DISCUSSION PAPER SERIES

No. 6450

RULES VERSUS DISCRETION IN LOAN RATE SETTING

Geraldo Cerqueiro, Hans Degryse
and Steven Ongena

FINANCIAL ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP6450.asp**

ISSN 0265-8003

RULES VERSUS DISCRETION IN LOAN RATE SETTING

Geraldo Cerqueiro, CentER, Tilburg University
Hans Degryse, CentER, Tilburg University, TILEC and K.U. Leuven
Steven Ongena, CentER, Tilburg University and CEPR

Discussion Paper No. 6450
September 2007

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **FINANCIAL ECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as a private educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions. Institutional (core) finance for the Centre has been provided through major grants from the Economic and Social Research Council, under which an ESRC Resource Centre operates within CEPR; the Esmée Fairbairn Charitable Trust; and the Bank of England. These organizations do not give prior review to the Centre's publications, nor do they necessarily endorse the views expressed therein.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Geraldo Cerqueiro, Hans Degryse and Steven Ongena

CEPR Discussion Paper No. 6450

September 2007

ABSTRACT

Rules versus Discretion in Loan Rate Setting*

We propose a heteroscedastic regression model to identify the determinants of the dispersion in interest rates on loans granted to small and medium sized enterprises. We interpret unexplained deviations as evidence of the banks' discretionary use of market power in the loan rate setting process. "Discretion" in the loan-pricing process is most important, we find, if: (i) loans are small and uncollateralized; (ii) firms are small, risky and difficult to monitor; (iii) firms' owners are older, and, (iv) the banking market where the firm operates is large and highly concentrated. We also find that the weight of "discretion" in loan rates of small credits to opaque firms has decreased somewhat over the last fifteen years, consistent with the proliferation of information-technologies in the banking industry. Overall, our results reflect the relevance in the credit market of the costs firms face in searching information and switching lenders.

JEL Classification: G21 and L11
Keywords: financial intermediation, loan rates, price discrimination and variance analysis

Geraldo Cerqueiro
Tilburg University
Finance Department
P.O. Box 90153
5000 LE, Tilburg
THE NETHERLANDS
Email: g.m.cerqueiro@uvt.nl

Hans Degryse
Tilburg University
Finance Department
P.O. Box 90153
5000 LE, Tilburg
THE NETHERLANDS
Email: h.degryse@uvt.nl

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=166058

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=121947

Steven Ongena
Tilburg University
Finance Department
P.O. Box 90153
5000 LE, Tilburg
THE NETHERLANDS
Email: steven.ongena@uvt.nl

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=144729

* We are grateful to Sigbjorn Atle Berg, Allen Berger, Pedro Bom, Vicente Cunãt, Thomas Gehrig, Florian Heider, José Liberti, Carlos Lourenço, Bertrand Melenberg, Lars Norden, Viorel Roscovan, Greg Udell and Bas Werker, as well as participants at the 2007 Federal Reserve Bank of Chicago's Conference on Bank Structure and Competition, the 2007 CEPR-ESSFM Meeting (Gerzensee), the 2007 Conference on Small Business Banking and Financing: A Global Perspective (Cagliari), the 2006 NAKE Research Day (Amsterdam) and seminars at CentER – Tilburg University and ISEG (Lisbon) for many valuable comments and suggestions. The authors gratefully acknowledge financial support from NWO-The Netherlands and FWO-Flanders. The paper was completed while the first author was visiting the Sveriges Riksbank (Stockholm), whose hospitality is gratefully acknowledged. Hans Degryse holds the TILEC – AFM Chair on Financial Regulation.

Submitted 21 August 2007

1. Introduction

Technological progress has shaped the evolution of the banking industry over the last few decades. In particular, new information and communication technologies assisted banks in processing and summarizing information about applicants in credit scores that are used in pricing credit. Regardless of the proliferation of credit scoring technologies based solely on quantitative information, bankers often rely on their experience and distrust the blind use of quantitative information only. Consequently, the final decisions concerning credit approvals and loan terms are then based on many different attributes, from which the experience and the judgment of the credit staff continue to play a significant role. But how banks then actually end up pricing loans to small and medium sized firms remains a largely unanswered question.

In this paper, we investigate how particular loan, firm, relationship and market characteristics affect a bank's propensity to rely on statistical methods ("rules") or subjective judgments ("discretion") in the loan rate setting process. We employ a heteroscedastic regression model to empirically examine how these relevant characteristics relate to the unexplained variance of a linear loan-pricing model. We interpret the higher predictive power of the loan-pricing model as evidence of the greater importance of "rules" in the loan rate setting process. Larger unexplained variance, on the other hand, is then associated with the prevalence of "discretion".

The unexplained variance of contracted loan rates essentially reflects heterogeneity in lending technologies. Under this view, "rules" and "discretion" represent the extremes of a continuum along which lending technologies can be classified according to the relative weight of objective and subjective elements in the loan prices. "Rules" is best figuratively described by a computer model that receives objective information about applicants as an input; the differences in output (i.e., loan rates) then stem exclusively from the quantifiable differences among borrowers. In contrast, "discretion" is a pure judgmental technology whereby loan rates are entirely set on subjective grounds. These subjective assessments reflect eventually market imperfections and the loan officers' information about the firms' operating environment. Granted, we cannot for certain determine whether the deviations from our loan-pricing model reflect subjective elements or pricing errors made by loan officers. Although some theoretical models support the existence in equilibrium of randomized pricing strategies (see e.g. Varian (1980)), subjective pricing is fully consistent with the *raison d'être* of banks. Moreover and more importantly, the deviations are not random, we find, but well explained by observable firm, bank, loan contract and banking market characteristics.

The above distinction between "rules" and "discretion" also clarifies the cogency of our methodology. While a computer model treats all applicants equally, conditional on their hard information, we can think of the loan officer's judgment as corresponding to a different

pricing model for each lender. In other words, while "rules" corresponds to a single pricing model, "discretion" translates into a multiplicity of different models. As a result, the explanatory power of our empirical pricing model decreases with the degree of heterogeneity in the original pricing models, i.e. with the relative weight of "discretion".

In a frictionless world there should be no room for "discretion". Loan rates then should only vary with verifiable information. However, more reasonable descriptions of the real world indicate that the dispersion of loan prices, and hence "discretion", characterizes the equilibrium of credit markets under asymmetric and imperfect information. The extent to which loan rates reflect the prevalence of "rules" or "discretion" depends primarily on the magnitude of switching costs or information search costs firms face in credit markets. Other potential explanations for "discretion" involve imperfectly competitive credit market structures and the existence of regulatory constraints such as fixed or capped loan rates.[1] These market imperfections determine the bargaining power banks have vis-à-vis firms and set the boundaries within which banks engage in discretionary loan-pricing practices. Recent organization theory has further emphasized the importance of bank structure for the nature and success of the lending technology being employed (Berger and Udell (2002) and Stein (2002)).[2]

The estimates from our heteroscedastic regression models generate four new robust findings.[3] First, banks price larger loans according to more objective criteria or "rules". This finding illustrates that a firm's incentives to increase its search intensity constrains a bank's ability to price discriminate. Second, several alternative proxies for the borrower risk and opaqueness are all unilaterally associated with a larger unexplained dispersion of loan rates. This result suggests that the weight of "discretion" is positively related to the switching costs firms face. Third, banks seemingly exploit the available public information about their borrowers, which we measure by the age of the firm's owner, for discretionary purposes. In contrast, banks seem to avoid price discrimination based on private information. We interpret this result as the outcome of a defensive strategy undertaken by banks to shield their informational advantage against potential competitors (as in Gan and Riddiough (2006)).

[1] See Klemperer (1995) for a review of the literature dealing with switching costs and von Thadden (2004) for example for a direct application to bank loan pricing. Stigler (1961) introduced information search costs. Degryse and Ongena (2007a) for example review the various sources of bank rents.

[2] This literature suggests that decentralized, small banks have a comparative advantage in small business lending, an activity that is often viewed as idiosyncratic and relationship-based. In practice, of course, loan officers may act independently of the formal hierarchical structure and have the latitude and incentives towards discretionary loan pricing. Bargaining skills of the borrower may then become a key determinant of the contracted loan price. Guttentag (2003) for example emphasizes the relevance of these skills for the charged rates in the U.S. mortgage market: "If the loan officer tabs you as unknowledgeable and timid, you will probably pay an "overage" -- a price above the price listed on the loan officer's price sheet. The lender and the loan officer usually share overages. If you are smart and forceful, on the other hand, you might get an underage - a price below the listed price."

[3] This robustness is largely explained by an important econometric result that states some form of independence between the first and second conditional moments in a regression model.

Finally, we find that "discretion" is most important in large and highly concentrated banking markets, and in periods of high liquidity.

We actually employ four different datasets in our analysis: the 1993, 1998 and 2003 Surveys of Small Business Finances (SSBF), as well as a dataset provided by an important Belgian bank. We find no significant disparities in results between the samples, suggesting that our results are not sample-specific, time-specific, nor driven by bank heterogeneity. However, we do find evidence of a decrease over the last fifteen years of the importance of "discretion" in small credits to opaque businesses. This result is consistent with the proliferation in the banking industry of credit scoring technologies for small business lending.

Empirical research on price dispersion is still very limited (see Dahlby and West (1986), Hortaçsu and Syverson (2004), and Sorensen (2000)) and this literature typically aims to test theories of information differentiation.[4] Our study is the first to empirically analyze the cross-sectional determinants of discretionary pricing in the market for business loans. While loan rate dispersion itself has been widely documented before,[5] no study so far (to the best of our knowledge) has identified the actual sources of this dispersion. Our motivation for this analysis goes beyond the empirical regularity that loan-pricing models often tend to fit the data rather poorly (see e.g. Petersen and Rajan (1994), and Berger and Udell (1995)). What is even more striking is the heterogeneity in model fits, a heterogeneity that seems to depend on the type of loans and borrowers in the sample that is being investigated.[6]

The credit market is a unique laboratory to study price dispersion, since frictions in the credit market are present on both the demand (firm uncertainty about how competitive is a loan offer) and the supply side (bank uncertainty about firm quality). Moreover, our variance analysis allows us to map the magnitude of the price dispersion on a continuum of lending technologies, which differ in the relative weight of subjective inputs or "discretion". As result, our study generalizes the conceptual framework proposed recently by Berger and Udell (2006). On the other hand, our analysis is ultimately mute on the hard versus soft information dichotomy introduced by Stein (2002). Indeed, we disregard the nature of the information and

[4] See for example Baye, Morgan and Scholten (2006) for a comprehensive review of this literature.

[5] Heffernan (2002) finds that the margin between the highest and lowest lending rates for U.K. mortgages is relatively small (0.45 percentage points), compared with the market for personal loans, where there is a range of 8.17 percentage points. Hassink and Van Leuvensteijn (2007) find that lending rates in the Dutch mortgage market are highly dispersed both across lenders (1 percentage point) and within lenders (0.4 percentage points), even after controlling for borrowers' characteristics and regions. Martín, Saurina and Salas (2005) detect substantial and persistent unexplained dispersion of retail loan rates in Spain, across banks and products. Degryse and Ongena (2005) analyze data from a large Belgian bank and report substantial variation in loan rates at the branch level.

[6] For example, Degryse and Ongena (2005) estimate the same loan-pricing model for two independent sub-samples: one with small loans (below $5,000) and another with large loans (above $50,000). The fit of their regressions is very different; the R^2's are 1% for small loans and 67% for large loans, respectively.

simply concentrate on how well each singular piece of information helps to explain variation in the observed loan rates.[7]

The rest of the paper proceeds as follows. Section 2 provides the theoretical background, while Section 3 presents and motivates the methodology. Section 4 discusses the primary dataset and Section 5 the results. Section 6 provides the robustness tests. Section 7 tests for temporal changes in our results and Section 8 concludes.

2. Theoretical Background

Loan rate dispersion is justified by the existence of market frictions of the sort that the economics and financial intermediation literature has extensively analyzed. We choose to discuss the theoretical predictions concerning the determinants of loan rate dispersion in three separate subsections: information asymmetry, search costs and bank organization. This categorization structures the theoretical exposition as well as the subsequent interpretation of our empirical results.

