001)	-0.002***	(.001)	-0.003***	(.001)	-0.002***	(.001)	-0.002***	(.001)	-0.002***	(.001)
>=11%, < 13%	-0.002**	(.001)	-0.002***	(.001)	-0.002***	(.001)	-0.001**	(.001)	-0.001**	(.001)	-0.002***	(.001)
>=13%, < 20%	-0.001	(.001)	-0.001	(.001)	-0.0013281	(.001)	-0.001	(.001)	-0.001*	(.001)	-0.001**	(.001)
Age												
18 to 23 years	0.012***	(.003)	0.004**	(.002)	0.009***	(.003)	0.012***	(.003)	0.011***	(.003)	0.003**	(.002)
23 to 25 years	0.010***	(.002)	0.003**	(.001)	0.008***	(.002)	0.010***	(.002)	0.010***	(.002)	0.003**	(.001)
25 to 30 years	0.008***	(.002)	0.003***	(.001)	0.006***	(.002)	0.008***	(.002)	0.006***	(.002)	0.002**	(.001)
30 to 45 years	0.006***	<0.001	0.003***	0.00077	0.005***	(.001)	0.006***	0.00091	0.004***	(.001)	0.002***	(.001)
45 to 50 years	0.005***	<0.001	0.003***	0.00067	0.004***	(.001)	0.005***	0.00082	0.004***	(.001)	0.002***	(.001)
50 to 60 years	0.004***	<0.001	0.003***	0.00051	0.003***	(.001)	0.004***	0.00059	0.003***	(.001)	0.002***	<0.001
Job Stability												
<= 2 years	0.005***	<0.001	0.003***	<0.001	0.003***	<0.001	0.005***	<0.001	0.004***	<0.001	0.003***	<0.001
Borrower Industry	Yes		Yes		Yes		Yes		Yes		Yes	
Time Fixed Effects	Yes		Yes		Yes		Yes		Yes		Yes	
Diagnostics												
Prob > X^2	0.000		0.000		0.000		0.000		0.000		0.000	
Log Pseud likelihood	-37,032.6 7		-34,393.717		-34,287.289		-34,849. 81		-33,664.989		-32,86975	
Pseudo R^2	5.26%		6.83%		5.93%		5.60%		8.81%		10.96%	
# of obser ations	1,068,000		1,041,291		1,041,291		1,041,291		1,041,291		1,0 1 291	
# of bank clusters	296		296		296		296		296		296	

Table 5

Private information from savings accounts and borrower defaults

This table presents the results of a probit regression. 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 relationships characteristics as a result of relationships via savings account (the existence of savings accounts and assets held in these accounts). All variables are defined in Appendix I. In model (2), the omitted relationship variable is assets > 2000 Euros. The coefficients for borrower industries (as described in Appendix I) as well as intercept and time fixed effects are not shown. Only the marginal effects are reported. Heteroscedasticity consistent standard errors clustered at the bank level are shown in parentheses. ***,**,* denote significance levels at the 1, 5 and 10 percent level, respectively.

	(1)		(2)	
Relationship Characteristics				
Scope of Relationships: Assets held in the Bank (Yes/No)				
Savings Accounts	-0.004***	<0.001		
Scope of Relationships: Assets held in the Bank (Amount)				
No Assets			0.005***	(.001)
< 50 EUR			0.004***	(.001)
>50 EUR , < 2000 EUR			0.001***	<0.001
Borrower Characteristics				
Log (Income)	-0.002***	<0.001	-0.002***	<0.001
Repayment Burden (% of Income)				
< 5%	-0.002***	(.001)	-0.002***	(.001)
>= 5%, < 11%	-0.002***	(.001)	-0.003***	(.001)
>=11%, < 13%	-0.002**	(.001)	-0.001***	(.001)
>=13%, < 20%	-0.001	(.001)	-0.001**	(.001)
Age				
18 to 23 years	0.013***	(.003)	0.011***	(.003)
23 to 25 years	0.011***	(.002)	0.010***	(.002)
25 to 30 years	0.008***	(.002)	0.007***	(.002)
30 to 45 years	0.006***	(.001)	0.005***	(.001)
45 to 50 years	0.005***	(.001)	0.004***	(.001)
50 to 60 years	0.004***	(.001)	0.004***	(.001)
Job Stability				
<= 2 years	0.004***	<0.001	0.003***	<0.001
Borrower Industry	Yes		Yes	
Time Fixed Effects	Yes		Yes	
Diagnostics				
Prob > X^2	0.000		0.000	
Log Pseudolikelihood	-36,840.39		-36,793.20	
Pseudo R^2	5.76%		.88%	
# of observations	1,068,000		1,068 00	
# of bank clu ters	296		296	

Table 6

Private information from prior consumer loans and borrower defaults

This table presents the results of a probit regression. 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 relationships characteristics based on repeat lending with different look-back windows. *Prior Loan within 2 yr (1yr) look-back* are dummy variables equal to 1 if the borrower was granted a loan within 2 years (1 year) prior to the current loan. # *Prior Loan Defaults* measures the number of loans the borrower defaulted on in the past and which were originated during our sample period. All variables are defined in Appendix I. The coefficients for borrower industries (as described in Appendix I) as well as intercept and time fixed effects are not shown. Only the marginal effects are reported. Heteroscedasticity consistent standard errors clustered at the bank level are shown in parentheses. ***, **, * denote significance levels at the 1, 5 and 10 percent level, respectively.

