The Sources of Value of Relationships: Screening, Monitoring and the Likelihood of Consumer Default

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

March 16, 2010

Abstract

This paper analyzes the importance of retail consumers' banking relationships for their loan defaults. We find that loans of retail customers, who have a relationship with their savings bank when they apply 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).We find that the banks' screening policy is geared to minimize default risk. We also find that banks' screening process is not deterministic but includes elements of subjective assessment or discretion which is associated with lower frequencies of defaults. Even after taking screening into account, relationships still have a first order impact in reducing borrower defaults. In other words, we identify a distinct value of existing relationships in screening based on discretionary behavior of banks as well as in monitoring.

We thank seminar participants at Georgia Tech and Mannheim for comments and suggestions.

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

1. Introduction

Understanding how banks make loans is important and has been at the forefront of the current financial crisis. An important question is how should the process of loan making by banks be regulated to minimize risks? For example, should the loan making process be entirely codified so that the potential for discretion does not exist, and loans are made based on hard, verifiable information collected by the bank? Allowing discretion could allow for the information obtained from relationship specific assets to be incorporated to improve the quality of loans made. This is an open question which is of interest to academicians, 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]

Our paper adds to this literature studying bank-depositor relationships. 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. 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 key contribution of our paper is that we examine the impact of different kinds

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

of relationships that existed prior to granting the loan in reducing default rates. Specifically, we show that relationships matter in various forms, scope and depth, and even simple transaction or savings accounts make a difference. A further contribution of this paper is that we examine the sources of value of relationships. In particular, we find that the banks' screening policy is geared to minimize default risk. We also find that banks' screening process is not deterministic but includes elements of subjective assessment or discretion which is associated with *lower* frequencies of defaults. Even after taking screening into account, relationships still have a first order impact in reducing borrower defaults. In other words, we identify a distinct value of existing relationships in screening based on discretionary behavior of banks as well as in monitoring. To the best of our knowledge, this result is new to the literature.

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. The interesting feature for our analysis is that the scoring system does provide a credit assessment of the applicant but only as a guideline. The final loan granting decision is made by each individual bank also 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. These data thus provide an ideal opportunity to investigate the sources of value of relationships particularly at reducing the likelihood of consumer default.

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. This result is consistent with relationships providing banks with a unique advantage in monitoring their borrowers.[4] 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 that determine relationships. To facilitate identification, we include an instrument that proxies for the availability of financial intermediaries to customers in their

[4] Our results are consistent with the literature on bank specialness, among others, Fama (1985), James (1987), Lummer and McConnell (1989), Billett et al. (1995) and Dahiya et al. (2003).

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? Or can banks use the private information that they receive from their frequent interactions with their customers from current and savings accounts or from prior loans to improve their monitoring process? Our prior results are suggestive of monitoring as the predominant source of value. However, our sample is censored as we can only observe the performance of loans when applicants receive credit. If banks screen and reject applicants based on subjective assessment which is not a deterministic function of observable risk characteristics, the estimates of our default model might be biased.

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. Moreover, we find that this correlation is negative, i.e. discretion in screening of loan applicants is associated with lower default rates. 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 in monitoring borrowers.

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)). 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. Similarly, Norden and Weber (2009) show that checking account activity provides valuable information for banks in predicting the default probability of small firms and their subsequent loan contract terms. Related to these two papers, Agarwal, Chomsisengphet, Liu, and Souleles (2009) document for consumer credit markets 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. First, 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. Second, we show that relationships matter in various forms, scope and depth. Third, instead of analyzing the behavior of one bank we examine the loan making decision of 296 different banks. Fourth, we identify both screening and monitoring as distinct sources of value of relationships.

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, Section 4 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.[5] These loans are usually given on an unsecured basis, i.e. without collateral, and it is not possible to sell or securitize these loans unless they default.[6] 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

[5] The sample thus does not comprise mortgage loans, checking accounts with credit lines, 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 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.