2.1. Asymmetry of Information

The severity of information asymmetries in the credit market may determine the nature of the lending technology that is being employed. The last decade witnessed an exponential increase in the adoption by banks of credit scoring models to evaluate loan applications, especially for commercial credits (Akhavein, Frame and White (2005)). However, a score may not be totally informative for the bank, especially if it approaches some "gray area". In this case, banks typically complement the score with other types of information (Hand and Healey (1997)), increasing their ability *ex post* to price discriminate with respect to the marginal borrowers (Berger, Frame and Miller (2005)).

The analysis of the Sharpe (1990) model by von Thadden (2004) embeds an explicit correspondence between the severity of the information asymmetry in the credit market and the dispersion of loan rates. Our simulations of von Thadden's mixed-strategy equilibrium indicate that the von Thadden (2004) model predicts that the variance of loan rates:[8] (i) increases non-monotonically in the proportion of "good" firms in the market (the maximum variance corresponds to a proportion of about one-half in which case the bank has minimum information about the quality of the pool); (ii) increases in the difference in quality between

[7] In fact, it is problematic to argue that the residuals of a empirical pricing model provide a good measure of the amount of soft information that is being used only, since in general: (i) the true model is unknown; (ii) the model omits relevant hard information; and (iii) several variables have a dual role (e.g., is the age of the firm capturing hard or soft information?).

[8] We simulate 300 cross-sectional samples of loan contracts from the equilibrium derived in von Thadden (2004), by randomizing the following structural parameters of the model: θ (proportion of "good" firms), *pH-pL* (difference in repayment probabilities between the "good" and the "bad" firm), and p (expected repayment probability in the credit market when no information about the firms is available). Then, we regress the log of the variance in loan rates obtained for each of the samples on the value of these parameters. All results reported are statistical significant at the 1% level and the R^2 of the regression is 78.3%.

the "good" and the "bad" firms; and, (iii) decreases in the average probability of loan repayment in the credit market.[9]

More general economic analyses can be found in Bester (1988, 1993). The price emerging from the bargaining process between buyer and seller in these models is determined by the search and switching costs (the buyer's "outside option"). Bester (1993) shows that *ceteris paribus* buyers with low switching costs face fixed prices, while buyers with high switching costs fall into a "haggling equilibrium" with negotiated prices. For a homogeneous product bargaining may actually dominate posting prices (Wang (1995)).

Studies that focus on the possibility of intertemporal risk sharing between firms and banks also generate broadly similar predictions. Petersen and Rajan (1995) for example show that in concentrated markets banks may optimally subsidize young (and hence opaque) firms in their early stages. Later on, banks recoup their losses by charging higher loan rates to the older, now locked-in firms. Kim, Kristiansen and Vale (2006) similarly show that interest rates, especially for riskier firms, follow a lifecycle pattern. Their prediction implies that loan rates for riskier firms exhibit larger unexplained deviations.[10]

Recent work by Gan and Riddiough (2006) demonstrates that banks may for strategic reasons avoid engaging in risk-based pricing in order to deter entry from competitors. Because the loan price conveys private information about the quality of the borrower, banks have an incentive to charge a uniform rate to higher credit-quality firms, and a risk-based rate to lower credit-quality firms. Consequently, loan rate dispersion will be an increasing function of the firms' riskiness in their model too.

To conclude all referred studies predict a positive correlation between firm opaqueness and the importance of "discretion" in loan pricing. This correlation stems from the high switching costs faced by an opaque firm, which result from the informational advantage a lending bank has over its competitors. A bank typically obtains this informational advantage over the course of its relationship with the firm, during which the bank learns more about the firm's creditworthiness and prospects (Sharpe (1990), Rajan (1992) and Boot and Thakor (1994)). Consequently, stronger ties with a firm imply less uncertainty for a lender, and hence less "discretion". On the other hand, this informational advantage also translates into higher switching costs, yielding the opposite prediction.

Naturally, the magnitude of switching costs also depends on the type of firm. Large firms for example are generally regarded as more transparent to outsiders. First, typically more

[9] Recent theoretical Industrial Organization papers also study the impact of customer recognition in an environment of dynamic competition on heterogeneity in product prices. These models show that firms may set different prices to qualitatively identical new and old customers. The price difference depends on the market structure, the degree of consumer and firm patience, and the firm's abilities to recognize new from existing consumers (see e.g. Villas-Boas (1999) or Fudenberg and Tirole (2000)).

[10] Petersen and Rajan (1995) and Kim, Kristiansen and Vale (2006) provide empirical evidence that supports these predictions. Machauer and Weber (1998) analyze how interest rate premia relate to the firms' credit ratings and find a positive, though non-monotone, relation between the level of risk and the dispersion of the charged premia.

information is publicly available on large firms. Second, most large firms have their financial records audited, so that this hard financial information provides a more precise and reliable indicator of its creditworthiness.[11] Third, large firms are more likely to engage multiple creditors, *ex ante* assuaging potential liquidity and holdup problems (Detragiache, Garella and Guiso (2000), von Thadden (1992), Ongena and Smith (2000)). Consequently, banks dealing with large clients are constrained to price loans uniformly or based on "rules". Discretionary deviations from such a rule-based equilibrium result in either unprofitable deals (if loan rates are set too low) or a higher probability of losing the (potential) client to a competitor (if loan rates are set too high).

2.2. Search Costs

It is well established in the economics literature that the extent of price dispersion may be due to the costly consumer search of information (Stigler (1961), Salop and Stiglitz (1982)). The search intensity, in turn, depends on the consumer's incentives such as the magnitude and frequency of the purchase. This intuitive mechanism has straightforward application to the credit market. In particular, it implies that both the level and variance of loan rates should decrease in the size of the loan. On the one hand, a bank can only expect to grant a loan if the loan rate offered is sufficiently competitive vis-à-vis rival banks. On the other hand, banks cannot persistently undercut the competitive spread as this may decrease the profitability of the loan to an undesirable level.

The search mechanism is similar to the presence of switching costs in generating a positive correspondence between loan rate dispersion and the magnitude of the underlying costs faced by the borrower. The magnitude of either the switching or search costs determines the degree of market power the bank has vis-à-vis the borrowers.[12] In turn, market power translates into higher dispersion of loan rates, which reflects a larger weight of "discretion" in the loan rate setting process.

2.3. Bank Organization

Recent organizational theories emphasize the importance of bank organization for the loan officers' incentives to produce soft information, a key component of relationship and judgmental lending. In particular, Berger and Udell (2002), and Stein (2002) suggest that decentralized, small banks have a comparative advantage in the production of soft

[11] Previous studies demonstrate that financial information is only a good predictor of payment performance for large firms (see e.g. Eisenbeis (1996)).

[12] The main difference is that search costs may be driven mostly by incentives such as the size of the loan, whereas switching costs are mainly a product of firm specific characteristics such as firm size and the stock of private information about the firm the bank possesses.

information.[13] This comparative advantage stems from the smaller vertical distance between allocation and control in a decentralized bank. Because soft information is costly to acquire and difficult to transmit to others, decentralization ameliorates both the incentive problem (Stein (2002)) and the agency problem (Berger and Udell (2002); see also Godbillon-Camus and Godlewski (2005)).[14] In sum, this line of research suggests that the loan underwriting practices of smaller banks employ more discretionary criteria.

For a given bank, Liberti and Mian (2006) show that the sensitivity of credit approval to hard information is larger for credit applications handled at higher hierarchical levels, i.e. levels that are more distant from the source of information. In addition, they show that the terms of large loans as well as applications from large firms are typically handled at higher hierarchical level within the lending bank. This confirms our prior that loan rates of both large loans and loans to large firms should exhibit substantially less unexplained dispersion, consistent with a lesser incidence of "discretion".

3. Econometric Methodology

In order to identify the determinants of the dispersion of loan rates, we employ the regression model with multiplicative heteroscedasticity proposed by Harvey (1976). The heteroscedastic version extends the linear regression model by also parametrizing the unexplained variance as function of exogenous covariates. We may think of the heteroscedastic regression model as comprising two equations – one to explain the mean of loan rates, and the other to explain the residual variance of loan rates. We will refer throughout the paper to the mean equation as the loan-pricing model, and concentrate our analysis on the variables in the variance equation, which determine the precision of the loan-pricing model.

A conventional formulation of the heteroscedastic model is given by:

$$y_i = X_i'\beta + u_i, \tag{1}$$

$$Log\sigma_i^2 = Z_i'\gamma, \tag{2}$$

with the identifying assumptions:

[13] Consistent with this view, Cole, Goldberg and White (2004) find that the lending decisions of large banks are more likely based on financial variables and that the decisions of small banks are more likely a function of the strength of the firm-bank relationship. Berger et al. (2005) also test for potential differences in the lending technology of small and large banks and find that large banks often lend to large, physically distant firms with good accounting records. Degryse, Laeven and Ongena (2007) link the number of visits (by borrowers or loan officers) to observed patterns of geographical reach of bank branches and spatial loan pricing.

[14] This implies that large and small banks should differ in their compensation schemes, as well as in their internal rules (that define, among other things, the pricing latitude to give loan officers).

$$E\left[u_i \mid X_i\right] = 0, \tag{3}$$

$$Var\left[u_i \mid Z_i\right] = \sigma_i^2 = \exp\{Z_i'\gamma\}, \tag{4}$$

where y_i is the dependent variable, X_i a vector of explanatory variables in the mean equation, u_i a disturbance term, σ_i^2 the residual variance, and Z_i a vector of explanatory variables in the variance equation.

Under the normality assumption, the conditional distribution of y_i is given by:

$$y_i \mid X_i, Z_i \xrightarrow{d} N\left(X_i'\beta, \exp\{Z_i'\gamma\}\right) \tag{5}$$

As a result, we obtain Maximum-Likelihood estimates in the heteroscedastic regression model by maximizing the following log-likelihood function with respect to β and γ:

$$LogL = \frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n} Z_i'\gamma - -\frac{1}{2}\sum_{i=1}^{n}\exp(-Z_i'\gamma)\left(y_i - X_i'\beta\right)^2 \tag{6}$$

An important aspect of this methodology is that the parameters in the mean (β) and variance (γ) equations are uncorrelated.[15] While counterintuitive, this corollary will prove fundamental in our analysis since it allows us to treat the two equations separately in terms of variable selection and interpretation of the results.

The interpretation of the parameters of interest (γ) is crucial in our analysis. Pick one variable in Z, say Z_k, and its respective parameter, γ_k. A positive γ_k indicates that the precision of the loan-pricing model decreases in Z_k. We interpret this result as evidence of a positive correlation between the variable Z_k and the weight of "discretion" in the loan rate setting process.

Next, we provide an example that should intuitively substantiate the adequacy of the methodology used in this study. Suppose we have detailed information on a sample of loan contracts, including observed characteristics of the loan and borrower. We line up (from smallest to largest) the existing observations according to one of these characteristics, say the loan amount; then, we perform rolling regressions of the loan rate on all available loan and

[15] Harvey (1976) provides a technical explanation. The underlying principle – what one could call orthogonality between the first and second moments in a conditional distribution – is at the center of several well-known results in regression analysis; for instance, this principle explains why the consistency of the coefficients (i.e. the slopes) in a linear regression model is not affected by heteroscedasticity or autocorrelation in the error term.

borrower characteristics using a fixed-width window. This procedure generates a sequence of parameter estimates and measures of fit relating to the rolling regressions. Our main interest lies in the relation between the explanatory power (or its reciprocal, the residual variance) of the loan-pricing model and the sorting variable (the loan amount). Figure 1 illustrates the outcome of the described method, when applied to our dataset.[16] The graph plots the sequence of the residual standard errors obtained in the rolling regressions against the loan amount (the sorting variable). Simple visual inspection discloses a negative relation between the unexplained variation of loan rates and the size of the loan. This result indicates that the loan-pricing model (i.e. "rules") gains predictive power as we move towards windows containing larger loans.[17] According to this evidence, we should expect in the variance equation a negative sign for the parameter associated with the loan amount variable.