	(1)		(2)		(3)	
Relationship Characteristics						
Prior Consumer Loans						
Prior Loan (Yes/No)	-0.003***	<0.001				
Prior Loan within 2 yr look-back			-0.003***	<0.001		
Prior Loan within 1 yr look-back					-0.003***	<0.001
Past Loan Performance						
# Prior Loan Defaults	0.022***	(.003)	0.022***	(.003)	0.022***	(.003)
Borrower Characteristics						
Log (Income)	-0.002***	<0.001	-0.002***	<0.001	-0.002***	<0.001
Repayment Burden (% of Income)						
< 5%	-0.002***	<0.001	-0.002***	<0.001	-0.002***	<0.001
>= 5%, < 11%	-0.002***	(.001)	-0.002***	(.001)	-0.002***	(.001)
>=11%, < 13%	-0.001**	(.001)	-0.001**	(.001)	-0.001**	(.001)
>=13%, < 20%	-0.001	(.001)	-0.001	(.001)	-0.001	(.001)
Age						
18 to 23 years	0.010***	(.002)	0.010***	(.002)	0.010***	(.002)
23 to 25 years	0.008***	(.002)	0.008***	(.002)	0.008***	(.002)
25 to 30 years	0.006***	(.001)	0.006***	(.001)	0.006***	(.001)
30 to 45 years	0.005***	(.001)	0.005***	(.001)	0.005***	(.001)
45 to 50 years	0.004***	(.001)	0.004***	(.001)	0.004***	(.001)
50 to 60 years	0.003***	<0.001	0.003***	<0.001	0.003***	<0.001
Job Stability						
<= 2 years	0.004***	<0.001	0.004***	<0.001	0.004***	<0.001
Borrower Industry	Yes		Yes		Yes	
Time Fixed Effects	Yes		Yes		Yes	
Diagnostics						
Prob > X^2	0.000		0.000		0. 00	
Log Pseudolikelihood	-31,947.6		-32,040.09		-32,222.08	
Pseudo R^2	18.27%		18.04%		17.57%	
# of observations	1,068,000		1,068,000		1,068,000	
# of bank clusters	296		296		296	

Table 7

Combinations of relationship measures and borrower defaults

This table presents the results of a probit regression. The dependent variable is a binary variable equal to 1 if the borrower defaults within the first 12 months after loan origination. Model (1) repeats the analysis from model (6) in Table 4 and model (2) adds whether or not the borrower also had a savings account. Model (3) considers whether borrowers had simultaneously checking and savings accounts at their bank. Model (4) adds whether or not the borrower had a prior loan during our sample period controlling for previous loan defaults to model specification (2). The coefficients for borrower industries (as described in Appendix I) as well as intercept, bank and time fixed effects are not shown. Only the marginal effects are shown. Heteroscedasticity consistent standard errors clustered at the bank level are shown in parentheses. ***, **, * denote significance levels at the 1, 5 and 10 percent level, respectively.

	(1)		(2)		(3)		(4)	
Relationship Characteristics								
Relationship Length								
Relationship <2years	0.011***	(.001)	0.010***	(.001)			0.007***	(.001)
Relationship >=2, <5years	0.006***	(.001)	0.005***	(.001)			0.004***	<0.001
Relationship >=5, <9years	0.003***	<0.001	0.003***	<0.001			0.002***	<0.001
Relationship >=9, <15years	0.001***	<0.001	0.001***	<0.001			0.001***	<0.001
Scope of Relationships: Cards & Checking Account Information								
Debit and Credit Card	-0.002***	<0.001	-0.002***	<0.001			-0.001***	<0.001
Debit Card	-0.002***	<0.001	-0.002***	<0.001			-0.001***	<0.001
Credit Card	-0.001**	<0.001	-0.001*	<0.001			-0.001***	<0.001
Scope of Relationships: Credit Line								
Used > 150%	0.003***	(.001)	0.003***	(.001)			0.003***	(.001)
Used > 120%, <=150%	-0.00002	<0.001	0.0001	<0.001			0.001	<0.001
Used > 100%, <=120%	-0.001*	<0.001	-0.0005	<0.001			-0.0001	<0.001
Used > 80%, <=100%	-0.002***	<0.001	-0.002***	<0.001			-0.001***	<0.001
Used > 20%, <=80%	-0.004***	<0.001	-0.004***	<0.001			-0.003***	<0.001
Used > 0%, <=20%	-0.003***	<0.001	-0.003***	<0.001			-0.002***	<0.001
Positive account balance	-0.005***	<0.001	-0.005***	<0.001			-0.004***	<0.001
Scope of Relationships: Assets held in the Bank (Amount)								
Checking Accounts					-0.003***	(.001)		
Savings Accounts			-0.002***	<0.001	-0.001	(.001)	-0.001***	<0.001
Savings Accounts x Checking Accounts					-0.002	(.002)		
Prior Consumer Loans								
Prior Loan (Yes/No)							-0.002***	<0.001
Past Loan Performance								
# Prior Loan Defaults							0.014***	(.002)

Table 7 (cont'd)

