Borrower-Lender Distance, Credit Scoring, and the Performance of Small Business Loans

Robert DeYoung*
Federal Reserve Bank of Chicago
Chicago, IL

Dennis Glennon*
Office of the Comptroller of the Currency
Washington, DC

Peter Nigro
Bryant University
Smithfield, RI

DRAFT
December 15, 2004

Abstract: Credit to small businesses is an important underpinning for job creation and macroeconomic growth. We develop a theoretical model of decision-making under risk and uncertainty in which agents (bank lenders) have imperfect information about loan applications, and also have imperfect ability to make decisions based on that information. The model yields a number of testable implications related to ongoing trends in small business lending: the increasing physical distance between borrowers and lenders that could exacerbate adverse selection problems; the implementation of small business credit scoring models that could mitigate these information problems; and the recent reductions in government loan guarantees that create poor incentives for small business lenders.

We test these implications for a sample of 33,531 loans to small businesses made under the SBA 7(a) loan program between 1983 and 2001. We believe this is the first study to test the impact of borrower-lender distance, credit-scoring models, and the tradeoff between these two phenomena on the probability of loan default.

Our *preliminary empirical results* offer substantial support for the predictions of our theory. We find that borrower-lender distance is positively associated with loan default, and that the adoption of credit-scoring dampens this relationship. However, we find that credit-scoring lenders experience higher default rates on average; this suggests that diversification and revenue-based benefits from running high-volume, credit-scored lending strategies offset the costs of higher expected default rates. We also find that more generous government loan guarantees, and more competitive local lending markets, are both associated with higher loan default rates.

* The views expressed here are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Chicago, the Federal Reserve System, the Office of the Comptroller of the Currency, or the U.S. Treasury Department. The authors are especially grateful to Dan McMillen for sharing his time and geographic mapping expertise, and to Scott Frame for graciously providing his data on credit-scoring.
DeYoung is the corresponding author: Economic Research Department, Federal Reserve Bank of Chicago, 230 South LaSalle Street, Chicago, IL 60604, voice: 312-322-5396, fax: 312-294-6262, email: robert.deyoung@frbchi.org.

1. Introduction

Over the past two decades, small business borrowers and their bank lenders have been traveling increasingly longer distances to reach each other. A number of studies have documented the increasing distance between lenders and their small-business borrowing customers (Cyrnak and Hannan 2000; Degryse and Ongena 2002; Petersen and Rajan 2002; Wolken and Rohde 2002; Brevoort and Hannan 2004). For example, in 2001 the median borrower-lender distance for newly originated business loans guaranteed by the Small Business Administration (SBA) was a little over 30 miles, more than five times the distance for the median SBA loan during the mid-1980s.

Two separate forces are typically invoked to explain this phenomenon: the ongoing consolidation of the banking industry and recent advances in technologies associated with bank lending. Over the past twenty years the number of commercial banks in the U.S. has declined by about 50 percent, falling from approximately 14,500 banks during the mid-1980s to less than 8,000 banks today. So while the geographic distribution of small business borrowers in the U.S. has remained relatively static, the number of small, locally focused community banks in close proximity to these businesses declined markedly.[1] Over roughly the same time period, improvements in information, communications, and financial technologies have allowed banks to assess the creditworthiness, and monitor the financial condition, of increasingly distant small business borrowers. In the most extreme case, lenders using credit-scoring models, on-line financial data, and securitized asset markets can originate and manage large portfolios of small business loans, with little need for the repetitive face-to-face contact and relationship-building usually associated with small business lending.

Because small businesses tend to be informationally opaque, they cannot access public capital markets and as a result have traditionally depended on local banking institutions for financing. All else

---

[1] It is less well-known that the number of commercial bank *branches* has actually increased by about 50% over time, from about 40,000 branches in the mid-1980s to over 67,000 branches today. Although this would suggest that the surviving banking companies have packed the geographic landscape more tightly and thus moved closer to potential small business customers. But this is not the case on average: banks typically build their new branches close to the existing branches of rivals in order to compete for retail deposits.

equal, increased geographic distance between small business borrowers and their lenders will exacerbate these information problems, potentially reducing the supply of credit to small businesses and dampening the pace of job creation in many local economies.[2] New underwriting technologies like credit-scoring, which rely more on quantifiable financial information and less on qualitative information gleaned via inter-personal relationships, can potentially mitigate these adverse informational effects of distance. Moreover, if these techniques can facilitate lending to distant small businesses an without increase in the probability of default, the likelihood of local-market entry by non-local lenders will increase and this should result in more favorable lending terms for small businesses. Government policy may also provide an offsetting effect. If necessary, subsidies to small business lenders – similar in spirit to the loan guarantees provided by the SBA – could be increased to provide an incentive for banks to maintain the flow of credit to distant small business borrowers. However, loan subsidies do not solve the information problem, and may even encourage poor underwriting and monitoring standards, likely resulting in higher default rates on small business loans.

This study attempts to extend our understanding of how increased borrower-lender distance potential worsens the performance of small business loans, and how new lending technologies and existing government subsidies may mitigate or exacerbate those effects. Following Heiner (1983, 1985), we develop a theoretical model of decision-making under risk and uncertainty in which agents (bank lenders) not only lack perfect information, but also lack the capacity to make perfect decisions given the quality of that information. The model yields a number of empirically testable implications: (1) An increase in information uncertainty (e.g., increased borrower-lender distance) yields reduced loan acceptance rates in equilibrium and increased loan defaults in the short-run; (2) an improvement in lender decision-making ability (e.g., implementation of credit-scoring models) yields increased loan acceptance

---

[2] Small businesses create a disproportionate percentage of new U.S. jobs, although estimates of the size of this effect vary. According to the SBA, small businesses "provide 75 percent of the net new jobs added to the economy" (www.sba.gov). Several other researchers, however, have noted that the SBA job creation estimates may suffer from a variety of conceptual, methodological and measurement issues (e.g, Davis, Haltiwanger and Shuh 1994,1996). No matter the true magnitude of the job creation effect, however, small firms play a major role in the

rates in equilibrium and reduced loan defaults in the short-run; and (3) an increase in expected performance gains (e.g., higher SBA loan guarantees) yields increased loan acceptance rates and increased loan defaults in equilibrium. We test these implications for a sample of 35,991 loans to small businesses originated under the SBA 7(a) loan program between 1983 and 2001.

To the best of our knowledge, this is the first study to test the impact of borrower-lender distance, credit-scoring models, and the tradeoff between these two phenomena on the probability of loan default. Another innovation is our use of lender-specific characteristics (e.g., lender size, lender loan quality) and the market-specific competitive environment (e.g., borrower local market concentration) to better explain the variance in loan default rates across lenders and markets.

Our preliminary empirical results offer substantial support for the predictions of our theory. Borrower-lender distance is positively associated with loan default, but the adoption of credit-scoring dampens or offsets this relationship. However, holding borrower-lender distance constant, the data suggest that credit-scoring lenders experience higher average default rates; since the lion's share of credit-scored small business loans are originated by large banks, this may indicate that benefits from running high-volume small business lending strategies (e.g., diversification, securitization fees) allow these lenders to accept higher expected default rates on individual loans. Finally, we find that more generous government loan guarantees and more competitive local lending markets are both associated with higher loan default rates.

2. Bank lending and distance

Economists typically characterize small businesses as "informationally opaque," meaning that the current financial condition and future financial performance of these firms can be assessed only by those (e.g., owners, managers, suppliers, bankers) with first-hand knowledge of the firm. This "opacity" precludes access to the public credit (e.g., bonds, commercial paper) markets, making small businesses dependent on private bank loans for debt financing – usually from lenders located close enough to make

---

economy.

frequent visits to screen for creditworthiness and monitor ongoing business activity. Given these informational needs, the market for small business lending tends to be local in scope. There is a growing body of academic work on small business lending and community banking; Berger and Udell (1998) and DeYoung, Hunter, and Udell (2004) provide reviews of the literature.

Recent improvements in communications technology (e.g., fax machines, the Internet) and greater information availability (e.g., credit bureaus) have facilitated information flows from borrower to lender, reducing to some extent the frequency of on-site visits by loan officers to screen and monitor small business loans. Similarly, credit scoring models and other automated credit analyses have allowed bank lenders to rely relatively less on non-quantifiable 'soft' information, such as a loan officer's first-hand knowledge of the borrower's business prospects and managerial abilities, and rely relatively more on 'hard' information such as the borrower's quantifiable financial condition (Stein 2002). All else equal, by reducing the cost of gleaning information from small business borrowers – i.e., by making these firms less informationally opaque – these new technologies are likely to increase the distance at which banks are able to profitably lend to small business borrowers.

A number of recent studies have found that the distance between banks and their small business borrowers has been increasing over the past two decades, suggesting that technological progress has allowed banks to screen and monitor their small business customers from greater distances. Cyrnak and Hannan (2000) found that that the average market share of small business loan originations made by U.S. banks outside their local market increased substantially between 1996 to 1998; this share of out-of-market loans more than doubled for banks located in urban markets and nearly doubled for banks located in rural markets. Wolken and Rohde (2002) used data from the 1993 and 1998 National Surveys of Small Business Finance (NSSBF) to show that the proportion of small business loans written by banks more than 30 miles away increased by about 13 percent, and the median distance between a small business borrower and its bank increased by about 11 percent, between these two survey dates. Petersen and Rajan (2002) estimated that the distances between U.S. banks and their small business borrowers increased on average by about 3 to 4 percent per year during the two decades leading up to 1993.[3] Degryse and Ongena (2002) used travel time to measure borrower-lender distance for loans made by a large supplier of small business credit in Belgium, and found that this distance increased during the 1990s, albeit only at a rate of about 9 seconds per year. Using detailed spatial data from commercial loans made under the Community Reinvestment Act (CRA), Brevoort and Hannan (2004) show that, within metropolitan areas, banks become less likely to make loans as borrower-lender distance increases; moreover, this effect appears to have grown stronger over time and as bank size declines, consistent with theories that increasing competition from large banking companies has caused small banks to focus more locally, where they arguably have the greatest informational advantages (e.g., Dell'Ariccia and Marquez, forthcoming).

Banking industry consolidation may also have contributed to increased borrower-lender distance. The number of U.S. banks has declined substantially over the past two decades, but the number of bank branch locations has increased. For example, in 1985 there were about 14,500 banks operating over 57,700 banking offices (about 4 offices per bank), but by 2000 there were about 8,300 banks operating over 72,400 banking offices (almost 9 offices per bank).[4] If the screening and monitoring of small business loans is performed largely by bank personnel stationed at main bank offices, then the decline in the number of banks will increase the effective distance between borrowers and lenders, and will reduce geographic competition in small business lending. However, if the screening and monitoring of small business loans is performed by bank personnel stationed at some of the new bank branch locations, then these changes in industry structure may signal an increasingly competitive environment in which banks compete with each other by, among other things, locating their loan officers closer to their small business customers in order to provide more convenient service.

