Two-Sided Matching and Spread Determinants in the Loan Market: An Empirical Analysis

Jiawei Chen*

January 13, 2005

Abstract

Participants on each side of the business loan market (banks and borrowing firms) rank participants on the other side based on their quality, and prefer to be matched with a partner that has a higher rank. As a result, who matches with whom is endogenously determined. Furthermore, it is shown that because of the endogenous matching, the regressors in the loan spread equation are correlated with the error term, so the OLS estimates are biased. To obtain unbiased estimates, a two-sided matching model in the loan market is developed to supplement the loan spread equation. In this model there is vertical heterogeneity on both sides of the market, and the matching of banks and firms is determined by the unique equilibrium outcome of a two-sided matching game. Bayesian inference is feasible using a Gibbs sampling algorithm that performs Markov chain Monte Carlo simulations. The empirical analysis uses a sample of 1369 U.S. loan facilities from 1996 to 2003. We find evidence of positive assortative matching of sizes in the market, and show that for agents on both sides of the market there are similar relationships between quality and size, which explains the positive assortative matching of sizes. We also find that banks' risk and firms' risk significantly affect their respective quality as viewed by the other side of the market. Bayesian estimates for the loan spread equation are markedly different from the simple OLS estimates, confirming that bias results from ignoring the two-sided matching process in the market.

*Department of Economics, Johns Hopkins University, Baltimore, MD 21218, USA. E-mail: jiawei@jhu.edu. I am grateful to my advisors Joseph Harrington and Matthew Shum for their encouragement and support. I also thank Ali Khan, Robert Moffitt, Tiemen Woutersen and seminar participants at Johns Hopkins University for their helpful comments and suggestions.

1 Introduction

In the business loan market, banks choose firms, firms choose banks, and who matches with whom is endogenously determined. That's because participants on each side of the market rank participants on the other side based on their quality, and prefer to be matched with one that has a higher rank.

To see that, let's first consider firms' choices. When a bank lends to a firm, the bank supplies credit to the firm. But much more than that, it also provides expert monitoring and endorsement effects based on reputation (see, for example, Diamond, 1984 and 1991), and there is ample empirical evidence that these "by-products" are important for firms. For instance, Billet, Flannery and Garfinkel (1995) and Johnson (1997) show that banks' monitoring ability and reputation have significant positive effects on borrowers' performance in the stock market.

The size of a bank, i.e., the amount of its total assets, plays an important role in firms' choices. For example, a firm cares about its lender's size because it has a strong influence on the firm's market value (Preece and Mullineaux, 1996). Furthermore, the size of a bank may place a constraint on its lending. If the firm is large, a small bank may not be able to provide the funds that the firm needs.[1] Even if the firm is currently small, as it grows this lending limitation can become binding, and this is taken into account when the firm is choosing a bank.[2] The size of a bank is also a proxy for its reputation, making it an important factor in the bank's quality.

On the other hand, large banks usually have more organizational layers and are more geographically dispersed than small banks, so they are generally less effective in processing and transmitting borrower information, especially "soft information", such as the leadership of the firm management and the morale of the staff. According to Williamson's (1967, 1988) theory of hierarchical control, the more complex is an organization, the more likely it is to have an information distortion problem be-

[1] Banks typically prefer well diversified loan portfolios, and would not lend to any single borrower a significant portion of their loanable funds.

[2] Absent switching costs, a small firm that grows large can simply switch to a larger bank if it feels its needs can not be satisfied. But borrowers face switching costs in changing lenders, so firms care about their banks' size even when they are small. See Hubbard, Kuttner and Palia (2002) for a discussion on such switching costs.

tween its successive hierarchies. Therefore, large banks typically face a more severe information distortion problem than small banks.[3] The complex and hierarchical organizational form of large banks also means that their officers will have less "formal authority" (Aghion and Tirole, 1997), so they will have weaker initiative or incentive to acquire information. What's more, Brickley, Linck and Smith (2003) observe that officers in small to medium-sized banks own higher percentages of their banks' stocks than their counterparts in large banks do. This creates stronger ownership incentive so these officers will devote more effort to the collection and processing of borrower information. They are also more likely to use this information in a way that best represents shareholders' interest. As a result, small to medium-sized banks have an advantage over large banks in their officers' information incentive. This shows the size of a bank has different effects on its quality as perceived by firms, and whether large banks or small banks are more attractive is determined by the net effect.

Since the benefits firms receive from banks are associated with banks' characteristics, firms generally prefer top-tier banks, i.e., banks with higher monitoring ability, better reputation, suitable size, and so on. Banks differ in these characteristics, and are ranked according to a composite index that combines them.

Now let's turn to banks' choices. At the same time when firms are ranking banks, they are themselves being ranked by banks. In making their lending decisions, loan officers in a bank typically credit screen the applicants (firms) and assess their credit worthiness. Only those who are considered credit worthy are granted loans (see, for example, Hauswald and Marquez (2002) for credit screening and Besanko and Thakor (1987) and Bhattacharya and Thakor (1992) for credit rationing). Firms with low leverage ratios (total debt/total assets) or high current ratios (current assets/current liabilities) are usually considered to be less risky and more credit worthy. Large firms also have an advantage here, because they generally have higher repaying ability, are more diversified in their assets, and are more likely to have well-

[3] It should be noted that when it comes to the processing of borrowers' "hard information", such as the information contained in their income statements, balance sheets, and credit ratings, large banks might do as well as or even better than small banks. See Berger et al. (forthcoming) for a discussion on this issue.

documented track record that can be used to convince the loan officers that they are credit worthy.

However, the large size of a firm can also have negative effects on its attractiveness. Because large firms have stronger financial needs, the loan made to a large firm usually has a larger amount and accounts for a higher percentage of the bank's assets, thus reducing the bank's diversification. Since banks prefer well diversified portfolios, the large size of the borrowing firm might be considered as unattractive. In addition, lending to a large firm means that the bank's control over the firm's investment decisions will be relatively small, which is undesirable.[4] As a result, the size of a firm has different effects on its quality as perceived by banks, and which size is most attractive depends on the relative magnitudes of these effects.

The above analysis shows there is endogenous two-sided matching in the loan market: banks choose firms, firms choose banks, and they prefer partners with a higher quality. As a result, firms with higher quality tend to match with banks with higher quality, and firms with lower quality tend to match with banks with lower quality.

Bias arises if we estimate the spread equation ignoring this matching process. To illustrate this let's assume for a moment that a bank's quality as viewed by borrowers is solely determined by its liquidity risk, and that a firm's quality as viewed by lenders is solely determined by its information costs.[5] Further assume that banks' liquidity risk, firms' information costs, and non-price loan characteristics such as maturity, loan size, and whether or not the loan is secured are determinants of loan spreads.[6] So the spread equation can be written as:

$$r_{ij} = \kappa L_i + \lambda I_j + N_{ij}'\alpha^N + \nu_{ij}, \tag{1}$$

where r_{ij} is the loan spread if bank i lends to firm j, L_i is bank i's quality (liquidity risk), I_j is firm j's quality (information costs), and N_{ij} is the non-price loan characteristics.

[4]See Rajan (1992) for a discussion on banks' control over borrowers' investment decisions.

[5]In reality, the notions of banks' and firms' qualities are multidimensional, but that does not change the conclusion that bias may arise if the matching process is not taken into account.

[6]The effects of banks' characteristics, firms' characteristics, and non-price loan characteristics on loan spreads are analyzed in Coleman, Esho and Sharpe (2002) and Hubbard, Kuttner and Palia (2002).

Since liquidity risk and information costs are not observed, the bank's ratio of cash to total assets and the firm's ratio of property, plant, and equipment (PP&E) to total assets are used as their proxies, respectively. Assume

$$L_i = \rho C_i + \eta_i, \text{ and} \tag{2}$$

$$I_j = \sigma P_j + \delta_j, \tag{3}$$

where C_i is bank i's ratio of cash to total assets, and P_j is firm j's ratio of PP&E to total assets. Now (1) becomes

$$r_{ij} = \kappa(\rho C_i + \eta_i) + \lambda(\sigma P_j + \delta_j) + N'_{ij}\alpha^N + \nu_{ij} \tag{4}$$

$$= \kappa\rho C_i + \lambda\sigma P_j + N'_{ij}\alpha^N + \kappa\eta_i + \lambda\delta_j + \nu_{ij}. \tag{5}$$

As a result, the error term in the spread equation contains η_i and δ_j, which affect the matching process through their impacts on the banks' quality (liquidity risk) and the firms' quality (information costs). Later in this paper it will become clear that the regressors in the loan spread equation are correlated with the error term, so OLS estimates of the loan spread equation are biased.[7]

In order to obtain unbiased estimates of the loan spread determinants, a many-to-one two-sided matching model in the loan market is developed and estimated in this paper. The model is a special case of the College Admissions Model, for which an equilibrium matching always exists (Roth and Sotomayor, 1990). Two-sided matching models prove to be a powerful tool in the analysis of various markets in which agents are divided into two sides and each chooses a partner or partners from the other side, such as the labor market, the marriage market, and the education market. There have also been a few studies on two-sided matching in financial markets. Sorensen (2003) applies a two-sided matching model to study the influence of venture capitalists on the companies in which they invest. Fernando, Gatchev, and Spindt (2004) study the matching between firms and their underwriters.

Bayesian inference about the parameters in the model is obtained using a Gibbs sampling algorithm that performs Markov chain Monte Carlo (MCMC) simulations.

[7]The endogeneity problem that comes from the use of proxies in the matching setting is recognized in the literature. For example, Ackerberg and Botticini (2002) describe the endogeneity problem introduced by the use of proxies in analyzing the matching process associated with agricultural contracts.

This empirical method is developed and utilized by Gelfand and Smith (1990), Albert and Chib (1993), and Geweke, Gowrisankaran, and Town (2003).

Sorensen (2003) is the first study that uses this method to estimate a two-sided matching model, and is the study closest to the present one. With the assumption of aligned preferences, i.e., the assumption that the valuations (joint surpluses) of the potential pairs of venture capitalists and companies determine the preferences for the agents on both sides of the market, he is able to guarantee the uniqueness of the equilibrium matching. He then proceeds to estimate the model using a Gibbs sampling algorithm, exploiting the fact that the unique equilibrium matching can be characterized by a set of inequalities regarding the valuations of all potential pairs. In the present study, the uniqueness of the equilibrium matching is obtained through the assumption that there is vertical heterogeneity on both sides of the loan market, and the unique equilibrium matching is characterized by a set of inequalities regarding the banks' and the firms' quality indexes.

The assumption of vertical heterogeneity means all banks have identical preference orderings over the firms and all firms have identical preference orderings over the banks, i.e., there is perfect sorting on both sides of the market. This ignores horizontal heterogeneity and precludes the possibility of a substantial joint surplus that can not be attributed to either agent's characteristics and is derived from some pair-specific characteristics, such as the bank's expertise in the borrower's industry and the distance between the agents' headquarters. If a substantial joint surplus of this type is introduced in the model and is not split between the pair according to a fixed parameter, then there might be multiple equilibria. We consider it an interesting avenue for future research to explore the effects of this type of joint surplus.

In Sorensen (2003), the error terms in the valuation equation are assumed to be independent random variables conditional on the exogenous variables. In the loan market, the fact that banks' and firms' unobserved quality is part of the error terms in the joint surplus equation means these error terms are correlated. Consequently, we deal with the error terms in banks' and firms' quality index equations directly and assume them to be independent random variables conditional on the exogenous variables. This treatment of the error terms enables us to estimate the effects of the agents' unobserved quality on the loan spreads.

Our empirical analysis uses a sample of 1369 U.S. loan facilities that involve 146 banks and 1007 firms from 1996 to 2003. We combine three datasets to obtain the necessary information for estimation. Information on loan characteristics comes from the DealScan database produced by Loan Pricing Corporation. Information on bank characteristics comes from the Reports of Condition and Income (the Call Reports) provided by the Federal Reserve Board. Finally, information on firm characteristics comes from the Compustat database, a product of Standard & Poor.

The empirical analysis finds that positive assortative matching of sizes is prevalent in the loan market: large banks tend to match with large firms and small banks tend to match with small firms. We then show that for agents on both sides of the market there are similar relationships between quality and size, which leads to similar size rankings for both sides and explains the positive assortative matching of sizes. Furthermore, it is found that banks' risk and firms' risk significantly affect their respective quality as viewed by the other side of the market. It is shown that because of the endogenous matching, the regressors in the loan spread equation are correlated with the error term, so the OLS estimates are biased. Bayesian estimates for the loan spread equation are markedly different from the simple OLS estimates, confirming that bias results from ignoring the endogenous two-sided matching process in the market.

The directions of the bias are as expected and are discussed in Section 5. In a nutshell, firms' unobserved quality is negatively correlated with the error term in the loan spread, so if a bank characteristic positively affects banks' quality, the OLS estimate of the coefficient on this characteristic in the loan spread equation will be biased downward, and vice versa. On the other hand, banks' unobserved quality is positively correlated with the error term in the loan spread, so if a firm characteristic positively affects firms' quality, the OLS estimate of the coefficient on this characteristic in the loan spread equation will be biased upward, and vice versa.

The remainder of the paper is organized as follows: Section 2 provides the specification of the model, Section 3 presents the empirical method for Bayesian inference, Section 4 describes the data used in the study, Section 5 presents and interprets the empirical results, and Section 6 concludes.

2 Model

The first component of our model is a spread equation, in which the loan spread is a function of the bank's characteristics, the firm's characteristics, and the non-price characteristics of the loan. A two-sided matching model in the loan market supplements the spread equation to permit non-random matching of banks and firms.

2.1 Spread Equation

We are interested in estimating the following spread equation:

$$r_{ij} = B_i'\alpha_1 + F_j'\alpha_2 + N_{ij}'\alpha_3 + \epsilon_{ij} \equiv U_{ij}'\alpha + \epsilon_{ij}, \tag{6}$$

where r_{ij} is the loan spread if bank i lends to firm j, B_i is a vector of bank i's characteristics that are proxies for its quality, F_j is a vector of firm j's characteristics that are proxies for its quality, N_{ij} represents the non-price loan characteristics, and ϵ_{ij} represents $N(0, \sigma_\epsilon^2)$ random variables that are independent conditional on the exogenous variables.[8] Equation 6 is the multi-dimensional version of equation 5, with B_i', F_j', α_1, α_2, α_3, and ϵ_{ij} corresponding to C_i, P_j, $\kappa\rho$, $\lambda\sigma$, α^N, and $\kappa\eta_i + \lambda\delta_j + \nu_{ij}$, respectively.