	(1)		(2		(3)		(4)	
Borrower Characteristics								
Log (Income)	-0.001***	<0.001	-0.001***	<0.001	-0.002***	<0.001	-0.001***	<0.001
Repayment (% of Income)								
< 5%	-0.002***	<0.001	-0.002***	<0.001	-0.002***	(.001)	-0.002***	<0.001
>= 5%, < 11%	-0.002***	(.001)	-0.002***	(.001)	-0.002***	(.001)	-0.002***	<0.001
>=11%, < 13%	-0.002***	(.001)	-0.002***	(.001)	-0.002**	(.001)	-0.001***	<0.001
>=13%, < 20%	-0.001**	(.001)	-0.001**	(.001)	-0.001	(.001)	-0.001**	<0.001
Age								
18 to 23 years	0.003**	(.002)	0.003**	(.002)	0.012***	(.003)	0.002**	(.001)
23 to 25 years	0.003**	(.001)	0.003**	(.001)	0.010***	(.002)	0.002**	(.001)
25 to 30 years	0.002**	(.001)	0.002**	(.001)	0.008***	(.002)	0.002**	(.001)
30 to 45 years	0.002***	(.001)	0.002***	<0.001	0.006***	(.001)	0.002***	<0.001
45 to 50 years	0.002***	(.001)	0.002***	<0.001	0.005***	(.001)	0.002***	<0.001
50 to 60 years	0.002***	<0.001	0.002***	<0.001	0.004***	(.001)	0.001***	<0.001
Job Stability								<0.001
<= 2 years	0.003***	<0.001	0.003***	<0.001	0.003***	<0.001	0.002***	<0.001
Borrower Industry	Yes		Yes		Yes		Yes	
Time Fixed Effects	Yes		Yes		Yes		Yes	
Diagnostics								
Prob > χ^2	0.000		0.000		0.000		0.000	
Log Pseudolikelihood	-32,869.785		-32,771.56		-36,807.12		-28,094.10	
Pseudo R^2	10.96%		11.23%		5.84%		23.90%	
Number of observations	1,041,291		1,041,291		1,041,291		1,041,291	
Number of bank clusters	296		296		296		296	

Table 8

Internal Ratings

This table presents the results of a probit regression. The dependent variable is a binary variable equal to 1 if the borrower defaults within the first 12 months after loan origination. The relationship variables for (i) transaction accounts, (ii) savings accounts and (iii) prior loans are defined as in the previous tables. All regression models use the internal ratings that form the basis for the credit decision of the banks to control for borrower risk. The coefficients for borrower industries (as described in Appendix I) as well as intercept, bank and time fixed effects are not shown. Only the marginal effects are shown. Heteroscedasticity consistent standard errors clustered at the bank level are shown in parentheses. ***,**,* denote significance levels at the 1, 5 and 10 percent level, respectively.

	(1)		(2)		(3)		(4)		(5)		(6)	
Relationship Characteristics												
Relationship Yes/No												
Relationship (Yes)	-0.003***	(.001)										
Relationship Length												
Relationship <2years			0.012***	(.002)								
Relationship >=2, <5years			0.007***	(.001)								
Relationship >=5, <9years			0.003***	(.001)								
Relationship >=9, <15years			0.002***	<0.001								
Scope of Relationships: Cards & Checking Account Information												
Debit and Credit Card					-0.003***	(.001)						
Debit Card					-0.003***	(.001)						
Credit Card					-0.002***	<0.001						
Credit Line (Yes)							-0.001***	(.001)				
Scope of Relationships: Assets held in the Bank												
Savings Accounts									-0.003***	<0.001		
Prior Consumer Loans												
Prior Loan (Yes/No)											-0.004***	(.001)

Table 8 (cont'd)

	(1)		(2)		(3)		(4)		(5)		(6)	
Internal Rating												
1	-0.008***	(.001)	-0.007***	(.001)	-0.007***	(.001)	-0.007***	(.001)	-0.007***	(.001)	-0.006***	(.001)
2	-0.006***	(.001)	-0.005***	(.001)	-0.005***	(.001)	-0.005***	(.001)	-0.005***	(.001)	-0.004***	<0.001
3	-0.006***	(.001)	-0.005***	(.001)	-0.005***	(.001)	-0.005***	(.001)	-0.006***	(.001)	-0.004***	<0.001
4	-0.006***	(.001)	-0.005***	(.001)	-0.006***	(.001)	-0.005***	(.001)	-0.006***	(.001)	-0.004***	<0.001
5	-0.006***	<0.001	-0.005***	<0.001	-0.006***	<0.001	-0.005***	<0.001	-0.006***	<0.001	-0.004***	<0.001
6	-0.006***	(.001)	-0.005***	(.001)	-0.006***	(.001)	-0.005***	(.001)	-0.005***	(.001)	-0.004***	<0.001
7	-0.005***	(.001)	-0.004***	<0.001	-0.005***	<0.001	-0.004***	<0.001	-0.005***	(.001)	-0.004***	<0.001
8	-0.005***	(.001)	-0.004***	<0.001	-0.004***	<0.001	-0.004***	<0.001	-0.004***	(.001)	-0.004***	<0.001
9	-0.004***	<0.001	-0.003***	<0.001	-0.004***	<0.001	-0.003***	<0.001	-0.004***	<0.001	-0.003***	<0.001
10	-0.003***	<0.001	-0.003***	<0.001	-0.003***	<0.001	-0.002***	<0.001	-0.003***	<0.001	-0.002***	<0.001
11	-0.002***	<0.001	-0.002***	<0.001	-0.002***	<0.001	-0.002***	<0.001	-0.002***	<0.001	-0.002***	<0.001
Borrower Industry	Yes		Yes		Yes		Yes		Yes		Yes	
Time Fixed Effects	Yes		Yes		Yes		Yes		Yes		Yes	
Diagnostics												
Prob > X^2	0.000		0.000		0.000		0.000		0.000		0.000	
Log Pseudolikelihood	-44,180.949		-41,179.000		-41,329.877		-41,922.810		-43,865.580		-37,598.160	
Pseudo R^2	4.56%		5.89%		5.55%		4.19%		5.24%		18.78%	
Number of observations	1,214,741		1,185,351		1,185,351		1,185,351		1,214,741		,214,741	
Number of bank clusters	298		298		298		298		298		298	