---

[3] Petersen and Rajan (2002) did not observe an actual time series of data. Rather, they constructed a synthetic time series based on cross-sectional data in the 1993 NSSBF. They observed time indirectly based on the age of the bank-borrower relationship in 1993, and found that borrowers with longer banking relationships tended to be located closer to their banks.

[4] Based on data from the Federal Deposit Insurance Corporation website, www.fdic.gov.

Although no previous study has tested whether increased distance between a bank and its borrowers affects loan performance, some studies have tested whether increased distance between a bank's head office and its affiliated offices affects affiliate performance. As banking organizations grow geographically, it may become more difficult for these banks to monitor the actions of local loan officers who are located increasingly further away from headquarters. Studying multi-bank organizations in the U.S. during the 1990s, Berger and DeYoung (2001) found that the ability of senior headquarters managers to control the efficiency of their affiliate banks dissipated rapidly as the distance to the affiliate increased. In a follow-up study, Berger and DeYoung (2004) found evidence consistent with the conjecture that technological advances between 1985 and 1998 helped mitigate these distance-related control problems.

The literature on cross-border bank performance also suggests a link between increased distance and poor bank performance. DeYoung and Nolle (1996), Chang, Hasan, and Hunter (1998) and Berger, DeYoung, Genay, and Udell (2000) all find that foreign affiliates operate less efficiently than domestic banks. A primary explanation is that cross-border expansion is hindered by differences in languages, cultures, and institutional structures between the home and host countries; however, after controlling for these factors, pure distance appears to limit cross-border expansion of banking companies (Buch 2003, Berger, Buch, DeLong, and DeYoung 2004).

3. Lending decisions under uncertainty: A conceptual framework

Underwriting loans and monitoring loans are information-intensive activities. Lenders require extensive information about the borrower, the local business environment, and the macro-economy, as well as the skills to analyze this information and make decisions based on that analysis. To model this process, it is important to recognize that neither the information observed by the lender, nor the lender's decision-making skills, are perfect. These imperfections can result in sub-optimal lending decisions and outcomes (e.g., borrower delinquency or default) that harm bank performance.

Conventional analysis of decision-making under risk and uncertainty applies a very narrow definition of uncertainty: agents know all possible events and the outcomes of those events, use all available information to assign probabilities to the likelihood of each of the events occurring, and given this information choose actions that maximize expected performance (e.g., profits or utility). Conventional analysis does not address the uncertainty that emerges when agents lack the capacity to make perfect decisions (i.e., incorrectly interpret information).[5] In the model developed here, decisions are a function of both the quality of the information available to the decision-maker and the quality of the decision-maker's interaction with that information. We borrow heavily from the theory of decision-making under uncertainty developed by Heiner (1983, 1985, 1985a, 1986) – a theoretical framework that extends the concept of uncertainty beyond that associated with risk alone.[6] Once our model is constructed, we use it to study the effects of changes in information technology, the distance between borrower and lender, and government loan guarantees.

3a. Two components of decision-making under uncertainty

We define uncertainty with respect to both imperfect information and imperfect decision skills. As a result of the uncertainty, not only are decision-makers exposed to errors due to random effects (i.e., risk) but also errors due to incorrect interpretation of information. In this section we outline a simple model of decision making under uncertainty in which decisions are conditional on the lender's ability to interpret the decision environment. This model leads to lender behavior that is based on rules that are rational, though not necessarily optimal.

We formalize the effect of decision errors on lending behavior using the following simple model of decision-making under uncertainty. Let S represent all relevant states of nature, embracing all possible combinations of local and national economic growth, interest rates, price movements, job market conditions, changes in asset values, and other external factors that effect general loan repayment performance. Let X represent all information available to the decision-maker about each state of nature S.

---

[5] Conventional analysis generally speaks of uncertainty as the absence of perfect knowledge or information. We expand the meaning of uncertainty to include the absence of perfect decision-making. This necessarily leads to different definitions for the terms "risk" and "uncertainty." Uncertainty in the context of this paper is closely related to the type of uncertainty outlined by Knight (1933).
[6] See also Beshouri and Glennon (1996) for an application of this analysis to bank lending decisions.

In the context of our discussion, X includes borrower-specific and loan-specific information, such as the applicant's financial profile, credit history, debt burden, collateral, cash flows, terms of the loan, etc. Finally, let A represent the set of actions available to the decision-maker. We will characterize A narrowly as the loan approval/denial decision, which the lender makes conditional on his imperfect information X and his imperfect decision skills. Thus, the lender may make inappropriate decisions, such as selecting action $\alpha \in A$ (loan approval) when conditions suggest that selecting action $\beta \in A\text{-}\alpha$ would increase performance (type II error), or selecting $\beta \in A\text{-}\alpha$ (loan denial) when conditions suggest that selecting $\alpha \in A$ would increase performance (type I error).

We represent the lender's *imperfect information* by the probabilities that the information in his possession either correctly or incorrectly identifies the true state of nature. Let $S^*_\alpha \subset S$ represent the subset of possible states of nature in which choosing action α is optimal, and let $X^*_\alpha \subset X$ represent the subset of information that signals to the lender that α is the best choice. Then let $r^X_\alpha = p(X^*_\alpha \mid S^*_\alpha)$ be the conditional probability that the information received (X^*_α) correctly signals that the optimal states of nature for selecting α exist and $w^X_\alpha = p(X^*_\alpha \mid S - S^*_\alpha)$ the conditional probability that this same information is received when non-optimal states of nature for selecting α exist; and let $\rho^X_\alpha = r^X_\alpha/w^X_\alpha$ measure the relative *reliability of the information*. As the information becomes more reliable, $r^X_\alpha \rightarrow 1$, $w^X_\alpha \rightarrow 0$, and $\rho^X_\alpha \rightarrow \infty$.

We represent the lender's *imperfect decision skills* by a decision function B(x) which maps information $x \in X$ into actions A.[7] The decision function incorporates the limitations placed on decision skills that lead to decision errors beyond those generally associated with risk (i.e. imperfect information). More formally, let $r^B_\alpha = p(B(x)=\alpha \mid X^*_\alpha) < 1$ be the conditional probability that action α is selected when optimal messages are received; let $w^B_\alpha = p(B(x)=\alpha \mid X - X^*_\alpha) > 0$ be the conditional probability that action

[7] Note that we separate the decision function B(x) from the performance function (e.g., profit maximization). This contrasts with conventional choice theory, in which the decision function is the performance function (e.g., lenders choose actions that maximize profits). It can be shown that the decision and performance functions are equivalent in the special case when decision skills are perfect (see Heiner 1985).

α is selected when non-optimal messages are received; and let $\rho^B_\alpha = r^B_\alpha/w^B_\alpha$ measure the relative *reliability of a lender's behavior* in responding to information. As lenders become more reliable at responding correctly to information received, $r^B_\alpha \to 1$, $w^B_\alpha \to 0$, and $\rho^B_\alpha \to \infty$.

We can jointly express the uncertainties due to imperfect information and imperfect decision skills in a single reliability ratio. Assume that the choice of α is correct ($s \in S^*_\alpha$). Then a lender can select α under two scenarios: if the analyst receives information that α is optimal ($x \in X^*_\alpha$) and correctly interprets this information ***or*** if he receives information that β is optimal ($x \in X-X^*_\alpha$) and incorrectly interprets this information. More formally, the joint conditional probability that α is the right choice is

$$\begin{aligned} r^{XB}_\alpha &= p(B(x)=\alpha \mid S^*_\alpha) \\ &= p(X^*_\alpha \mid S^*_\alpha)\, p(B(x)=\alpha \mid X^*_\alpha) + p(X-X^*_\alpha \mid S^*_\alpha)\, p(B(x)=\alpha \mid X-X^*_\alpha) \\ &= r^X_\alpha r^B_\alpha + (1 - r^X_\alpha)\, w^B_\alpha. \qquad (1) \end{aligned}$$

Similarly, the joint conditional probability that α is the wrong choice is

$$w^{XB}_\alpha = w^X_\alpha r^B_\alpha + (1 - w^X_\alpha)\, w^B_\alpha. \qquad (2)$$

The ratio of the joint conditional probabilities that α is the right relative to the wrong choice (i.e., equations 1 and 2) is the *joint reliability ratio*:

$$\rho^{XB}_\alpha = \frac{r^{XB}_\alpha}{w^{XB}_\alpha} = \frac{r^X_\alpha r^B_\alpha + (1 - r^X_\alpha)\, w^B_\alpha}{w^X_\alpha r^B_\alpha + (1 - w^X_\alpha)\, w^B_\alpha} = \frac{r^X_\alpha (\rho^B_\alpha - 1) + 1}{w^X_\alpha (\rho^B_\alpha - 1) + 1}. \qquad (3)$$

This ratio illustrates that uncertainty due to imperfect information (r^X_α and w^X_α) and uncertainty due to imperfect decision-making skills (ρ^B_α) are interactive in determining ρ^{XB}_α. As information becomes more perfect (i.e., $r^X_\alpha \to 1$, $w^X_\alpha \to 0$), $\rho^{XB}_\alpha \to \rho^B_\alpha$; and, as decision-making skills become more perfect (i.e., $r^B_\alpha \to 1$, $w^B_\alpha \to 0$), $\rho^{XB}_\alpha \to r^X_\alpha/w^X_\alpha = \rho^X_\alpha$.

The assumption that lenders do not always know the effect of their actions on performance is critical to our argument: it implies that lenders restrict their behavior until they are 'reasonably' confident they will gain from selecting a particular action instead of reacting optimally (i.e., selecting the action that maximize expected profits). Because of the uncertainty associated with the inability to recognize the benefits, lenders develop, in practice, underwriting guidelines that impose restrictions on the loan officers to respond to information that is difficult to interpret. These guidelines generally include threshold values for specific underwriting ratios, pricing sheets, and other well developed "rules-of-thumb" that reduce the discretion of loan officers. For example, bank policies may restrict loan officers from taking applications or approving loans unless the borrower has collateral in excess of some fixed percentage of loan value, the borrower's operating earnings exceed interest expenses by some fixed multiple, or the borrower resides within the bank's local lending area. In the latter case, restricting lending to geographic areas most familiar to the loan analyst may increase the reliability of information and/or reduce the likelihood of analyst decision errors, and thus increase the likelihood the bank will benefit (i.e., increase profits) from the analyst decisions.

Relying on rules, however, will inevitably lead to decision errors. These are reflected in the construction of the conditional probabilities in equations (1) and (2). The conditional probability that α is the right choice can be expressed in terms of type I errors (incorrectly rejecting a loan application), and the conditional probability that α is the wrong choice can be expressed in terms of type II errors (incorrectly accepting a loan application). Defining the probabilities of type I and type II errors as $t_I = p(B(x)\neq\alpha \mid S^*_\alpha)$ and $t_{II} = p(B(x)=\alpha \mid S\text{-}S^*_\alpha)$, respectively, then it follows that $r^{XB}_\alpha = 1\text{-}t_I$ and $w^{XB}_\alpha = t_{II}$.