4. Sample Description

The primary dataset used in this study is the 1993 National Survey of Small Business Finances (NSSBF), a survey conducted by the Federal Reserve Board and the Small Business Administration. The 1993 NSSBF collected data for the fiscal year 1993 for a nationally representative sample of 4,637 for profit, non-governmental, non-agricultural businesses with fewer than 500 employees. The dataset provides a detailed look at these firms – their characteristics and their use of credit and other financial services. We focus, in particular, in the firms' most recent borrowing experiences during the sample period (1990-1994), which include data on the characteristics of the borrower and of the lender, as well as the contracted loan terms. We restrict our analysis to the sample of 1,695 firms that provided information on their most recent loans (about 36% of the total sample). We dropped 31 observations pertaining to firms that did not report all the required information and 39 observations related to loans granted by non-financial firms. We end up with a final sample of 1,625 observations.[18]

The NSSBF is ideally suited for our purposes, for several reasons. First, the sample is quite heterogeneous with respect to the types of borrowers and loan contracts. Despite its focus on small businesses, the sample also contains sizable loans and loans granted to reasonably large firms, both of which we presume to exhibit a more transactional nature. The coexistence of different types of borrowers and loans ensures a diversity of loan-pricing technologies in our sample. While this form of heterogeneity is typically problematic in

[16] We delay until the next section the description of our data.
[17] Remember that the results in Degryse and Ongena (2005) (their table VI, p. 256) we reported in footnote 5 were similar to this example.
[18] Cole, Goldberg and White (2004) argue that the observations corresponding to loans granted before 1993 – the year in which firm specific data is collected – may be potentially endogenous. Unlike them, we do not drop these observations (200 in total) since our model estimates a large number of parameters. Our results, however, do not change when we drop these observations.

empirical work, the essence of our methodology is precisely to identify the determinants of this diversity. In particular, we associate unexplained loan rate dispersion to a discretionary or judgmental pricing model. For that reason, we choose not to restrict the sample analyzed to particular categories of loans (e.g. lines of credit) or borrowers.

Second, because it has comprehensive information on loan contracts and firm characteristics, the NSSBF permits us to exhaustively control for a significant share of the variation in loan rates due to "rules". This point is particularly relevant since we interpret the magnitude of unexplained deviations from our loan-pricing model as a measure of the banks' use of "discretion".

Finally, data is also available for the subsequently conducted surveys (SSBF 1998 and SSBF 2003). Since a consistent definition and a majority of identical questions are used across the three surveys, we are able to carry out a temporal analysis of our results over a relatively large time span (15 years).

Table 1 presents the variables used in our study, along with definitions and descriptive statistics. We now turn to a detailed description and motivation for each of these variables.

4.1. Interest Rate Variables

The dependent variable is the interest rate on the firm's most recent loan we label as the *Loan Rate*. On the right hand side, we include the variable *Prime Rate* to control for changes in the underlying cost of capital in the economy.[19] Most firms in our sample pay a premium on top of the prime rate. However, 71 firms in our sample face loan rates equal to the current prime rate, and 91 firms enjoy loan rates below the prime rate.

4.2. Loan Characteristics

We employ the dummy variable *Floating* to control for differences in level between fixed and variable interest rates. *Ln(Loan Amount)* is the log of the amount of the loan in dollars.[20] The size of the loan provides a tacit measure of the bargaining power the firm has in setting loan terms (the loan rate, in particular), since the loan amount is strongly correlated with the firm's incentives to search for potentially better terms (Stigler (1961)). *Ln(Loan Maturity)* is the log of the loan repayment duration and proxies for the term risk. *Collateral* is a dummy variable indicating whether the loan is secured by collateral. *Collateral* may represent the potential risk faced by the lender in the loan contract, as well as signal the quality of the firm (Bester (1985), Besanko and Thakor (1987)).

[19] We obtain the prime rate from the Federal Reserve Economic Database (FRB of St. Louis).

[20] We used the logarithmic transformation in all continuous variables, except the financial ratios, in order to assuage scale problems in our regressions. The logarithmic transformation is particularly convenient in our case because we specify an exponential form of heteroscedasticity.

4.3. Firm Owner Characteristics

The dummies *Proprietorship*, *Partnership* and *Corporation* control for legal and governance aspects of the firm. In addition, we employ *Corporation* (that includes both regular and S-type corporations) as a measure of firm size. The size of the firm measures the lender's access to reliable financial information about the firm (Eisenbeis (1996)). Liberti and Mian (2006) find evidence that firm size as well as loan size determine the hierarchical level within the bank that must approve the terms of the loan. In addition, these variables correlate with the size of the bank. As a result, *Corporation* and *Ln(Loan Amount)* should also capture between, as well as within, differences in banks' organizational structures (Berger and Udell (2002), Stein (2002)).

Minority indicates whether the firm is owned by members of a minority groups (African-American, Asian or Native). Our motivation to include *Minority* follows from the evidence that minority groups tend to be more adverse to looking around for the best deals and to bargaining (Black, Boehm and DeGennaro (2003)).

Ln(Owner's Age) is the log of the age of the firm's owner in years as of year-end 1993. Petersen and Rajan (1994) find that the reputation of the firm's owner is more important than that of the business in predicting loan rates. We correct *Ln(Owner's Age)* by the length of the firm-bank relationship to avoid spurious correlation between these two variables. Following Petersen and Rajan (1994) and Berger and Udell (1995), we interpret the age of the firm (owner) as the amount of public information available about the firm.[21]

4.4. Firm Accounting Information

If financial statement lending was the dominant transactional technology employed by banks in the sample period we analyze (1990-1994), then firms' accounting figures should have a determinant role in explaining loan rates.[22] *Ln(Assets)* is the log of the book value of the firm's assets. We also include the following accounting identities for the fiscal year 1993, deflated by total assets: sales, profits, inventories, accounts receivable, accounts payable, total amount of loans and total debt. In addition, we include a variable that captures the firm's reliance on trade credit (*Trade Credit Use*). Petersen and Rajan (1994) suggest that trade credit usage measures to what extent firms are credit rationed by financial institutions, hence affecting a bank's perception of the quality of the firm.

[21] We obtain similar results when we replace the age of the owner by the age of the firm.

[22] Uchida, Udell and Yamori (2007) notice that financial statement lending is a benchmark technology that banks may combine with other technologies in their underwriting method. Financial statement lending was probably the dominant transactional technology in US in the early 90s, right before the advent of Small Business Credit Scoring (SBCS). In fact, the largest provider of external models, Fair Isaac, introduced its first SBCS model in 1995.

4.5. Credit History

Previous studies find that the history of the principal (e.g., Eisenbeis (1996), Berger, Frame and Miller (2005)) and third-party mercantile ratings (Kallberg and Udell (2003)) are very strong predictors of payment performance of small-business loans. Unfortunately, the 1993 NSSBF does provide neither such credit-risk measures of the firm nor extensive owner-specific information.[23] The 1993 NSSBF does include, however, some information regarding the credit history of the firm and owner that we use. *Bankrupt* indicates whether the firm or its principal owner has declared bankruptcy within the past seven years; *Owner (Firm) Delinquent* indicates whether the owner (firm) has been 60 or more days delinquent on personal (business) obligations within the past 3 years; and *Judgments* is a dummy that equals one if any judgments have been rendered against the principal owner within the past 3 years. We further summarize the credit history of the owner and firm in the variable *Clean Record*. *Clean Record* is a dummy that equals one when all previous variables (*Bankrupt, Owner Delinquent, Firm Delinquent* and *Judgments*) equal zero. In addition, we include *IRS Problem* – a variable indicating whether the firm reported to have had problems with IRS regulation or penalties during the last year. This variable should also provide useful information about the credit ratings of the firms.

4.6. Relationship Characteristics

Boot (2000) argues that relationship lending, ultimately a judgmental lending technology, is a multi-dimensional concept. Accordingly, we control in our pricing-model for three dimensions of the nature of the firm-bank relationship. *Ln(Duration)* is the log of duration in years of the relationship the firm has had with the lender. There is ample evidence in the literature that the duration of the firm-bank relationship, a common measure of the stock of information the bank acquired, affects credit terms (see e.g. Berger and Udell (1995), and Petersen and Rajan (1994, 1995)). *Main Bank* indicates whether the lender is the firm's primary source of financial services, capturing the scope of the relationship. *Personal* is a dummy variable indicating whether the firm's most frequent method of conducting business with the lender is in person, or face-to-face.

4.7. Competition/Location Measures

The variable *Concentrated* indicates whether the Herfindahl-Hirschman Index (HHI) in the deposits market of the MSA or county where the firm's headquarters office is located is

[23] To test for the potential bias in our results due to the omission of relevant owner and credit-specific information we estimate two different specifications of the heteroscedastic model, using the 1998 SSBF. The first specification is similar to the one we use in this paper, while the second includes four additional variables (the firm's credit score as obtained from Dun&Bradstreet, the owner's level of education, the value of the owner's residence and the owner's net worth *sans* residence). We find that the two specifications provide almost identical fits and estimates for the variance equation.

greater than 1,800.[24] The banking market structure should have both a direct and indirect effect on the distribution of loan rates. On the one hand, greater market concentration increases the banks' bargaining power with respect to borrowers. On the other hand, it should also influence the nature of the firm-bank relationship (Boot and Thakor (2000)), as well as the likelihood that banks engage in intertemporal risk sharing (Petersen and Rajan (1995), Kim, Kristiansen and Vale (2006)).

The variable *Ln(Distance)* is the log of the distance between the firm's main office and the lending institution's office. Degryse and Ongena (2005) demonstrate that banks engage in spatial price discrimination. Petersen and Rajan (2002) and Berger et al. (2005) document that the lender-borrower relation becomes increasingly impersonal as their physical distance grows. *Personal* controls for the alternative transmission channel through which the firm-bank distance might affect loan rates. Finally, we include the variable *MSA*, which indicates whether the firm is located in a Metropolitan Statistical Area.

4.8. Other Control Variables

We additionally include several sets of dummy variables that control for other potential sources of information in the loan-pricing model. This information may relate to the purpose of the loan (*Loan Types*), census region (*Regions*), firm industry (*SIC Codes*), year of approval (*Loan Approval Year*) and type of lending institution (*Lender Type*).

4.9. Variable Selection in the Variance equation

We need to employ a large number of regressors in the mean equation, disregarding potential collinearity problems, for its role is to predict loan rates as accurately as possible. In the variance equation, on the other hand, we seek to understand how variables associated to particular market imperfections affect the remaining dispersion of loan rates. Although it is technically possible to employ the same set of covariates in both equations, the above argument should justify *per se* that we perform variable selection in the variance equation. An additional reason for not including many variables in the variance equation is that collinearity among regressors and the ensuing decrease in the degrees of freedom might prevent us from identifying (in a statistical sense) the principal forces affecting the dispersion of loan rates.

Next we describe the following variable selection in the variance equation, which is rooted in the theoretical discussion in Section 2. The variables *Collateral*, *Minority*, *Corporation*, *Clean Record* and *IRS Problem* proxy for the firm's opacity *ex ante*. In contrast, *Ln(Duration)* and *Ln(Distance)* measure how easy it is for the bank to monitor the firm, i.e. the firm's opacity *ex post*. The theory suggests a positive correlation between firm opaqueness and the importance of "discretion", which results from the informational advantage banks

[24] Unfortunately, this is the only measure of banking market concentration available in the 1993 NSSBF.

have over competitors. The equilibrium derived in von Thadden (2004) for example predicts that the variance of loan rates is an increasing function of the severity of information asymmetries in the credit market, i.e. in the uncertainty pertaining to the quality of the pool of borrowers. This informational advantage, which is greater vis-à-vis more opaque firms, may enhance a bank's ability to price discriminate, as well as generate unpredictable lifecycle patterns for interest rates (Petersen and Rajan (1995), Kim, Kristiansen and Vale (2006)).

The loan amount should be a crucial input in the firm's decision of how much to invest in information acquisition, i.e. how much time devoted to "price-shop" (Stigler (1961)). Because of this latent increase in competition among potential lenders, larger loans should be priced more homogeneously. In addition, we expect the variables *Ln(Loan Amount)* and *Corporation* to be strongly correlated with bank size. As a result, these variables should capture potential differences in organizational structure. Stein (2002) and Berger and Udell (2002) suggest that small banks are better at collecting and handling soft information than their larger counterparts. Because these soft elements are unobservable and easily manipulated, credit decisions by small banks should reflect a greater weight of "discretion".

The variables *Ln(Owner's Age)* and *Ln(Duration)* capture, respectively, the amount of public and private information available to the bank about the firm. These variables should relate to more "discretion" in the loan-pricing process, since added information sharpens a bank's ability to price discriminate. In contrast, banks may strategically conceal their private information by setting loan rates merely on basis of publicly observable signals (Gan and Riddiough (2006)).