Table 9

External Ratings

This table presents the results of a probit regression. The dependent variable is a binary variable equal to 1 if the borrower defaults within the first 12 months after loan origination. The relationship variables for (i) transaction accounts, (ii) savings accounts and (iii) prior loans are defined as in the previous tables. External ratings are dummy variables with the external rating category 1 representing the highest borrower quality class, and rating categeory 8 (omitted) representing the lowest quality borrower class, respectively. The coefficients for borrower industries (as described in Appendix I) as well as intercept, bank and time fixed effects are not shown. Only the marginal effects are shown. Heteroscedasticity consistent standard errors clustered at the bank level are shown in parentheses. ***,**,* denote significance levels at the 1, 5 and 10 percent level, respectively.

	(1)		(2)		(3)		(4)		(5)		(6)	
Relationship Characteristics												
Relationship Yes/No												
Relationship (Yes)	-0.006***	(.002)										
Relationship Length												
Relationship <2years			0.005***	(.001)								
Relationship >=2, <5years			0.003***	(.001)								
Relationship >=5, <9years			0.001**	<0.001								
Relationship >=9, <15years			0.0001	(.001)								
Scope of Relationships: Cards & Checking Account Information												
Debit and Credit Card					-0.002***	<0.001						
Debit Card					-0.002***	<0.001						
Credit Card					-0.0004	<0.001						
Credit Line (Yes)							-0.001***	<0.001				
Scope of Relationships: Assets held in the Bank												
Savings Accounts									-0.001***	<0.001		
Prior Consumer Loans												
Prior Loan (Yes/No)											-0.002***	<0.001
External Rating												
1	-0.003***	(.001)	-0.002***	<0.001	-0.002***	<0.001	-0.003***	<0.001	-0.003***	<0.001	-0.001***	<0.001
2	-0.003***	<0.001	-0.002***	<0.001	-0.002***	<0.001	-0.003***	<0.001	-0.003***	<0.001	-0.001***	<0.001
3	-0.002***	(.001)	-0.002***	<0.001	-0.002***	<0.001	-0.002***	<0.001	-0.002***	<0.001	-0.001***	<0.001
4	-0.002***	<0.001	-0.002***	<0.001	-0.002***	<0.001	-0.002***	<0.001	-0.002***	<0.001	-0.001***	<0.001
5	-0.002***	<0.001	-0.001***	<0.001	-0.001***	<0.001	-0.002***	<0.001	-0.002***	<0.001	-0.001***	<0.001
6	-0.002***	<0.001	-0.001***	<0.001	-0.001***	<0.001	-0.001***	<0.001	-0.002***	<0.001	-0.001***	<0.001
7	-0.002***	<0.001	-0.002***	<0.001	-0.002***	<0.001	-0.002***	<0.001	-0.002***	<0.001	-0.001***	<0.001

Table 9 (cont'd)

	(1)		(2)		(3)		(4)		(5)		(6)	
Log (Income)	-0.002***	<0.001	-0.001	(.001)	-0.001	(.001)	-0.001	(.001)	-0.001***	<0.001	-0.0003	<0.001
Repayment Burden (% of Income)												
< 5%	-0.002***	(.001)	-0.002***	(.001)	-0.002***	(.001)	-0.002***	(.001)	-0.002***	(.001)	-0.001***	<0.001
>= 5%, < 11%	-0.003***	(.001)	-0.002***	(.001)	-0.002***	(.001)	-0.002***	(.001)	-0.002***	(.001)	-0.001***	<0.001
>=11%, < 13%	-0.002**	(.001)	-0.002***	(.001)	-0.002***	(.001)	-0.002***	(.001)	-0.001**	(.001)	-0.001***	<0.001
>=13%, < 20%	-0.002	(.001)	-0.002***	(.001)	-0.002***	(.001)	-0.002***	(.001)	-0.001*	(.001)	-0.001***	<0.001
Age												
18 to 23 years	0.003	(.002)	0.001	(.001)	0.001	(.002)	0.002	(.002)	0.003	(.002)	0.002	(.001)
23 to 25 years	0.005	(.003)	0.003*	(.002)	0.003	(.002)	0.005	(.003)	0.005	(.003)	0.003	(.002)
25 to 30 years	0.002	(.002)	0.001	(.002)	0.001	(.002)	0.002	(.002)	0.002	(.002)	0.002	(.002)
30 to 45 years	0.002	(.001)	0.001	(.001)	0.002	(.001)	0.002*	(.001)	0.002	(.001)	0.002*	(.001)
45 to 50 years	0.002	(.002)	0.002	(.002)	0.002	(.002)	0.002	(.002)	0.002	(.002)	0.002	(.001)
50 to 60 years	0.003*	(.001)	0.001*	(.001)	0.001	(.001)	0.002	(.001)	0.003*	(.001)	0.002	<0.001
Job Stability												
<= 2 years	0.002***	<0.001	0.001**	<0.001	0.001***	<0.001	0.001***	<0.001	0.004***	<0.001	0.003***	<0.001
Borrower Industry	Yes		Yes		Yes		Yes		Yes		Yes	
Time Fixed Effects	Yes		Yes		Yes		Yes		Yes		Yes	
Diagnostics												
Prob > X^2	0.000		0.000		0.000		0.000		0.000		0.000	
Log Pseudolikelihood	-2,063.117		-1,835.231		-1,831.411		-1,857.582		-2,075.121		-1,545.024	
Pseudo R^2	9.44%		9.92%		10.11%		8.82%		9.18%		32.29%	
Number of bservations	86,628		83,151		83,151		83,151		86,628		86,628	
Number of ba k clusters	73		73		73		73		73		73	