[6] 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.

compliance with the Basel II requirements.[7] 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.[8] Our data includes flags for each of these default events and the associated date. 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.

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. The interesting feature for our analysis is that the scoring system does provide a credit assessment of the applicant but only as a guideline. The final loan granting decision is made by each individual bank also taking into account its respective ability and

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

[8] 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.

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.

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. 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. The loan repayment in percent of the borrower's income amounts to more than 20% only for 6.6% of the borrowers, for all the other borrowers it is less than 20%. Most borrowers work in the service industry and have been in their current job for more than two years.

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

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

A. Univariate Results

In the first step, the performance of consumer loans is analyzed in a univariate framework, and it is tested 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%, as shown in Table 1. This means that the

difference in default rates between new and existing customers is more than 1.5 times higher than the average default rate. Furthermore, 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 with no checking account 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 first 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. As this analysis does not control for other important factors and in particular for the loan applicants' credit risk, the link between relationships and loan defaults is tested in a multivariate framework in the next section.

B. Multivariate Results

We now turn to the multivariate analysis and take into account additional control factors. Our analysis proceeds in two steps. In the next section, we analyze whether existing relationships reduce the default probability of consumer loans even after controlling for several borrower characteristics. 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. Panel Regressions

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 line 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[9] 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.[10] 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 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 with a tenure of less than two years in their current job default 0.3% to 0.5% more often than customers with longer tenure in their current job. 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 pure 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. Model (3) takes into account the intensity of a relationship by analyzing the impact of different combinations of

[9] "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.

[10] All variables are defined in Appendix I.

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.

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 > 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. 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.

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.[11] *# 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.

[11] 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).

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, our results suggest that relationships that existed 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.

B.2. Bank-depositor relationships and default probabilities: Instrumental variables

The previous results establish a correlation between having prior relationships and default rates. However, one could argue that relationships may proxy for customer risk characteristics. For example, relationship customers might be less risky than new customers based on characteristics we are not able to control for in our model. If this was indeed the case, it would be less clear to what extent our results are driven by relationships rather than risk. We address this point in two ways. First, it is important to note that some of our relationships proxies are not likely to immediately proxy for risk, e.g. the existence of a savings account. Second, we address this point more formally using a simultaneous equation model in which we augment the main probit equation (default model) with an additional probit equation that explains what factors influence relationships (relationship model). We use a bivariate Probit model as both default and the existence of a relationship are binary variables. We instrument for relationships and in particular the existence of a checking account, which is usually the first relationship that a customer builds with a bank. For better identification, we require an instrument that predicts whether or not a borrower has a relationship with the bank but not default.

The instrument we propose measures the availability of financial intermediaries to customers in their region.[12] This variable is constructed using the number of branches of all banks that are located throughout Germany. The underlying intuition is that a customer is more likely to have a checking account with a savings bank if fewer other banks operate in the savings bank's region. More precisely, the instrument proxies for the concentration of banks in that region as measured by the Herfindahl-Hirschmann Index (HHI). We collect data for each 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 and use this information to calculate the HHI for each region. Appendix 3 provides more information about these banks, the total number of branches in Germany and the average number of branches in each district. The mean HHI is 0.22, the minimum HHI is 0.12 and the maximum HHI 0.45, respectively. Our key identifying assumption is that the availability of financial intermediaries influences the initiation and existence of a bank-depositor relationship but does not explain the default behavior of subsequently issued loans.

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.[13]

The results of the bivariate Probit model are presented in Table 8. The relationship model includes all control variables as shown in the previous models and further includes our instrument (*Log(HHI)*) to account for the fact that relationship customers might be less risky. The first column reports the results. The coefficient of the instrument confirms our expectation

[12] 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.

[13] 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.

that an increase in the HHI leads to a significantly higher likelihood of a checking account relationship, i.e. the fewer other banks operate in the same region as a savings bank, the more likely it is that a customer has a relationship with that savings bank. The second column reports the results from the default model (the third column the marginal effects). Most importantly, the result for the existence of a checking account in the second step does not differ from the results before: Customers with an existing checking account still have a significantly 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 term 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 customer risk or other characteristics.