Prior studies show that the bank's monitoring ability and risk, as well as the firm's risk and information costs are important determinants of the loan spread.[9] Since these aspects of the bank and the firm are not observed, we include a number of proxies for them in the spread equation.

Bank's Monitoring Ability. According to the hold-up theory described in Rajan (1992) and Diamond and Rajan (2000), a bank that has superior monitoring ability can use its skills to extract higher rents. Moreover, Leland and Pyle (1977),

[8] Hubbard, Kuttner and Palia (2002) suggest that (6) can be interpreted as a reduced form equation that results from the interaction of loan demand and loan supply, with B_i and F_j representing exogenous shifters.

[9] See, for example, Coleman, Esho and Sharpe (2002) and Hubbard, Kuttner and Palia (2002).

Diamond (1984, 1991) and Allen (1990) show that banks' monitoring plays a unique role in firms' operation and provides value to them. Therefore, *ceteris paribus*, we expect a bank that has higher monitoring ability to charge a higher spread.

A bank's salaries-expenses ratio, defined as the ratio of salaries and benefits to total operating expenses, is included in B as a proxy for its monitoring ability. Coleman, Esho and Sharpe (2002) argue that monitoring activities are relatively labor-intensive, and salaries reflect the staff's ability and performance in these activities. So a higher salaries-expenses ratio indicates a higher share of resources invested in the monitoring activities, and/or more competent monitoring staff.[10]

Bank's Risk. A bank's risk comes from two sources: inadequate capital and low liquidity. Hubbard, Kuttner and Palia (2002) suggest that a lower capital-assets ratio reduces a bank's ability to extract repayment, therefore lowering the recovery rate in default and forcing the bank to charge higher loan spreads. Furthermore, a bank that has a higher liquidity (or a lower liquidity risk) is better able to meet the credit or cash needs of its borrowers, so *ceteris paribus*, it will charge a higher spread.

A bank's capital-assets ratio is included in B as a proxy for its capital adequacy, whereas its ratio of cash to total assets is included as a proxy for its liquidity risk. The size of a bank (its total assets) is also a proxy for its risk, since a larger bank is likely to be better diversified and have a lower risk (see Coleman, Esho and Sharpe, 2002).

Firm's Risk. Proxies for a firm's risk include the leverage ratio (total debt/total assets), the current ratio (current assets/current liabilities), and the size of the firm.

Risk is positively associated with the leverage ratio, so a firm that has a higher leverage ratio will be charged a higher spread, all else being equal. The implication of the current ratio is just the opposite. Firms with higher current ratios are considered less risky, and typically get a lower spread on their loans. Furthermore, due to the

[10]If a bank adopts more advanced information technologies, the size of its staff can be reduced. However, for labor intensive activities such as monitoring, the decrease in the number of staff members is not expected to be significant. Furthermore, the salary rates for the staff members that are proficient in the new technologies are likely to be higher, so the impact of a bank's IT development on its total salaries is unclear.

diversification effects of increasing firm size, firm risk is negatively associated with firm assets, and a larger firm can usually get a loan with a lower spread.

Firm's Information Costs. It is commonly accepted that smaller firms pose larger information asymmetries and are associated with higher information costs, because they typically lack well-documented track records. So the size of a firm, which has been included above, is also a proxy for information costs.

Another proxy for a firm's information costs is the ratio of property, plant, and equipment (PP&E) to total assets, which indicates the relative significance of tangible assets in the firm. A firm that has relatively more tangible assets poses smaller information asymmetries. Consequently it can borrow at a lower spread, all else being equal.

Non-Price Loan Characteristics. Non-price loan characteristics are included on the RHS of the spread equation as control variables. They are maturity (in months), natural log of the loan facility size, purpose dummies such as "acquisition" and "recapitalization", type dummies such as "a revolver credit line with duration shorter than one year", and a secured status dummy. The construction of these variables are presented in Section 4.

2.2 Two-Sided Matching Model

If we have a random sample of bank-firm matches and observe the loan spreads for the matched pairs, then estimating (6) using OLS would be just fine, since ϵ_{ij} represents $N(0, \sigma_\epsilon^2)$ random variables that are independent conditional on the exogenous variables. However, as argued in the introduction, the sample we have is not a random sample and the bank-firm matching process is correlated with the spread equation. To obtain unbiased estimates, a two-sided matching model is developed to supplement the spread equation and address the sample selection problem resulting from the non-random matching between banks and firms:

$$r_{ij} = B_i'\alpha_1 + F_j'\alpha_2 + N_{ij}'\alpha_3 + \epsilon_{ij} \equiv U_{ij}'\alpha + \epsilon_{ij}, \ r_{ij}, \ N_{ij} \text{ observed iff } H_{ij} = 1 \quad (7)$$

$$H_i^b = B_i'\beta + \eta_i, \quad (8)$$

$$H_j^f = F_j'\gamma + \delta_j, \quad (9)$$

$$H_{ij} = \begin{cases} 1 & \text{if bank } i \text{ lends to firm } j, \\ 0 & \text{otherwise,} \end{cases} \quad (10)$$

where H_i^b is the quality index of bank i, H_j^f is the quality index of firm j, and H_{ij} is the choice indicator. η_i and δ_j are mean zero random variables that are independent conditional on the exogenous variables, and are allowed to be correlated with ϵ_{ij}.

In the two-sided matching model, whether H_{ij} equals one or zero is determined by both banks' choices and firms' choices, and the outcome corresponds to the unique equilibrium matching (defined below), which depends on the H_i^b's and the H_j^f's.

It is worth noting that in the loan market, the two-sided matching process among banks and firms takes place before the loan spreads are determined. For example, it is documented in Miller and Bavaria (2003) and Yago and McCarthy (2004) that during the second half of the 1990s, market-flex language became common in the business loan market, which allows the pricing of a loan to be determined after the loan agreement is made. During the matching process, banks and firms do not know exactly what the loan spreads would be, but they do take into account the expectation of the spreads. We will return to this issue when we discuss banks' and firms' preferences.

Agents, Quotas and Matches. Let I_t and J_t denote, respectively, the sets of banks and firms in market t, where $t = 1, 2, ...T$. I_t and J_t are finite and disjoint. In what follows, the market subscript t is sometimes dropped to simplify the notation.

In the empirical implementation of this model, a market is specified to contain all the firms that borrow during the same half-year and the banks that lend to them. During a half-year, it is reasonable to assume that a borrowing firm borrows only once, since in such a short period of time it is likely that its financial needs can be satisfied by a single loan, whereas borrowing multiple loans would substantially increase the administrative costs, such as the costs associated with the negotiation process.[11]

On the other hand, a bank sometimes lends to multiple firms during a half-year period. A bank's lending activity is restricted in two ways. First, loan assessment, approval, monitoring, and review processes are relatively labor-intensive, and a bank's lending activity is restricted by the amount of resources that is available for these processes, e.g., the number of its loan officers.[12] Consequently, the number of loans that a bank can make during a given half-year is limited.[13]

Second, the total amount of loans a bank can make may be constrained by the availability of insured deposits, the primary source of funds for bank lending (Jayaratne and Morgan, 2000). Moreover, according to the 1988 Basel Capital Accord, banks are required to hold a minimum overall capital equal to 8% of their risk-weighted assets, with all business loans included in the full weight category (Repullo and Suarez, forthcoming). Assuming a bank's overall capital is fixed in a given half-year, this also sets an upper bound on the total amount of loans a bank can make.

Jayaratne and Morgan (2000) find evidence that the deposit constraint on bank lending operates only on small banks whose assets are less than $100 million, while the larger banks appear to be unconstrained. In the dataset used in the present study, these small banks account for less than 1% of all banks. Furthermore, banks' capital-assets ratios have been substantially increased since the early 1990's. For example, Coleman, Esho and Sharpe (2002) compare the loan dataset (1987-1992) used in Hubbard, Kuttner and Palia (2002) to the loan dataset (1994-1999) used in their own study and find that banks' capitalization is significantly improved in the second period. What's more, in Coleman, Esho and Sharpe's study, the average of banks' capital-assets ratios is 7.2%, whereas in the loan dataset (1996-2003) examined by the present study, that average is 8.6%. This shows that banks'

[11] In the case of multiple banks participating in the same loan, the lead arranger is considered as the lending bank. We return to this issue in Section 4.

[12] See Cole, Goldberg and White (2001) for a description on the loan approval process, and Blackwell and Winters (1997) for a discussion on the loan review process.

[13] In the long run, the limit on the number of loans that a bank can make during a half-year can change, since the bank can hire or lay off loan officers if needed.

capitalization has been further improved, and that the lending constraint that comes from inadequate capital becomes less of a concern. Therefore, in the present study the limit on the total amount of loans is assumed to be non-binding, whereas the limit on the number of loans is assumed to be binding. These assumptions simplify the empirical implementation and make the model tractable. The possible impacts of these assumptions on the empirical results are discussed in Sections 5 and 6.

We assume in market t, bank i can lend to q_{it} firms, whereas firm j can borrow from only one bank, so the model is a special case of the many-to-one two-sided matching model, also known as the College Admissions Model (Roth and Sotomayor, 1990). q_{it} is known as the *quota* of bank i in the matching literature, and we assume in an equilibrium (defined below) each bank uses up its quota.

The set of all potential loans (known as *matches* in the matching literature) is given by $M_t = I_t \times J_t$. A *matching*, μ_t, is a set of matches such that $(i, j) \in \mu_t \Longleftrightarrow$bank i and firm j are matched in market t.

Let $\mu_t(i)$ denote the set of firms that borrow from bank i in market t, and let $\mu_t(j)$ denote the set of banks that lend to firm j in market t, which is a singleton. We then have

$$(i, j) \in \mu_t \Longleftrightarrow j \in \mu_t(i) \Longleftrightarrow i \in \mu_t(j) \Longleftrightarrow \{i\} = \mu_t(j) \tag{11}$$

Preferences. The matching of banks and firms is determined by the equilibrium outcome of a two-sided matching game. In such a game, the payoff firm j receives if it borrows from bank i is H_i^b, whereas the payoff bank i receives if it lends to the firms in the set $\mu_t(i)$ is $\sum_{j\in\mu_t(i)} H_j^f$, i.e., the quality indexes of firms are assumed to be *additive.*[14] Consequently, each bank prefers firm j to firm j' iff $H_j^f > H_{j'}^f$, and each firm prefers bank i to bank i' iff $H_i^b > H_{i'}^b$. Furthermore, the quality indexes are assumed to be distinct so there are no "ties" and agents will never be indifferent between relevant loans.

Note that at the time when the agents choose partners, they do not know exactly what the loan spread will be, so they take into account the expectation of the

[14]The additivity of firms' quality indexes implies that banks' preferences over sets of firms are *responsive*, i.e., for any two sets that differ in only one firm, a bank prefers the set containing the more preferred firm. See Roth and Sotomayor (1990) for a discussion on the responsiveness assumption.

spread, which is a function of the characteristics of the agents. Assuming that the bank's characteristics and the firm's characteristics are linear and separable in the expected spread, this consideration will be reflected in the quality indexes, and so these indexes can be thought of as the "spread-adjusted" quality indexes.

Under the above specification there is vertical heterogeneity on both sides of the loan market: all banks have identical preference orderings over the firms and all firms have identical preference orderings over the banks, i.e., each side of the market has a common ranking over the other side of the market. Vertical heterogeneity is commonly assumed in economic applications. For example, Wong (2003) assumes that in the marriage market, men and women are ranked by the other side of the market based on their marriage indexes, which are determined by their wage and education. Therefore, all women have a common preference ordering over men, and all men have a common preference ordering over women. Other examples of vertical heterogeneity include: the market for lawyers in which they are ranked according to their quality as determined by their experience, academic performance, and the quality of their law schools (Spurr, 1987); the market for workers in which they are ranked according to their productivity (Oi, 1983); and so on. Vertical heterogeneity on both sides of the loan market guarantees that the equilibrium matching is unique. This issue will be discussed below.

Notice that the joint surplus for the pair of bank i and firm j is

$$\begin{aligned} s_{ij} &= H_i^b + H_j^f \\ &= B_i'\beta + F_j'\gamma + \eta_i + \delta_j \\ &= B_i'\beta + F_j'\gamma + \omega_{ij}. \end{aligned} \tag{12}$$

Two features of the joint surplus are worth mentioning. First, the error term ω_{ij} consists of η_i and δ_j. As a result, $cov(\omega_{ij}, \omega_{ij'}) \neq 0$ and $cov(\omega_{ij}, \omega_{i'j}) \neq 0$, $\forall i \neq i'$, $j \neq j'$. I.e., the ω_{ij}'s are not independent variables. This means in order to write down the likelihood function in the estimation stage, we should deal with η_i and δ_j directly, rather than deal with their sum ω_{ij}.

Second, we have assumed that the source of any part of the joint surplus is either the bank or the firm. In a more general model, the joint surplus would include an additional component, one that can not be attributed to either agent's characteristics. Let $c_{ij} = c(W_{ij}) + \zeta_{ij}$ denote this component, where W_{ij} represents pair-specific characteristics, such as the bank's expertise in the borrower's industry and the distance between the agents' headquarters.[15] The division of c_{ij} between the pair would be endogenous, i.e., bank i would receive $\lambda_{ij} c_{ij}$ and firm j would receive $(1 - \lambda_{ij}) c_{ij}$, where $\lambda_{ij} \in [0, 1]$ is endogenously determined. For instance, λ_{ij} could be determined by the bargaining process between the bank and the firm. As long as the magnitudes of the c_{ij}'s are not large enough to change the preference orderings, we would still have vertical heterogeneity on both sides of the market.[16] For example, if all quality indexes are distinct integers while the c_{ij}'s are at most 0.5 in absolute value, then the preference orderings are entirely determined by the quality indexes.

Equilibrium Matching. A matching is considered an equilibrium if it is stable, i.e., if there is no blocking coalition of agents. A coalition of agents is blocking if they prefer to deviate from the current matching and form new matches among them. The set of equilibria corresponds to the core of a cooperative game.