Table 10

Endogeneity of Relationships (Bivariate Probit Models)

This table presents the results of a bivariate probit regression. The variables used are the same as in the previous models, as well as two additional instruments. Log(Branches/Population) is the natural logarithm of the number of branches of each bank as a percentage of inhabitants of the particular region the bank is headquartered. Log (HHI) is the natural logarithm of the Hirshman-Herfindahl Index. The number of branches of each particular bank is used to construct the HHI. In the diagnostic section we test the H0 that the contemporaneous correlation between the error terms of the 1st stage and 2nd stage equation are uncorrelated. ***, **, * denote significance levels at the 1, 5 and 10 percent level, respectively.

	Probit		Bivariate Probit Models							
			Panel A: Relationship Equation				Panel B: Default Equation			
			(1)		(2)		(1)		(2)	
	Default		Relationships		Relationships		Default	Marginal Effects	Default	Marginal Effects
Relationship (Yes)	-0.006***	(.001)	N/A		N/A		-.814***	(.311)	-0.025	-0.772**	(.317)	-0.022
Borrower Characteristics												
Log (Income)	-0.002***	<0.001	-0.063***	(.007)	-0.062***	(.007)	-.161***	(.01)	-0.002	-0.161***	(.01)	-0.002
Repayment (% of Income)												
< 5%	-0.002***	(.001)	-0.069***	(.01)	-0.067***	(.011)	-.165***	(.025)	-0.001	-0.165***	(.025)	-0.001
>= 5%, < 11%	-0.003***	(.001)	0.038***	(.007)	0.041***	(.007)	-0.125***	(.014)	-0.001	-0.126***	(.014)	-0.001
>=11%, < 13%	-0.002**	(.001)	0.003	(.011)	0.005	(.012)	-0.074***	(.022)	-0.001	-0.074***	(.022)	-0.001
>=13%, < 20%	-0.001	(.001)	-0.021**	(.008)	-0.018**	(.009)	-0.026***	(.016)	0.000	-0.026	(.016)	0.000
Age												
18 to 23 years	0.012***	(.003)	-0.276***	(.019)	-0.276***	(.021)	0.386***	(.04)	0.006	0.387***	(.04)	0.006
23 to 25 years	0.010***	(.002)	-0.258***	(.019)	-0.258***	(.021)	0.356***	(.041)	0.006	0.356***	(.041)	0.006
25 to 30 years	0.008***	(.002)	-0.302***	(.017)	-0.303***	(.018)	0.289***	(.038)	0.004	0.289***	(.038)	0.004
30 to 45 years	0.006***	<0.001	-0.214***	(.015)	-0.214***	(.017)	0.314***	(.035)	0.004	0.315***	(.035)	0.004
45 to 50 years	0.005***	<0.001	-0.153***	(.017)	-0.153***	(.018)	0.255***	(.038)	0.003	0.256***	(.038)	0.003
50 to 60 years	0.004***	<0.001	-0.086***	(.016)	-0.086***	(.017)	0.216***	(.035)	0.003	0.216***	(.035)	0.003
Job Stability												
<= 2 years	0.005***	<0.001	-0.192***	(.007)	-0.192***	(.007)	0.303***	(.011)	0.004	0.303***	(.014)	0.004
Instruments												
Log(Branches / Population)			0.041***	(.009)	0.036***	(.006)	N/A		N/A	N/A		N/A
Log(HHI)					0.026**	(.013)				N/A		N/A
Diagnostics												
Wald Test (H0: ρ=0)							0.165	(.115)		0.185	(.118)	
Number of observation	1,068,000						880,966			880,966		
X^2(df)							10,857			10,848		
Prob > X^2							<0.0001			<0.0001		

Table 11

Default probabilities and sample selection

This table presents the results from a Heckman selection model with probit models in both selection and outcome equation. 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 whether or not the loan applicant was approved. The control variables used are the same as in the previous models, as well as two additional instruments. Log(Branches/Population) is the natural logarithm of the number of branches of each bank as a percentage of inhabitants of the particular region the bank is headquartered. 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. All variables are defined in Appendix I. ***,**,* denote significance levels at the 1, 5 and 10 percent level, respectively.