B.3. 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 (Bayes et al., 1989). If the banks' screening process is not deterministically determined 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. There are at least two subjective elements important in our empirical setting: first, the credit scores and internal ratings only provide a guideline for the 296 banks in our sample. The final loan granting decision is based on each bank's willingness and ability to take risks. Second,

the loan officers in the banks have some discretion in approving or rejecting loan applications, i.e. "soft" information is important in screening applicants and is a deterministic factor of the sample selection problem. The banks and loan officers have their individual selection rules which can potentially bias our result.

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 which 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 Van and Van Pragg (1981)).

In the selected sample,

$$E\lfloor y_{ij} \mid z_{ij} = 1 \rfloor = \beta' x_{ij} + \rho\sigma_{\varepsilon}\lambda(\gamma' w_{ij})$$

The model is estimated using MLE. The explanatory variables in the loan granting and default equation are identical., We add (*Log(HHI))* as instrument to the selection equation for identification.[14] This variable 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), and Zarutskie (2006) 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 is 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.

In a first step, we use whether or not the borrower had a checking account at the time she applied for the loan as proxy for relationships. The results are reported in Table 9. The loan approval decision (column (1)) and default decision (column (2)) are reported together with the marginal effects of the default decision (column (3)). 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 subjective information or discretion is clearly important in the loan granting process.

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 would be consistent with a bank's screening policy being geared to minimize default rates. We find that all variables that are significant in both the

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

approval and default model carry the opposite signs, which is consistent with this interpretation and in line with what we would expect.

The estimate of the correlation that maximizes the likelihood is negative and significant suggesting that discretion at the loan application stage is associated with lower frequencies of default. In other words, screening matters! It matters based on objective criteria or credit scores as explained above, but, more importantly, it matters because subjective elements in the screening policy reduce borrower default rates. The coefficient of the relationship variable is still negative and significant. Moreover, correcting for sample selection increased the magnitude of the coefficient. Having a relationship with the bank reduces the probability of default by 3.2%, ceteris paribus. That is, monitoring is an important determinant of default rates.

As Table 9 shows that sample selection is important, we estimate our earlier regressions again accounting for selection using the same approach as outlined above. Table 10 reports the results. Panel A of Table 10 presents the results from the default model, Panel B the results from the loan granting model and Panel C the marginal effects from the default model, respectively. The diagnostic section in Panel B reports the correlation between the two models and its p-value.

Model (1) repeats the results from Table 9 for comparison. Model (2) analyzes the impact of relationship duration on default rates. Consistent with our earlier results, default rates decrease for longer relationships and are lowest for customers with relationships of more than 15 years (which is the omitted group). Default rates also decrease for customers with a more extensive usage of credit and debit cards, as can be seen from model (3). Customers who have both a credit and a debit card exhibit a particularly low default rate, ceteris paribus. Borrowers with neither debit nor credit card are omitted. The same holds for customers with an existing credit line. The results in model (4), panel C, show that these customers have a 3.1% lower likelihood to default. Model (5) analyzes the impact of the amount of assets that customers hold with the savings banks. As before, the more assets customers hold with their savings bank, the less likely are they to default on their loans. Finally, model (6) tests whether the existence of a prior loan affects the default rates of the current loan. The empirical evidence points to a 0.1% lower default rate for customers who have had a loan with the savings bank in the past.

Taken together, we find that relationship customers are less likely to default compared to non-relationship customers, even after controlling for the potential selection bias. These results suggest that monitoring is an important determinant for the reduction of defaults in consumer loan.

Panel B reports the results for the selection model. The dependent variable is a dummy variable that takes a value of 1 if the bank approves a customer's loan application. Similar to Table 9, we add (*Log(HHI))* as instrument to the selection equation for identification. We do not report the estimates of the control variables for brevity.