3b. A joint reliability condition

The introduction of imperfect decision skills increases the complexity of the decision making process required to improve performance (i.e., increase profits). However, we can derive a decision rule based on the joint reliability of the information available to the lender and the lender's ability to use that information correctly that indicates when flexibility to select an action is more likely to improve rather than worsen performance. As outlined by Heiner (1985a) let $g^e_\alpha = p^s_\alpha\, r^{XB}_\alpha\, g_\alpha$ be the expected gain from correctly selecting α, where $p^s_\alpha = p(S^*_\alpha)$ is the unconditional probability that α is the correct choice (i.e., $s\in S^*_\alpha$) and $g_\alpha = \pi(\alpha;S^*_\alpha)$ is the performance gain from correctly selecting α. Let $l^e_\alpha = (1\text{-}p^s_\alpha)\, w^{XB}_\alpha\, l_\alpha$ be the expected loss from incorrectly selecting α, where $l_\alpha = \pi(\alpha;S\text{-}S^*_\alpha)$ is the performance loss from incorrectly selecting α. The lender will benefit from selecting α if the expected gain exceeds the expected loss, that is, if $p^s_\alpha\, r^{XB}_\alpha\, g_\alpha > (1\text{-}p^s_\alpha)\, w^{XB}_\alpha\, l_\alpha$. Rearranging terms, we derive a reliability condition that must be satisfied in order for an agent to benefit from selecting a specific action α under uncertainty (Heiner, 1986a):

$$\rho^{XB}_\alpha = \frac{r^{XB}_\alpha}{w^{XB}_\alpha} > \frac{l_\alpha}{g_\alpha} \frac{(1 - p^s_\alpha)}{p^s_\alpha} = T_\alpha \,. \tag{4}$$

The inequality has a straightforward interpretation: the lender will approve a loan application (i.e., select α) only when the joint reliability of the lender's information and ability to use that information (ρ^{XB}_α) exceeds some minimum expected performance bound (T_α) necessary to improve expected lender performance. It is intuitive that this minimum bound is determined by the expected relative return, l_α/g_α, and the inverse of the odds that the conditions for correctly selecting α exists, $(1\text{-}p^s_\alpha)/p^s_\alpha$. In equilibrium, the lender is more likely to approve a loan application for high values of r^{XB}_α, p^s_α and g_α and for low values of w^{XB}_α and l_α.

We illustrate these results using the unit-probability box in Figure 1. Moving from bottom-to-top in the box causes $r^{XB}_\alpha \rightarrow 1$, increasing the conditional probability that α is correctly chosen (fewer type I errors occur). Moving from left-to-right causes $w^{XB}_\alpha \rightarrow 1$, increasing the conditional probability that α is incorrectly chosen (more type II errors). Thus, in the extreme northwest corner of the box where $r^{XB}_\alpha = 1$ and $w^{XB}_\alpha = 0$ (i.e., an infinite joint reliability ratio ρ^{XB}_α), the lender has both perfect information and perfect decision skills, and as a result makes no type I or type II errors (i.e., $t_I = t_{II} = 0$). More realistically, however, lenders have both imperfect information and imperfect decision skills; under these circumstances the type I and type II error rates are determined by the frequency with which the lender selects α. We proceed in our analysis under the following reasonable assumption: for any given frequency of selecting α, lenders will first consider those loan applications with the most complete and most easily interpretable information, before moving on to considering applications with less complete and/or less easily interpretable information.

Lenders that never select α are located at the southwest corner of the box, where the probability $r^{XB}{}_{\alpha}$ of selecting α when conditions are correct is zero and the probability $w^{XB}{}_{\alpha}$ of selecting α when conditions are incorrect is also zero. In this extreme case, the probability of type I error $t_I = r^{XB}{}_{\alpha} - 1 = 1$ and the probability of type II error $t_{II} = w^{XB}{}_{\alpha} = 0$. Lenders that select α sometimes, but infrequently, will be located at a point like X. These lenders will make very few type II errors (incorrectly accepting a loan application) but will commit a large number of type I errors (incorrectly rejecting a loan application), and thus will have a very high joint reliability ratio $\rho^{XB}{}_{\alpha} = r^{XB}{}_{\alpha}/w^{XB}{}_{\alpha} = (1\text{-}t_I)/(t_{II})$. Selecting α more frequently requires lenders to consider applications with increasingly less complete or less easily interpretable information, causing the conditional probability $w^{XB}{}_{\alpha}$ that α is chosen incorrectly to increase quickly relative to the conditional probability $r^{XB}{}_{\alpha}$ that α is chosen correctly. The resulting decrease in the joint reliability ratio $\rho^{XB}{}_{\alpha} = r^{XB}{}_{\alpha}/w^{XB}{}_{\alpha}$ with increasing values of α is represented by points Y and Z, and is consistent with increasing probability of type II errors and decreasing probability of type I errors. In the extreme situation in which a lender always selects α, $r^{XB}{}_{\alpha}$ will be one ($t_I = 0$) and $w^{XB}{}_{\alpha}$ will also be one ($t_{II} = 1$), or the northeast corner (1,1) of the box.

The curved line passing through points X, Y, and Z in Figure 1 is the *reliability ratio curve* (RRC), the locus of all attainable joint reliability ratios $r^{XB}{}_{\alpha}/w^{XB}{}_{\alpha}$ for a given level of information uncertainty and decision-making skills.[8] The concave shape requires the joint reliability ratio $\rho^{XB}{}_{\alpha}$ to decrease (i.e., $w^{XB}{}_{\alpha}$ to increase relative to $r^{XB}{}_{\alpha}$) from left-to-right along the RRC, consistent with an increase in the frequency with which lenders choose α as described above. All else equal, moving from left-to-right along the RRC increases the likelihood of type II errors, and thus leads to higher rates of (ex post) loan default.

[8] The reliability ratio curve is also more commonly referred to as the Receiver Operating Characteristics (ROC) curve used in the signal-detection literature (Green and Swet, 1974) and in the credit scoring and risk measurement literature (Stein, 2003; and Engelmann, et al. 2003). The concave slope of the curve is consistent with most empirical studies of behavior in the signal-detection experiments and the assumption that a likelihood-ratio criteria underlies the decision rule (Green and Swets, 1974; and Heiner, 1986a).

The slope of the ray extending from the origin in Figure 1 is the *minimum expected performance bound* (T_α) defined above, i.e., the joint reliability ratio above which selecting α is expected to improve performance, and below which choosing α is expected to reduce performance. The equilibrium in Figure 1 is quite clear: by condition (4), the lender will approve loan applications to the left of point Y along the RRC and will reject loan applications to the right of point Y along the RRC.

3c. Testable comparative static results

We can use the analysis outlined in the unit-probability box to illustrate three fundamental comparative static results: (i) an increase in information uncertainty, e.g., which might occur with an increase in borrower-lender distance; (ii) an improvement in decision skills, e.g., which might occur with the application of credit-scoring models; and (iii) an increase in the gain or reduction in the loss associated with selecting α, e.g., which might occur with increased interest rate margins due to reductions in local market competition, or reduced loan losses due to increased government loan guarantees.

In Figure 2 we demonstrate the effect of an increase in information uncertainty (e.g., greater borrower-lender distance) on the selection of α. Starting in equilibrium at point A and holding decision-making skills constant, greater information uncertainty pushes all attainable joint reliability ratios further from the northwest corner, making the RRC less concave. A smaller percentage of loan applications now exceed the minimum expected performance bound T_α. Faced with less certain information, lenders will select α less frequently (moving from point A to point C), in effect trading more type I errors in exchange for fewer type II errors. Note that even though lenders now approve fewer loans, the expected loan default rate will not change in equilibrium because points A and C both satisfy the same joint reliability condition implied by T_α.[9] However, if a lender does not immediately recognize that information quality has deteriorated, it will continue to apply its existing "rule-of-thumb" underwriting guidelines to the new loan applications in the short-run: the frequency of α will not change and the lender will simply move

---

[9] This follows from the definitions $t_I = 1\text{-}r^{XB}{}_\alpha$ and $t_{II} = w^{XB}{}_\alpha$ and the condition that $T_\alpha = r^{XB}{}_\alpha / w^{XB}{}_\alpha$ is fixed. As a result, the proportion of non-defaulted loans to defaulted loans remains fixed.

from point A to a point like B.[10] As a result, the loan default rate will increase because the lender will be making more type I errors *and* more type II errors. Because new loans season slowly, it may be several years before the lender receives this performance feedback and responds by adjusting its lending policies to point C.

In Figure 3 we demonstrate the effect of an improvement in decision-making skills (e.g., adopting credit-scoring models) on the selection of α. Starting in equilibrium at point A and holding information quality constant, improved decision-making skills pushes all attainable joint reliability ratios closer to the northwest corner, resulting in a more concave RRC curve. A greater percentage of loan applications now exceed the minimum expected performance bound T_α, so lenders will select α more frequently (moving from point A to point E), in effect trading fewer type I errors in exchange for more type II errors. Again, because the joint reliability condition corresponding to T_α is unchanged at the margin, the loan default rate is unchanged in equilibrium. However, in practice the lenders may not immediately recognize the increased accuracy of their decisions: in the short-run, the frequency of α will not change and the lender will simply move from point A to a point like D, with the loan default rate decreasing as the lender makes fewer type I errors *and* fewer type II errors. It may be several years before the lender receives this performance feedback and responds by adjusting its lending policies to point E.

In Figure 4 we demonstrate the effects of increases in expected performance gains g_α and/or reductions in expected performance losses l_α (e.g., higher government loan guarantees) on the selection of α. Starting in equilibrium at point A and holding both information uncertainty and decision-making skills constant, an increase in expected performance gains (and/or a reduction in expected performance losses) reduces the ratio of expected loan returns l_α/g_α , causing a downward rotation of the minimum performance bound $T_\alpha \rightarrow T_\alpha'$ (see equation 4). The lender will select α more frequently (moving from point A to point F) so that the joint reliability ratio $\rho^{XB}{}_\alpha$ remains consistent with the lower minimum

[10] Points A and B represent identical values of α because they lie equidistant between the endpoints of their respective RRCs.

performance bound T_{α}', in effect trading fewer type I errors in exchange for more type II errors. Because the joint reliability condition for loan approval is now less stringent (for fixed levels of information certainty and decision-making skills), the expected loan default rate will increase. Note that the increase in loan supply (reduced type I errors) in this scenario is consistent with the expansion of credit access that motivates subsidized lending programs, and the increased expected default rate is consistent with loan guarantees to encourage lenders to make increased amounts of these loans.