	Selection Models									
	Panel A: Selection Equation				Panel B : Default Equation					
	(1)		(2)		(1)			(2)		
	Access		Access		Default		Marginal Effects	Default		Marginal Effects
Relationship (Yes)	0.511***	(.013)	0.513***	(.013)	-0.297***	(.028)	-0.004	-0.293***	(.027)	-0.004
Borrower Characteristics										
Log (Income)	0.158***	(.007)	0.155***	(.007)	-0.155***	(.011)	-0.002	-0.154***	(.011)	-0.002
Repayment (% of Income)										
< 5%	0.410***	(.017)	0.399***	(.017)	-0.160***	(.025)	-0.001	-0.160***	(.025)	-0.001
>= 5%, < 11%	0.283***	(.008)	0.271***	(.008)	-0.123***	(.014)	-0.001	-0.122***	(.014)	-0.001
>=11%, < 13%	0.067***	(.012)	0.054***	(.012)	-0.073***	(.022)	-0.001	-0.072***	(.022)	-0.001
>=13%, < 20%	-0.003	(.009)	-0.015	(.009)	-0.025	(.016)	0.000	-0.025	(.016)	0.000
Age										
18 to 23 years	-0.965***	(.025)	-0.963***	(.025)	0.376***	(.042)	0.006	0.371***	(.041)	0.006
23 to 25 years	-0.620***	(.026)	-0.619***	(.026)	0.355***	(.041)	0.006	0.353***	(.041)	0.006
25 to 30 years	-0.556***	(.024)	-0.554***	(.024)	0.291***	(.038)	0.004	0.290***	(.038)	0.004
30 to 45 years	-0.414***	(.024)	-0.413***	(.024)	0.315***	(.035)	0.004	0.315***	(.035)	0.004
45 to 50 years	-0.284***	(.025)	-0.281***	(.025)	0.256***	(.038)	0.003	0.256***	(.038)	0.003
50 to 60 years	-0.165***	(.024)	-0.163***	(.024)	0.216***	(.035)	0.003	0.216***	(.035)	0.003
Job Stability										
<= 2 years	-0.413***	(.007)	-0.413***	(.007)	0.303***	(.014)	0.004	0.301***	(.014)	0.004
Instruments										
Log(Branches / Population)	0.062***	(.005)	0.086***	(.006)	N/A		N/A	N/A		N/A
Log(HHI)			-0.138***	(.014)				N/A		N/A
Diagnostics										
Wald Test (H0: ρ=0)					0.137	(.132)		0.192	(.131)	
Number of uncensored observations					880,966			880,966		
Number of censored observations					20,205			20,205		
X^2(df)					7,035			7,094		
Prob > X^2					< 0001			<0.0001		

Table 12

Default probabilities and sample selection: Alternative relationship proxies

This table presents the results from a Heckman selection model with probit models in both selection and outcome equation. 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 whether or not the loan applicant was approved. The control variables used are the same as in the previous models, as well as an additional instrument. Log(Branches/Population) is the natural logarithm of the number of branches of each bank as a percentage of inhabitants of the particular region the bank is headquartered. All variables are defined in Appendix I. In model (1), the omitted relationship variable is *Relationships > 15 years*; in (2) borrowers without a debit and credit card (*No Cards*) are omitted; in (4) customers with assets above 2,000 Euros are omitted (*>=2000EUR*). Coefficient and marginal effects are shown, except for borrower characteristics as well as intercept and time fixed effects. Panel A reports the outcome equation, Panel B the selection equation. Only the coefficients are shown in Panels A and B. Heteroscedasticity consistent standard errors clustered at the bank level are shown in parentheses. ***, **, * denote significance levels at the 1, 5 and 10 percent level, respectively.

Panel A: Outcome Equation (Dependent Variable: Default)

	(1)			(2)			(3)			(4)			(5)		
	Default		Marg. Eff.	Default		Marg. Eff.	Default		Marg. Eff.	Default		Marg. Eff.	Default		Marg. Eff.
Relationship Characteristics															
Relationship Length															
Relationship <2years	0.536***	(.038)	0.012												
Relationship >=2, <5years	0.387***	(.03)	0.007												
Relationship >=5, <9years	0.232***	(.026)	0.003												
Relationship >=9, <15years	0.110***	(.025)	0.001												
Scope of Relationships: Cards & Checking Account Information															
Debit and Credit Card				-0.476***	(.039)	-0.003									
Debit Card				-0.323***	(.025)	-0.003									
Credit Card				-0.090***	(.038)	-0.001									
Credit Line (Yes)							-.455***	(.027)	-0.009						
Scope of Relationships: Assets held in the Bank (Amount)															
No Assets										0.481***	(.033)	0.007			
< 50 EUR										0.347***	(.032)	0.005			
>50 EUR , < 2000 EUR										-0.178***	(.032)	0.002			
Prior Consumer Loans															
Prior Loan (Yes/No)													-.0822***	(.022)	-0.001

Panel B: Selection Equation (Dependent Variable: Access)

	(1)		(2)		(3)		(4)		(5)	
	Default	**Marg. Eff.**	**Default**	**Marg. Eff.**	**Default**	**Marg. Eff.**	**Default**	**Marg. Eff.**	**Default**	**Marg. Eff.**
Log(Branches/Population)	0.093*** (.012)	0.002	0.108*** (.012)	0.002	0.112*** (.012)	0.002	0.0984*** (.012)	0.002	-0.113*** (.012)	0.064
Relationship Characteristics										
Relationship Length										
Relationship <2years	-0.681*** (.018)	-0.025								
Relationship >=2, <5years	-0.512*** (.016)	-0.012								
Relationship >=5, <9years	-0.283*** (.016)	-0.004								
Relationship >=9, <15years	-0.134*** (.016)	0.000								
Scope of Relationships: Cards & Checking Account Information										
Debit and Credit Card			0.753*** (.025)	0.012						
Debit Card			0.568*** (.012)	0.016						
Credit Card			0.551*** (.026)	0.009						
Credit Line (Yes/No)					0.249*** (.019)	0.024				
Scope of Relationships: Assets held in the Bank (Amount)										
Assets (Yes)										
No Assets							-0.578*** (.018)	-0.013		
< 50 EUR							-0.209*** (.019)	-0.001		
>50 EUR , < 2000 EUR							0.038** (.019)	0.006		
Prior Consumer Loans										
Prior Loan (Yes/No)									.0245*** (.012)	0.054
Diagnostics										
Wald Test (H0: ρ=)	-0.189	(.167)	-0.086	(.67)	-0.165	(.307)	0.001	(.994)	-0.027	(.82)
Number of uncensored observations	398,900		398,900		398,900		398,900		398,900	
Number of censored observations	10,007		10,007		10,007		10,007		10,007	
X^2(df)	989.53		952.25		1094.12		64.86		840.77	
Prob > X^2	<0.0001		<0.0001		<0.0001		<0.0001		<0.0001	