Finally, the variables *MSA* and *Concentrated* control for differences in banking market size and structure. One would intuitively expect more "discretion" in more concentrated markets (Petersen and Rajan (1995)). In addition, it seems also plausible that larger banking markets are characterized by a greater diversity of underwriting procedures and by higher search costs (Varian (1980) and McAfee (1983)). As a result, we expect both *MSA* and *Concentrated* to relate to more loan rate dispersion.

5. Empirical Results – NSSBF 1993

We estimate a heteroscedastic linear regression model to analyze the determinants of the dispersion of loan rates. The mean equation specifies a linear pricing model that extracts from loan rates that information (or variation) pertaining to "rules". In the variance equation, we identify the factors affecting the residual variance of loan rates, i.e. the "discretion".

While our interest lies mainly in the parameters of the variance equation, we also report the estimates of the mean equation. Columns I and II of Table 2 report the coefficients of the mean and variance equations, respectively. The dependent variable is the interest rate on the

firm's most recent loan, in basis points (bp).[25] In the mean equation we employ all variables described in Table 1 with the exception of *Clean Record*.[26,27] In the variance equation we refine the variable selection as motivated in the previous section. We turn now to the discussion of our empirical results. We briefly summarize the results of the mean equation and then present, in detail, those pertaining to the variance equation.

5.1. Mean Equation

The average firm in our sample obtains an 8.4% loan rate.[28] Consistent with the results in Petersen and Rajan (1994 and 1995) we find that loan rates are relatively insensitive to changes in the cost of capital for banks.

Larger loans benefit from lower interest rates. This result probably reflects the borrowers' efforts to get the best possible deals concerning large loans, as well as the dilution of contractual and operational fixed costs. Moreover, larger firms obtain more favorable rates. This result suggests that banks are able to extract larger rents or that they perceive higher risk in smaller, informationally opaque firms. All the effects mentioned are statistically significant at the 1% level.

Consistent with moral hazard theories we find that collateralized loans pay on average 31 bp more than unsecured loans. The magnitude of this effect lies below the range estimated by Brick and Palia (2007), who employed the same dataset and proposed a set of instruments to identify the impact of *Collateral* on loan rates.[29]

Firms with healthier balance sheets and superior track records seem to enjoy lower loan rates, although the majority of the estimated coefficients are statistically insignificant at conventional levels. Regarding credit history, we emphasize that firms whose owners have been delinquent on personal obligations pay a premium of about 78 bp. This finding confirms that the history of the owner is a strong predictor of payment performance of small-business loans (Eisenbeis (1996)). The estimate obtained for *Minority* suggests that firms owned by minority groups pay a premium of 33 bp. We cannot determine, however, whether this result

[25] Our results are virtually unaltered when we use the credit spread, defined as the loan rate minus the prime rate, as the dependent variable.

[26] We use the composite variable *Clean Record*, rather than its components, in the variance equation in order to identify an unambiguous effect of a firm's credit history (our measure of firm credit risk) on the residual variance of loan rates, and to conserve degrees of freedom. In the mean equation, though, we employ their respective constituents (*Bankrupt*, *Owner Delinquent*, *Firm Delinquent*, and *Judgments*).

[27] Some of the variables in the mean equation, namely those related to dimensions of the loan contract other than the loan rate, raise endogeneity concerns. Our results do not change when we drop these variables (namely *Ln(Loan Amount)*, *Collateral*, *Ln(Loan Maturity)* and *Floating*) from the mean equation.

[28] Because we center all explanatory variables, the constants of the model have a direct interpretation – they represent the expected loan rate (mean equation) and the residual variance (variance equation) for the average firm in our sample.

[29] The omission of relevant covariates can also explain the insignificant effect of *Collateral* on *Loan Rate* that Brick and Palia (2007) obtain in their OLS regression. Our results are not directly comparable to theirs, because Brick and Palia (2007) employ only lines of credit (L/Cs). Nevertheless, we note that our results remain unaltered if we restrict our sample to L/Cs.

is due to differences in negotiation skills, differences in the willingness to search for better loan terms, or whether it reflects the practice of race-based discrimination.[30]

The remaining estimates are particularly inaccurate. We justify the non-statistical significance of many potentially important predictors of loan rates with the inability of our model to identify such a large number of parameters with a rather limited sample size.

The (unadjusted) R^2 of the loan-pricing model is 25%. This value is substantially higher than what other studies that employ data from the NSSBF obtain, which is probably due to the greater number of explanatory variables we include in our regressions.[31] The distinctive aspect in our study is that we recognize the explanatory power of the loan-pricing model to depend on certain characteristics of the loan contract, the borrower, and the lender. These characteristics reflect the nature of the lending technology adopted by the bank in the underwriting process, and hence determine how accurately "rules" predict observed loan rates.

5.2. Variance Equation

We now explicitly investigate how the fit of the loan-pricing model depends on loan contract, borrower, lender and market structure characteristics. Positive coefficients in the variance equation indicate larger unexplained deviations (i.e., a poorer fit of the mean equation). In turn, large unexplained deviations are consistent with the banks' adoption of loan underwriting procedures that rely more heavily on subjective inputs. In sum, we associate large deviations with the predominance of "discretion" in the loan-pricing process and small deviations with the prevalence of "rules".

The coefficient of *Ln(Loan Amount)* is negative and highly significant. We provide several interpretations for this finding. First, larger loans incite firms to prospect the credit market for the best possible deal, as the potential savings in interest payments outweigh the costs of searching information and switching lenders. In turn, banks anticipate this latent increase in competition by employing more standardized criteria or "rules" in the underwriting process of large loans. Second, there is evidence that applications involving larger loans are generally appraised by higher hierarchical levels within the lending bank, levels that soft information, and hence "discretion" may hardly reach (Liberti and Mian (2006)). On the other hand, loan amount may strongly correlate with the size of the lending bank, which we do not control for (yet, as we deal with size in robustness tests using another

[30] There is evidence that unexplained differences in loan rates between African American- and white male-owned firms exist (Cavalluzzo, Cavalluzzo and Wolken (2002)).

[31] For instance, Petersen and Rajan (1994) employ 1,389 observations from the NSSBF 1988 and obtain a R^2 of 14.5%; Berger and Udell (1995) employ a sample of lines of credit from the same survey and obtain a R^2 of 9.5%. It is unreasonable to 'blame' the poor fits of the empirical loan-pricing models entirely on an insufficient number of controls or observations. For instance, Brick and Palia (2007) employ 80 covariates and obtain a R^2 of 11%; furthermore, Degryse and Ongena (2005) analyze a sample of 15,044 loans and obtain an R^2 of 22%, even though their regression contains 83 covariates.

dataset). Consequently, the negative coefficient of *Ln(Loan Amount)* may also indicate that larger banks rely more heavily on automated and standardized decisions mechanisms (Akhavein, Frame and White (2005), Frame, Srinivasan and Woosley (2001)).

The negative sign of *Corporation* matches the evidence that smaller firms are intrinsically riskier from a bank's perspective. Either because the information provided by small firms is less reliable (Eisenbeis (1996)) or subject to faster depreciation (Chan, Greenbaum and Thakor (1986)), banks have incentives to collect proprietary information and afterwards arbitrarily capitalize on their informational monopoly.[32]

The coefficients of *Minority* and *IRS Problem* are positive while the coefficients of *Collateral* and *Clean Record* are negative, all being statistically significant. We pool these variables in our analysis because they all proxy for the borrower opaqueness.[33, 34] In particular, this finding points towards the notion that the dispersion of loan rates increases in the risk of the borrower. This result is consistent with the market equilibrium analyzed in von Thadden (2004), which predicts a negative relation between the quality of the pool of borrowers and the variance of the loan rates. Petersen and Rajan (1995) and Kim, Kristiansen and Vale (2006)) predict that banks' incentives to intertemporally cross-subsidize firms increase with the severity of informational problems they face. In addition, riskier firms that obtained credit probably underwent a more comprehensive screening process, endowing banks with more soft information about these firms. These two mechanisms may amplify unpredicted deviations from an objective loan-pricing model and hence explain our results. Alternatively, the positive relation between firm risk and "discretion" may also reflect a bank's incentive to disguise the true credit quality of high-quality borrowers through pool pricing (Gan and Riddiough (2006)).

The negative estimate obtained for *Ln(Duration)* indicates a greater prevalence of "rules" when there are strong ties between the firm and the lending bank. This seems to contradict the conventional wisdom that additional private information about a firm enhances a bank's ability to discretionary shift loan prices.[35] However, the closeness of their relationship also determines a lender's ability to monitor the firm, through the nonborrowing side and the personal proximity of their relation, as well as the accuracy and relevance of the information that is available.[36] As a result, we interpret this finding in line with our other measures of

[32] Off course, corporations tend to prefer large banks, so that the abovementioned arguments pertaining to *Ln(Loan Amount)* apply.
[33] As suggested before, the positive sign of *Collateral* in the pricing equation indicates that this variable is signaling borrower risk, as predicted by moral hazard theories.
[34] See Phelps (1972) for a formal argument with respect to the correlation between credit quality and race.
[35] We added in a separate model the variables *Main* and *Personal* to investigate whether the other dimensions of the firm-bank relationship had a dissimilar impact on the residual variance. We obtained negative coefficients for these two variables and maintained a negative estimate for Duration, reinforcing the rather paradoxical result of a negative link between the strength of relationship and the importance of "discretion".
[36] For instance, Mester, Nakamura and Renault (2006) show that the information provided in checking accounts helps the bank to monitor commercial borrowers and to assess sudden changes in credit risks.

borrower opaqueness, i.e. as evidence that the strength of the firm-bank relationship ameliorates the bank's perception of the borrower's risk.[37]

Age of the firm (or its owner's) may measure the firm's public transparency (e.g. Berger and Udell (1995), Petersen and Rajan (1994, 1995)). We obtain a positive and significant coefficient for *Ln(Owner's Age)*, suggesting that banks are able to discriminate loan prices by exploiting public information about borrowers.[38] Given this result, one may pertinently question why banks do not seem to employ private information about firms in loan prices as well. The model by Gan and Riddiough (2006) offers a possible explanation. Banks have an incentive not to disclose private information about their clients in order to keep the informational barriers. In contrast, discriminating loan prices based on public information should not provide incentives to competitors to enter the market since that information is available to all participants. Consequently, an additional explanation for the negative coefficient of *Ln(Duration)* is the bank's incentive to stick to uniform pricing vis-à-vis their long-standing clients to shield its informational advantage vis-à-vis competitors.

Concerning market structure, we obtain positive estimates for the variables *MSA* and *Concentrated*. Petersen and Rajan (1995) assert that higher levels of banking market concentration provide the necessary incentives for banks to engage in intertemporal risk sharing. In addition, the positive sign of *MSA* indicate that loans rates are more dispersed in larger markets (as predicted in Varian (1980), and Carlson and McAfee (1983)). We also find that *Ln(Distance)* is associated with larger unpredicted deviations. Because distance decreases a bank's competence to monitor the firm, this result is consistent with the view that distance represents an additional risk factor for the bank.[39]

5.3. Economic Significance of Estimates

We analyze the economic significance of the coefficients in the variance equation in terms of their effects on the R^2 of the mean equation.[40] The constant measures the residual

[37] The model in von Thadden (2004) predicts that the variance of loan rates, besides decreasing in the quality of the pool of borrowers, it also decreases on the uncertainty concerning the distribution of the borrowers' quality (specified in the model by the spread in success rates between high and low quality firms). Accordingly, our results may be simply capturing this link – stronger ties between firm and bank means less uncertainty for the bank (even if the firm turns out to be a "lemon").

[38] If *Ln(Owner's Age)* provides a signal (either positive or negative) about the creditworthiness of the firm's owner and if banks use this public information to price loans, "good" firms will get lower rates while "bad" firms will get higher rates. Consequently, the unexplained variance increasing in *Ln(Owner's Age)* may indicate that banks use public information to price loans given to the "good" and "bad" types (that coexist in our sample). Alternatively, older managers may face higher search costs, for instance because they are more reluctant or less skillful in using modern search technology available on the internet (Brown and Goolsbee (2002)). The positive sign of *Ln(Owner's Age)* in the mean equation is in line with this view.