4. Empirical implementation

The theoretical model generates four empirically tractable hypotheses regarding loan default rates, as illustrated in Figures 2, 3, and 4 and discussed above.[11] We empirically test these three hypotheses using data on SBA 7(a) loans originated between 1983 and 2001 and a discrete-time hazard model of loan default. This framework is an empirical analog to the semi-parametric Cox proportional hazard model (Allison 1990, Shumway 2001, Brown and Goetzmann 1995, Deng 1995). It allows us to test each of our hypotheses about the likelihood of loan default conditional on: the effects of time since loan origination; the effects of borrower-specific, lender-specific, and market-specific conditions at the time of loan origination that remain constant during the life of the loan; and the effects of macroeconomic and other conditions that may vary over the life of the loan (so-called time-varying covariates).

4a. Data sources and descriptive statistics

We test our three hypotheses using a random sample of 35,991 SBA 7(a) loans originated by 5,552 qualified SBA program lenders between 1983 and 2001. The SBA 7(a) loan program provides loan guarantees for small business firms that are otherwise unable to access credit through conventional means. This program is the U.S. government’s primary policy tool for addressing the credit availability

---

[11] The theoretical model also generates a number of hypotheses about loan supply and credit access. Briefly stated, loan supply (the frequency of selecting α) will increase as: borrower information becomes more certain; lender decision-making skills improve; and expected returns from lending improve, ceteris paribus. Testing these hypotheses empirically requires a full-blown loan supply-loan demand model, and is beyond the scope of this study, which focuses on loan performance rather than credit access.

concerns of small businesses; for example, in 1999 the SBA provided over $10 billion of guarantees on more than 43,000 small business loans. Loan guarantees are provided to eligible businesses through qualified financial institutions (mainly but not exclusively commercial banks) that select the firms to receive loans, initiate SBA involvement, underwrite the loans within SBA program guidelines, and monitor and report back to the SBA the progress of these loans. Under the 7(a) program, the SBA shares all loan losses *pro rata* with the lending institution (i.e., the SBA does not take a first-loss position), based on the remaining outstanding balances at the time of default and the contractual guarantee percentage stipulated by the SBA at the time of the loan. Because lenders share in the losses, they have (perhaps reduced) incentives to screen for creditworthiness and to set appropriate loan interest rates and contract terms.

The lender typically holds and services the loan until maturity; however, there is also a secondary market for the guaranteed portion of these loans, and this market facilitates the securitization of portfolios of credit-scored SBA loans. Loans in arrears more than sixty days may be put back to the SBA in exchange for a payment equal to the guaranteed portion of the principal plus delinquent interest.

Table 1 displays annual trend data between 1983 and 2001 for each of the central concepts in our experiment: SBA loan default rates; borrower-lender distance (our proxy for information uncertainty); lender use of credit-scoring models (our proxy for decision-making ability); and the percentage of the loan guaranteed by the SBA (our proxy for expected performance gains or losses).

Default rates for SBA loans tended to decline during our sample period, from a high of around 27% for loans originated in 1984, to a low of around 6% for loans originated in 2001. A substantial amount of the decline shown in the last few rows of Table 1 is caused by right-censoring in the data (i.e., recently originated loans that have not yet matured past an age at which they are most likely to default); still, loan default rates fell to near 12% in 1993 before beginning to increase in the mid-1990s. Declining default rates may be associated with improved macroeconomic conditions during these years – or they may also be associated with improvements in the SBA loan program itself over time. For example, Table 1 shows that on average the SBA guarantee percentage has declined over time, from around 85% for loans originated during the mid-1980s, to around 70% for loans originated at the end of our sample period. By reducing the value of the lender's put option, a lower guarantee should increase lenders' incentives to carefully screen and monitor loans.

The average distances between SBA lenders and their small business customers increased markedly toward the end of our sample period. Between 1983 and 1993, the median borrower-lender distance fluctuated in a tight band between 6 miles and 8 miles, but began accelerating soon after that, reaching 10 miles by 1996; 20 miles by 1999; and 30 miles by 2001. The mean borrower-lender distance increased even more dramatically – from around 30 miles in the mid-1990s to almost 100 miles by 1997; 200 miles by 2000; and over 300 miles by 2001. Although the naturally truncated distribution of this variable (which cannot be less than zero) dictates that its mean exceed its median, this large of a discrepancy may indicate that lenders using credit-scoring models to originate large volumes of small business loans are doing so at greater distances than lenders using more traditional underwriting methods. This is consistent with the available data on the use of credit-scoring models by U.S. banking companies, which indicates that lenders did not begin using these models to screen small business loan applications in large numbers until the mid-1990s – about the time that the distribution of borrower-lender distance began to grow increasingly skewed.

4b. A discrete-time hazard modeling approach

We use a discrete-time hazard framework to empirically test the loan-default hypotheses outlined above. Consistent with all empirical approaches based on hazard functions, we measure the likelihood that loan i (i = 1,2,…,N) originated at time t = 0 will default during some time period t > 0 (t= 1,2,…T), given that it has not defaulted up until that time. More specifically, the discrete-time hazard approach requires us to report our data in an 'event history' format: a series of binary variables $D_i(1),...D_i(T)$, where

$D_i(t)=1$ if loan i defaults during time period t, and $D_i(t)=0$ otherwise.[12] These N separate event histories for each loan i are 'stacked' one on top of the other, resulting in a column of zeros and ones having $\sum_{i=1}^{N} T$ rows. This event-history data design permits a hazard model to be estimated using qualitative dependent variable (e.g., logit or probit) techniques. We define D^*_{it} as a latent index value that represents the unobserved propensity of loan i to default during time period t, conditional on covariates X and W:

$$\begin{aligned} D^*_{it} &= \mathbf{X_i}\beta + \mathbf{W_{it}}\gamma + \varepsilon_{it} \\ &= \mathbf{Z}\phi + \varepsilon_{it} \end{aligned} \tag{5}$$

where $\mathbf{X}$ is a vector of time-invariant covariates, $\mathbf{W}$ is a vector of time-varying covariates, β and γ are the corresponding vectors of parameters to be estimated, and ε is an error term assumed to be distributed as standard logistic. We write (5) more compactly using $\mathbf{Z} = [\mathbf{X},\mathbf{W}]$ and $\phi = \begin{bmatrix} \beta \\ \gamma \end{bmatrix}$ to represent the full set of time-invariant and time-varying covariates and parameters, respectively. We further define:

$$D_{it} = 0 \text{ if } D^*_{it} \leq 0$$

$$D_{it} = 1 \text{ if } D^*_{it} > 0$$

so that the probability that $D_{it} = 1$ (that is, the probability that loan i defaults during period t conditional on having survived until period t, or the hazard rate) is given by:

$$\text{prob}(D^*_{it} > 0) = \text{prob}(\mathbf{Z}\phi + \varepsilon > 0)$$

$$\text{prob}(D^*_{it} > 0) = \text{prob}(\varepsilon > -\mathbf{Z}\phi)$$

$$\text{prob}(D_{it} = 1) = \Lambda(\mathbf{Z}\phi) \tag{6}$$

where $\Lambda(\cdot)$ is the logistic cumulative distribution function. We estimate equation (6) using standard binomial logit techniques. Based on the construction of the data, we refer to this empirical approach as a

[12] Measuring time in quarters, the event history $D_i(1),...D_i(t),...D_i(T)$ for a 3-year loan will be five zeros followed by a one (0,0,0,0,0,1) if that loan defaults in the sixth quarter after it was originated, but will be a string of twelve zeros if that loan does not default. Loans that are prepaid prior to their contractual maturity, or right-censored loans (still performing but not yet mature at the end of our sample period), are also represented by strings of zeros.

'stacked-logit' model. The stacked-logit is a very flexible approach compared to most other multivariate hazard function models: in addition to allowing for time-varying covariates on the right-hand-side of the logit model, this approach does not require us to impose any parametric restrictions (e.g., a Weibull distribution) on the loan default distribution (the hazard function).

4c. Regression specification and hypothesis tests

We estimate the stacked-logit model using a longitudinal data set with 545,297 loan-quarter observations comprised of 33,531 individual SBA loans originated by 5,552 unique lenders between 1983 and 2001. The model includes the following variables:

$$\begin{aligned}\Pr[\, D_i(t)=1|\mathbf{Z}\,] = \Lambda[\, & \ln DISTANCE_i,\ SCORER_{ij},\ \ln DISTANCE_i{*}SCORER_{ij},\ SBA\%_i,\ SCORINGEXP_{ij}, \\ & MATURITY_i,\ NEWFIRM_i,\ FIRMSIZE_i,\ HHI_i,\ URBAN_i,\ JOBGROWTH_i(t), \\ & BANKSIZE_{ij},\ CHARGEOFFS_{ij},\ RESERVES_{ij}\,;\ \phi\,] \qquad (7)\end{aligned}$$

where i indexes the loan and j indexes the lender. Variables bearing a (t) are time varying, i.e., they may have a different value for each quarter t in the event history of loan i. All other variables are valued as of the quarter in which the loan was originated. Table 2 shows definitions, summary statistics, and data sources for each of the variables specified in (7).

The three main test variables DISTANCE, SCORER, and SBA% correspond, respectively, to the theoretical implications illustrated above in Figures 2, 3, and 4. DISTANCE is our proxy for information uncertainty. It is equal to the distance measured in miles "as the crow flies" between the Zip Code centroid of the small business borrower and the Zip Code centroid of the lending office (which may or may not be the bank's head office). We used *Maptitude* geographic mapping software to calculate this distance. Recognizing that the cost-per-mile of travel is decreasing in distance (i.e., time and cost economies of scale in distance), we specify this variable in natural logs. Because our model captures *short-run* (quarterly) changes in loan default $D_i(t)$ with respect to changes in loan-specific conditions, we expect a positive coefficient on lnDISTANCE in our estimations, consistent with the discussion accompanying Figure 2 above.

SCORER is our proxy for the lender's decision-making ability. SCORER is a binary variable equal to one if the lender uses credit-scoring to screen at least some of its small business loan applications. Unfortunately, SBA does not identify if loans were originated using a credit scoring tool. We generate the data for SCORER based on a survey of the 200 largest U.S. bank holding companies performed in 1997 by Akhavein, Frame, and White (2001). There are limitations to this approach. For example, we cannot specify SCORER for the final four years of our 1983-2001 data period, and we also must assume that SCORER = 0 for banking companies and their affiliates too small to be included in the survey. While neither of these limitations is desirable, they may not be as problematic as they at first seem: the first limitation can be addressed simply by dropping loans originated after 1997, while the second limitation is unlikely to be meaningful insofar as credit scoring small business loans was almost exclusively a large bank activity prior to 1998.