Table 13

Private Information and Borrower Default Incentives

This table presents the results of a probit regression. The dependent variable is a binary variable equal to 1 if the borrower defaults within the first 12 months after loan origination. Behavioral Score measures the private information from transaction account behavior of the borrower prior to the loan origination date. 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 Appendix I. The coefficients for borrower industries (as described in Appendix I) as well as intercept, bank and time fixed effects are not shown. Only the marginal effects are shown. ***,**,* denote significance levels at the 1, 5 and 10 percent level, respectively.

	(1)		(2)		(3)		(4)		(5)		(6)	
Account Behavior												
Behavioral Score	-0.004***	<0.001	-0.003***	<0.001	-0.004***	<0.001	-0.003***	<0.001	-0.002***	<0.001	-0.001***	<0.001
Relationship Characteristics												
Relationship Yes/No												
Relationship (Yes)	-0.002***	(.001)										
Relationship Length												
Relationship <2years			0.012***	(.001)							0.007***	<0.001
Relationship >=2, <5years			0.007***	(.001)							0.003***	<0.001
Relationship >=5, <9years			0.004***	(.001)							0.002***	<0.001
Relationship >=9, <15years			0.002***	<0.001							0.001***	<0.001
Scope of Relationships: Cards & Checking Account Information												
Debit and Credit Card					-0.003***	<0.001					-0.002***	<0.001
Debit Card					-0.003***	<0.001					-0.002***	<0.001
Credit Card					-0.001	<0.001					-0.0003	<0.001
Scope of Relationships: Credit Line												
Credit Line (Yes)							-0.013***	(.002)				
Used > 150%									-0.00005	<0.001	0.0005	<0.001
Used > 120%, <=150%									-0.001***	<0.001	-0.0007**	<0.001
Used > 100%, <=120%									-0.002***	<0.001	-0.001***	<0.001
Used > 80%, <=100%									-0.003***	<0.001	-0.002***	<0.001
Used > 20%, <=80%									-0.005***	(.001)	-0.003***	<0.001
Used > 0%, <=20%									-0.003***	<0.001	-0.002***	<0.001
Positive account balance									-0.006***	<0.001	-0.004***	<0.001
Scope of Relationships: Assets held in the Bank (Amount)												
No Assets											0.005***	<0.001
< 50 EUR											0.004***	<0.001
>50 EUR , < 2000 EUR											0.002***	<0.001
Prior Consumer Loans												
Prior Loan (Yes/No)											-0.001***	<0.001
Diagnostics												
Prob > X^2	0.000		0.000		0.000		0.000		0.000		0.000	
Log Pseudolikelihood	-11957.164		-11221.88		-11239.53		-11292.89		-10968.86		-10,529	
Pseudo R^2	5.66%		7.37%		7.22%		6.78%		9.45%		13.08%	
Number of observations	400,773		393,601		393,601		393,601		393,601		393,601	
Number of bank clusters	168		168		168		168		168		168	

Table 14

Private Information and Borrower Default Incentives: Selection Models

This table presents the results from a Heckman selection model with probit models in both selection and outcome equation. 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 whether or not the loan applicant was approved. Behavioral Score measures the private information from transaction account behavior of the borrower prior to the loan origination date. The control variables used are the same as in the previous models, as well as an additional instrument Log(Branches/Population) is the natural logarithm of the number of branches of each bank as a percentage of inhabitants of the particular region the bank is headquartered. Both coefficients and marginal effects are reported. All variables are defined in Appendix I. The coefficients of borrower characteristics, borrower industries and time fixed effects are not shown. ***,**,* denote significance levels at the 1, 5 and 10 percent level, respectively.

	Selection Models				
	1st stage (Selection Equation)		2nd stage (Default Equation)		
	(1)		(1)		
	Access		Default		Marginal Effects
Behavioral Score	0.113***	(.012)	-0.0003***	(.046)	-0.002
Relationship (Yes)	0.303***	(.028)	-0.161***	(.046)	-0.004
Borrower Characteristics	Yes		Yes		
Borrower Industry	Yes		Yes		
Time Fixed Effects	Yes		Yes		
Instruments					
Log(Branches / Population)	0.113***	(.012)	N/A		N/A
Diagnostics					
Wald Test (H0: ρ=0)			-0.001	(.649)	
Number of uncensored observations			398,900		
Number of censored observations			10,007		
X^2(df)			843		
Prob > X^2			<0.001		

Table 15

Private Information and Borrower Default Incentives: Sample selection with alternative relationship proxies

This table presents the results from a Heckman selection model with probit models in both selection and outcome equation. 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 whether or not the loan applicant was approved. The control variables used are the same as in the previous models, as well as an additional instrument. Log(Branches/Population) is the natural logarithm of the number of branches of each bank as a percentage of inhabitants of the particular region the bank is headquartered. All variables are defined in Appendix I. In model (1), the omitted relationship variable is *Relationships > 15 years*; in (2) borrowers without a debit and credit card (*No Cards*) are omitted; in (4) customers with assets above 2,000 Euros are omitted (*>=2000EUR*). Coefficient and marginal effects are shown, except for borrower characteristics as well as intercept and time fixed effects. Panel A reports the outcome equation, Panel B the selection equation. Coefficients and marginal effects are shown in Panels A and B. ***,**,* denote significance levels at the 1, 5 and 10 percent level, respectively.