[39] The firm-bank distance may also affect the dispersion of loan rates through the nature of their relationship (Hauswald and Marquez (2006)). In fact, the coefficient *Ln(Distance)* nearly halves when we add *Main* and *Personal*, suggesting that this alternative channel plays a substantial role.

[40] Two assumptions are implicit in this analysis: we compute the changes in R^2 conditional on all covariates and on the sample (implying that we treat the total variance of loan rates as exogenous). Under these assumptions, the

variance of the loan-pricing model for the average firm in the sample. Its estimated value implies a mean deviation from the predicted loan rate of 77 bp, or a R^2 of 55% in the mean equation. Because this R^2 should represent a lower bound of the true importance of "rules" in loan rates, this result suggests that "rules" have substantially more weight than "discretion" for the average borrower. The coefficient of *Ln(Loan Amount)* indicates that an increase in the loan amount from $25,000 (the 25th percentile) to $550,000 (the 75th percentile) induces a nearly six-fold increase in the fit of the pricing model. An increase in the distance separating the firm from the bank from 1 mile (the 25th percentile) to 13 miles (the 75th percentile) more than doubles the R^2 of the pricing model; however, going from 1 mile to 304 miles (the 95th percentile) increases the R^2 by a factor of four. The economic effects of the other continuous variables, *Ln(Duration)* and *Ln(Owner's age)*, are comparatively modest, ranging in magnitude close to the effects of the discrete variables (i.e. with elasticities smaller than one).

Another insightful exercise is to assess the dispersion of loan rates for different types of borrowers. For instance, consider an unsecured loan of $25,000 granted to a single business (firm A).[41] Firm A's track record contains recent business delinquencies and it has conducted business with the lender for three years. Suppose there exists also a corporation (firm B) that is granted a $550,000 secured loan. In contrast to firm A, firm B enjoys a clean legal record and it has had a relationship with the lender for 13 years. According to our estimates, the expected loan rates for firms A and B are 9.3% and 8.1%, respectively. In addition our estimates translate into R^2s of the estimated pricing model of 1% for firm A and 81% for firm B. The residual deviations implied by these R^2s indicate that, with 95% confidence, firm A could face any loan rate in the range from 5.1% to 13.5%, whereas the range for firm B is substantially narrower: 6.3% to 9.9%. In short, these results illustrate the contrast between the intense use of "discretion" by the bank in the loan pricing process (firm A) and a standardized loan-pricing model (firm B).

6. Robustness Tests

6.1. Analysis of Extreme Residuals

Because we focus on the unexplained loan rate spreads paid by the firms, a pertinent question is to what extent the estimated firm-specific variances may reflect asymmetric, or one-sided, deviations. In particular, we seek to identify the factors associated with large upward and downward deviations from the "rules". For this purpose we propose to analyze separately positive and negative residuals. Our empirical strategy is as follows. We estimate a

---

interpretation of the coefficients in the variance equation is straightforward: for a continuous (discrete) variable, it denotes the relative change in the R^2 resulting from a relative (discrete) change in the corresponding variable.

[41] In the subsequent analysis we set the continuous variables to their 25th or 75th percentiles, as explicitly mentioned or implied in the text. The omitted variables take the value of zero.

quantile regression of *Loan Rate* on the set of explanatory variables used in the loan-pricing model.[42] Then, importing the classic terminology from Salop and Stiglitz (1977), we generate the dummy variables *Rip-off* / *Bargain*, which indicate whether the estimated disturbance is larger / smaller than one standard deviation above / below the sample mean of the residuals.[43] Finally, we estimate *logit* regressions of the variables *Rip-off* and *Bargain* on the same set of variables we employ in the variance analysis (see Column II in Table 2).

Table 3 displays the results. Consistent with our previous findings, *Ln(Loan Amount)* and *Corporation* decrease the probability of observing large deviations, in particular the probability of a rip-off. In contrast, the variables *Minority, IRS Problem* and *Ln(Distance)* increase the probability of both a bargain and a rip-off. *Minority* has a stronger effect on a downwards deviation, which is consistent with a bank's willingness to subsidize opaque firms (Petersen and Rajan (1995)). Conversely, banks are substantially more likely to add a mark-up to a firm that has a fiscal delinquency in its track record. We are unable to draw conclusions on the remaining results, as they are particularly noisy and mixed.

6.2. Relevance of the Information in the Loan Granting Process

We admittedly saturated our loan-pricing model with a great number of covariates, raising the concern that some of that information may be irrelevant from the bank's perspective. We cannot directly test this assertion, for the true pricing model is unknown to us. Nevertheless, we propose to test this assertion indirectly. If the information contained in our variables is relevant for pricing purposes, then it should predict relatively well the banks' decisions of granting or denying credit to applicants. We estimate the probability of a firm being granted a loan as function of the set of variables employed in the loan-pricing model. In other words, we replicate a multivariate logistic credit-scoring model.[44] Studies focusing on scoring methods suggest that financial institutions extensively use multivariate logistic models to estimate default probabilities (Altman and Saunders (1998), Hand and Henley (1997)). Using the 50% probability as the cut-off decision point, our model predicts correctly 88% of the observed outcomes, clearly reflecting the relevance of the information used. We believe our results are potentially strengthened if banks use more conservative cut-offs in the acceptance/rejection decision.[45]

---

[42] The quantile (or median) regression has the advantage over the linear (or mean) regression model that, while not affecting the residual variance, it is robust to skewness in the distribution of the residuals. For instance, if positive deviations from the true pricing model are more sizeable than the negative ones, then a mean regression model will underestimate positive deviations and overestimate the negative ones.

[43] The sample mean of residuals of the quantile regression is positive (about 17 bp), suggesting that banks are more likely to shift loan rates upwards than downwards.

[44] Naturally, we drop all variables related to loan characteristics. The loan amount is replaced by the amount requested. The total number of observations is 1,916, distributed into 1,625 accepted loans and 291 credit denials.

[45] As suggested before, banks may reduce risk by bringing other types of information into the credit approval decision, especially when the applicants' scores are in the "grey area" (Hand and Henley (1997)).

6.3. Sample Selection

The analysis we made so far disregards the fact that banks may use the same subjective elements in both the approval/rejection and the loan pricing decisions. We question how our results change when we take into account this sample selection bias. For this purpose, we combine the heteroscedastic regression model with a heteroscedastic probit model (the selection equation). Because we include the same variables in both variance equations, we allow for two sources of correlation between the two error terms: the variables in the variance equation and a parameter (ρ) measuring the degree of sample selection bias.[46] Our results indicate the absence of selection bias as ρ is negligible (both economically and statistically). We also find that our variables in the variance equation have very low explanatory power in the credit approval model, suggesting that "discretion" mainly occurs through loan-pricing.

6.4. Specification of the Pricing Model

We question to what extent the linearity assumption in our loan-pricing model may be biasing our results. As a robustness test, we estimate the pricing model and obtain the fitted values, i.e. the predicted loan rates; then, we re-estimate the model including in the mean equation the quadratic and cubic terms of the fitted values.[47] If the potential non-linearities in the model are biasing our results, then the introduction of these extra terms should significantly alter the estimates in the variance equation. The estimates obtained (not tabulated) indicate that is not the case. Despite the non-linear terms being statistically significant in the mean equation, the results in the variance equation are virtually unaffected.[48]

6.5. Discontinuities in "Rules" versus "Discretion"

The application of "discretion" itself may be inherently discontinuous. For example, various loan and firm types may be assigned to different loan officers operating under unique pricing rules (see Liberti and Mian (2006)). To test for the presence of such discontinuities we decompose each continuous variable into a linear spline with two breaks.[49] We then estimate a heteroscedastic regression using the splines instead of the original variables in the variance equation.

---

[46] In particular, the covariance between the two equations is given by $\rho \exp(-0.5\gamma_0 + \gamma' Z)$, where ρ is the correlation coefficient between the two equations measuring the degree of sample selection bias, and where γ_0 is the constant term in the variance equation.

[47] In spirit similar to the Ramsey (1969)'s Specification (or Reset) Test.

[48] We also estimate the model employing linear splines (at ten equally spaced knots) in the mean equation in place of the original (continuous) variables. Apart from a substantial increase in the coefficient of *Ln(Owner's Age)*, the results in the variance equation remain virtually unaltered.

[49] Each continuous variable is mapped into the following three mutually exclusive categories: small (if the value of the variable is in the lowest tercile), large (if the value is in the largest tercile) and medium (for intermediate values). Our results are qualitatively similar when we impose four breaks (yielding five distinct categories) instead of two.

We find evidence of a pronounced discontinuity in the loan amount on "discretion". In particular, we obtain a sharp negative effect for small loans (below $47,000), a moderate negative effect for large loans (above $325,000), and surprisingly no effect for medium-sized loans (between $47,000 and $325,000). We conjecture that the strong negative effect for small loans reflects search efforts by firms. The break occurring at mid-sized loans followed by the negative effect for large loans may capture either a discontinuity at the hierarchical level within the lending banks or a discontinuity across banks of different sizes.

Our previous findings suggest that "discretion" decreases over the duration of the relationship between the firm and the bank. We confirm this result for short (less than 3 years) and long (more than 10 years) durations, but we find the opposite effect for medium lengths (between 3 and 10 years).[50] Concerning *Ln(Distance)*, our results indicate that the positive effect of this variable on "discretion" is constant but confined to firms located further than two miles from the lender. That is, *Ln(Distance)* does not affect "discretion" when the firm is within a two mile radius of the bank. Finally, we find no discontinuities regarding the age of the owner.

6.6. Bank Heterogeneity

A major drawback of using the NSSBF 1993 dataset is that it comprises several unidentified banks. This is problematic as banks can differentiate themselves along service dimensions and operating costs. Fortunately, we have access to an alternative dataset that allows us to control for bank heterogeneity. This dataset, which has been used in Degryse and Van Cayseele (2000), and Degryse and Ongena (2005, 2007b), consists of a large number of loans made by one important Belgian bank.[51] Despite being less comprehensive than the NSSBF in terms of firm-specific information, we find several attractive features in the Belgian dataset. First, and most of all, it allows us to include branch fixed-effects in our regressions to control for branch heterogeneity within the single lending bank. Second, it permits us to test the sensitivity of our results with respect to the NSSBF sample design. In particular, the potential differences in financial, geographical and cultural landscapes between the U.S. and Belgium provide a serious challenge to the scope of our results.

We estimate a heteroscedastic regression model using all the information available in the Belgian dataset that we employed previously in the regressions with the NSSBF sample (i.e., the variables described in Table 1). Due to data limitations in the Belgian sample, we are only able to explain the variance as function of the variables *Ln(Loan Amount)*, *Collateral*,

[50] By allowing a greater number of intervals in the spline we actually see that the initial break occurs for relationships lasting between one and one-and-a-half years. In the medium term the loan officer may have achieved her maximum informational advantage. Internal mandated rotation and outside banks' observing the repeated lender – borrower interactions might ultimately weaken this informational advantage.

[51] We refer to these studies for a detailed description of the dataset.

Corporation, Ln(Duration), *Concentrated* and *Ln(Distance)*.[52] Aside from *Collateral*, the results (not tabulated) confirm our previous findings, as the estimates are nearly identical in terms of sign and statistical significance to those we obtain with the NSSBF sample (see Table 2, Column II).[53] Besides highlighting that bank heterogeneity is not driving our earlier results, this robustness check demonstrates that their scope goes beyond the specificity of the sample analyzed.

6.7. Industry Heterogeneity

Banks may adopt different loan pricing strategies or "rules" for different industries. We run a heteroscedastic regression for each of four broadly defined industry groups: construction (SIC1 – division C, 186 observations), manufacturing (SIC2 and SIC3: 274 observations), trade (SIC5: 502 observations) and services (SIC6, SIC7 and SIC8: 571 observations).[54] The pooled sample results largely carry through to these industry-specific regressions, though the new estimates display larger standard errors. We report, however, the following discrepancies in the variance equation, which are likely related to composition differences across these samples. First, *Ln(Amount)* seems not to be a determinant of "discretion" in the manufacturing sector.[55] Second, *Concentrated* has a negative sign and is statistically significant in the construction and trade sectors.

7. Has "Discretion" Varied Over Time?

We now seek to detect temporal changes in the banks' loan-pricing behavior. In particular, we investigate whether the weight of "discretion", reflected in the greater unexplained dispersion of loan rates, has changed over time.