We expect the coefficient on SCORER to be negative, consistent with our discussion accompanying Figure 3 regarding the *short-run* change in loan default rates upon the initial implementation of improved information technology.[13] A related variable developed as a proxy for lenders' credit-scoring experience is SCORINGEXP, which equals the number of cumulative quarters since a lender began using credit-scoring models for small business loans. Our theoretical discussion predicts a positive coefficient on SCORINGEXP: as improved default rates (negative coefficient on SCORER) are revealed to lenders over time, experienced credit-scorers will begin to accept a larger portion of their loan applications (i.e., choose α more frequently), pushing loan default rates back toward their pre-credit-scoring levels (a positive coefficient on SCORINGEXP).

However, lenders that implemented credit-scoring models may have simultaneously altered their lending practices in other ways. For example, credit-scoring lenders may have also increased their loan approval rates (choosing α more frequently), because the diversification benefits accompanying a high-volume program of credit-scored and securitized small business loans made higher default rates more acceptable. In this case – not anticipated in our theoretical model which does not include loan portfolio effects – we would find a positive coefficient on SCORER. In this scenario we have no a priori expectation for the sign on SCORINGEXP.

SBA% is defined as the percentage of the outstanding balances guaranteed by the SBA and paid to the lending bank in the case of loan default. It is our empirical proxy for the "reduction in expected performance losses" concept from our theoretical model; in this case a reduction in credit losses due to a government guarantee. We expect a positive coefficient on this variable, as clearly predicted by our theory model – see the discussion that accompanies Figure 4.

We include a variety of loan-specific and lender-specific variables to control for variation in loan default not related to our main hypotheses. MATURITY3 and MATURITY7 are binary variables equal to one if the loan has a 3-year or 7-year contractual maturity (the excluded binary variable is MATURITY15).[14] Glennon and Nigro (2005) find that loan default rates vary significantly for SBA loans of different maturities. NEWFIRM is a binary variable equal to one if the borrower is 3-years old or less. We expect loans to these firms to default more frequently given the relative financial fragility of young firms. FIRMSIZE is equal to the number of full-time equivalent employees at the borrowing firm. We have no a priori expectations regarding the relationship between this variable and loan default because a large number of employees could indicate either economic robustness or input inefficiency.

HHI is the Herfindahl index calculated using the deposit-shares of banks and thrifts in the borrowing firm's local geographic market (an MSA or a rural county) and URBAN is a binary variable equal to one if the borrower is located in an MSA. HHI varies inversely with lending competition, and given that HHI tends to be substantially lower in urban areas than in rural areas, URBAN varies positively

---

[13] Because lender decision-making ability and information uncertainty are symmetric and offsetting in our theory model, we specify SCORER interactively with DISTANCE in some of our estimations.

with lending competition. We have no a priori expectations regarding the relationships between these market structure variables and loan default because competition has potentially offsetting price and output effects: by pushing down loan interest rates competition reduces the expected performance gains to lending, resulting in reduced lending at the margin (lower frequency of α) and thus lower default rates; but competition also creates incentives to increase output, resulting in more loans at the margin (higher frequency of α) and thus higher default rates. We specify HHI and URBAN interactively in our model.

JOBGROWTH_ORIG is the annualized rate of employment growth in the 3-digit SIC industry of the borrower within the borrower's home state during the quarter in which the loan was originated. JOBGROWTH_EVENT is the annualized rate of employment growth during the portion of the event history that has already passed as of time t. We would expect strong local and industry economic conditions to be negatively associated with loan default.

BANKSIZE is the total assets of the lending bank. We have no a priori expectations regarding the relationship between bank size and loan default: large size may reduce the probability of loan default by giving lenders access to more resources for evaluating loan applications and monitoring loans; or it may increase the probability of loan default if large size or organizational bureaucracy interferes with the bank-borrower relationships needed to acquire soft (non-quantifiable) information on borrow creditworthiness. CHARGEOFFS is the ratio of loans charged-off to total loans at the lending bank. We expect a positive relationship between this variable and loan default probability: a high value for this ratio likely indicates either poor loan underwriting practices or a policy of making high-risk loans. RESERVES is the ratio of loan loss reserves to total loans at the lending bank. We expect a negative relationship between this variable and loan default probability (holding loan charge-offs constant) if high loan loss reserves indicates a risk-averse lending policy.

Finally, to capture the anticipated concave shape of the loan default hazard function, we include a sixth-order polynomial of loan age, i.e., the number of quarters since loan origination. This flexible

[14] SBA loans can have various maturities: our sample is comprised of guaranteed loans with 3-year maturities (typically for working capital); 7-year maturities (the most common, often used to purchase equipment); and 15-year

specification is found by Glennon and Nigro (2005) to provide a good approximation of loan seasoning effects.[15]

5. Results

We estimate our discrete-time hazard model (7) for 33,531 SBA 7(a) loans originated between 1983 and 2001 for which we had full information, and for a smaller set of 23,143 loans originated between 1983 and 1997, a time period that better corresponds with our limited data on credit scoring (SCORER). The results from the full sample and subsample estimations are reported in Tables 3 and 4, respectively. To evaluate our three main hypotheses we constructed marginal default effects (reported at the bottom of the tables) based on the coefficient estimates in each regression as follows: we calculated the derivatives with respect to the three main test variables (lnDISTANCE, SCORER, and SBA%); evaluated these derivatives separately for each loan-quarter observation in our data; and took the unweighted averages of these evaluated derivatives across all observations. Greene (1997) shows that this procedure is preferred to the standard method of derivatives evaluated at the sample means, and that the two approaches are equivalent in large samples.

5a. Main hypothesis tests

We find a positive and statistically significant association between the probability of loan default and borrower-lender distance, although the effect is economically small. Based on the estimates in column [1], a doubling of borrower-lender distance (from 54 miles to 108 miles) is associated with only about a 2 percent increase in the probability of default (from 0.93% to 0.95%) at the means of the data.[16] The sign and magnitude of this result is robust across the eight regressions displayed in Tables 3 and 4. Moreover, this result is consistent with our theoretical model: the <u>positive</u> marginal effect is consistent

---

maturities (typically real estate loans).

[15] In alternate regressions, we included time fixed effects dummies to control for changes in broad economic, regulatory, and technological conditions not captured by the other right-hand-side variables. The results were qualitatively unchanged.

[16] The calculation is as follows: .000278*ln2 = .0001927, which is equal to approximately 2% of the mean default rate of .0093 shown in Table 1.

with the theoretical prediction that increased information uncertainty (i.e., increased DISTANCE) yields higher loan default rates in the short-run, while the small magnitude of this effect is suggestive of the long-run theoretical equilibrium in which lenders' reactions to increased information uncertainty yield an unchanged rate of loan default.

We also find a positive and statistically significant association between the probability of loan default and banking companies that credit score at least some of their small business loans. On average, loans originated by credit-scoring banks are between 15 percent (based on estimates from column [1]) and 24 percent (based on estimates from column [6]) more likely to default than average.[17] This result is not consistent with our theoretical model, which predicted that improved decision-making skills (i.e., SCORER=1) would reduce loan default rates in the short-run, and leave them unchanged in long-run equilibrium.

There is more than one way to interpret this unexpected result. For instance, if banks are applying credit scoring models in a fashion that reduces their decision-making abilities (that is, substituting an automated credit-scoring model that uses imperfect 'hard' information for an underwriting methodology that relies on the superior judgments of loan officers using 'soft' information), then these results are quite consistent with the short-run predictions of our theory. Or it may be the case that banks applying credit scoring technology are knowingly generating higher expected defaults (for the average loan) in exchange for some offsetting benefit related to credit scoring that we do not capture in our model, for example: better risk-pricing that generates higher interest revenues; larger loan volumes that generate portfolio diversification benefits; or the ability to securitize small business loans which generates loan origination fees, loan servicing fees, and allows the bank to use its scarce capital more efficiently.[18]

The regression specifications in columns [3], [4], [7], and [8] include the interaction term lnDISTANCE*SCORER, allowing us to test another prediction of our theoretical model: that improved

[17] For example, for column [1] the calculation is as follows: .001460*1 = .001460, which increases the mean default shown in Table 1 from .0093 to .0108, or about 16%.

decision-making ability (proxied here by SCORER) can potentially offset increased information uncertainty (proxied here by DISTANCE, and manifested in the data by higher default rates). The estimated coefficient on the interaction term is negative and statistically significant in all four regressions, indicating a default-rate tradeoff between borrower-lender distance and credit scoring, consistent with the theory. We have calculated this tradeoff for the data in column [3]: the marginal effect of lnDISTANCE(SCORER=0) equals 0.0003315 while the marginal effect of lnDISTANCE(SCORER=1) equals –0.0000903. Thus, for a doubling of borrower-lender distance, the probability of loan default increases by about 2.5 percent for lenders that do not credit score their small business loans, but *decreases* by about 0.7 percent for lenders that do use credit scoring for some of their small business loans. The net effect of doubling distance *plus* using credit scoring is essentially an economic wash: the probability of loan default declines on net from 0.930 percent to 0.924 percent. (Related to credit scoring, the credit-scoring experience variable SCOREXP is not statistically significant in any of the regressions.)

The results with respect to SBA% are consistent with our theoretical predictions that reductions in expected performance losses (i.e., higher loan guarantees) will yield higher loan default rates. On average, a ten percentage point increase in the loan guarantee rate (from 81% to 91%) is associated with between a 5 percent (based on estimates in column [5]) and a 9 percent (based on estimates in column [1]) increase in the probability of loan default.[19] This result has the most direct implications for small business lending policy, as it links greater government subsidies to poorer loan performance.

5b. Other results

The remainder of the coefficients reported in Tables 3 and 4 have reasonable signs and most are statistically significant. Small business loans with shorter maturities (MATURITY3 and MATURITY7) are more likely to default than loans with longer maturities (15-years). This may have to do with the nature of the amount and type of collateral at stake (long-term loans tend to be larger, and are secured by

[18] We acknowledge another possibility: lenders with historically high loan default rates may be more likely to adopt credit scoring models. We plan to investigate this possibility in future versions of the paper.

[19] For example, for column [1] the calculation is as follows: .008102*.10 = .0008102, which increases the mean default shown in Table 1 from .0093 to .0101, or about 9%.

land and buildings) or the fact that long-term loans are more likely to be securitized and hence the lender has reputational capital at stake.[20] Borrowers that are less than 3-years old at loan origination (NEWFIRM) were more likely to default on their loans than more mature small businesses. Holding these age effects constant, we find some weak evidence that larger borrowers (FIRMSIZE) were more likely to default.