The results for the relationship variables show that existing customers have a significantly higher likelihood of receiving a loan than new customers. The coefficients for these relationship variables are highly significant. For example, the results in model (2) suggest that customers with a shorter relationship have a much smaller probability of receiving a loan than customers with a longer relationship. In particular, customers with a relationship of less than two years have a significantly lower probability of receiving a loan than customers with a relationship of more than 15 years, which serve as the benchmark group. Model (3) provides evidence that the usage of credit and debit cards significantly increases customers' chances of receiving a loan. Likewise, model (4) shows that an existing credit line significantly increases these chances. Finally, the findings in model (5) and model (6) document that customers' chances of receiving a loan significantly increase with the amount of assets that they keep in their savings bank. Taken together, the results suggest that bank-depositor relationships increase the likelihood that these customers receive a loan. Their acceptance rates in loan decisions significantly increase with the length and the intensity of their relationships with their savings banks.

4. 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 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). While the result is particularly pronounced for those types of relationships from which the bank can generate the most valuable information, in particular the existence and usage of credit lines, and the usage of debit and credit cards, our results suggest that relationships even in the form of simple transaction and savings account are economically important.

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 not only 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 in the monitoring process even after the improvement in the initial screening process. The results in this paper highlight the benefits of bank relationships and empirically identify two sources for these benefits. 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] 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.

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

[4] 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.

[5] 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.

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

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

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

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

[10] 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.

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

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

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

[14] 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.

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

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

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

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

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

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

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

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

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

[24] 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.

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

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

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

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

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

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

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

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

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

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

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

[36] 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 (Continuous)	Relationship length (continuous) is measured as the difference between the application date and the opening date of the first checking account with the bank in days.
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)	
Assets (continuous)	The amount of money held in accounts other than checking accounts (e.g. savings account, brokerage account) measured in Euro.
< 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.
Instrument	
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%.
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.

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 throughout 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
Job / Industry								
Unemployed	1,068,000	0.007	0.086	0	0	0	0	1
Service Sector	1,068,000	0.238	0.426	0	0	0	0	1
Public Sector	1,068,000	0.134	0.341	0	0	0	0	1
Retired	1,068,000	0.116	0.321	0	0	0	0	1
Construction Work	1,068,000	0.057	0.232	0	0	0	0	1
Other[2)]	1,068,000	0.239	0.426	0	0	0	0	1
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	87,251	4.256	2.359	1	2	4	6	8

1) Number of observations is number of loan applications.

2) "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. 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)

	(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 Pseudolikelihood	-37,032.687		-34,393.717		-34,727.289		-34,849.281		-33,664.989		-32,869.785	
Pseudo R^2	5.26%		6.83%		5.93%		5.60%		8.81%		10.96%	
# of observations	1,068,000		1,041,291		1,041,291		1,041,291		1,041,291		1,041,291	
# 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%		5.88%	
# of observations	1,068,000		1,068,000	
# of bank clusters	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		0.000	
Log Pseudolikelihood	-31,947.60		-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)

	(2)		(3)		(1)		(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 > X^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

The effect of relationships on default rates: Instrumental variables

This table presents the results of a bivariate Probit regression. The variables used are the same as in the previous models, as well as an additional instrument. Log (HHI) is the natural logarithm of the Hirshman-Herfindahl Index which measures the competition among banks. The number of branches of each particular bank is used to construct the HHI. Heteroscedasticity consistent standard errors are shown in parentheses. ***, **, * denote significance levels at the 1, 5 and 10 percent level, respectively.