The literature fails to propose a clear-cut prediction regarding how the importance of "rules" and "discretion" may have evolved over the last two decades. It is widely recognized that the exponential increase in information processing capabilities has provided banks with innovative and low-cost tools to support their credit decisions. Less obvious is that the effective influence of these technologies in the banks' credit decisions has increased over time. For instance, Berger, Frame and Miller (2005) argue that the new technologies may act as a substitute for or complement to existing lending technologies. Accordingly, one could

[52] Unfortunately, the Belgian dataset does not provide credit history information on the firm, accounting information, information about the owner nor does it specify the financial services the bank provides to the firm. As a result, we cannot include in the variance equation the variables *Minority*, *Ln(Owner's Age)*, *Clean Record, IRS Problem* and *MSA*. In addition, in the Belgian sample *Ln(Distance)* is the shortest traveling time to the lender (in minutes).

[53] The coefficient of *Collateral* is positive and highly significant. Because we lack proxies for firm risk in the variance equation (such as the legal record), it may well be that *Collateral* is standing for credit risk.

[54] Our sample contains only 17 observations from the primary sector (agriculture and mining) and 75 observations from utilities (SIC4), hence we disregard these sectors in our analysis. To conserve degrees of freedom we estimate the heteroscedastic model with the same set of variables in the mean and variance equations (i.e., those variables we previously selected for the variance equation, as described in Section 4.9).

[55] The median loan amount in the manufacturing sector is about four times the total sample average.

expect an increasing or a decreasing role over time for "discretion".[56] In our subsequent empirical analysis, we seek to identify which of these effects dominates. Specifically, we test in a panel sample for the existence of a trend in the unexplained variation of loan rates.

7.1. Panel Sample

We construct a fully rotating panel by merging the 1993 NSSBF (1,625 observations, from 1990 to 1994), the 1998 SSBF (708 observations, from 1996 to 2000) and the 2003 SSBF (1,568 observations, from 2001 to 2005). The resulting panel sample spans 16 years, from 1990 to 2005. A consistent definition and a majority of identical questions used across all three surveys permits an analysis of changes over time.[57] Table 4 reports sample statistics for each of the three samples. We opt to provide independent statistics for each sample as the unfolded panel allows us to assess major changes in the sample compositions across surveys, as well as spot temporal correlations.

We start by analyzing the differences between the 1998 SSBF and the other samples.[58] This period is characterized by a severe contraction in credit supply, reflected in the higher cost of capital and smaller loans, when compared to the 1993 NSSBF and 2003 SSBF. There is also an unusual higher incidence of transactional lending, signaled by the low percentage of L/Cs (nearly half of that observed in the other samples), the low fraction of loans granted by commercial banks (not tabulated), and the atypical values displayed by the relationship variables (shorter durations, higher incidence of impersonal relations and lower incidence of cross-selling). In addition, borrowing firms in the 1998 SSBF are on average smaller, more profitable, leveraged and riskier. Concerning market structure, firms in the 1998 SSBF operate in more concentrated markets and in rural areas, and borrow largely from distantly located lenders.

By comparing the 1993 NSSBF to the 2003 SSBF we identify three clear trends in the data, irrespective of the referred fluctuations in the cost of capital surrounding 1998. First, there is an increase in loan maturity. Second, there is a steep decrease in the incidence of loans secured by collateral. Finally, the dispersion of loan rates has increased over time, which is not to say that "discretion" has increased over time. In the next section we formally test for the presence of such a trend in the data.

[56] One could also argue that the proliferation of the internet led to a decrease in search costs, and hence of "discretion". However, this may have not been the case initially (mid- to late-nineties). First, internet banking is a rather recent phenomenon, and second, there is evidence that internet in early times actually led to an increase in price dispersion (Brown and Goolsbee (2002)).

[57] *IRS Problem* is the only variable not present in the three surveys.

[58] The 1998 sample has a rather odd composition, when compared to the 1993 and 2003 surveys, probably due to the global financial crisis surrounding 1998. Our results are virtually unaffected when we remove the 1998 survey from our panel sample.

7.2. Results

Table 5 displays the results from estimating a heteroscedastic regression model with the panel sample. We just present the variance equation estimates to conserve on space. Column I employs a specification identical to the one we used to derive our cross-sectional results (see Column II, Table 2).[59] The panel results, apart from the drop in both magnitude and significance of *Clean Record* and *Ln(Duration)*, confirm our previous findings with the 1993 NSSBF. We note, however, a substantial increase in the overall explanatory power of the model, as the R^2 of the variance-weighted linear regression (i.e. the loan-pricing model weighted by the estimated variances) leaps from 28% to 50%. This increase in explanatory power seems to be a product of the increase in the number of observations.[60]

In Column II we add the temporal dimension to our analysis by including in the variance equation the variables *Prime Rate* and *Year*. *Prime Rate* controls for fluctuations in banks' risk shifting incentives. Consistent with the view in Rajan (2006), the negative coefficient for *Prime Rate* indicates that banks tend to take more risks in periods of high liquidity (i.e. low interest rates). The variable *Year* captures a time trend in the residual variance. We obtain a positive and statistically significant trend, which we interpret as tentative evidence of an increase over time in the weight of "discretion".

Despite suggestive, these results may simply reflect structural changes in market conditions occurring during the period analyzed. We have in mind, in particular, the growing evidence that banks were increasingly lending to riskier (Berger, Frame and Miller (2005)), more distantly located firms, and on a more transactional basis (Petersen and Rajan (2002)). Our cross-sectional results (see Column II in Table 2) and the predictions in von Thadden (2004) imply that these referred changes could explain *per se* a positive trend in the unexplained variance.

To account for potential changes in market conditions we interact every variable in the variance equation with *Year*. Column III displays the results. The previous estimates in the variance equation are virtually unaffected by the inclusion of the interaction terms, which simply add a dynamic dimension to our cross-sectional results. For instance, it seems that the positive effect of *Clean Record* and *Concentrated* on the residual variance (i.e. "discretion") has increased over time. In contrast, loans secured by collateral were during this period increasingly priced according to "rules". Consistent with the concerns expressed in Rajan (2006), our estimates also suggest that banks' risk shifting incentives have intensified over time.

[59] We employ in the mean equation the same variables we used in our cross-sectional analysis (see Table 1) plus a full set of year dummies (16) to control for aggregate fluctuations.

[60] We test this assertion by estimating separately a regression model for each sample. We do not observe significant differences in explanatory power in the three samples. This result leads us to conclude that the increase in the number of observations is probably driving the increase in explanatory power of the loan-pricing model.

Because we center all covariates, the variable *Year* captures a time trend in the residual variance for the sample average firm.[61] Its negative coefficient suggests a declining role for "discretion" over the sample period for the average firm, though the effect is not statistically significant.[62] But the interaction terms reveal an interesting asymmetry this result masks. In particular, it seems that the negative time trend is more pronounced for small firms with weaker credit histories that obtain small loans, and nearly inexistent for large firms borrowing large amounts. In particular, we obtain a statistically and economically significant decrease over time of the weight of "discretion" for a small firm that borrows $10,000 from a new lender, consistent with the proliferation of credit scoring models for micro-loans (Berger, Frame and Miller (2005)) and the findings in Brevoort (2006) and DeYoung, Frame, Glennon, McMillen and Nigro (2007). In contrast, we cannot reject the hypothesis that the role of "discretion" in underwriting decisions concerning large loans and loans granted to large, transparent firms has remained steady over the last fifteen years.[63]

8. Conclusion

In this paper, we propose a heteroscedastic linear regression model to analyze the unexplained dispersion of loan rates. Our variance analysis allows us to infer about the nature of the lending technology that banks employ in the loan granting process. Larger unexplained deviations from the loan-pricing model are consistent with banks engaging in discretionary loan-pricing (a judgmental technology). On the other hand, we interpret smaller deviations as the predominance of "rules" (a fully transparent lending technology). By parametrizing the unexplained variance of the loan-pricing model, we are able to identify the factors associated with "rules" and "discretion". We believe this study is unique in that it combines within the same empirical framework a wide array of theoretical developments in the financial economics literature on the role of informational frictions on market performance.

Consistent with theories based on costly search of information, we find that the weight of "discretion" in loan prices decreases in the size of the loan. In contrast, loan rate dispersion increase in various measures of the borrower opaqueness, in particular firm size, credit history and the bank's effort to monitor the firm. We interpret this result as evidence that the switching costs faced by firms are an important source of market power for the lending bank.

[61] Computing the marginal effect of *Year* for a median firm yields a similar result.
[62] The 2003 SSBF contains a substantial number of renewals of lines of credit. Renewal decisions may rely more on old information than decisions on new applications, introducing additional noise in our static loan-pricing model. Indeed, we obtain a sharper negative trend for "discretion" when we drop these renewals. The other results are virtually unaffected.
[63] These temporal results may be partially masked by structural changes in the legal environment in the U.S. banking industry, i.e., the 1999 Gramm-Leach-Bliley Act, that occurred during our sample period. We include an indicator variable for loans granted after 1999 in the variance equation (of the model in Column III). This inclusion reinforces our previous results as the negative trend for "discretion" becomes steeper. Interestingly, we find that the residual variance is 57% larger for loans granted after 1999, this effect being statistically significant at the 1% level.

Our results also reveal that the weight of “discretion” increases with the age of the firm owner, as well as with the size and level of concentration in the banking market.

We perform several robustness tests regarding the specification of the loan-pricing model, and compared our primary results (NSSBF 1993) with those obtained in a dataset pertaining to a large Belgian bank. We found minor disparities, suggesting that our results are not sample-specific, time-specific, or driven by differences in banks’ technological and organizational structures. Finally, we construct a panel sample with the 1993, 1998 and 2003 (N)SSBF and show that the significance of “discretion” decreased over the period analyzed (1990-2005) for small loans to opaque firms. Changes in market conditions rooted in the development of information and communication technologies seem to explain the decrease in loan rate dispersion during this period.

References

Akhavein, J. Frame, W., and White, L. (2005) The Diffusion of Financial Innovations: An Examination of the Adoption of Small Business Credit Scoring by Large Banking Organizations. *Journal of Business*, 78, 577-596.

Altman, E. and Saunders, A. (1998) Credit Risk Measurement: Developments over the Last 20 Years. *Journal of Banking and Finance*, 21, 1721-1742.

Baye, M. R., Morgan, J. and Scholten, P. (2006) Information, Search, and Price Dispersion. *In Handbook of Economics and Information Systems* (ed. T. Hendershott), Elsevier, Amsterdam

Berger A., Frame, W. and Miller, N. (2005) Credit Scoring and the Availability, Price, and Risk of Small Business Credit. *Journal of Money, Credit, and Banking*, 37, 191-222.

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

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

Berger, A. and Udell, G. (2002) Small Business Credit Availability and Relationship Lending: The Importance of Bank Organizational Structure. *Economic Journal*, 112, 32-53.

Berger. A. and Udell, G. (2006) A More Complete Conceptual Framework for SME Finance. *Journal of Banking and Finance*, 30, 2945-2966.

Besanko, D., and Thakor, A. (1987) Collateral and Rationing: Sorting Equilibria in Monopolistic and Competitive Credit Markets. *International Economic Review*, 28, 671-689.

Bester, H. (1985) Screening vs. Rationing in Credit Markets with Imperfect Information. *American Economic Review*, 75, 850-855.

Bester, H. (1988) Bargaining, Search Costs and Equilibrium Price Distributions. *Review of Economic and Statistics*, 55, 201-214.

Bester, H. (1993) Bargaining versus Price Competition with Quality Uncertainty. *American Economic Review*, 83, 278-288.

Black, H., Boehm, T. and DeGennaro, R. (2003) Is There Discrimination in Mortgage Pricing? The Case of Overages. *Journal of Banking and Finance*, 27, 1139-1165.

Boot, A. (2000) Relationship Banking: What do we Know? *Journal of Financial Intermediation*, 9, 7-25.

Boot, A., and Thakor, A. (1994) Moral Hazard and Secured Lending in an Infinitely Repeated Credit Market Game. *International Economic Review*, 35, 899-920.

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

Brick, I., and Palia, D. (2007) Evidence of Jointness in the Terms of Relationship Lending. *Journal of Financial Intermediation*, 16, 452-476.

Brevoort, K. (2006) An Empirical Examination of the growth in Out-of-Market Commercial Lending: The Changing Competitive Landscape and the Role of Asymmetric Information, Mimeo, Federal Reserve Board.