We find evidence that lender competition is associated with higher loan defaults. The coefficient on HHI is never statistically significant (suggesting that increased concentration has no effect on loan default in rural markets), but the interaction term HHI*URBAN carries a large negative, statistically significant coefficient throughout (indicating that increased concentration as associated with lower default rates in urban markets). The latter result is consistent with two non-mutually exclusive explanations. First, standard microeconomic theory predicts that firms with market power will reduce output (i.e., low loan approval rates); hence only relatively "safe" loans will be originated and default rates will be low. Second, Petersen and Rajan (1998) argue that small business borrowers in concentrated markets have high switching costs, and as a result lenders are more likely to cultivate valuable relationships with these clients (e.g., careful monitoring that minimizes loan default rates) because they represent a relatively long stream of future cash flows. Although the coefficient on the URBAN dummy is positive and (marginally) significant, the impact of URBAN evaluated at the means of the data (i.e., $\beta_{HHI*URBAN}$*HHI + β_{URBAN}) remains negative, consistent with lower loan default rates in the typically more competitive urban markets. (We note that this evidence of higher loan default rates in competitive markets does not necessarily mean that competition is welfare-reducing, because a larger total number of loans are likely to be originated in competitive markets.)

Robust economic activity at the time of loan origination (JOBGROWTH_ORIG) and during the life of the loan (JOBGROWTH_EVENT) are both associated with reductions in loan defaults. Large

---

[20] The average 15-year loan in our sample was $253,000, compared to $131,000 for 7-year loans and $57,000 for 3-year loans. About 26 percent of the 15-year loans in our sample were sold by the original lender, compared to about 18 percent of 7-year loans and less than 2 percent of 3-year loans.

lenders (BANKSIZE) have lower loan default rates, perhaps because these lenders can afford to attract and retain high-quality staff that specialize in underwriting and monitoring loans – or perhaps because small banks have a limited array of financial products and hence face greater pressure to grow their loan portfolios. All else equal, banks that have written off large amounts of bad loans in the recent past (CHARGEOFFS) are more likely to originate small business loans that eventually default (perhaps indicating a high tolerance for credit risk), and banks with high levels of loan loss reserves (RESERVES) are less more likely to originate small business loans that eventually default (perhaps indicating a low tolerance for insolvency risk).

6. Conclusions and implications for policy

Numerous studies have documented that small business borrowers and their bank lenders have moved further apart over the past two decades. There are (at least) three phenomena at work here: banking industry consolidation greatly reduced the number of bank lenders; communications technology allowed financial information to flow more cheaply and accurately across longer distances; and new decision-making technologies permit lenders to analyze this information at arms-length from their borrowers. We model these phenomena using a theoretical model of decision-making under risk and uncertainty, and this framework yields testable implications about the impact of borrower-lender distance, credit-scoring technologies, and government loan subsidies on the performance of small business loans. We test these implications for a random sample of 33,531 small business loans made under the SBA 7(a) loan program between 1983 and 2001. We believe this is the first study to test the impact of borrower-lender distance, credit-scoring models, and the tradeoff between these two phenomena on the probability of loan default.

We stress that our empirical results are preliminary. Nonetheless, we find substantial support in the data for the predictions of our theoretical model. Loan defaults increase with borrower-lender distance, and the adoption of credit-scoring models dampens this relationship. On average, lenders that use credit-scoring models experience higher default rates than those that do not, which suggests that benefits from high-volume, credit-scored and securitized lending strategies (e.g., diversification, fee generation, recycling of equity capital) offset the costs of higher expected default rates. More generous government loan guarantees are associated with higher loan default rates, as are competitive conditions in local lending markets. These last two findings have obvious implications for public policy: both illustrate the tradeoff between policies that increase the quantity of small business credit (providing government subsidies, encouraging market competition) and the information problems resulting in inefficient allocation of resources to uncreditworthy borrowers.

References

Akhavein, J.D., W.S. Frame, and L.J. White. 2001. "The Diffusion of Financial Innovations: An Examination of the Adoption of Small Business Credit Scoring by Large Banking Organizations," working paper, Federal Reserve Bank of Atlanta and Stern School of Business, New York University.

Berger, A.N., C. Buch, G. DeLong, and R. DeYoung. 2004. "The Comparative Advantages of Nations at Exporting Financial Institutions Management via M&As," *Journal of International Money and Finance* 23(3): 333-366.

Berger A.N. and R. DeYoung. 2001. "The Effects of Geographic Expansion on Bank Efficiency," Journal of Financial Services Research 19: 163-184.

Berger A.N. and R. DeYoung. 2004. "Technological Progress and the Geographic Expansion of the Banking Industry," Federal Reserve Bank of Chicago, Working Paper 2002-07 (updated in 2004).

Berger, A.N., R. DeYoung, H. Genay, and G.F. Udell. 2000. "Globalization of Financial Institutions: Evidence from Cross-Border Banking Performance." *Brookings-Wharton Papers on Financial Services* 3: 23-158.

Berger A.N. and G.F. Udell. 1998. "The economics of small business finance: The roles of private equity and debt markets in the financial growth cycle," *Journal of Banking and Finance* 22: 613-673.

Beshouri, C. and D. Glennon. "CRA as Market Development or Tax: An Analysis of Lending Decisions and Economic Development," *Proceedings of a Conference on Bank Structures and Competition,* Federal Reserve Bank of Chicago, (1996) 556-585.

Brevoort, K.P., and T.H. Hannan. 2004. "Commercial Lending and Distance: Evidence from Community Reinvestment Act Data," Federal Reserve System Board of Governors, FEDS working paper 2004-24.

Buch, C.M. 2003. "Information or Regulation: What is Driving the International Activities of Commercial Banks?" *Journal of Money, Credit, and Banking* 35(6): 851-870.

Chang, C.E., I. Hasan, and W.C. Hunter. 1998. "Efficiency of Multinational Banks: An Empirical Investigation," *Applied Financial Economics* 8(6): 1-8.

Cyrnak A. and T. Hannan. 2000, "Non-Local Lending to Small Business," Board of Governors, manuscript.

Degryse, H., and S. Ongena. 2002. "Distance, Lending Relationships, and Competition," University of Tilburg, working paper.

Dell'Ariccia, G. and R. Marquez. "Information and Bank Credit Allocation," *Journal of Financial Economics*, forthcoming.

Davis, S., J.Haltiwanger, and S.Schuh (1994), "Small Business and Job Creation: Disecting thte Myth and Reassessing the Facts," *Labor Markets, Employment, and Job Creation*, edited by L. Solomon abd A. Leoinson, Boulder, CO: Westview Press, 13-32.

Davis, S., J.Haltiwanger, and S.Schuh. 1996. *Job Creation and Destruction*, Cambridge, MA, MIT Press.

DeYoung, R., W.C. Hunter, and G.F. Udell. 2004. "The Past, Present, and Probable Future for Community Banks," *Journal of Financial Services Research* 25(2/3): 85-133.

DeYoung, R. and D.E. Nolle. 1996. "Foreign-Owned Banks in the U.S.: Earning Market Share or Buying It?" *Journal of Money, Credit, and Banking* 28(4): 622-636.

Engelmann, B., E. Hayden, and D. Tashe. "Measuring the Discriminative Power of Rating Systems," Deutsche Bundesbank, Discussion Paper Series 2: Banking and Financial Supervision, No. 01/2003, (2003).

Glennon, D. and P. Nigro. 2005. "An Analysis of SBA Small Business Loan Defaults by Maturity Structure." Office of the Comptroller of the Currency, working paper.

Green, D.M. and J.A. Swets. *Signal Detection Theory and Psychophysics*, John Wiley and Sons, Inc. New York, NY, (1974).

Greene, W.H. 1997. *Econometric Analysis*., 3rd edition. Upper Saddle River: Prentice Hall.

Heiner, Ronald A. "The Origin of Predictable Behavior," *American Economic Review,* vol. 73, (1983): 560-93.

Heiner, Ronald A. "Origin of Predictable Behavior: Further Modeling and Applications," *American Economic Review*, vol. 75, (May, 1985): 391-96.

Heiner, Ronald A. "Uncertainty, Signal Detection Experiments, and Modeling Behavior," in R. Langlois, ed. The New Institutional Economics, New York: Cambridge University Press, (1985a).

Heiner, Ronald A. "The Economics of Information when Decisions are Imperfect," in Alan J. and Heather W. MacFadyen, ed. Economic Psychology: Intersections in Theory and Application, Elsevier Science Publications BV (North-Holland), (1986): 291-350.

Knight, R.H. *Risk, Uncertainty, and Profit.* Boston: Houghton-Mufflin, (1933).

Petersen, M.A. and R.G. Rajan. 1998.

Petersen, M.A. and R.G. Rajan. 2002. "The Information Revolution and Small Business Lending: Does Distance Still Matter?" *Journal of Finance*, forthcoming.

Rose, P.S., W.L. Scott. 1984. "Heterogeneity in Performance Within the Bank Holding Company Sector: Evidence and Implications, *Journal of Economics and Business* 36: 1-14.

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

Stein, R.M. "Power, Profitability, and Prices: Why Powerful Models Increase Profits and How to Define a Lending Cutoff If You Must," Technical Report #021223, Moody's KMV, NY (2003).

Wolken, J.D., and Rohde, D. 2002. "Changes in the Location of Small Businesses' Financial Service Suppliers between 1993 and 1998." Federal Reserve Board internal memorandum.

Table 1
Random sample of 35,991 SBA 7(a) loans to small businesses between 1983 and 2001.
Mean data by year.

Year	**# of loans**	**% of loan originated in year that:**		**percentage of outstanding loan balances guaranteed by the SBA**	**borrower-lender distance in miles**		**# of banking companies using credit scoring for small business loans**
		defaulted	**prepaid**		**mean**	**median**	
1983	23	26.09	52.17	88.96	17.34	7.96	0
1984	848	27.59	55.07	86.94	31.38	6.12	0
1985	676	23.37	56.66	87.37	28.59	6.25	0
1986	1,011	23.84	51.83	85.31	29.59	6.01	0
1987	1,016	22.15	49.41	84.65	43.30	6.85	0
1988	880	23.75	47.84	84.20	29.75	6.58	0
1989	1,011	19.39	64.89	84.56	30.10	6.92	0
1990	1,068	20.97	62.92	84.84	29.77	6.60	0
1991	1,166	15.27	61.66	84.76	30.04	7.81	1
1992	1,722	13.59	63.12	84.54	29.44	7.21	3
1993	1,801	12.10	59.47	84.56	28.30	7.60	5
1994	2,697	16.69	54.99	83.92	32.98	8.32	9
1995	4,627	18.65	53.04	86.27	41.06	8.74	22
1996	2,945	19.22	48.63	79.14	61.54	9.81	51
1997	3,469	17.67	43.13	77.46	94.63	11.93	62
1998	3,366	13.40	33.39	74.41	171.42	15.76	62
1999	3,194	8.74	27.80	70.51	171.58	20.07	62
2000	3,155	8.69	17.34	70.43	205.53	19.14	62
2001	1,316	5.62	9.57	68.84	323.81	31.14	62
1983-2001 (weighted averages)	*35,991*	*17.53%*	*48.64%*	*79.74%*	*92.62 miles*	*11.93 miles*	--