	Coefficents				Marginal Effects
	Relationships		**Default**		**Default**
Relationship (Yes)	N/A		-0.706***	(.238)	-0.023
Borrower Characteristics					
Log (Income)	-0.07***	(.007)	-0.132***	(.009)	-0.002
Repayment (% of Income)					
< 5%	-0.064***	(.01)	-0.227***	(.022)	-0.002
>= 5%, < 11%	0.040***	(.007)	-0.213***	(.012)	-0.003
>=11%, < 13%	0.005	(.011)	-0.143***	(.02)	-0.002
>=13%, < 20%	-0.018**	(.008)	-0.082***	(.014)	-0.001
Age					
18 to 23 years	-0.256***	(.019)	0.499***	(.034)	0.012
23 to 25 years	-0.255***	(.019)	0.448***	(.034)	0.010
25 to 30 years	-0.300***	(.017)	0.392***	(.032)	0.008
30 to 45 years	-0.215***	(.015)	0.380***	(.03)	0.006
45 to 50 years	-0.140***	(.017)	0.282***	(.032)	0.005
50 to 60 years	-0.082***	(.016)	0.259***	(.03)	0.004
Job Stability					
<= 2 years	-0.163***	(.007)	0.270***	(.011)	0.004
Log(HHI)	0.118***	(.009)	N/A		N/A
Diagnostics					
	Wald Test: ρ=0.15573 (p=0.1172)				
# of observations	1,036,901				
	$X^2(76)$=6,921.75				
	Prob > X^2=0.000				

Table 9

Default probabilities and sample selection: Screening and Monitoring

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 (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. 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.

	Coefficents				**Marginal Effects**
	Accepted		**Default**		**Default**
Relationship (Yes)	0.547***	(.012)	-0.521***	(.013)	-0.036
Borrower Characteristics					
Log (Income)	0.151***	(.007)	-0.168***	(.006)	-0.007
Repayment (% of Income)					
< 5%	0.381***	(.016)	-0.335***	(.015)	-0.010
>= 5%, < 11%	0.252***	(.008)	-0.253***	(.008)	-0.009
>=11%, < 13%	0.035***	(.012)	-0.079***	(.012)	-0.003
>=13%, < 20%	-0.034***	(.008)	-0.004	(.009)	0.000
Age					
18 to 23 years	-0.894***	(.022)	0.853***	(.022)	0.078
23 to 25 years	-0.547***	(.023)	0.573***	(.022)	0.041
25 to 30 years	-0.484***	(.021)	0.502***	(.021)	0.031
30 to 45 years	-0.353***	(.02)	0.408***	(.02)	0.019
45 to 50 years	-0.241***	(.022)	0.292***	(.021)	0.015
50 to 60 years	-0.128***	(.021)	0.212***	(.02)	0.010
Job Stability					
<= 2 years	-0.394***	(.007)	0.394***	(.007)	0.022
Log(HHI)	-0.177***	(.009)	N/A		N/A
Diagnostics					
	Wald Test: ρ= -0.97571 (p<0.001)				
# of censored observations	23,060				
# of uncensored observations	1,036,901				
	$X^2(38)=17740.59$				
	Prob > X^2=0.000				

Table 10

The effect of relationships on default rates: 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. The control variables used are the same as in the previous models, as well as an additional instrument. Log (HHI) is the natural logarithm of the Hirshman-Herfindahl Index which measures the competition among banks. The number of branches of each particular bank is used to construct the HHI. All variables are defined in Appendix I. In model (2), the omitted relationship variable is *Relationships > 15 years*; in (3) borrowers without a debit and credit card (*No Cards*) are omitted; in (5) customers with assets above 2,000 Euros are omitted (*>=2000EUR*). The coefficients for borrower characteristics as well as intercept and time fixed effects are not shown. Panel A reports the outcome equation, Panel B the selection equation, and Panel C the marginal effects of the variables in the outcome equation, respectively. 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)	(6)
Relationship Characteristics						
Relationship Yes/No						
Relationship (Yes)	-0.521** (.013)					
Relationship Length						
Relationship <2years		0.518*** (.011)				
Relationship >=2, <5years		0.326*** (.01)				
Relationship >=5, <9years		0.137*** (.01)				
Relationship >=9, <15years		0.010*** (.009)				
Scope of Relationships: Cards & Checking Account Information						
Debit and Credit Card			-0.554*** (.01)			
Debit Card			-0.490*** (.007)			
Credit Card			-0.402*** (.016)			
Credit Line (Yes)				-0.474*** (.009)		
Scope of Relationships: Assets held in the Bank (Amount)						
No Assets					0.346*** (.01)	
< 50 EUR					0.128*** (.012)	
>50 EUR , < 2000 EUR					-0.072*** (.012)	
Prior Consumer Loans						
Prior Loan (Yes/No)						-0.030*** (.008)
Borrower Characteristics	Yes	Yes	Yes	Yes	Yes	Yes
Borrower Industry	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Table 10 (cont'd)