Brown, J. and Goolsbee, A. (2002) Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry. *Journal of Political Economy*, 110, 481-507.

Carlson, J. A. and McAfee, R. P. (1983) Discrete Equilibrium Price Dispersion. *Journal of Political Economy*, 91, 480-493.

Cavalluzzo, K., Cavalluzzo, L. and Wolken, J. (2002) Competition, Small Business Financing, and Discrimination: Evidence from a New Survey. *Journal of Business*, 75, 641-680.

Chan, Y., Greenbaum, S. and Thakor, A. (1986) Information Reusability, Competition and Bank
Asset Quality. *Journal of Banking and Finance*, 10, 255-276.

Cole, R., Goldberg, L. and White, L. (2004) Cookie-Cutter versus Character: The Micro Structure of Small Business Lending by Large and Small Banks. *Journal of Financial and Quantitative Analysis*, 39, 227-252.

Dahlby, B. and West, D. S. (1986) Price Dispersion in an Automobile Insurance Market. *The Journal of Political Economy*, 94, 418-438.

Degryse, H., and Van Cayseele, P. (2000) Relationship Lending within a Bank-Based System: Evidence from European Small Business Data. *Journal of Financial Intermediation*, 9, 90-109.

Degryse, H., Laeven, L. and Ongena, S. (2007) The Impact of Organizational Structure and Lending Technology on Banking Competition. Mimeo, Tilburg University.

Degryse, H. and Ongena, S. (2005) Distance, Lending Relationships, and Competition. *Journal of Finance*, 60, 231-266.

Degryse, H. and Ongena, S. (2007a) Competition and Regulation in the Banking Sector: A Review of the Empirical Evidence on the Sources of Bank Rents. Mimeo, Tilburg University.

Degryse, H. and Ongena, S. (2007b) The Impact of Competition on Bank Orientation. *Journal of Financial Intermediation*, 16, 399-424.

Detragiache, E., Garella, P. G. and Guiso, L. (2000) Multiple versus Single Banking Relationships: Theory and Evidence. *Journal of Finance*, 55, 1133-1161.

DeYoung, R., Frame W. S., Glennon D., McMillen D. P., and Nigro P. J. (2007), Commercial Lending Distance and Historically Underserved Areas, Working Paper, Federal Reserve Bank of Atlanta.

Eisenbeis, R. (1996) Recent Developments in the Application of Credit-Scoring Techniques to the Evaluation of Commercial Loans. *IMA Journal of Mathematics Applied in Business & Industry*, 7, 271-290.

Frame, W., Srinivasan, A., Woosley, L. (2001) The effect of credit scoring on small business lending. *Journal of Money, Credit, and Banking*, 33, 813-825.

Fudenberg, D. and Tirole, J. (2000), Customer Poaching and Brand Switching, *Rand Journal of Economics*, 31, 634-657.

Gan, J. and Riddiough, T. (2007) Monopoly and Information Advantage in the Residential Mortgage Market. *Forthcoming in Review of Financial Studies*.

Godbillon-Camus, B. and Godlewski, C. (2005) Credit Risk Management in Banks: Hard Information, Soft Information and Manipulation. Mimeo, University of Strasbourg.

Guttentag, J. (2003) Referrals: Who Can You Trust? Retrieved November 14, 2006, from http://finance.yahoo.com/loan/mortgage/how_to_shop_for_a_loan/article/101523/Referrals:_Who_Can_You_Trust.

Hand, D. and Henley, W. (1997) Statistical Classification Methods in Consumer Credit Scoring: A Review. *Journal of the Royal Statistical Society*, 160, 523–541.

Harvey, A. (1976) Estimating Regression Models with Multiplicative Heteroscedasticity. *Econometrica*, 44, 461-465.

Hassink, W. and Leuvensteijn, M. (2007) Measuring Transparence in the Dutch Mortgage Market. *De Economist,* 155, 23-47.

Hortaçsu, A. and Syverson, C. (2004) Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds. *Quarterly Journal of Economics*, 119, 403-456.

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

Heffernan, S. (2002) How Do UK Financial Institutions Really Price Their Banking Products? *Journal of Banking and Finance*, 26, 1997-2016.

Kallberg, J. and Udell, G. (2003) The Value of Private Sector Business Credit Information Sharing: The US Case. *Journal of Banking and Finance*, 27, 449-469.

Kim, M., Kristiansen, E. and Vale, B. (2006) What Determines Banks' Market Power? Akerlof versus Herfindahl. Mimeo, Norges Bank.

Klemperer, P. (1995) Competition When Consumers Have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade. *Review of Economic Studies*, 62, 515–539.

Liberti, J. and Mian, A. (2006) Initiative, Incentives and Soft Information. How Does Delegation Impact the Role of Bank Relationship Managers? Mimeo, Kellogg School of Management.

Machauer, A. and Weber, M. (1998) Bank Behavior Based on Internal Credit Ratings of Borrowers. *Journal of Banking and Finance*, 22, 1355-1383.

Mester, L., Nakamura, L. and Renault, M. (2006) Transaction Accounts and Loan Monitoring. *Review of Financial Studies*, 20, 529-556.

Ongena, S. and Smith, D. (2000) What Determines the Number of Bank Relationships? Cross-Country Evidence. *Journal of Financial Intermediation*, 9, 26-5.

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

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

Petersen, M. and Rajan., R. (2002) The Information Revolution and Small Business Lending: Does Distance Still Matter? *Journal of Finance*, 57, 2553-2570.

Phelps, E. (1972) The Statistical Theory of Racism and Sexism. *American Economic Review,* 62, 659-661.

Rajan, R. (1992) Insiders and Outsiders: The Choice between Informed and Arm's-Length Debt. *Journal of Finance*, 47, 191-212.

Rajan, R. (2006) Has Financial Development Made the World Riskier? *European Financial Management*, 12, 499-533.

Ramsey, J. (1969) Tests for Specification Errors in Classical Linear Least Squares Regression Analysis. *Journal of the Royal Statistical Society*, 32, 350-371.

Salop, S. and Stiglitz, J. (1977) Bargain and Rip-offs: A Model of Monopolistically Competitive Price Dispersion. *Review of Economic Studies*, 44, 493-510.

Salop, S. and Stiglitz, J. (1982) The Theory of Sales: A Simple Model of Equilibrium Price Dispersion with Identical Agents. *American Economic Review,* 72, 1121-1130.

Sharpe, S. A. (1990) Asymmetric Information, Bank Lending and Implicit Contracts: A Stylized Model of Customer Relationships. *Journal of Finance*, 45, 1069-1087.

Sorensen, A. T. (2000) Equilibrium Price Dispersion in Retail Markets for Prescription Drugs. *Journal of Political Economy*, 108, 833-850.

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

Stigler, G. (1961) The Economics of Information. *Journal of Political Economy*, 69, 213-225.

Uchida, H., Udell, G. and Yamori, N. (2007) SME Financing and the Deployment of Technologies in Japan. RIETI Discussion Paper.

Varian, H. R. (1980) A Model of Sales. *American Economic Review*, 70, 651-659.

Villas-Boas, J. M. (1999) Dynamic Competition with Customer Recognition. *Rand Journal of Economics*, 30, 604-631.

Von Thadden, E. L. (1992) The Commitment of Finance, Duplicated monitoring and the Investment Horizon, Working Paper in Financial Markets, ESF-CEPR.

von Thadden, E. L. (2004) Asymmetric Information, Bank Lending, and Implicit Contracts: The Winner's Curse. *Finance Research Letters*, 1, 11-23.

Wang, R. (1995) Bargaining versus Posted-Price Selling. *European Economic Review*, 39, 1747-1764.

Figure 1
Moving-Window Regressions

This figure plots the root mean squared error (RMSE) of the moving-window regressions against the log of the loan amount. First, we line up the observations by the loan amount (from smallest to largest), and then perform moving-window regressions of the loan rate on the set of variables described in table 1, using a fixed window size of 500 observations. The figure plots the sequence of RMSEs obtained in the regressions against the log of the loan amount. The dataset used is the 1993 NSSBF.

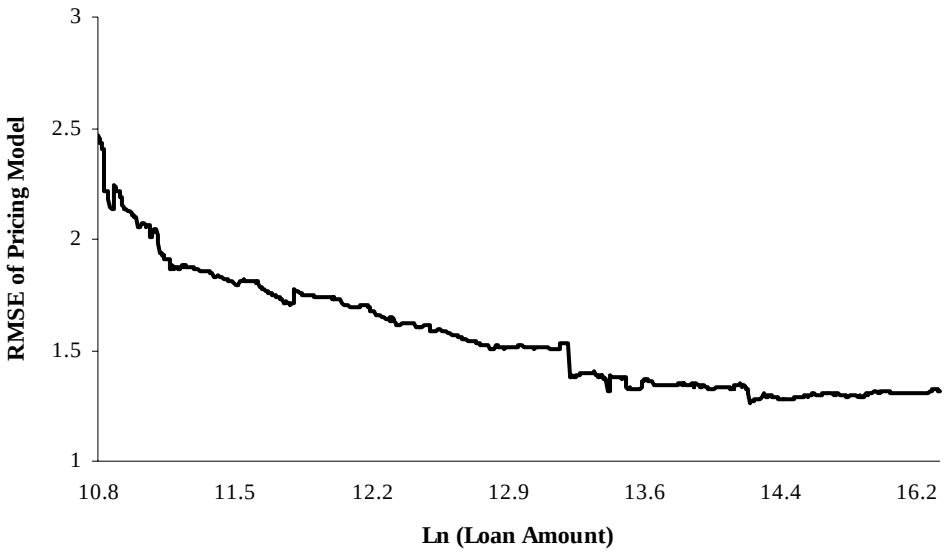

Table 1 – Descriptive Statistics, 1993 NSSBF

The table defines the variables employed in the empirical specifications and provides some sample statistics: mean, median (Med), standard deviation (S.d.), minimum (Min) and maximum (Max). The statistics do not take into account the 1993 NSSBF sample weights. The number of observations is 1,625. Other variables used in our regressions (dummies): type of loan (5), 2-digit SIC codes (9), year of loan (4), census regions (8) and type of lender (8). We do not present descriptive statistics for these variables to conserve on space.

Variable	Description	Mean	Med.	S.d.	Min	Max
Interest Rate Variables						
Prime Rate	Prime rate (%)	6.5	6	0.72	6	10
Loan Characteristics						
Loan Rate	Loan rate (%) [Dependent Variable]	8.4	8	2.1	0	32
Loan Amount	Loan size ($million)	1.1	0.1	4.7	0.00	100
Loan Maturity	Loan maturity (years)	3.3	1	4.4	0.08	30
Collateral	=1 if loan is collateralized	0.72	1		0	1
Floating	=1 if floating loan rate	0.59	1		0	1
Firm Characteristics						
Organization Type						
Proprietorship	=1 if firm is a proprietorship	0.18	0		0	1
Partnership	=1 if firm is a partnership	0.07	0		0	1
S-Corporation	=1 if firm is a S-corporation	0.29	0		0	1
Regular Corporation	=1 if firm is a regular corporation	0.46	0		0	1
Owner Characteristics						
Owner's Age	Age of firm's primary owner at start of relation with lender	42	41	11	2	81
Minority	=1 if firm is owned by minority race	0.09	0		0	1
Firm Accounting Information						
Assets	Total assets ($million)	33	7	96	0.008	2,400
Sales	= sales / assets	4.4	2.7	6.9	0	92
Profits	=profits / assets	0.41	0.1	2.3	-31	43

Table 1 (cont.)