Formally, μ_t is an equilibrium matching in market t iff there does not exist $\tilde{I} \subset I_t$, $\tilde{J} \subset J_t$ and $\tilde{\mu}_t \neq \mu_t$ such that $\tilde{\mu}_t(i) \subset \tilde{J} \cup \mu_t(i)$ and $\sum_{j \in \tilde{\mu}_t(i)} H_j^f > \sum_{j \in \mu_t(i)} H_j^f$ for all $i \in \tilde{I}$, and $\tilde{\mu}_t(j) \in \tilde{I}$ and $H^b_{\tilde{\mu}_t(j)} > H^b_{\mu_t(j)}$ for all $j \in \tilde{J}$.

The stability concept used in the above definition is group stability. A related stability concept is pair-wise stability. A matching is pair-wise stable if there is no blocking bank-firm pair. For the College Admissions Model, Roth and Sotomayor (1990) prove that pair-wise stability is equivalent to group stability, and that an equilibrium always exists. Since the present model is a special case of the College Admissions Model, these results carry over to the present model. Furthermore, Eeckhout (2000, Corollary 3) shows that in a one-to-one two-sided matching model, the equilibrium matching is unique if there is vertical heterogeneity on both sides of the market. Appendix A shows that this sufficient condition for uniqueness also applies to many-to-one two-sided matching models. Since in the present model this condition is satisfied, the equilibrium matching is unique.

[15] See Stomper (2004) for a discussion on banks' industry expertise and Coval and Moskowitz (2001) for a discussion on the importance of physical distance for information gathering.

[16] So the equilibrium matching would still be unique. See below.

Similar to Sorensen (2003), the unique equilibrium matching here can be characterized by a set of inequalities. These inequalities are constructed based on the fact that there is no blocking bank-firm pair for the equilibrium matching. Consider an arbitrary matching in market t, μ_t. Suppose bank i and firm j are not matched in μ_t. (i,j) is a blocking pair iff $H_j^f > \min_{j' \in \mu_t(i)} H_{j'}^f$ and $H_i^b > H_{\mu_t(j)}^b$. So (i,j) is not a blocking pair iff $H_j^f < \min_{j' \in \mu_t(i)} H_{j'}^f$ or $H_i^b < H_{\mu_t(j)}^b$. Equivalently, (i,j) is not a blocking pair iff $H_j^f < \overline{H}_{ji}^f$ and $H_i^b < \overline{H}_{ij}^b$, where

$$\overline{H}_{ji}^f = \begin{cases} \min_{j' \in \mu_t(i)} H_{j'}^f & if\ H_i^b > H_{\mu_t(j)}^b \\ \infty & otherwise, \end{cases} \tag{13}$$

and

$$\overline{H}_{ij}^b = \begin{cases} H_{\mu_t(j)}^b & if\ H_j^f > \min_{j' \in \mu_t(i)} H_{j'}^f \\ \infty & otherwise. \end{cases} \tag{14}$$

Now suppose bank i and firm j are matched in μ_t. Bank i or firm j is part of a blocking pair iff $H_j^f < \max_{j' \in f(i)} H_{j'}^f$ or $H_i^b < \max_{i' \in f(j)} H_{i'}^b$, where $f(i)$ is the set of firms that do not currently borrow from bank i but would prefer to do so, and $f(j)$ is the set of banks that do not currently lend to firm j but would prefer to do so. These two sets contain the feasible deviations of the agents, and are given by

$$f(i) = \{j \in J_t \backslash \mu_t(i) : H_i^b > H_{\mu_t(j)}^b\}, \text{ and} \tag{15}$$

$$f(j) = \{i \in I_t \backslash \mu_t(j) : H_j^f > \min_{j' \in \mu_t(i)} H_{j'}^f\}. \tag{16}$$

Therefore, neither bank i nor firm j is part of a blocking pair iff $H_j^f > \underline{H}_{ji}^f$ and $H_i^b > \underline{H}_{ij}^b$, where $\underline{H}_{ji}^f = \max_{j' \in f(i)} H_{j'}^f$ and $\underline{H}_{ij}^b = \max_{i' \in f(j)} H_{i'}^b$.

Let μ_t^e denote the (unique) equilibrium matching in market t. The above analysis leads to the following characterization of the equilibrium matching:

$$\mu_t = \mu_t^e \iff H_i^b \in (\underline{H}_i^b, \overline{H}_i^b), \forall i \in I_t \text{ and } H_j^f \in (\underline{H}_j^f, \overline{H}_j^f), \forall j \in J_t, \tag{17}$$

where

$$\underline{H}_i^b = \max_{j \in \mu_t(i)} \underline{H}_{ij}^b, \tag{18}$$

$$\overline{H}_i^b = \min_{j \notin \mu_t(i)} \overline{H}_{ij}^b, \tag{19}$$

$$\underline{H}_j^f = \underline{H}_{j,\mu_t(j)}^f, \text{ and} \tag{20}$$

$$\overline{H}_j^f = \min_{i \notin \mu_t(j)} \overline{H}_{ji}^f. \tag{21}$$

This characterization of the equilibrium matching will be used in the Bayesian inference method described in Section 3.

3 Empirical Method

Two-sided matching in the loan market presents numerical challenges when it comes to estimation. Maximum likelihood estimation requires integrating a highly nonlinear function over thousands of dimensions, which can not be accomplished analytically. Several simulation based methods are possible candidates for inference, which include the simulated maximum likelihood estimator, the method of simulated moments estimator, and Bayesian inference using the Gibbs sampler. Geweke, Keane and Runkle (1994) show that Bayesian inference based on the Gibbs sampler outperforms the other methods. The estimation of the present two-sided matching model in the loan market uses this method.

Based on the distributions of the error terms and the prior distributions of the parameters, we use Markov-chain Monte Carlo (MCMC) techniques to iteratively simulate the quality indexes conditional on the data and the parameters, and the parameters conditional on the data and the quality indexes. The objective is to recover the joint posterior distribution of the parameters and the quality indexes. The projection of the joint posterior distribution on the parameters gives the posterior distributions of the parameters, Bayesian inference is then derived from these posterior distributions.

The method described above has been applied to estimation in a number of studies. For example, Albert and Chib (1993) use it in a binary probit model, Geweke, Keane and Runkle (1997) use it in a multinomial probit model, Geweke, Gowrisankaran and Town (2003) extend it to a multinomial probit model with a dichotomous outcome variable, and Sorensen (2003) applies it to a two-sided matching model with a dichotomous outcome variable.

3.1 Distributions of Error Terms

In the quality index equations, β and σ_η, as well as γ and σ_δ, are jointly unidentified since scaling the parameters and the standard deviations by any positive constant leaves agents' preferences unaffected. Adding or subtracting a constant from H_i^b or H_j^f does not affect the preferences, either. Therefore, σ_η and σ_δ are set to one to fix the scale of the parameters, and the constant is excluded from the RHS of the quality index equations.

As discussed above, the endogeneity problem arises because of the use of proxies in the spread equation and in the quality index equations. Consequently the error terms in these equations are correlated. A convenient treatment of the correlation is to work with the population regression of ϵ_{ij} on η_i and δ_j:

$$\epsilon_{ij} = \kappa\eta_i + \lambda\delta_j + \nu_{ij}, \tag{22}$$

$$cov(\eta_i, \nu_{ij}) = 0, \tag{23}$$

$$cov(\delta_j, \nu_{ij}) = 0, \tag{24}$$

where ν_{ij} represents independent $N(0, \sigma_\nu^2)$ error terms conditional on the exogenous variables. This implies that $cov(\epsilon_{ij}, \eta_i) = \kappa$, $cov(\epsilon_{ij}, \delta_j) = \lambda$, and $\sigma_\epsilon^2 = \kappa^2 + \lambda^2 + \sigma_\nu^2$.

3.2 An Additional Assumption

If the signs of β, γ, κ and λ are all reversed, the outcome would be observationally equivalent to the original one. To see that, let's consider a new set of quality indexes $\tilde{H}_i^b = -H_i^b = B_i'(-\beta) - \eta_i$ and $\tilde{H}_j^f = -H_j^f = F_j'(-\gamma) - \delta_j$. We have $\epsilon_{ij} = (-\kappa)(-\eta_i) + (-\lambda)(-\delta_j) + \nu_{ij}$. This new set of quality indexes are the opposite of the original ones and the preference orderings are just upside down. Referring to the algorithm described in Appendix A that forms the unique equilibrium matching, we see that the new equilibrium matching will be exactly the same as the old one. That means two sets of parameters, (β, γ, κ, λ) and ($-\beta$, $-\gamma$, $-\kappa$, $-\lambda$), are both admissible given the same observed (endogenous and exogenous) variables, and the empirical method will not be able to distinguish between them.

Consequently, in order for the model to be identified, the sign of one of the parameters in β, γ, κ and λ must be specified. In the present study, this assumption is $\lambda < 0$, which is consistent with our prior belief that firms' unobserved risk or unobserved information costs are positively correlated with loan spreads, since higher risk or higher information costs correspond to lower firm quality. It should be noted that $\lambda < 0$ is not a restriction. If the truth is $\lambda > 0$, we would obtain a set of estimates of the quality index equations in which all the good factors have negative coefficients and all the bad factors have positive coefficients. The task of telling which of the two exactly opposite scenarios is reasonable is performed by the researcher, not the empirical method.

3.3 Prior Distributions

In addition to the normal distributions of error terms, we assume that all the parameters except σ_ν^2 have independent normal prior distributions (truncated normal prior distribution for λ) that include all reasonable parameter values well within their supports.

The prior distributions of α, β, γ, and κ are $N(\bar{\alpha}, \bar{\Sigma}_\alpha)$, $N(\bar{\beta}, \bar{\Sigma}_\beta)$, $N(\bar{\gamma}, \bar{\Sigma}_\gamma)$, and $N(\bar{\kappa}, \bar{\sigma}_\kappa^2)$, respectively, whereas the prior distribution of λ is $N(\bar{\lambda}, \bar{\sigma}_\lambda^2)$ truncated from above at 0. In the estimation the means of these prior distributions are zero vectors (scalar 0 for $\bar{\kappa}$ and $\bar{\lambda}$), and the variance-covariance matrices are $10I$, where I is a corresponding identity matrix (scalar 1 for $\bar{\sigma}_\kappa^2$ and $\bar{\sigma}_\lambda^2$). These normal prior distributions imply that the conditional posterior distributions of these parameters will also be normal (truncated normal for λ). For any parameter, the variance of the prior distribution is at least 233 times the variance of the posterior distribution, which shows the information contained in the Bayesian inference is substantial.[17]

The prior distribution of $1/\sigma_\nu^2$ is gamma,[18] $1/\sigma_\nu^2 \sim G(a, b)$, $a, b > 0$, i.e.,

$$p(\frac{1}{\sigma_\nu^2} \mid a, b) = \frac{1}{\Gamma(a) b^a} (\frac{1}{\sigma_\nu^2})^{a-1} \exp(-\frac{1}{b\sigma_\nu^2}). \tag{25}$$

In the estimation $a = 2.5$ and $b = 1$. This is again a diffuse prior.

[17] Since we do not have strong prior beliefs about the parameters, diffuse priors are chosen. McCulloch and Rossi (1994) show that for diffuse priors the choice of the locations and the variances of the priors is not critical. For the present study, we tried larger variances and other changes in the prior distributions, and the estimates are left virtually unchanged.

[18] Thus the prior distribution of σ_ν^2 is inverted gamma. See Casella and Berger (2002) for details. The gamma distribution or its multivariate generalization, the Wishart distribution, are common choices of prior distribution for the inverse of a variance or a variance-covariance matrix. See, for example, McCulloch and Rossi (1994).

3.4 Conditional Posterior Distributions

In the model, the exogenous variables are B_i, F_j, and N_{ij}, which are abbreviated as X_{ij}. The observed endogenous variables are r_{ij} (the loan spread) and H_{ij} (the choice indicator). The quality indexes are H_i^b and H_j^f. The parameters are α, β, γ, κ, λ, and $1/\sigma_\nu^2$, which are abbreviated as θ.

In market t, let X_t, r_t, μ_t and H_t^* represent the above variables, where μ_t embodies all the H_{ij}'s in the market and H_t^* denotes all the (banks' and firms') quality indexes in the market.

We obtain the joint density of the endogenous variables and the quality indexes conditional on the exogenous variables and the parameters as follows:

$$
\begin{aligned}
p_t(r_t, \mu_t, H_t^* \mid X_t, \theta) &= \mathbf{1}[\mu_t = \mu_t^e \mid H_t^*] \\
\times \prod\nolimits_{(i,j)\in\mu_t} &\left\{ \frac{1}{\sqrt{2\pi\sigma_\nu^2}} \exp\left[-\frac{(r_{ij} - U_{ij}'\alpha - \kappa(H_i^b - B_i'\beta) - \lambda(H_j^f - F_j'\gamma))^2}{2\sigma_\nu^2} \right] \right\} \\
\times \prod\nolimits_{i\in I_t} &\left\{ \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(H_i^b - B_i'\beta)^2}{2} \right] \right\} \times \prod\nolimits_{j\in J_t} \left\{ \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(H_j^f - F_j'\gamma)^2}{2} \right] \right\},
\end{aligned}
\tag{26}
$$

where $\mathbf{1}[.]$ is the indicator function and $\mu_t = \mu_t^e$ indicates that μ_t is the equilibrium matching, which is characterized by (17). The product of these densities for $t = 1, 2, ..., T$ gives the joint density $p(r, \mu, H^* \mid X, \theta)$ for the variables in all the markets.

From Bayes' rule, the density of the posterior distribution of H^* and θ conditional on the data (the data augmented conditional posterior distribution) is

$$
\begin{aligned}
p(H^*, \theta \quad | \quad & X, r, \mu) = p(\theta) \times p(r, \mu, H^* \mid X, \theta)/p(r, \mu \mid X) \\
&= C \times p(\theta) \times p(r, \mu, H^* \mid X, \theta) \\
&= C \times p(\theta) \times \prod_{t=1}^{T} p_t(r_t, \mu_t, H_t^* \mid X_t, \theta),
\end{aligned}
\tag{27}
$$

where $p(\theta)$ represents the prior densities of the parameters and $p(r, \mu \mid X)$ is the probability of observing r and μ conditional on the exogenous variables. Here and below C (and C') is a generic normalization constant.

Successful application of the Gibbs sampling algorithm requires simple conditional posterior distributions of the quality indexes and the parameters from which random numbers can be generated at low computational cost. These conditional posterior distributions are obtained by examining the kernel of the conditional posterior densities and are described in Appendix B. It is shown that the conditional posterior distributions are truncated univariate normal for H_i^b, H_j^f, and λ, multivariate normal for α, β, and γ, univariate normal for κ, and gamma for $1/\sigma_\nu^2$.