Table 2
Summary statistics for variables used in estimation of equation (7). Random sample of 33,531 SBA 7(a) loans to small businesses between 1983 and 2001 for which we have full information.

variable name	definition	data source	full sample 1983-2001		subsample 1983-1997	
			mean	std	mean	std
Default	Dependent Variable. = 1 if loan i defaulted at time t.	SBA	0.0093	0.0964	0.0095	0.0971
DISTANCE	= straight-line distance in miles between borrower and the lending office of the lending institution.	SBA, Call Report	54.14	246.14	26.34	93.57
lnDISTANCE	= natural log of DISTANCE.	SBA, Call Report	2.1050	1.9104	1.8717	1.7760
SCORE	= 1 if lender used credit scoring models to evaluate at least some of its small business loans at t=0.	Akhavein, Frame, and White (2001)	0.1130	0.3167	0.0476	0.2130
SCOREXP	= number of quarters since lender began using credit scoring for small business loans.	Akhavein, Frame, and White (2001)	1.3588	4.3703	0.3002	1.5727
SBA%	= percentage of outstanding loan balance guaranteed by the SBA.	SBA	0.8108	0.0941	0.8338	0.0721
MATURITY3	= 1 for 3-year loans.	SBA	0.0835	0.2767	0.0720	0.2585
MATURITY7	= 1 for 7-year loans.	SBA	0.6278	0.4834	0.6073	0.4883
NEWFIRM	= 1 if borrower is 3-years old or less.	SBA	0.3101	0.4625	0.3002	0.4584
FIRMSIZE	= number of full-time employees at borrowing firm.	SBA	12.78	114.31	13.36	119.33
HHI	= deposit-based Herfindahl index in local market of the borrower.	FDIC Summary of Deposits	0.1977	0.1198	0.2046	0.1236
URBAN	= 1 if borrower is located in a Metropolitan Statistical Area.	Call Report	0.8039	0.3971	0.7840	0.4115
JOBGROWTH_ORIG	= percent employment growth in the borrower's industry and home state during the three quarters prior to t=0.	BEA	0.0052	0.0283	0.0059	0.0288
JOBGROWTH_EVENT	= percent employment growth in the borrower's industry and home state during the life of the loan (0 to t).	BEA	0.0147	0.0116	0.0141	0.0118
BANKSIZE	= lender assets (billions of 2001 $).	Call Report	12.97	2.36	12.60	2.09
CHARGEOFFS	= ratio of lender's loan charge-offs to assets.	Call Report	0.0039	0.0072	0.0042	0.0076
RESERVES	= ratio of lender's loan loss reserves to assets (x100).	Call Report	1.6655	0.9069	1.7146	0.9673
LOANAGE	= number of quarters since loan origination.	SBA	11.34	8.88	12.52	9.36
Number of loan-quarters		SBA	545,297		433,120	
Number of unique loans		SBA	33,531		23,413	
Number of unique lenders		SBA	5,552		4,921	

Table 3

Results from discrete-time hazard model estimation (stacked-logit) estimation of equation (7). Dependent variable is loan default. Random sample of 33,531 SBA 7(a) loans to small businesses between 1983 and 2001 for which we have full information. All variable definitions are displayed in Table 2. * and ** indicate statistical significance at the 5 percent and 1 percent levels, respectively.

	[1]		[2]		[3]		[4]	
Variable	coefficient	std. err.	coefficient	std. err.	coefficient	std. err.	coefficient	std. err.
Intercept	-5.5136**	0.1954	-8.9312**	0.2554	-5.4766**	0.1978	-8.9345**	0.2574
lnDISTANCE	0.0299**	0.00822	0.0316**	0.00825	0.0378**	0.00894	0.0392**	0.00896
SCORE	0.1572**	0.0559	0.1667**	0.0558	0.3457**	0.1037	0.2685*	0.1052
SCORE*lnDISTANCE					-0.0481*	0.0225	-0.0527*	0.0226
SCOREXP					-0.00261	0.00714	0.00768	0.00725
SBA%	0.8719**	0.1799	0.7571**	0.1814	0.8031**	0.1868	0.7559**	0.1885
MATURITY3	0.6033**	0.0597	0.6368**	0.0614	0.6035**	0.0599	0.6313**	0.0616
MATURITY7	0.6525**	0.0387	0.606**	0.0398	0.6544**	0.0387	0.6075**	0.0398
NEWFIRM	0.2108**	0.0295	0.216**	0.0295	0.2094**	0.0295	0.2145**	0.0295
FIRMSIZE	0.000156*	0.000077	0.000161*	0.000078	0.000156*	0.000077	0.000163*	0.000078
HHI	0.1291	0.2087	0.1389	0.2085	0.1307	0.2086	0.139	0.2084
HHI*URBAN	-0.9189**	0.2745	-0.9393**	0.2745	-0.9328**	0.2745	-0.9444**	0.2745
URBAN	0.1689*	0.0772	0.17*	0.0772	0.165*	0.0772	0.1674*	0.0772
JOBGROWTH_ORIG	-1.995**	0.5499	-1.7157**	0.5472	-1.9905**	0.5495	-1.706**	0.5472
JOBGROWTH_EVENT	-5.2659**	1.1722	-5.6093**	1.1693	-5.2465**	1.1732	-5.5636**	1.17
BANKSIZE	-0.0288**	0.00799	-0.027**	0.008	-0.0279**	0.00806	-0.0276**	0.00806
CHARGEOFFS	5.6842**	1.843	5.9577**	1.8383	5.8056**	1.8419	5.9964**	1.839
RESERVES	-0.0671**	0.0181	-0.0726**	0.0181	-0.0695**	0.0181	-0.0738**	0.0182
LOANAGE			1.2836**	0.0765			1.2825**	0.0764
$LOANAGE^2$			-15.6394**	1.2423			-15.6138**	1.2421
$LOANAGE^3$			90.5828**	9.231			90.3838**	9.2293
$LOANAGE^4$			-270.4**	33.7569			-269.7**	33.7472
$LOANAGE^5$			397**	58.6547			395.9**	58.6317
$LOANAGE^6$			-2.2609**	0.3852			-2.254**	0.385
N	545297		545297		545297		545297	
D=1	5120		5120		5120		5120	
D=0	540177		540177		540177		540177	
Marginal effects:								
lnDISTANCE	.000278		.0002927		.0002976		.0003049	
SCORER=0					*.0003315*			
SCORER=1					*-.0000903*			
SCORER	.001460		.0015461		.0021992		.0015192	
SBA%	.008102		.0070206		.0074625		.0070094	

Table 4

Results from discrete-time hazard model estimation (stacked-logit) estimation of equation (7). Dependent variable is loan default. Subsample of 23,413 SBA 7(a) loans to small businesses between 1983 and 1997 for which we have full information. All variable definitions are displayed in Table 2. * and ** indicate statistical significance at the 5 percent and 1 percent levels, respectively.

	[5]		[6]		[7]		[8]	
Variable	coefficient	std. err.	coefficient	std. err.	coefficient	std. err.	coefficient	std. err.
Intercept	-5.134**	0.2371	-8.6953**	0.3049	-5.1341**	0.2375	-8.695**	0.3052
lnDISTANCE	0.0282**	0.00936	0.0283**	0.00938	0.0343**	0.00969	0.0344**	0.0097
SCORE	0.2014**	0.0754	0.2053**	0.0755	0.4284**	0.1578	0.4264**	0.1578
SCORE*lnDISTANCE					-0.0963*	0.0375	-0.0978**	0.0377
SCOREXP					0.00728	0.0184	0.009	0.0184
SBA%	0.5186*	0.2363	0.5903*	0.2376	0.5042*	0.2369	0.5763*	0.2382
MATURITY3	0.615**	0.0687	0.6101**	0.0711	0.6154**	0.0687	0.6108**	0.0711
MATURITY7	0.7208**	0.0421	0.6603**	0.0435	0.7217**	0.0421	0.6615**	0.0435
NEWFIRM	0.2194**	0.033	0.2207**	0.0331	0.2179**	0.0331	0.2192**	0.0331
FIRMSIZE	0.000136	0.000086	0.000146	0.000086	0.000141	0.000085	0.000152	0.000086
HHI	0.0425	0.2227	0.0484	0.2224	0.0461	0.2225	0.0516	0.2223
HHI*URBAN	-1.0515**	0.2997	-1.053**	0.2995	-1.0607**	0.2998	-1.0617**	0.2996
URBAN	0.1811*	0.0836	0.1826*	0.0836	0.1798*	0.0835	0.1813*	0.0835
JOBGROWTH_ORIG	-1.4802*	0.6004	-1.4578*	0.5974	-1.4819*	0.6008	-1.4589*	0.5977
JOBGROWTH_EVENT	-4.2703**	1.299	-4.2292**	1.3006	-4.2208**	1.3005	-4.183**	1.3021
BANKSIZE	-0.0363**	0.00902	-0.0379**	0.00905	-0.0361**	0.00904	-0.0377**	0.00906
CHARGEOFFS	6.142**	1.9489	6.4489**	1.9388	6.1933**	1.9486	6.5006**	1.9386
RESERVES	-0.0753**	0.0193	-0.0769**	0.0193	-0.0766**	0.0193	-0.0783**	0.0193
LOANAGE			1.3095**	0.0862			1.3089**	0.0862
$LOANAGE^2$			-15.9791**	1.3679			-15.9689**	1.3678
$LOANAGE^3$			91.8342**	9.9585			91.758**	9.957
$LOANAGE^4$			-270.8**	35.7922			-270.5**	35.7862
$LOANAGE^5$			392.3**	61.3037			391.9**	61.2917
$LOANAGE^6$			-2.2053**	0.3978			-2.2027	0.3977
N	433120		433120		433120		433120	
D=1	4123		4123		4123		4123	
D=0	428997		428997		428997		428997	
Marginal effects:								
lnDISTANCE	.0002655		.0002728		.0002657		.0002726	
SCORER	.0018968		.0022900		.0019300		.0022436	
SBA%	.0048836		.0047483		.0055486		.0054165	