Panel B: Selection Equation (Dependent Variable: Access)

	(1)		(2)		(3)		(4)		(5)		(6)	
Instrument												
Log(HHI)	-0.177***	(.009)	-0.191***	(.009)	-0.119***	(.007)	-0.168***	(.009)	-0.194***	(.01)	-0.164***	(.009)
Relationship Characteristics												
Relationship Yes/No												
Relationship (Yes)	0.547***	(.012)										
Relationship Length												
Relationship <2years			-0.441***	(.011)								
Relationship >=2, <5years			-0.262***	(.01)								
Relationship >=5, <9years			-0.081***	(.009)								
Relationship >=9, <15years			0.024***	(.009)								
Scope of Relationships: Cards & Checking Account Information												
Debit and Credit Card					0.572***	(.011)						
Debit Card					0.523***	(.007)						
Credit Card					0.475***	(.018)						
Credit Line (Yes/No)							0.446***	(.009)				
Scope of Relationships: Assets held in the Bank (Amount)												
Assets (Yes)									-0.290***	(.009)		
No Assets									-0.003	(.01)		
< 50 EUR									0.196***	(.01)		
>50 EUR , < 2000 EUR												
Prior Consumer Loans												
Prior Loan (Yes/No)											0.071***	(.008)
Borrower Characteristics	Yes		Yes		Yes		Yes		Yes		Yes	
Borrower Industry	Yes		Yes		Yes		Yes		Yes		Yes	
Time Fixed Effects	Yes		Yes		Yes		Yes		Yes		Yes	
Diagnostics												
	Wald Test: ρ= -0.97571 (p<0.001)		Wald Test: ρ= -0.96539 (p<0.001)		Wald Test: ρ= -0.99451 (p<0.001)		Wald Test: ρ= -0.98028 (p<0.001)		Wald Test: ρ= -0.96387 (p<0.001)		Wald Test: ρ= -0.97884 (p<0.001)	
# of censored observations	23,060		23,060		23,060		23,060		23,060		23,060	
# of uncensored observations	1,036,901		1,036,901		1,036,901		1,036,901		1,036,901		1,036,901	
	$X^2(38)$=17740.59		$X^2(41)$=18697.80		$X^2(40)$=24463.91		$X^2(38)$=19908.09		$X^2(40)$=16886.07		$X^2(38)$=17110.71	
	Prob > X^2=0.000		Prob > X^2=0.000		Prob > X^2=0.000		Prob > X^2=0.000		Prob > X^2=0.000		Prob > X^2=0.000	

Table 10 (cont'd)

Panel C: Outcome Equation - Marginal Effects						
	(1)	(2)	(3)	(4)	(5)	(6)
Relationship Characteristics						
Relationship Yes/No						
Relationship (Yes)	-0.036***					
Relationship Length						
Relationship <2years		0.033***				
Relationship >=2, <5years		0.017***				
Relationship >=5, <9years		0.006***				
Relationship >=9, <15years		0.0003				
Scope of Relationships: Cards & Checking Account Information						
Debit and Credit Card			-0.016***			
Debit Card			-0.020***			
Credit Card			-0.011***			
Credit Line (Yes)				-0.031***		
Scope of Relationships: Assets held in the Bank (Amount)						
No Assets					0.0160***	
< 50 EUR					0.005***	
>50 EUR , < 2000 EUR					-0.003***	
Prior Consumer Loans						
Prior Loan (Yes/No)						-0.001***