3.5 Gibbs Sampling Algorithm

Our objective is to draw Bayesian inference about the parameters by approximating their respective posterior distributions. This is achieved using a Gibbs sampling algorithm (Geman and Geman, 1984; Gelfand and Smith, 1990; Geweke, 1999) with data augmentation (Tanner and Wong, 1987; Albert and Chib, 1993).

In the algorithm, the parameters and the quality indexes are partitioned into blocks. First, each of the parameter vectors α, β, γ, κ, λ, and $1/\sigma_\nu^2$ is a block. In addition, each quality index is a block. In market t the number of quality indexes is equal to the number of agents, which is $|I_t| + |J_t|$. So together we have $\sum_{t=1}^{T}(|I_t| + |J_t|) + 6$ blocks.

In each iteration of the algorithm, each block is simulated conditional on all the others according to the data augmented conditional posterior distribution (equation 27) described in the last subsection. This method relies on the fact that under weak conditions widely satisfied by econometric models (including this one), if we cyclically simulate from each of the conditional distributions, the resulting sequence of draws converge in distribution to the joint distribution.[19] For example, in order to draw from the joint distribution of $\theta = (\theta_1, \theta_2)$ with only the knowledge of the conditional distributions of $\theta_1 \mid \theta_2$ and $\theta_2 \mid \theta_1$, we could start from any $(\theta_1^0,\, \theta_2^0)$ in the support of the joint distribution and draw cyclically from the conditional distributions. The sequence $\{(\theta_1^1 \mid \theta_2^0,\, \theta_2^1 \mid \theta_1^1),\, (\theta_1^2 \mid \theta_2^1,\, \theta_2^2 \mid \theta_1^2),\, (\theta_1^3 \mid \theta_2^2,\, \theta_2^3 \mid \theta_1^3),\, ...\}$ converges in distribution to the joint distribution of θ, where $\theta_1^1 \mid \theta_2^0$ denotes a draw from the conditional distribution of $\theta_1 \mid \theta_2^0$, and so on.[20] Since the draws converge to a unique stationary Markov chain, this method is a Markov chain Monte Carlo

[19]A sufficient condition for convergence is that the chain is defined by a strictly positive transition kernel, such as the product of truncated normal, normal, and gamma distributions in the present model, so that the chain is irreducible and aperiodic, which guarantees that it has a unique stationary distribution. See, for example, Tierney (1991) and McCulloch and Rossi (1994).

[20]This example is discussed in McCulloch and Rossi (1994).

(MCMC) method.[21]

After discarding a number of initial draws to allow the distributions to "burn-in", the rest of the draws are used for Bayesian inference. The posterior means of the parameters are approximated by $\frac{1}{m}\sum_{l=1}^{m}\theta^m$, where θ^m is the m'th draw from the conditional posterior distributions, and the posterior standard deviations are approximated by $\left[\frac{1}{m}\sum_{l=1}^{m}\left(\theta^m - \frac{1}{m}\sum_{l=1}^{m}\theta^m\right)^2\right]^{1/2}$. The quality indexes are also simulated in each iteration, but their simulated values are only used to obtain the conditional posterior distributions of the other quality indexes and the parameters.

The conditional posterior distributions are multivariate normal for α, β, and γ, univariate normal for κ, and gamma for $1/\sigma_\nu^2$. Simulations from these distributions are straightforward using Matlab. The conditional posterior distributions for the quality indexes and λ are truncated univariate normal. We use inverse c.d.f. sampling algorithm to draw from such distributions. If $\phi \sim N(\mu_\phi, \sigma_\phi^2)$ truncated to $(\underline{\phi}, \overline{\phi})$, then $\tilde{\phi} = \Phi^{-1}(u)$ is an appropriate random draw, where $u \sim U(\Phi(\underline{\phi}), \Phi(\overline{\phi}))$ and Φ is the c.d.f. of $N(\mu_\phi, \sigma_\phi^2)$. See Devroye (1986) for details on inverse c.d.f. sampling.

In our study, the starting values for the quality indexes are zeros, and the starting values for the parameters are the modes of their prior distributions. Different starting values have a minimal impact on the estimation results.

4 Data

We obtain the data used in this study from three sources. Information on loans (spread and non-price characteristics) comes from the DealScan database produced by Loan Pricing Corporation. To obtain information on bank characteristics, we match the banks in DealScan to those in the Reports of Condition and Income (the Call Reports) from the Federal Reserve Board. To obtain information on firm characteristics, we match the firms in DealScan to those in the Compustat database, a product of Standard & Poor.

[21]The key reference for general state space Markov chains is Nummelin (1984). Tierney (1991) contains an excellent discussion on Gibbs sampling and other related Markov chain methods.

4.1 Sample Construction

The DealScan database contains detailed information on over 131,000 loan facilities and high-yield bonds worldwide dating back to 1988. Its primary source of data is Securities and Exchange Commission filings, although data from large loan syndicators and the Loan Pricing Corporation's own staff of reporters are also collected. For each loan facility, DealScan reports pricing information (interests and fees), information on non-price loan characteristics such as maturity, secured status, the purpose of the loan, and the type of the loan, as well as the identities of the borrower and the lender(s).

We restrict our attention to loan facilities between U.S. banks and U.S. firms from 1996 to 2003. These loan facilities constitute our sample and are divided into sixteen markets, with each market containing all the lending banks and all the borrowing firms in the same half year: January to June or July to December.

Data on banks' and firms' characteristics are from the quarter prior to the beginning of the market. For example, for a loan facility in the first half year of 1996, we obtain bank data and firm data from the Call Report and the Compustat database corresponding to the fourth quarter of 1995. For a loan facility in the second half year of 1996, data from the second quarter of 1996 are used.

A loan facility has to satisfy the following criteria to be included in the sample:

1. Data on basic loan characteristics, such as spread, maturity, and secured status, are not missing in the DealScan database.

2. If there is more than one lender, one lead arranger is specified.[22]

3. The firm borrowed only once in this market.

4. The lending bank can be matched to a bank in the Call Report of the relevant quarter. We lose the observation if no matching bank name can be found or multiple banks with the same name are found.

[22] When there are multiple lenders, the characteristics of the lead arranger are the most relevant for our analysis. For example, Angbazo, Mei and Saunders (1998) document that in syndicated loans, the administrative, monitoring, and contract enforcement responsibilities lie primarily with the lead arranger.

5. Data on basic bank characteristics, such as total assets and total equity capital, are not missing in the Call Reports.

6. The borrowing firm can be matched to a firm in the Compustat database.[23]

7. Data on basic firm characteristics, such as total assets, total debt, and current liabilities, are not missing in the Compustat database.

The remaining sample consists of 1369 loan facilities associated with 146 banks and 1007 firms.[24] Figure 1 contains plots of the number of banks and the number of firms in each market. It is shown that the number of banks in each market is relatively stable, while the number of firms exhibits a slightly upward trend. Note that the number of firms in each market is also the number of loan facilities in each market, since in our sample each firm borrowed only once in the relevant market.

4.2 Variables

Corresponding to the two components of the model (the spread equation and the two-sided matching model) there are two observed endogenous variables: the loan spread and the matching of banks and firms.

Consistent with prior studies, we use the All-In Spread Drawn, or AIS, reported in the DealScan database as the loan spread. The AIS is expressed as a markup over the London Interbank Offering Rate (LIBOR). It equals the number of basis points (a basis point is 1/100 of a percentage point) in the coupon spread plus the annual fee plus any one-time fee divided by the loan maturity. When the coupon spread is priced off another benchmark interest rate, the following differentials are used by the Loan Pricing Corporation to convert the coupon spread into LIBOR terms: Prime=255 basis points, Cost of Funds=0 basis points, Commercial Paper=3 basis points, T-Bills=-34 basis points, Federal Funds=0 basis points, Money Market Rate=0 basis points, Bankers' Acceptance=-18 basis points, and CDs=-6 basis points. Since a number of exogenous variables used in our analysis are expressed

[23]The matching processes in 4 and 6 are carried out conservatively in the sense that a loan facility is included in the sample only when the names can be matched unambiguously.

[24]Some banks and some firms participated in more than one market. The numbers of banks in each market add up to 455, and the numbers of firms in each market add up to 1369.

in percentage points, we divide the AIS by 100 to obtain r_{ijt}. Figure 2 plots the weighted average loan spread in percentage points for each market.

The matching of banks and firms in each market, μ_t, is given by the matched agents' names recorded in our loan facilities data.

The RHS of the spread equation includes a constant, year dummies, and three groups of exogenous variables: bank characteristics, firm characteristics, and non-price loan characteristics.

In the first group, a bank's salaries-expenses ratio, defined as the ratio of salaries and benefits to total operating expenses, is used as a proxy for the bank's monitoring ability. A bank's capital-assets ratio, defined as the ratio of total equity capital to total assets, is used as a proxy for the banks's capital adequacy. Furthermore, a bank's ratio of cash to total assets is included as a proxy for its liquidity risk. Finally, we also include four size dummies as proxies for a bank's risk: $bank\_size2$, $bank\_size3$, $bank\_size4$, and $bank\_size5$. $bank\_size2_i$ equals 1 if bank i has \$5 billion to \$13 billion assets (roughly the 20th and the 40th percentiles), and equals 0 otherwise. $bank\_size3_i$ equals 1 if bank i has \$13 billion to \$32 billion assets (roughly the 40th and the 60th percentiles), and equals 0 otherwise. $bank\_size4$, and $bank\_size5$ are defined analogously, with the cut-off point at \$76 billion (roughly the 80th percentile).

In the second group, we include a firm's leverage ratio (total debt/total assets), current ratio (current assets/current liabilities), and four size dummies, $firm\_size2$, $firm\_size3$, $firm\_size4$, and $firm\_size5$, as proxies for the firm's risk. $firm\_size2 \sim 5$ are defined in a way analogous to $bank\_size2 \sim 5$, with cut-off points at \$65 million, \$200 million, \$500 million, and \$1,500 million.[25] The ratio of property, plant, and equipment (PP&E) to total assets is used as a proxy for the firm's information costs.

In the last group, we include such non-price loan characteristics as maturity (in months), natural log of facility size, purpose dummies, type dummies, and a secured status dummy. The 26 specific purposes reported in DealScan are combined into five categories: acquisition (acquisition lines and takeover), general (corporate purposes

[25] These size dummies enable us to detect the nonlinear relationships between bank/firm sizes and the loan spread. This will be discussed in section 5.

and working capital), miscellaneous (capital expenditure, equipment purchase, IPO related finance, mortgage warehouse, project finance, purchase hardware, real estate, securities purchase, spinoff, stock buyback, telecom build-out, and trade finance), recapitalization (debt repayment/debt consolidation/refinancing and recapitalization), and other (all other purposes). Consequently, we have four purpose dummies, each representing one of the first four categories above. In addition, there are three categories of types: revolver/line < 1 year (i.e., a revolving credit line whose duration is less than one year), revolver/line ≥ 1 year (i.e., a revolving credit line whose duration is greater than or equal to one year), and other (all other types). As a result, we include two type dummies, each indicating one of the first two categories. A secured status dummy is also included, which equals one if the loan facility is secured and equals zero otherwise.

The RHS variables in the bank quality index equation are the bank characteristics, whereas the RHS variables in the firm quality index equation are the firm characteristics. Neither equation includes a constant on its RHS because adding or subtracting a constant from the quality indexes does not affect the agents' preferences.

Nominal variables used in this study (Bayesian inference and OLS) that are not expressed as a ratio (bank assets, firm assets, and facility size) are deflated using the GDP (Chained) Price Index (formerly GDP Deflator).[26] The year 2000 is used as the base year. All ratios are expressed in percentage points.

Table 1 provides the definitions and sources of the variables. Summary statistics of these variables are presented in Table 2.

5 Findings

In this section, we first present evidence that indicates positive assortative matching of sizes is prevalent in the loan market: large banks tend to match with large firms and small banks tend to match with small firms. We then show that for agents on both sides of the market there are similar relationships between quality and

[26] Obtained from the Historical Tables in the Budget of the United States Government (Fiscal Year 2005) disseminated by the U.S. Government Printing Office.

size: after controlling for other factors, the medium-sized agents are regarded as having the highest quality, followed by the largest agents, while the smallest agents are regarded as having the lowest quality. This leads to similar size rankings for both sides, which explains the positive assortative matching of sizes. Furthermore, the empirical results suggest that agents' risk substantially affect their respective attractiveness as viewed by the other side of the market. Bayesian estimates for the loan spread equation are markedly different from the simple OLS estimates, confirming that bias results from ignoring the two-sided matching process in the market. Finally, the effects of various bank characteristics, firm characteristics, and non-price loan characteristics on loan spreads are examined.

5.1 Positive Assortative Matching of Sizes

It is recognized in the literature that large banks tend to lend to large firms, whereas small banks tend to lend to small firms. See, for example, Hubbard, Kuttner and Palia (2002) and Berger et al. (forthcoming). To verify this positive assortative matching of sizes, two OLS regressions using the matched pairs are run: the bank's size on the firm's characteristics and the firm's size on the bank's characteristics. The results are reported in Tables 3 and 4. It is shown that the bank's size and the firm's size are strongly positively correlated. The coefficient on the firm's size in the first regression and the coefficient on the bank's size in the second regression are both positive and have t statistics at about 20. This indicates that there is indeed positive assortative matching of sizes in the loan market.

Figure 3 provides further evidence. The figure depicts the proportion of loans for each combination of bank-firm size groups. For example, the height of the column at (2, 3) represents the proportion of loans that are between a bank that is in the second bank size group (total assets between the 20th and the 40th percentiles) and a firm that is in the third firm size group (total assets between the 40th and the 60th percentiles). A clear pattern is observed in the figure: the highest columns are mostly on the main diagonal (from (1,1) to (5,5)), whereas the columns far off the main diagonal (e.g., (1,5) and (5,1)) are rather short. The figure illustrates that most of the loans are between a bank and a firm that have similar size positions on their respective side: large banks are matched with large firms, and small banks are matched with small firms.

The next subsection explores the reasons for this observed pattern by looking at the Bayesian inference results for the quality index equations. Non-size factors that affect the agents' quality indexes are also examined.