Variable	Description	Mean	Med.	S.d.	Min	Max
Inventories	= inventories / assets	0.18	0.1	0.23	0	1
Accounts Receivable	= accounts receivable / assets	0.22	0.16	0.22	0	1.1
Accounts Payable	= accounts payable / assets	0.15	0.09	0.2	0	3.5
Total Loans	= total loans / assets	0.46	0.33	0.75	0	20
Leverage	= total liabilities / assets	0.7	0.58	0.89	0.00	21
Trade Credit Use	% of purchases on trade credit	62	80	41	0	100
Firm Credit History						
Bankrupt	=1 if firm declared bankruptcy in past 7 years	0.02	0		0	1
Owner Delinquent	=1 if owner has been 60 or more days delinquent on personal obligations within past 3 years	0.07	0		0	1
Firm Delinquent	=1 if firm has been 60 or more days delinquent on business obligations within past 3 years	0.17	0		0	1
Judgments	=1 if any judgments have been rendered against the principal owner within the past 3 years	0.04	0		0	1
Clean Record	=1 if owner has clean record (the four previous variables equal zero)	0.78	1		0	1
IRS Problem	=1 if firm had IRS problems or penalties in the past year	0.34	0		0	1
Relationship Characteristics						
Duration	Duration of relationship with the lender (years)	7.9	5	7.7	0	53
Main Bank	=1 if loan granted by primary bank	0.73	1		0	1
Personal	=1 if firm mainly conducts business with lender in person	0.65	1		0	1
Competition/Location Measures						
Concentrated	=1 if HHI>1800 (firm's headquarters office)	0.5	1		0	1
MSA	=1 if firm located in MSA	0.77	1		0	1
Distance	Distance to lender (miles)	53	3	213	0	2,608

Table 2
Results of Heteroscedastic Regression with 1993 NSSBF

The table lists the coefficients and standard errors (S.e.) for the mean equation (Column I) and variance equation (Column II) from a heteroscedastic regression with *Loan Rate* (in basis points) as the dependent variable. The explanatory variables in both the mean and variance equation, which we describe in Table 1, are demeaned. We obtain the estimates by Maximum Likelihood. The symbols *, ** and *** denote significance at the 10, 5 and 1% level, respectively. Some measures of fit are provided. χ^2 (k) is the statistic for the joint test that all coefficients except the intercept are zero, for k degrees of freedom. Pseudo-R^2 is defined as 1-Likelihood(just a constant)/Likelihood (full model). The VWLS (variance-weighted least squares) R^2 is obtained by using the inverse of the estimated variances in the heteroscedastic model as weights in the corresponding linear regression model.

Variable	I – Mean Equation		II – Variance Equation	
	β	S.e.	γ	S.e.
Constant	842.2***	4.2	10.10***	0.04
Interest Rate Variables				
Prime Rate	50.8***	7.0		
Loan characteristics				
Ln(Loan Amount)	-10.1***	3.4	-0.27***	0.02
Ln (Loan Maturity)	0.2	4.4		
Collateral	31.4***	9.2	-0.18**	0.08
Floating Rate	-33.6***	8.8		
Firm characteristics				
Corporation			-0.24***	0.09
Owner characteristics				
Ln(Owner's Age)	8.6	9.6	0.39***	0.13
Minority	32.6*	17.1	0.34***	0.13
Accounting information				
Ln(Assets)	-14.2***	3.7		
Sales	-1.3*	0.8		
Profits	2.0	2.1		
Inventories	-25.6	19.0		
Accounts Receivable	-13.3	19.5		
Accounts Payable	15.3	29.2		
Total Loans	-4.6	17.4		
Total Debt	16.3	16.2		
Trade Credit Use (%)	-0.1	0.1		
Credit history				
Bankrupt	-2.6	30.7		
Owner Delinquent	77.2***	20.3		
Firm Delinquent	15.1	11.6		
Judgments	24.4	23.0		

Table 2 (cont.)

Variable	I – Mean Equation		II – Variance Equation	
	β	S.e.	γ	S.e.
Clean Record			-0.25***	0.09
IRS Problem	7.7	7.8	0.16**	0.07
Relationship characteristics				
Ln(Duration)	-2.4	5.2	-0.12**	0.05
Main Bank	15.7	10.1		
Personal	5.5	8.5		
Competition/Location				
Concentrated	5.9	7.8	0.10	0.08
MSA	4.7	9.6	0.18**	0.09
Ln(Distance)	3.2	2.8	0.10***	0.02
Other Controls				
Loan Type (5)	Yes**			
Firm Organization Type (3)	Yes**			
SIC Codes (9)	Yes			
Loan Approval Year (4)	Yes**			
Regions (8)	Yes**			
Lender Type (8)	Yes***			
Number of observations		1,625		
Number of parameters		74		
χ^2-Statistic		980.4		
Pseudo-R^2 (%)		4.46		
VWLS R^2 (%)		28.26		

Table 3
Analysis of Extreme Residuals

The table lists the marginal effects (in %) from fitting a logit model to estimate the probabilities of observing a large positive residual (*Rip-off*) and a large negative residual (*Bargain*). The residuals originate from a quantile regression of *Loan Rate* on the set of variables presented in Table 1. *Rip-offs* (*Bargains*) correspond to residuals one standard deviation above (below) the sample mean. Bootstrapped standard errors (1000 replications) are provided. The symbols *, ** and *** denote significance at the 10, 5 and 1% level, respectively. The dataset used is the 1993 NSSBF.

Variable	**Prob. of Rip-off**		**Prob. of Bargain**	
	Mfx (%)	**S.e.**	**Mfx (%)**	**S.e.**
Loan Characteristics				
Ln(Loan Amount)	-2.2***	0.3	-1.4***	0.3
Collateral	-0.7	1.4	0.3	1.3
Firm Characteristics				
Corporation	-2.2	1.6	-1.3	1.5
Owner Characteristics				
Minority	2.3	2.4	3.4	2.3
Ln(Owner's Age)	-0.5	2.1	0.8	2.4
Credit History				
Clean Record	1.0	1.4	0.0	1.4
IRS Problem	3.0**	1.4	1.6	1.3
Relationship Characteristics				
Strength of Relationship	0.4	0.8	-1.6**	0.8
Competition/Location Measures				
Concentrated	-1.4	1.3	-1.0	1.2
MSA	0.0	1.6	-1.4	1.7
Ln(Distance)	0.9**	0.4	1.1***	0.3
Number of Observations	1,625		1,625	
Pseudo-R^2 (%)	7.3		7.1	

Table 4
Descriptive Statistics – (N)SSBF 1993, 1998 and 2003

The table provides some sample statistics (mean, median and standard deviation) for the 1993 NSSBF, 1998 SSBF and 2003 SSBF. The number of observations is 1625 for the 1993 NSSBF, 708 for the 1998 SSBF and 1568 for the 2003 SSBF. We remit to Table 1 the definition of the variables.

Variable	Mean			Median			Standard Deviation		
	1993	1998	2003	1993	1998	2003	1993	1998	2003
Interest Rate Variables									
Prime Rate	6.52	8.22	4.39	6	8.25	4.22	0.72	0.33	0.74
Loan Characteristics									
Loan Rate	8.4	9	5.7	8	9	5.5	2.1	2.2	2.5
Ln(Loan Amount)	12	11	12	12	11	12	2.1	1.9	1.9
Ln(Loan Maturity)	0.6	1	1.2	0	1.1	0.95	1	1.2	0.72
Collateral	0.72	0.65	0.56	1	1	1			
Floating Rate	0.59	0.33	0.55	1	0	1			
Loan Type									
Line of credit	0.6	0.3	0.62	1	0	1			
Capital lease	0.02	0.06	0.01	0	0	0			
Mortgage	0.09	0.12	0.10	0	0	0			
Vehicle loan	0.08	0.18	0.11	0	0	0			
Equipment loan	0.09	0.19	0.10	0	0	0			
Other	0.12	0.15	0.06	0	0	0			
Firm Characteristics									
Organization Type									
Proprietorship	0.17	0.26	0.16	0	0	0			
Partnership	0.07	0.07	0.07	0	0	0			
S-Corporation	0.29	0.34	0.45	0	0	0			
Corporation	0.46	0.34	0.32	0	0	0			
Owner Characteristics									
Ln(Owner's Age)	3.7	3.7	3.7	3.7	3.8	3.8	0.3	0.31	0.36
Minority	0.09	0.10	0.06	0	0	0			
Accounting Information									
Ln(Assets)	13	13	14	13	13	14	2.1	2.2	2.1
Sales	4.6	78	5.3	2.7	2.8	2.8	9.5	1,880	15
Profits	0.35	68	0.66	0.10	0.16	0.12	3	1,766	4.6
Inventories	0.18	0.16	0.16	0.08	0.04	0.03	0.23	0.23	0.22
Accounts Receivable	0.22	0.20	0.23	0.16	0.12	0.14	0.22	0.24	0.25

Table 4 (cont.)

Variable	Mean			Median			Standard Deviation		
	1993	1998	2003	1993	1998	2003	1993	1998	2003
Accounts Payable	0.15	0.32	0.14	0.09	0.07	0.06	0.20	2.96	0.29
Total loans	0.46	1.4	0.79	0.33	0.37	0.34	0.75	13	3.8
Leverage	0.7	1.8	1	0.58	0.68	0.60	0.89	14	3.9
Trade Credit Use (%)	62	61	63	80	80	80	41	41	40
Credit history									
Bankrupt	0.02	0.01	0.01	0	0	0			
Owner Delinquent	0.07	0.10	0.06	0	0	0			
Firm Delinquent	0.17	0.20	0.19	0	0	0			
Judgments	0.04	0.02	0.02	0	0	0			
Clean Record	0.78	0.76	0.77	1	1	1			
Relationship Characteristics									
Ln(Duration)	1.9	1.4	2	1.8	1.4	2.1	0.84	0.92	0.87
Main Bank	0.73	0.57	0.74	1	1	1			
Personal	0.65	0.55	0.61	1	1	1			
Competition/Location									
Concentrated	0.50	0.53	0.47	1	1	0			
MSA	0.77	0.74	0.77	1	1	1			
Ln(Distance)	1.9	2.4	2	1.4	1.8	1.6	1.6	2.1	1.6

Table 5
Results of Heteroscedastic Regression with Panel Sample

The table lists the coefficients (Coeff.) and standard errors (S.e.) for the variance equation from a heteroscedastic regression of *Loan Rate* on the set of variables described in Table 1. We obtain the panel sample by merging the 1993, 1998 and 2003 (N)SSBF samples. *Year* is the year in which the loan was granted. We remit to Table 1 the description of the remaining variables. We obtain the estimates by Maximum Likelihood. The symbols *, ** and *** denote significance at the 10, 5 and 1% level, respectively. We provide some measures of fit. Pseudo-R^2 is defined as 1-Likelihood(just a constant)/Likelihood (full model). The VWLS (variance-weighted least squares) R^2 is obtained by using the inverse of the estimated variances in the heteroscedastic model as weights in the corresponding regression model.

Variable	Panel Sample (I)		Panel Sample (II)		Panel Sample (III)	
	Coeff.	S.e.	Coeff.	S.e.	Coeff.	S.e.
Constant	10.45***	0.02	10.42***	0.02	10.4***	0.02
Loan Characteristics						
Ln(Loan Amount)	-0.20***	0.01	-0.23***	0.01	-0.22***	0.01
Collateral	-0.39***	0.05	-0.27***	0.05	-0.30***	0.05
Firm/Owner Characteristics						
Corporation	-0.11**	0.06	-0.13**	0.06	-0.14**	0.06
Minority	0.45***	0.08	0.46***	0.08	0.41***	0.09
Ln(Owner's Age)	0.15*	0.08	0.10	0.08	0.12	0.08
Clean Record	-0.05	0.05	-0.09	0.05	-0.07	0.05
Relationship Characteristics						
Ln(Duration)	0.00	0.03	-0.05	0.03	-0.06**	0.03
Competition/Location						
Concentrated	0.15***	0.05	0.14***	0.05	0.12**	0.05
MSA	0.23***	0.06	0.22***	0.06	0.24***	0.06
Ln(Distance)	0.15***	0.01	0.14***	0.01	0.13***	0.01
Temporal Variables						
Prime Rate			-0.06***	0.02	-0.03	0.02
Year (Time Trend)			0.04***	0.01	-0.08	0.08
Interaction Terms						
Year × Ln(Loan Amount)					0.01***	0.00
Year × Collateral					-0.03**	0.01
Year × Corporation					0.01	0.01
Year × Minority					0.02	0.02
Year × Ln(Owner's Age)					0.00	0.02
Year × Clean Record					0.06***	0.01
Year × Ln(Duration)					0.01	0.01
Year × Concentrated					0.02**	0.01
Year × MSA					0.03**	0.01
Year × Ln(Distance)					0.01*	0.00
Year × Prime Rate					-0.01**	0.01
Number of observations (N)	3,901		3,901		3,901	
Number of parameters	85		87		98	
Pseudo-R2 (%)	5.4		5.6		5.7	
VWLS-R2 (%)	49.8		48.6		46.7	