Panel A: Outcome Equation (Dependent Variable: Default)

	(1)			(2)			(3)			(4)			(5)			(6)		
	Default		**Marg. Eff.**	**Default**		**Marg. Eff.**	**Default**		**Marg. Eff.**	**Default**		**Marg. Eff.**	**Default**		**Marg. Eff.**	**Default**		**Marg. Eff.**
Tramsaction Account Behavior																		
Behavioral Score	-0.004***	<.001	0.005	-0.004***	<.001	0.004	-0.003***	<.001	-0.003	-0.005***	<.001	0.005	-0.004***	<.001	0.005	-0.0002***	<0.001	0.002
Relationship Characteristics																		
Relationship Length																		
Relationship <2years	0.536***	(.038)	0.012													0.428***	(.034)	0.007
Relationship >=2, <5years	0.387***	(.03)	0.007													0.302***	(.029)	0.004
Relationship >=5, <9years	0.232***	(.026)	0.003													0.185***	(.026)	0.002
Relationship >=9, <15years	0.110***	(.025)	0.001													0.089***	(.026)	0.001
Scope of Relationships: Cards & Checking Account Information																		
Debit and Credit Card				-0.476***	(.039)	-0.003										-0.367***	(.039)	-0.002
Debit Card				-0.323***	(.025)	-0.003										-0.259***	(.024)	-0.002
Credit Card				-0.090***	(.038)	-0.001										-0.052	(.038)	0.000
Credit Line (Yes)							-.455***	(.027)	-0.009							-0.492***	(.028)	-0.008
Scope of Relationships: Assets held in the Bank (Amount)																		
No Assets										0.481***	(.033)	0.007				0.452***	(.036)	0.006
< 50 EUR										0.347***	(.032)	0.005				0.379***	(.035)	0.005
>50 EUR , < 2000 EUR										-0.178***	(.032)	0.002				0.230***	(.035)	0.002
Prior Consumer Loans																		
Prior Loan (Yes/No)													-.0822***	(.022)	-0.001	-0.131***	(.023)	-0.001

Table 15 (cont'd)

Panel B: Selection Equation (Dependent Variable: Access)

	(1)			(2)			(3)			(4)			(5)			(6)		
	Default		**Marg. Eff.**	**Default**		**Marg. Eff.**	**Default**		**Marg. Eff.**	**Default**		**Marg. Eff.**	**Default**		**Marg. Eff.**	**Default**		**Marg. Eff.**
Log(Branches/Population)	0.093***	(.012)	0.003	0.108***	(.012)	0.003	0.112***	(.012)	0.004	0.0984***	(.012)	0.003	-0.113***	(.012)	0.004	0.085***	(.013)	0.002
Behavioral Score	0.0004***	<.001	0.010	0.0003***	<.001	0.006	0.0003***	<.001	0.007	0.0004***	<.001	0.010	0.0003***	<.001	0.010	0.0002***	<.001	0.004
Relationship Characteristics																		
Relationship Length																		
Relationship <2years	-0.681***	(.018)	-0.040													-0.486***	(.02)	-0.017
Relationship >=2, <5years	-0.512***	(.016)	-0.024													-0.352***	(.018)	-0.010
Relationship >=5, <9years	-0.283***	(.016)	-0.010													-0.187***	(.017)	-0.004
Relationship >=9, <15years	-0.134***	(.016)	-0.004													-0.104***	(.017)	-0.002
Scope of Relationships: Cards & Checking Account Information																		
Debit and Credit Card				0.753***	(.025)	0.011										0.648***	(.026)	0.008
Debit Card				0.568***	(.012)	0.013										0.523***	(.013)	0.010
Credit Card				0.551***	(.026)	0.008										0.522***	(.027)	0.006
Credit Line (Yes/No)							0.249***	(.019)	0.010							0.237***	(.02)	0.006
Scope of Relationships: Assets held in the Bank (Amount)																		
No Assets										-0.578***	(.018)	-0.023				-0.444***	(.019)	-0.012
< 50 EUR										-0.209***	(.019)	-0.006				-0.139***	(.02)	-0.003
>50 EUR , < 2000 EUR										0.038**	(.019)	0.001				0.080***	(.021)	0.002
Prior Consumer Loans																		
Prior Loan (Yes/No)													.0245***	(.012)	0.001	0.046***	(.013)	0.001
Diagnostics																		
Wald Test (H0: ρ=0)	-0.189		(.167)	-0.086		(.67)	-0.165		(.307)	0.001		(.994)	-0.027		(.82)	-0.110		0.441
Number of uncensored obs.	398,900			398,900			398,900			398,900			398,900			398,900		
Number of censored obs.	10,007			10,007			10,007			10,007			10,007			10,007		
X^2(df)	989.53			952.25			1094.12			964.86			840.77			1372.79		
Prob > X^2	<0.0001			<0.0001			<0.0001			<0.0001			<0.0001			<0.0001		