Figure 1

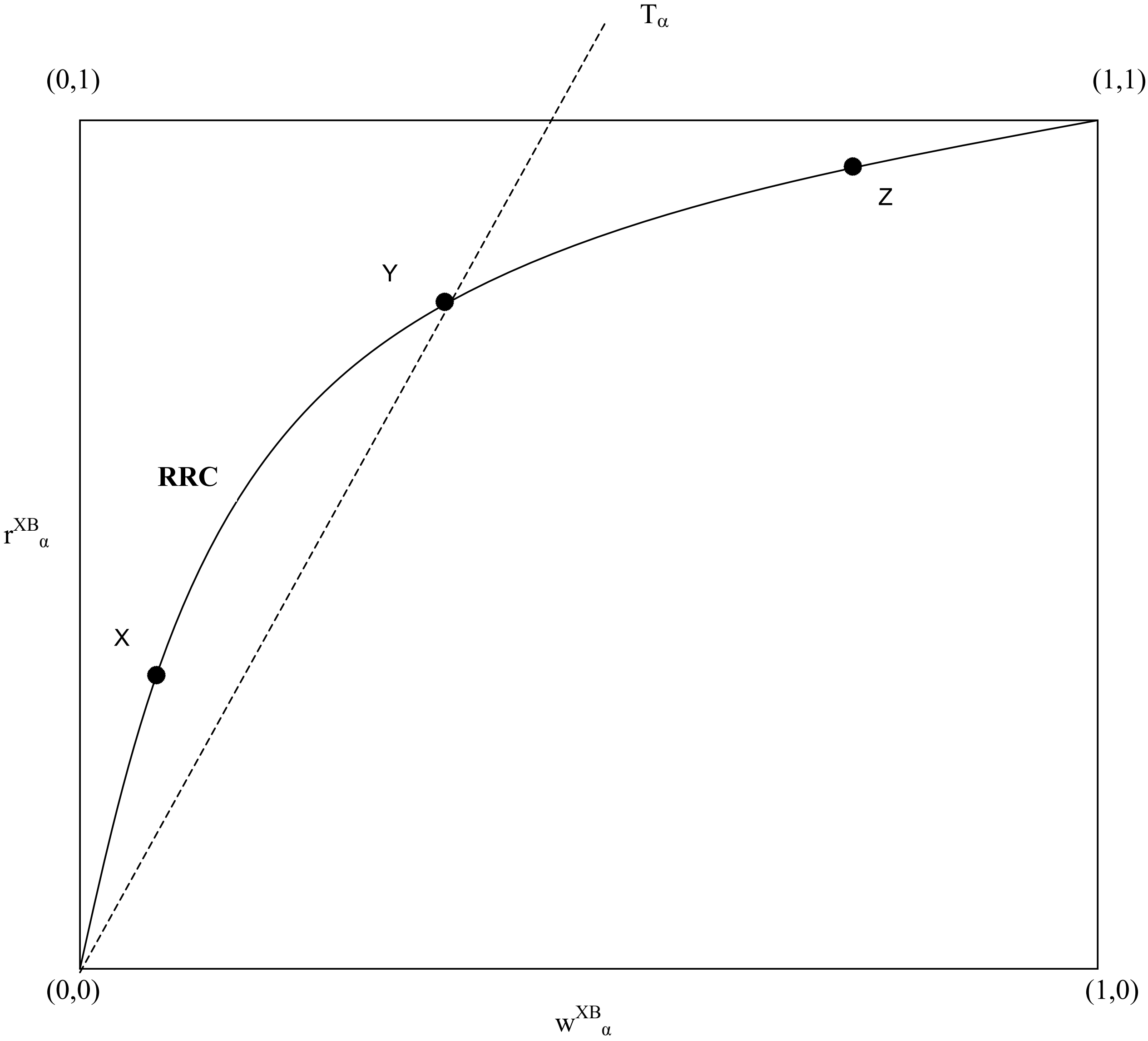

Figure 2

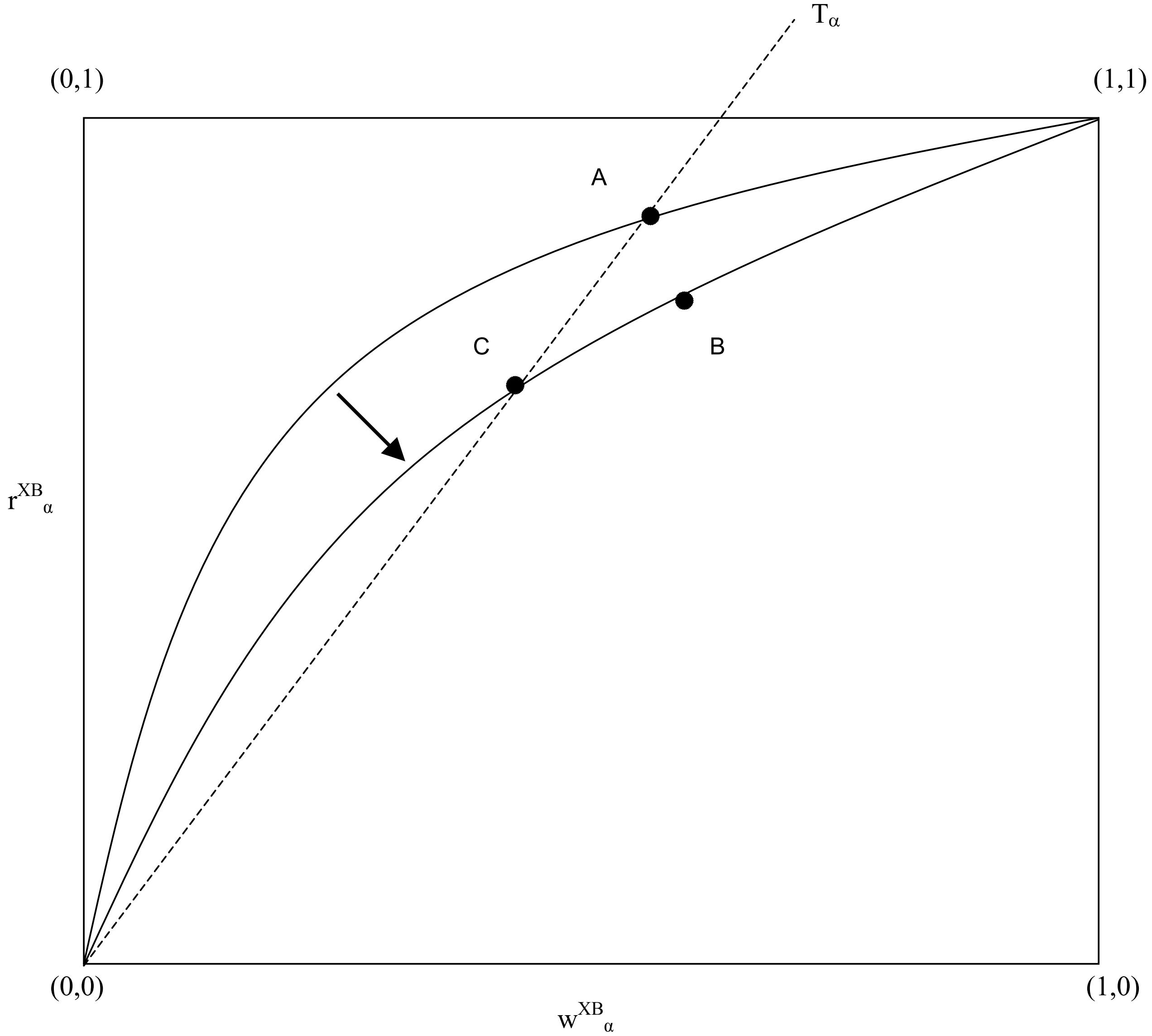

Figure 3

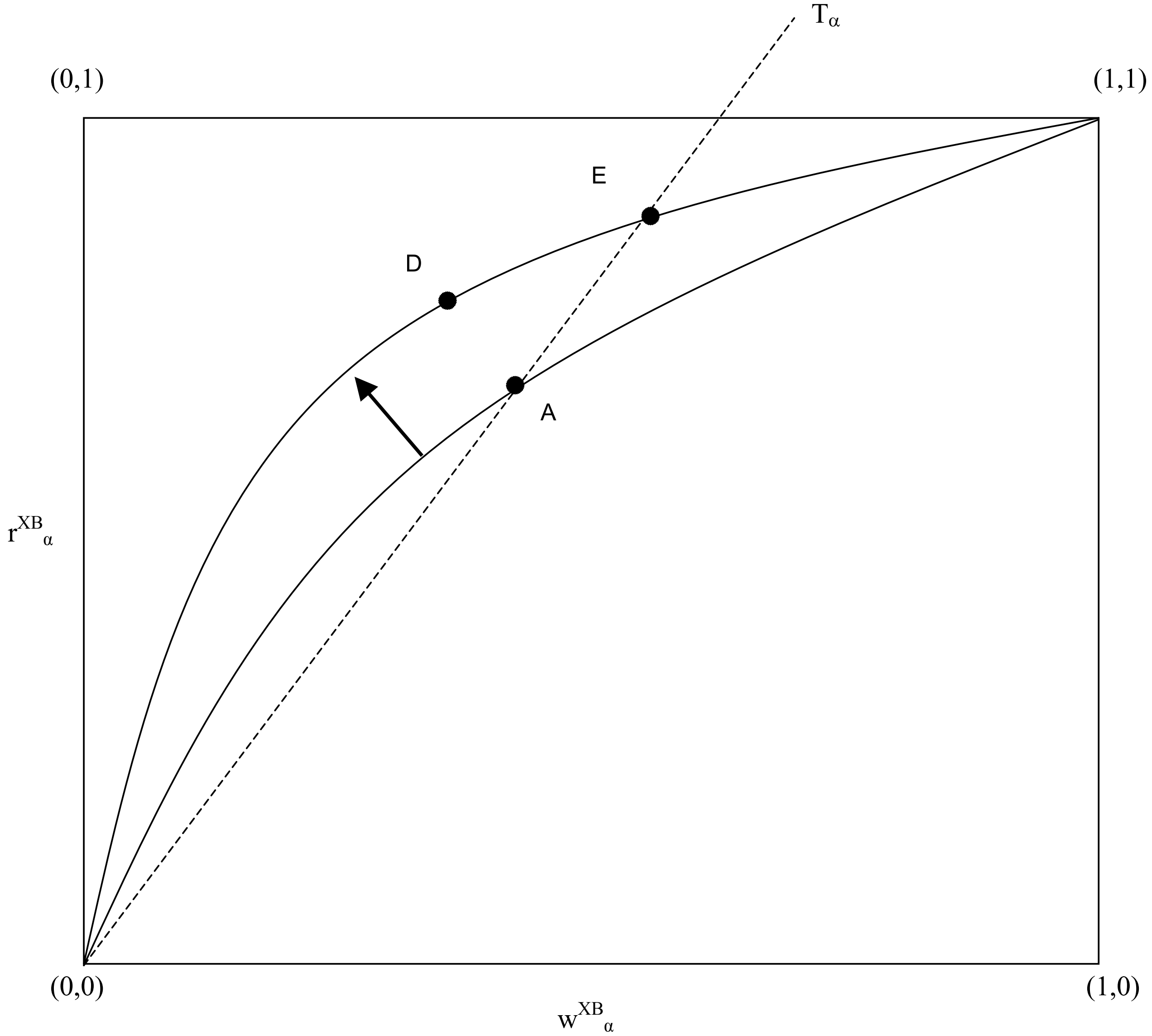

Figure 4