5.2 Factors in the Quality Indexes

Table 5 reports the posterior means and standard deviations of the coefficients in the quality index equations (8 and 9).[27]

Sizes of the Agents. The first observation is that all of the bank size groups and the firm size groups have positive coefficients and most of them are significant. This indicates that on both sides of the market, the group of the smallest agents (with assets below the 20th percentile), which is the omitted group, is considered as the worst in terms of quality. On the lenders' side, this result is consistent with the hypothesis that the smallest banks suffer from severe lending constraints and low reputation that are associated with their small sizes. Even though they can better collect and process borrower information because they have less organizational layers and their staff have stronger information incentive, this advantage is clearly outweighed by the disadvantages and as a result, they are the least attractive partners.

On the borrowers' side, the positive coefficients of the size groups show that the smallest firms are considered as the least credit worthy. This supports the view that the smallest firms have the lowest repaying ability, are the least diversified in their assets, and lack well-documented track record to prove their credit worthiness. Even though lending to these smallest firms means the lenders can better diversify their assets and have more control over the borrowers' investment decisions, these benefits are not strong enough to make these smallest firms attractive.

A closer look at the coefficients further reveals that on both sides of the market, it is the medium-sized agents that have the highest quality. On the lenders' side, banks who have assets between the 40th and the 80th percentiles (group 3 and group 4)

[27]All Bayesian results reported (Tables 5 and 6) are based on 20,000 iterations from which the initial 2,000 draws are discarded according to convergence diagnostics. Using Matlab 6.5, these iterations took 52 hours on a computer running Windows XP with a 1.3 GHZ processor.

are the most attractive. On the borrowers' side, firms with assets between the 40th and the 60th percentiles (group 3) are regarded as having the highest quality. The groups of largest agents (with assets above the 80th percentile) are less attractive than the medium-sized agents, but are better than the smallest agents.

As the size of a bank increases, it has greater lending capacity and its reputation is higher, which makes it more attractive. On the other hand, large banks typically have more severe information distortion problems, and their loan officers have weaker incentive in collecting and processing borrower information. For the group of the largest banks, these negative effects outweigh the banks' advantage over the medium-sized banks in terms of their lending capacity and reputation.

Similarly, as the size of a firm increases, its repaying ability increases, its assets are more diversified, and it can better provide the information that is needed to convince the loan officers that it is credit worthy. However, the group of the largest firms are less attractive than the medium-sized firms because lending to the largest firms means the bank will be less diversified in its assets and its control over the firms' investment decisions will be much weaker, and these disadvantages of the largest firms dominate their advantages over the medium-sized firms.

It should be noted that the negative effect of a firm's large size on its quality is probably underestimated, since our model assumes that the limit on the number of loans a bank can make is binding whereas the limit on the total amount of loans is non-binding. If we take into account that sometimes the binding limit is on the total amount of loans, then lending to a large firm should be less attractive: the large size of the loan means that the bank would probably have to sacrifice a number of other lending opportunities in order to lend to this large firm, and the bank's diversification of assets would be further impaired.

Since the relationship between a bank's quality and its size is similar to the one between a firm's quality and its size, the size rankings for both sides of the loan market are quite similar: from the highest quality to the lowest quality, the size ranking is 4-3-2-5-1 for the banks and 3-4-2-5-1 for the firms, where the numbers represent the size groups. All else being equal, the medium-sized agents have higher quality than the largest agents, which in turn have higher quality than the smallest agents.

This finding explains the positive assortative matching of sizes: medium-sized banks lend to medium-sized firms because both groups are the top candidates on their respective sides. After these agents get matched, the remaining agents face restricted choice sets. Among these agents, the largest banks and the largest firms are the top candidates, so they are matched together. Finally, the smallest banks and the smallest firms are the lowest in quality, so they have no choice but to match with each other.[28]

Other Factors in the Quality Indexes. The effects of the non-size variables on the quality indexes are also reported in Table 5. We find that both banks' risk and firms' risk substantially affect the agents' quality as viewed by the other side of the market.

On the banks' side, the coefficient on the ratio of cash to total assets is positive and significant. This shows that firms prefer banks with high liquidity to those with low liquidity, reflecting the significantly negative impact of banks' liquidity risk on their quality. The coefficients on the salaries-expenses ratio and the capital-assets ratio are both positive, which is consistent with the hypothesis that banks with higher monitoring ability and/or higher capital adequacy are more attractive. These two coefficients are, however, insignificant, suggesting that in the current sample the influence of these two ratios on banks' quality is not substantial.

On the firms' side, the current ratio has a positive and significant coefficient. This finding supports the view that a firm's quality is negatively related to its risk, for which the current ratio is a proxy. Banks prefer to lend to firms with higher current ratios, because these are the firms that are less risky. The other two non-size variables both have the expected signs: the coefficient on the leverage ratio is negative, which indicates that firms with higher leverage ratios are riskier, so they are less attractive to the banks; the coefficient on the ratio of PP&E to total assets has a positive sign, suggesting that firms with relatively more tangible assets pose smaller information asymmetries and have higher quality. The fact that these coefficients are insignificant indicates that in the current sample these two ratios are not important concerns of banks when they rank the borrowers.

[28]In general, as long as the rankings of a certain attribute of the agents are similar for both sides of the market, we can expect to observe positive assortative matching of this attribute.

5.3 Loan Spread Determinants

κ, the covariance between the error term in the loan spread equation and the error term in the bank quality index equation, is found to be significant (Table 5). This constitutes strong evidence that the matching process is correlated with the loan spread determination and should not be ignored. We conclude that the H_{ij}'s are correlated with the ϵ_{ij}'s and that the estimates for the loan spread equation from the simple OLS regression that fails to take account of the two-sided matching process are biased. To see that, let's rewrite the spread equation as follows, noting that each firm borrows only once in a market:

$$r_j = H_j' B\alpha_1 + F_j'\alpha_2 + N_j'\alpha_3 + \epsilon_j \tag{28}$$

$$= H_j' B\alpha_1 + F_j'\alpha_2 + N_j'\alpha_3 + \kappa H_j'\eta + \lambda\delta_j + \nu_j, \tag{29}$$

where r_j is the spread that firm j pays, $H_j = (H_{1j}, H_{2j}, ..., H_{Ij})'$ is a vector of choice indicators, $B = (B_1, B_2, ...B_I)'$ is a matrix of bank characteristics, N_j is the non-price loan characteristics associated with the loan for firm j, $\eta = (\eta_1, \eta_2, ..., \eta_I)'$ is a vector of error terms in the bank quality index equation, and ν_j represents $N(0, \sigma_\nu^2)$ random variables that are independent conditional on the exogenous variables. If H_j were in fact independent of ϵ_j-as it would be if firms were randomly matched to banks-then H_j would be exogenous in (28), and the OLS estimates would be unbiased. However, a significant κ shows H_j is correlated with ϵ_j, thus the OLS estimates are biased.

A comparison between the Bayesian estimates (Table 6) and the OLS estimates (Table 7) of the loan spread equation confirms the existence of the bias. These two sets of results are markedly different: the average percentage bias, defined as the average of $\left|(\hat{\theta}_{OLS} - \hat{\theta}_{Bayesian})/\hat{\theta}_{Bayesian}\right|$ across all estimated coefficients, is 23%, where $\hat{\theta}_{OLS}$ represents the OLS estimates and $\hat{\theta}_{Bayesian}$ represents the Bayesian estimates. Below we first examine the directions of the bias, and then analyze three groups of loan spread determinants according to the Bayesian results.

Directions of the Bias. First, it is worth noting that in Table 5, κ, the covariance between the error term in the loan spread equation and the error term in the bank quality index equation, is positive, whereas λ, the covariance between the error term in the loan spread equation and the error term in the firm quality index equation, is negative.[29]

For banks, the unobserved quality has two components that operate in opposite directions in terms of their effects on the loan spread: unobserved monitoring ability and unobserved risk. If the first component dominates, then the unobserved bank quality will be positively correlated with the loan spread: banks with higher unobserved monitoring ability have higher unobserved quality and will charge higher loan spreads, all else being equal. On the other hand, if the second component dominates, then the unobserved bank quality will be negatively correlated with the loan spread: banks with lower unobserved risk have higher unobserved quality and will charge lower loan spreads, all else being equal. The positive sign of κ shows that the unobserved monitoring ability dominates the unobserved risk and is the main component in banks' unobserved quality. This result may reflect that the proxy for banks' monitoring ability (the salaries-expenses ratio) is not as good as the proxies for banks' risk.

For firms, the unobserved quality has two components that operate in the same direction in terms of their effects on the loan spread: unobserved risk and unobserved information costs. Firms with either higher unobserved risk or higher unobserved information costs have lower unobserved quality and are charged higher loan spreads, all else being equal. The negative sign of λ is consistent with this relationship.

Given the signs of κ and λ, the directions of the bias in the OLS estimates of the loan spread equation are as expected. There are five variables in the group of bank characteristics that have a significant impact on banks' quality: the ratio of cash to total assets and the four size group dummies. All these variables positively affect banks' quality. Now take the ratio of cash to total assets for example. Suppose all firms are identical except that they have different unobserved risk, and consider two banks that differ only in terms of their ratios of cash to total assets. The bank with a higher ratio has a higher quality, so it will be matched with a firm that has a lower unobserved risk. Since λ is negative, the lower unobserved risk of the firm means the spread that this bank charges will have a smaller unobserved component. This gives

[29]Recall that in Section 3, the sign of λ is assumed (to be negative) in order for the model to be identified.

the econometrician that performs the simple OLS regression a wrong perception and makes the OLS estimate biased downward. Similarly, the coefficients on the four bank size group dummies in the loan spread equation are all biased downward in the OLS regression.

In the firm quality index equation, three variables have significant coefficients: the current ratio, $firm\_size3$, and $firm\_size4$. All these variables positively affect firms' quality. Since κ is positive, by an argument analogous to the one presented above, the OLS estimates of the coefficients on these variables in the loan spread equation should all be biased upward. This is exactly what happens.

The following analysis of the loan spread determinants are based on the Bayesian estimates.

Bank Characteristics. Most of the estimated coefficients on the bank characteristics have the expected signs and many of them are statistically significant.

The first finding is that the salaries-expenses ratio has a positive and significant coefficient. This confirms that banks with superior monitoring ability indeed charge higher loan spreads, all else being equal.

The coefficients on the capital-assets ratio and the ratio of cash to total assets are both insignificant, reflecting that for the sample that is being analyzed, banks' capital adequacy risk and liquidity risk do not have a substantial impact on the loan spreads charged.

The coefficients on the bank size group dummies are all negative and most of them are significant, supporting the hypothesis that larger banks are likely to have better diversified assets and hence a lower risk, so they will charge lower spreads. Furthermore, these coefficients exhibit a downward trend, as expected. The coefficients on $bank\_size4$ and $bank\_size5$ are substantially lower than those on $bank\_size2$ and $bank\_size3$ (-0.27 and -0.28 versus -0.18 and -0.14). This shows that compared to the group of the smallest banks, banks with assets between the 20th and the 60th percentiles charge loan spreads that are lower by around 15 basis points, whereas banks with assets above the 60th percentile charge loan spreads that are lower by nearly 30 basis points.

Firm Characteristics. Among the three non-size firm characteristics, two have significant coefficients: the leverage ratio and the current ratio. The coefficient on the leverage ratio is positive, whereas the coefficient on the current ratio is negative. Since a higher leverage ratio represents a higher borrower risk and a higher current ratio represents a lower borrower risk, these results give support to the hypothesis that firms with higher risk are charged higher loan spreads.

The coefficient on the ratio of PP&E to total assets is insignificant, suggesting that it does not affect borrowers' costs of funds in a substantial way.

On the other hand, the firm size group dummies all have negative and significant coefficients. This is consistent with the hypothesis that large firms in general are less risky and are associated with lower information costs, so they are charged lower loan spreads. Moreover, these coefficients also exhibit a downward trend. For example, compared to firms in group 1 (firms with assets below the 20th percentile), firms in group 2 (firms with assets between the 20th and the 40th percentiles) are charged loan spreads that are lower by 22 basis points, whereas firms in group 5 (firms with assets above the 80th percentile) are charged loan spreads that are lower by 46 basis points.

Non-Price Loan Characteristics. The estimated coefficients on non-price loan characteristics reported in Table 6 show that only a small number of these characteristics have a substantial impact on the loan spreads. In fact, out of the nine variables in this group only three have statistically significant coefficients. They are the natural log of facility size, revolver/line>=1 year dummy, and secured dummy.

Strahan (1999) and Dennis, Nandy and Sharpe (2000) predict that loans with longer maturity are likely to be associated with borrowers that are less risky, so these loans would have lower spreads. Their prediction is not supported in the dataset being analyzed. The coefficient on maturity is insignificant. This result suggests that in the dataset, the link between borrowers' risk and maturity is weak. To test that, an OLS regression is run regressing maturity on firm characteristics (the regression results are not shown). The adjusted R-squared is merely 0.0560.[30]

[30]Using a DealScan loan dataset that covers the period from 1987 to 1992, Hubbard, Kuttner and Palia (2002) find a negative and significant coefficient on maturity. They consider this result as slightly anomalous, and attribute it to either a correlation between unobserved borrower characteristics and maturity or the presence of fixed up-front costs.

On the other hand, the negative and significant coefficient on the natural log of facility size indicates that there are, in a sense, economies of scale in bank lending. The processes of loan approval, monitoring, and review are relatively labor-intensive. After controlling for bank characteristics and firm characteristics, the labor costs associated with these processes do not increase proportionally when the size of the loan increases. As a result, the larger is the loan, the lower the per-dollar labor costs will be, hence the lower spread.

While none of the purpose dummies has a significant coefficient, one of the type dummies does appear to matter. The revolver/line>=1 year dummy has a negative coefficient that is significant at the 10% level. Since revolving credit lines whose durations are greater than or equal to one year are by far the most common loan type (accounting for 67% of all loans), this result may reflect that other types of loans are relatively nonstandard or even custom-made, and banks charge higher loan spreads on them in order to compensate for the extra administrative costs associated with their nonstandard nature.

Finally, consistent with the findings in Angbazo, Mei, and Saunders (1998), Strahan (1999), and Dennis, Nandy, and Sharpe (2000), we find a positive and significant coefficient on the secured dummy. Since it is the riskier borrowers who are required to secure their loans with collateral, this result suggests that these borrowers' higher firm-specific risk is not sufficiently reduced by the collateral so they are charged higher loan spreads.

6 Conclusion

This paper investigates two-sided matching and spread determinants in the loan market. Using a sample of 1369 U.S. loan facilities that involve 146 banks and 1007 firms from 1996 to 2003, we find evidence of positive assortative matching of sizes in the market: large banks tend to match with large firms and small banks tend to match with small firms. We then show that for agents on both sides of the market there are similar relationships between quality and size, which leads to similar size rankings for both sides and explains the positive assortative matching of sizes. We also find that banks' risk and firms' risk significantly affect their respective quality as viewed by the other side of the market.

Bayesian estimates for the loan spread equation are markedly different from the simple OLS estimates, confirming that bias results from ignoring the two-sided matching process in the market. Furthermore, it is found that banks with superior monitoring ability tend to charge higher loan spreads, while large banks tend to charge lower loan spreads. On the other hand, firms with higher risk are likely to be charged higher loan spreads, whereas large firms are generally charged lower spreads.

Two restrictive assumptions made in the present study imply avenues for future research. First, we have assumed that in a given market, the limit on a bank's total amount of loans is non-binding, whereas the limit on its number of loans is binding. While this assumption is not without reasonableness given the features of the time period being analyzed and makes the problem well suited to attack with a many-to-one two-sided matching model, it underestimates the effects of loan sizes on banks' lending activities. In particular, if the limit on the total amount of loans is binding, the size of the borrower and the size of the loan would have greater importance when a bank makes its lending decisions, since a loan made to a large firm is typically large which means the bank would probably have to sacrifice a number of other lending opportunities. It would then be desirable to have a more general model in which it is endogenously determined whether the binding limit is on the total amount of loans or on the number of loans.

Second, in the present study vertical heterogeneity on both sides of the market is assumed. This assumption guarantees the uniqueness of the equilibrium matching so that we can write down the joint density of the endogenous variables and the quality indexes, which is the basis of our estimation. This assumption, however, comes at the cost of precluding the possibility of a substantial joint surplus that can not be attributed to either agent's characteristics and is derived from some pair-specific characteristics, such as the bank's expertise in the borrower's industry and the distance between the agents' headquarters. Since such a joint surplus likely exists in the loan market, we consider it an interesting topic for future research to allow for it. If a substantial joint surplus of this type is introduced in the model and is not split between the pair according to a fixed parameter, then there might be multiple equilibria, even though the existence of an equilibrium is still guaranteed. In that case, more advanced empirical methods will be needed for the estimation.

References

[1] Ackerberg, D. A. and M. Botticini (2002), "Endogenous Matching and the Empirical Determinants of Contract Form", *Journal of Political Economy*, Vol. 110:3, 564-91.

[2] Aghion, P. and J. Tirole (1997), "Formal and Real Authority in Organizations", *Journal of Political Economy*, Vol. 105, 1-29.

[3] Albert, J. and S. Chib (1993), "Bayesian Analysis of Binary and Polychotomous Data, *Journal of the American Statistical Association*, Vol. 88, 669-79.

[4] Allen, F. (1990), "The Market for Information and the Origin of Financial Intermediation", *Journal of Financial Intermediation*, Vol. 1, 3-30.

[5] Angbazo, L., J. Mei and A. Saunders (1998), "Credit Spreads in the Market for Highly Leveraged Transaction Loans", *Journal of Banking and Finance*, Vol. 22, 1249-82.

[6] Berger, A. N., N. H. Miller, M. A. Peterson, R. G. Rajan and J. C. Stein (forthcoming), "Does Function Follow Organizational Form? Evidence from the Lending Practices of Large and Small Banks", *Journal of Financial Economics*, forthcoming.

[7] Besanko, D. and A. V. Thakor (1987), "Collateral and Rationing: Sorting Equilibria in Monopolistic and Competitive Credit Markets", *International Economic Review*, Vol. 28:3, 671-89.

[8] Bhattacharya, S. and A. V. Thakor (1992), "Contemporary Banking Theory", *Journal of Financial Intermediation*, Vol. 3, 2-50.

[9] Billet, M., M. Flannery and J. Garfinkel (1995), "The Effect of Lender Identity on a Borrowing Firm's Equity Return", *Journal of Finance*, Vol. 50:2, 699-718.

[10] Blackwell, D. W. and D. B. Winters (1997), "Banking Relationships and the Effect of Monitoring on Loan Pricing", *Journal of Financial Research*, Vol. 20:2, 275-89.

[11] Brickley, J. A., J. S. Linck and C. W. Smith (2003), "Boundaries of the Firm: Evidence from the Banking Industry", *Journal of Financial Economics*, Vol. 70, 351-83.

[12] Casella, G. and R. L. Berger (2002), *Statistical Inference*, 2nd Edition, Pacific Grove, CA : Thomson Learning.

[13] Cole, R. A., L. G. Goldberg and L. J. White (2001), "Cookie-Cutter versus Character: The Micro Structure of Small Business Lending by Large and Small Banks", mimeo, New York University.

[14] Coleman, A. D. F., N. Esho, and I. G. Sharpe (2002), "Do Bank Characteristics Influence Loan Contract Terms?", APRA working paper 2002-01.

[15] Coval, J. and T. Moskowitz (2001), "The Geography of Investment: Informed Trading and Asset Prices", *Journal of Political Economy*, Vol. 109:4, 811-41.

[16] Dennis, S., D. Nandy and I. G. Sharpe (2000), "The Determinants of Contract Terms in Bank Revolving Credit Agreements", *Journal of Financial and Quantitative Analysis*, Vol. 35:1, 87-110.

[17] Devroye, L. (1986), *Non-Uniform Random Variate Generation*, New York: Springer-Verlag.

[18] Diamond, D. (1984), "Financial Intermediation and Delegated Monitoring", *Review of Economic Studies*, Vol. 51, 393-414.

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

[20] Diamond, D. and R. Rajan (2000), "A Theory of Bank Capital", *Journal of Finance*, Vol. 55:6, 2431-65.

[21] Eeckhout, J. (2000), "On the Uniqueness of Stable Marriage Matchings", *Economics Letters*, Vol. 69, 1-8.

[22] Fernando, C. S., V. A. Gatchev, and P. A. Spindt (2004), "Wanna Dance? How Firms and Underwriters Choose Each Other", mimeo, University of Oklahoma.

[23] Gelfand, A. and A. Smith (1990), "Sampling-Based Approaches to Calculating Marginal Densities", *Journal of the American Statistical Association*, Vol. 85, 398-409.

[24] Geman, S. and D. Geman (1984), "Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, 721-41.

[25] Geweke, J. (1999), "Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication" (with discussion and rejoinder), *Econometric Reviews*, Vol. 18, 1-126.

[26] Geweke, J., G. Gowrisankaran, and R. Town (2003), "Bayesian Inference for Hospital Quality in a Selection Model", *Econometrica*, Vol. 71:4, 1215-39.

[27] Geweke, J., M. Keane, and D. Runkle (1994), "Alternative Computational Approaches to Inference in the Multinomial Probit Model", *Review of Economics and Statistics*, Vol. 76, 609-32.

[28] Geweke, J., M. Keane, and D. Runkle (1997), "Statistical Inference in the Multinomial Multiperiod Probit Model", *Journal of Econometrics*, Vol. 80, 125-65.

[29] Hauswald, R. and R. Marquez (2002), "Competition and Strategic Information Acquisition in Credit Markets", mimeo, University of Maryland.

[30] Hubbard, R. G., K. N. Kuttner, and D. N. Palia (2002), "Are There Bank Effects in Borrowers' Costs of Funds? Evidence from a Matched Sample of Borrowers and Banks", *Journal of Business*, Vol. 75, 559-81.

[31] Jayaratne, J. and D. P. Morgan (2000), "Capital Market Frictions and Deposit Constraints at Banks", *Journal of Money, Credit, and Banking*, Vol. 32:1, 74-92.

[32] Johnson, S. (1997), "The Effects of Bank Reputation on the Value of Bank Loan Agreements", *Journal of Accounting, Auditing and Finance*, Vol. 24, 83-100.

[33] Leland, H. and D. Pyle (1977), "Informational Asymmetries, Financial Structure and Financial Intermediation", *Journal of Finance*, Vol. 32:2, 371-87.

[34] McCulloch, R. and P. E. Rossi (1994), "An Exact Likelihood Analysis of the Multinomial Probit Model", *Journal of Econometrics*, Vol. 64, 207-40.

[35] Miller, S. and S. Bavaria (2003), "A Guide to the U.S. Loan Market", pdf copy, Standard & Poor's.

[36] Nummelin, E. (1984), *General Irreducible Markov Chains and Non-Negative Operators*, Cambridge: Cambridge University Press.

[37] Oi, W. (1983), "Heterogeneous Firms and the Organization of Production", *Economic Inquiry*, Vol. 21, 247-70.

[38] Preece, D. and D. Mullineaux (1996), "Monitoring, Loan Renegotiation and Firm Value: The Role of Lending Syndicates", *Journal of Banking and Finance*, Vol. 20, 577-93.

[39] Rajan, R. (1992), "Insiders and Outsiders: The Choice Between Informed and Arms-Length Debt", *Journal of Finance*, Vol. 47, 1367-406.

[40] Repullo, R. and J. Suarez (forthcoming), "Loan Pricing under Basel Capital Requirements", *Journal of Financial Intermediation*, forthcoming.

[41] Roth, A. and M. Sotomayor (1990), *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*, Econometric Society Monograph Series, Cambridge University Press.

[42] Sorensen, M. (2003), "How Smart is Smart Money? An Empirical Two-Sided Matching Model of Venture Capital", mimeo, Stanford University.

[43] Spurr, S. (1987), "How the Market Solves an Assignment Problem: The Matching of Lawyers with Legal Claims", *Journal of Labor Economics*, Vol. 5:4, 502-32.

[44] Stomper, A. (2004), "A Theory of Banks' Industry Expertise, Market Power, and Credit Risk", mimeo, University of Vienna.

[45] Strahan, P. E. (1999), "Borrower Risk and the Price and Nonprice Terms of Bank Loans", Staff Report No 90, Federal Reserve Bank of New York.

[46] Tanner, T. and W. Wong (1987), "The Calculation of Posterior Distributions by Data Augmentation", *Journal of the American Statistical Association*, Vol. 82, 528-49.

[47] Tierney, L. (1991), "Markov Chains for Exploring Posterior Distributions", Technical Report No. 560, University of Minnesota.

[48] Williamson, O. E. (1967), "The Economics of Defense Contracting: Incentives and Performance", in R. McKean (ed.), *Issues in Defense Economics*, New York: Collier Macmillan.

[49] Williamson, O. E. (1988), "Corporate Finance and Corporate Governance", *Journal of Finance*, Vol. 43, 567-91.

[50] Wong, L. Y. (2003), "Structural Estimation of Marriage Models", *Journal of Labor Economics*, Vol. 21:3, 699-727.

[51] Yago, G. and D. McCarthy (2004), "The U.S. Leveraged Loan Market: A Primer", pdf copy, Milken Institute.

A Uniqueness of the Equilibrium Matching

The model described in Section 2 is a special case of the College Admissions Model, for which the existence of an equilibrium matching is proved in Roth and Sotomayor (1990).

A new feature of the model is that there is vertical heterogeneity on both sides of the market: all banks have identical preference orderings over the firms and all firms have identical preference orderings over the banks. Eeckhout (2000, Corollary 3) shows that in a one-to-one two-sided matching model, the equilibrium matching is unique if there is vertical heterogeneity on both sides of the market. Below it is shown that this sufficient condition for uniqueness also applies to many-to-one two-sided matching models.

Re-index banks and firms so that $i \succ_j i'$, $\forall i > i'$, $\forall j$, and $j \succ_i j'$, $\forall j > j'$, $\forall i$, where $i \succ_j i'$ denotes that firm j prefers bank i to bank i' and $j \succ_i j'$ denotes that bank i prefers firm j to firm j'. Consider the following J-step algorithm that forms matching μ:

Step 1: firm J borrows from bank I.

Step 2: if bank I's quota is at least two, firm $J-1$ borrows from bank I, too, otherwise it borrows from bank $I-1$.

Step 3: if bank I's quota is at least three, firm $J-2$ borrows from bank I, too, otherwise: if the combined quota of bank I and bank $I-1$ is at least three, firm $J-2$ borrows from bank $I-1$, otherwise it borrows from bank $I-2$.

And so on.

First, μ is an equilibrium matching. Suppose not, then there exists at least one blocking pair. Call it (i', j'). We have $i' > \mu(j')$ and $j' > \min\{j : j \in \mu(i')\}$. But then there is a contradiction, since according to the algorithm if $i' > \mu(j')$ then $j'' > j'$ $\forall j'' \in \mu(i')$, so $j' > \min\{j : j \in \mu(i')\}$ can not be satisfied.

Second, the equilibrium matching is unique. Suppose not, then there exists $\tilde{\mu} \neq \mu$ such that $\tilde{\mu}$ is also an equilibrium matching. There is at least one match that is in μ but not in $\tilde{\mu}$. Now consider the first step in the algorithm that forms a match that is not in $\tilde{\mu}$. Call this match (i', j'). It follows that $\min\{j : j \in \tilde{\mu}(i')\} < j'$ and that $\tilde{\mu}(j') < i'$ (since all of the matches formed in the earlier steps are in both μ

and $\tilde{\mu}$). Therefor $(i',\, j')$ is a blocking pair for $\tilde{\mu}$, a contradiction.

B Conditional Posterior Distributions

Let ϕ be a scalar parameter or a vector of parameters. If ϕ has a density

$$p(\phi \mid .) = C \times \exp\left[-\frac{1}{2}(\phi' M\phi + 2\phi' N + C')\right], \tag{30}$$

where M is a scalar or a matrix and N is a scalar or a vector, then $\phi \sim N(-M^{-1}N, M^{-1})$. Therefore, we obtain the conditional posterior distributions of H_i^b, H_j^f, α, β, γ, κ, and λ by examining their respective conditional posterior densities.

Let H^b, H_{-i}^b, H^f, and H_{-j}^f denote all the bank quality indexes, all the bank quality indexes except H_i^b, all the firm quality indexes, and all the firm quality indexes except H_j^f, respectively. According to the data augmented conditional posterior distribution (equation 27), the conditional posterior density of H_i^b is

$$\begin{aligned} p(H_i^b \mid H_{-i}^b, H^f, X, r, \mu, \theta) &= C \times \mathbf{1}[\underline{H}_i^b < H_i^b < \overline{H}_i^b] \times \exp\left[-\frac{(H_i^b - B_i'\beta)^2}{2}\right] \\ &\times \prod\nolimits_{j\in\mu_t(i)} \exp\left[-\frac{(r_{ij} - U_{ij}'\alpha - \kappa(H_i^b - B_i'\beta) - \lambda(H_j^f - F_j'\gamma))^2}{2\sigma_\nu^2}\right]. \end{aligned} \tag{31}$$

This implies that the conditional posterior distribution of H_i^b is $N(\hat{H}_i^b, \hat{\sigma}^2_{H_i^b})$ truncated to the interval $(\underline{H}_i^b, \overline{H}_i^b)$, where

$$\hat{H}_i^b = B_i'\beta + \frac{\kappa \sum_{j\in\mu_t(i)}[r_{ij} - U_{ij}'\alpha - \lambda(H_j^f - F_j'\gamma)]}{\sigma_\nu^2 + \kappa^2 q_{it}}, \tag{32}$$

and

$$\hat{\sigma}^2_{H_i^b} = \frac{\sigma_\nu^2}{\sigma_\nu^2 + \kappa^2 q_{it}}. \tag{33}$$

Similarly, the conditional posterior density of H_j^f is

$$\begin{aligned} p(H_j^f \mid H^b, H_{-j}^f, X, r, \mu, \theta) &= C \times \mathbf{1}[\underline{H}_j^f < H_j^f < \overline{H}_j^f] \times \exp\left[-\frac{(H_j^f - F_j'\gamma)^2}{2}\right] \\ &\times \exp\left[-\frac{(r_{\mu_t(j),j} - U_{\mu_t(j),j}'\alpha - \kappa(H_{\mu_t(j)}^b - B_{\mu_t(j)}'\beta) - \lambda(H_j^f - F_j'\gamma))^2}{2\sigma_\nu^2}\right] \end{aligned} \tag{34}$$

This implies that the conditional posterior distribution of H_j^f is $N(\hat{H}_j^f, \hat{\sigma}^2_{H_j^f})$ truncated to the interval $(\underline{H}_j^f, \overline{H}_j^f)$, where

$$\hat{H}_j^f = F_j'\gamma + \frac{\lambda[r_{\mu_t(j),j} - U_{\mu_t(j),j}'\alpha - \kappa(H_{\mu_t(j)}^b - B_{\mu_t(j)}'\beta)]}{\sigma_\nu^2 + \lambda^2}, \tag{35}$$

and

$$\hat{\sigma}^2_{H^f_j} = \frac{\sigma^2_\nu}{\sigma^2_\nu + \lambda^2}. \tag{36}$$

The conditional posterior distribution of α is $N(\hat{\alpha}, \hat{\Sigma}_\alpha)$, where

$$\hat{\Sigma}_\alpha = \left\{ \bar{\Sigma}_\alpha^{-1} + \sum_{t=1}^{T} \sum_{(i,j)\in\mu_t} \frac{1}{\sigma^2_\nu} U_{ij} U'_{ij} \right\}^{-1}, \tag{37}$$

and

$$\hat{\alpha} = -\hat{\Sigma}_\alpha \left\{ -\bar{\Sigma}_\alpha^{-1}\bar{\alpha} - \sum_{t=1}^{T} \sum_{(i,j)\in\mu_t} \frac{1}{\sigma^2_\nu} U_{ij}(r_{ij} - \kappa(H^b_i - B'_i\beta) - \lambda(H^f_j - F'_j\gamma)) \right\}. \tag{38}$$

The conditional posterior distribution of β is $N(\hat{\beta}, \hat{\Sigma}_\beta)$, where

$$\hat{\Sigma}_\beta = \left\{ \bar{\Sigma}_\beta^{-1} + \sum_{t=1}^{T} \sum_{i\in I_t} \frac{\sigma^2_\nu + \kappa^2 q_{it}}{\sigma^2_\nu} B_i B'_i \right\}^{-1}, \tag{39}$$

and

$$\hat{\beta} = -\hat{\Sigma}_\beta \left\{ -\bar{\Sigma}_\beta^{-1}\bar{\beta} + \sum_{t=1}^{T} \left[\sum_{(i,j)\in\mu_t} \frac{\kappa}{\sigma^2_\nu} B_i(r_{ij} - U'_{ij}\alpha - \kappa H^b_i - \lambda(H^f_j - F'_j\gamma)) - \sum_{i\in I_t} H^b_i B_i \right] \right\}. \tag{40}$$

The conditional posterior distribution of γ is $N(\hat{\gamma}, \hat{\Sigma}_\gamma)$, where

$$\hat{\Sigma}_\gamma = \left\{ \bar{\Sigma}_\gamma^{-1} + \sum_{t=1}^{T} \sum_{j\in J_t} \frac{\sigma^2_\nu + \lambda^2}{\sigma^2_\nu} F_j F'_j \right\}^{-1}, \tag{41}$$

and

$$\hat{\gamma} = -\hat{\Sigma}_\gamma \left\{ -\bar{\Sigma}_\gamma^{-1}\bar{\gamma} + \sum_{t=1}^{T} \left[\sum_{(i,j)\in\mu_t} \frac{\lambda}{\sigma^2_\nu} F_j(r_{ij} - U'_{ij}\alpha - \kappa(H^b_i - B'_i\beta) - \lambda H^f_j) - \sum_{j\in J_t} H^f_j F_j \right] \right\}. \tag{42}$$

The conditional posterior distribution of κ is $N(\hat{\kappa}, \hat{\sigma}^2_\kappa)$, where

$$\hat{\sigma}^2_\kappa = \left\{ \frac{1}{\bar{\sigma}^2_\kappa} + \sum_{t=1}^{T} \sum_{i\in I_t} \frac{q_{it}(H^b_i - B'_i\beta)^2}{\sigma^2_\nu} \right\}^{-1}, \tag{43}$$

and

$$\hat{\kappa} = -\hat{\sigma}^2_\kappa \left\{ -\frac{\bar{\kappa}}{\bar{\sigma}^2_\kappa} - \sum_{t=1}^{T} \sum_{(i,j)\in\mu_t} \frac{(r_{ij} - U'_{ij}\alpha - \lambda(H^f_j - F'_j\gamma))(H^b_i - B'_i\beta)}{\sigma^2_\nu} \right\}. \tag{44}$$

The conditional posterior distribution of λ is $N(\hat{\lambda}, \hat{\sigma}^2_\lambda)$ truncated from above at 0, where

$$\hat{\sigma}^2_\lambda = \left\{ \frac{1}{\bar{\sigma}^2_\lambda} + \sum_{t=1}^{T} \sum_{j\in J_t} \frac{(H^f_j - F'_j\gamma)^2}{\sigma^2_\nu} \right\}^{-1}, \tag{45}$$

and

$$\hat{\lambda} = -\hat{\sigma}_\lambda^2 \left\{ -\frac{\bar{\lambda}}{\bar{\sigma}_\lambda^2} - \sum_{t=1}^{T} \sum_{(i,j)\in\mu_t} \frac{(r_{ij} - U'_{ij}\alpha - \kappa(H_i^b - B'_i\beta))(H_j^f - F'_j\gamma)}{\sigma_\nu^2} \right\}. \tag{46}$$

Let $n = \sum_{t=1}^{T} |J_t|$ denote the total number of loans in all the markets. The conditional posterior density of $1/\sigma_\nu^2$ is

$$\begin{aligned} p(\frac{1}{\sigma_\nu^2} \quad | \quad & H^*, X, r, \mu, \alpha, \beta, \gamma, \kappa, \lambda) = C \times (\frac{1}{\sigma_\nu^2})^{a-1} \exp(-\frac{1}{b\sigma_\nu^2}) \\ & \times \prod_{t=1}^{T} \prod_{(i,j)\in\mu_t} \left\{ \frac{1}{\sigma_\nu} \exp\left[-\frac{(r_{ij} - U'_{ij}\alpha - \kappa(H_i^b - B'_i\beta) - \lambda(H_j^f - F'_j\gamma))^2}{2\sigma_\nu^2} \right] \right\} \\ = \quad & C \times (\frac{1}{\sigma_\nu^2})^{a-1+n/2} \\ & \times \exp\left\{ -\frac{1}{\sigma_\nu^2} \left[\frac{1}{b} + \frac{1}{2} \sum_{t=1}^{T} \sum_{(i,j)\in\mu_t} (r_{ij} - U'_{ij}\alpha - \kappa(H_i^b - B'_i\beta) - \lambda(H_j^f - F'_j\gamma))^2 \right] \right\} \end{aligned} \tag{47}$$

This implies that the conditional posterior distribution of $1/\sigma_\nu^2$ is $G(\hat{a}, \hat{b})$, where

$$\hat{a} = a + \frac{n}{2}, \tag{48}$$

and

$$\hat{b} = \left[\frac{1}{b} + \frac{1}{2} \sum_{t=1}^{T} \sum_{(i,j)\in\mu_t} (r_{ij} - U'_{ij}\alpha - \kappa(H_i^b - B'_i\beta) - \lambda(H_j^f - F'_j\gamma))^2 \right]^{-1}. \tag{49}$$

Table 1. Variable Definitions and Sources

VARIABLE	DEFINITION	SOURCE
Dependent Variable		
Loan Spread[1]	All-In Spread Drawn above LIBOR/100	DealScan
Independent Variables		
Bank Characteristics		
Salaries-Expenses Ratio[1]	Salaries and Benefits/Total Operating Expenses	Call Reports
Capital-Assets ratio[1]	Total Equity Capital/Total assets	Call Reports
Ratio of Cash to Total Assets[1]	Cash/Total assets	Call Reports
Bank_Size2	Dummy equal to 1 if the bank has $5 billion to $13 billion assets	Call Reports
Bank_Size3	Dummy equal to 1 if the bank has $13 billion to $32 billion assets	Call Reports
Bank_Size4	Dummy equal to 1 if the bank has $32 billion to $76 billion assets	Call Reports
Bank_Size5	Dummy equal to 1 if the bank has more than $76 billion assets	Call Reports
Firm Characteristics		
Leverage Ratio[1]	Total Debt/Total Assets	Compustat
Current Ratio[1]	Current Assets/Current Liabilities	Compustat
Ratio of Property, Plant, and Equipment to Total Assets[1]	PP&E/Total Assets	Compustat
Firm_Size2	Dummy equal to 1 if the firm has $65 million to $200 million assets	Compustat
Firm_Size3	Dummy equal to 1 if the firm has $200 million to $500 million assets	Compustat
Firm_Size4	Dummy equal to 1 if the firm has $500 million to $1,500 million assets	Compustat
Firm_Size5	Dummy equal to 1 if the firm has more than $1,500 million assets	Compustat
Non-Price Loan Characteristics		
Maturity	Loan Facility Length in Months	DealScan
Natural Log of Facility Size[2]	Log(Tranche Amount)	DealScan
Acquisition	Dummy equal to 1 if specific purchase is Acquisition	DealScan
General	Dummy equal to 1 if specific purchase is General	DealScan
Miscellaneous	Dummy equal to 1 if specific purchase is Miscellaneous	DealScan
Recapitalization	Dummy equal to 1 if specific purchase is Recapitalization	DealScan
Revolver/Line < 1 Yr.	Dummy equal to 1 if the loan is a revolving credit line with duration < 1 year	DealScan
Revolver/Line >= 1 Yr.	Dummy equal to 1 if the loan is a revolving credit line with duration ≥ 1 year	DealScan
Secured	Dummy equal to 1 if the loan is secured	DealScan

[1] Expressed in percentage points.
[2] Deflated using the GDP (Chained) Price Index.

Table 2. Summary Statistics

Variable	Number of Observations	Mean	Standard Deviation	Minimum	Maximum
Loan Spread	1369	1.8883	1.1953	0.15	10.80
Salaries-Expenses Ratio	455	24.6241	8.5873	3.1698	58.8139
Capital-Assets ratio	455	8.5533	2.5788	4.6505	32.2950
Ratio of Cash to Total Assets	455	6.9896	4.3929	0.0033	44.2286
Bank Assets ($ Million)	455	72311	124220	15.9774	625256
Leverage Ratio	1369	25.8408	23.1567	0	194.7757
Current Ratio	1369	226.0087	229.2540	7.7253	3167.5310
Firm Assets ($ Million)	1369	1807	6327	1.0579	172828
Ratio of PP&E to Total Assets	1369	31.0707	25.2299	0	95.7851
Maturity	1369	32.9094	22.9736	2	280
Facility Size ($ Million)	1369	192.5351	491.9396	0.1954	10202
Acquisition	1369	0.0964	0.2953	0	1
General	1369	0.4624	0.4988	0	1
Miscellaneous	1369	0.0446	0.2064	0	1
Recapitalization	1369	0.2871	0.4526	0	1
Revolver/Line < 1 Yr.	1369	0.0599	0.2374	0	1
Revolver/Line >= 1 Yr.	1369	0.6698	0.4704	0	1
Secured Status	1369	0.6560	0.4752	0	1

Table 3. OLS: Bank Size on Firm Characteristics

Coefficient	Mean	Std. Dev.
Constant	8.1109	0.1815***
Leverage Ratio	0.0037	0.0022*
Current Ratio	-0.0002	0.0002
Ratio of PP&E to Total Assets	-0.0053	0.0020***
Natural Log of Firm Assets	0.5039	0.0263***

1. The dependent variable is the natural log of the bank's total assets.
2. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 4. OLS: Firm Size on Bank Characteristics

Coefficient	Mean	Std. Dev.
Constant	1.1664	0.3757***
Salaries-Expenses Ratio	-0.0326	0.0051***
Capital-Assets ratio	0.0330	0.0218
Ratio of Cash to Total Assets	0.0009	0.0117
Natural Log of Bank Assets	0.4699	0.0236***

1. The dependent variable is the natural log of the firm's total assets.
2. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 5. Bayesian Inference: Posterior Means and Standard Deviations (Quality Index Equations)

	Mean	Std. Dev.
Bank Quality Index		
Salaries-Expenses Ratio	0.0017	0.0058
Capital-Assets ratio	0.0031	0.0189
Ratio of Cash to Total Assets	0.0258	0.0117**
Bank_Size2	0.4070	0.1496***
Bank_Size3	0.5275	0.1461***
Bank_Size4	0.5380	0.1495***
Bank_Size5	0.3009	0.1466**
Firm Quality Index		
Leverage Ratio	-0.0005	0.0013
Current Ratio	0.0003	0.0001**
Ratio of PP&E to Total Assets	0.0002	0.0012
Firm_Size2	0.1140	0.0830
Firm_Size3	0.2902	0.0867***
Firm_Size4	0.1654	0.0873*
Firm_Size5	0.1134	0.0880
κ	0.1717	0.0876**
λ	-0.1034	0.0848
$1/\sigma_v^2$	1.5131	0.0593***

1. The dependent variables are the quality indexes.
2. Posterior means and standard deviations are based on 20,000 draws from the conditional posterior distributions, discarding the first 2,000 as burn-in draws.
3. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Table 6. Bayesian Inference: Posterior Means and Standard Deviations (Loan Spread Equation)

	Mean	Std. Dev.
Constant	1.6019	0.2072***
Salaries-Expenses Ratio	0.0183	0.0033***
Capital-Assets ratio	0.0151	0.0119
Ratio of Cash to Total Assets	0.0015	0.0068
Bank_Size2	-0.1757	0.0986*
Bank_Size3	-0.1414	0.1060
Bank_Size4	-0.2685	0.1007***
Bank_Size5	-0.2785	0.0845***
Leverage Ratio	0.0091	0.0011***
Current Ratio	-0.0004	0.0001***
Ratio of PP&E to Total Assets	-0.0007	0.0010
Firm_Size2	-0.2211	0.0799***
Firm_Size3	-0.3240	0.0968***
Firm_Size4	-0.2696	0.1072**
Firm_Size5	-0.4579	0.1315***
Maturity	0.0002	0.0011
Natural Log of Facility Size	-0.1978	0.0250***
Acquisition	0.0415	0.1068
General	0.0250	0.0896
Miscellaneous	0.2008	0.1345
Recapitalization	0.0555	0.0950
Revolver/Line < 1 Yr.	0.1626	0.1057
Revolver/Line >= 1 Yr.	-0.1126	0.0585*
Secured	0.8834	0.0570***

1. The dependent variable is the loan spread.
2. Posterior means and standard deviations are based on 20,000 draws from the conditional posterior distributions, discarding the first 2,000 as burn-in draws.
3. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.
4. Dummies for years 1997-2003 are included on the RHS of the spread equation.

Table 7. OLS Estimates

	Coef.	Std. Err.
Constant	1.6721	0.1985***
Salaries-Expenses Ratio	0.0183	0.0031***
Capital-Assets ratio	0.0153	0.0113
Ratio of Cash to Total Assets	-0.0011	0.0061
Bank_Size2	-0.2078	0.0893**
Bank_Size3	-0.2065	0.0928**
Bank_Size4	-0.3391	0.0838***
Bank_Size5	-0.3184	0.0755***
Leverage Ratio	0.0089	0.0011***
Current Ratio	-0.0004	0.0001***
Ratio of PP&E to Total Assets	-0.0007	0.0010
Firm_Size2	-0.2097	0.0789***
Firm_Size3	-0.2826	0.0929***
Firm_Size4	-0.2465	0.1056**
Firm_Size5	-0.4416	0.1308***
Maturity	0.0001	0.0011
Natural Log of Facility Size	-0.1947	0.0255***
Acquisition	0.0296	0.1081
General	0.0269	0.0906
Miscellaneous	0.2024	0.1346
Recapitalization	0.0451	0.0958
Revolver/Line < 1 Yr.	0.1628	0.1063
Revolver/Line >= 1 Yr.	-0.1105	0.0588*
Secured	0.8957	0.0569***

1. The dependent variable is the loan spread.
2. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.
3. Dummies for years 1997-2003 are included on the RHS of the spread equation.

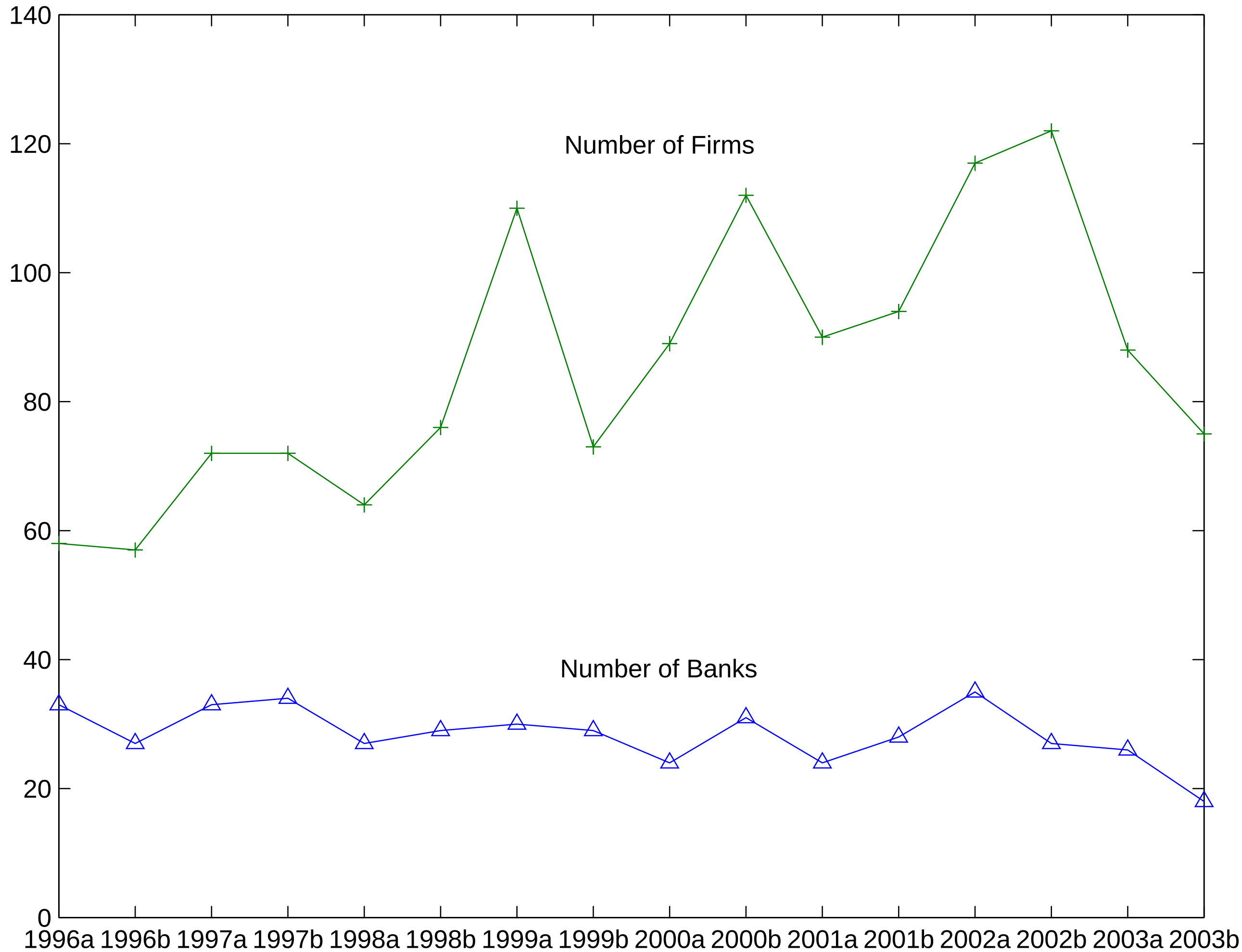

Figure 1. Number of Banks and Number of Firms in Each Market

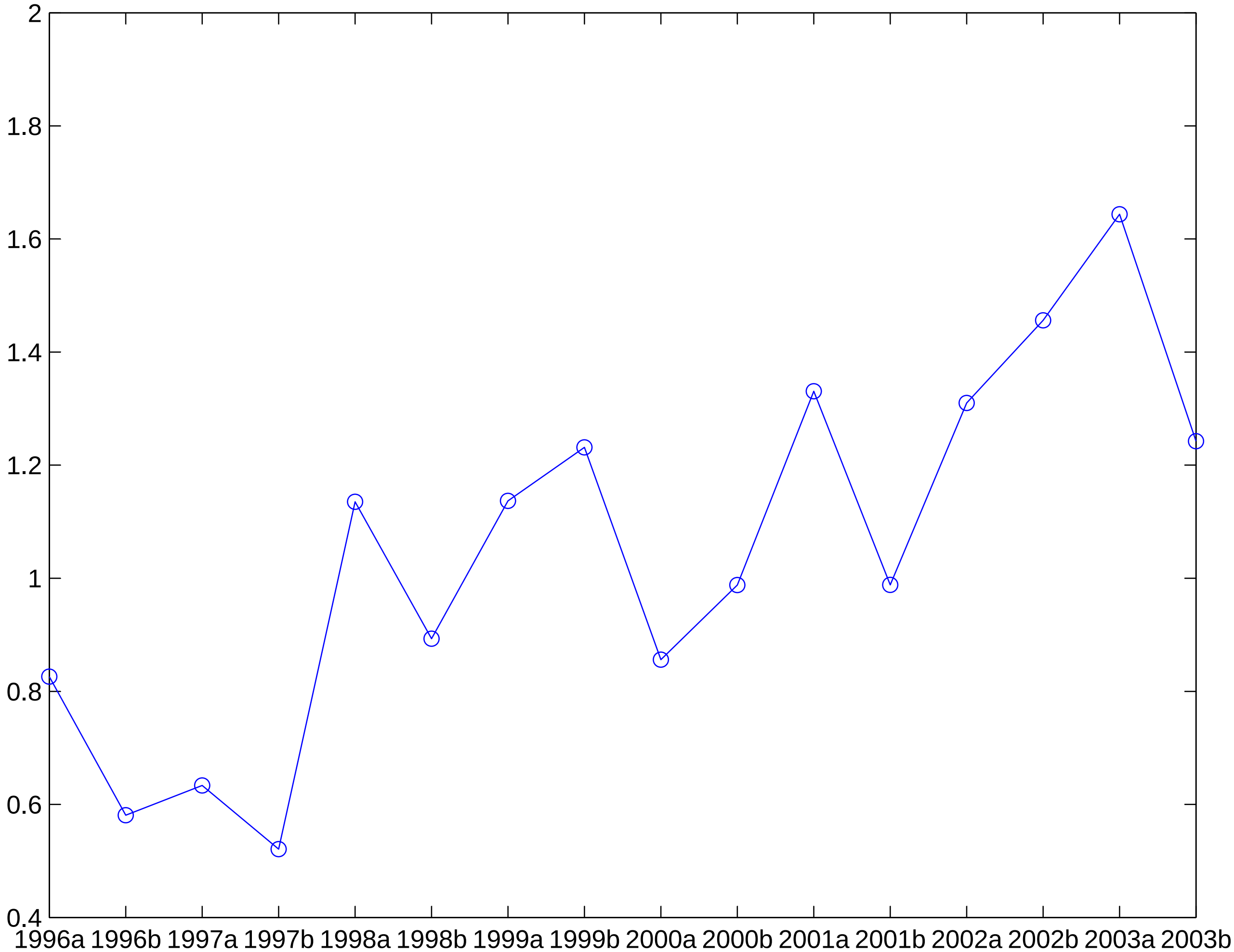

Figure 2. Weighted Average Loan Spread in Percentage Points

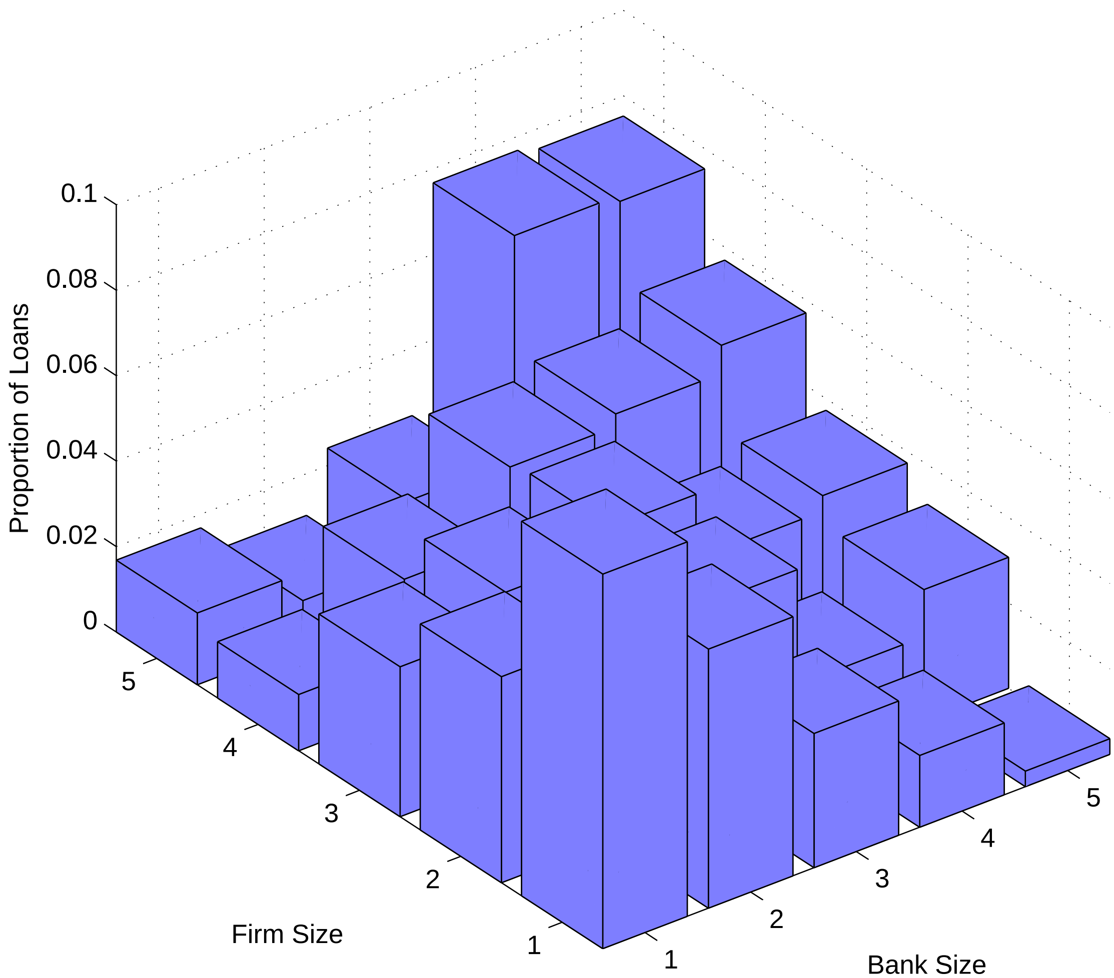

Figure 3. Proportions of Loans for Different Combinations of Size Groups