Available online at www.sciencedirect.com
ScienceDirect
J. Finan. Intermediation 17 (2008) 113–143
Journal of Financial Intermediation
www.elsevier.com/locate/jfi

Borrower–lender distance, credit scoring, and loan performance: Evidence from informational-opaque small business borrowers ☆

Robert DeYoung [a,*], Dennis Glennon [b], Peter Nigro [c]

[a] *University of Kansas, Lawrence, KS, USA*
[b] *Office of the Comptroller of the Currency, Washington, DC, USA*
[c] *Bryant University, Smithfield, RI, USA*

Received 6 June 2006

Available online 21 September 2007

Abstract

Over the past decade, the distances between small businesses and their bank lenders have increased substantially, as increasing numbers of bank lenders have implemented credit-scoring models to evaluate the creditworthiness of small businesses. These developments are antithetical to the traditional small business lending process, which emphasizes local proximity, bank–borrower relationships, and qualitative information. We theoretically model and empirically test whether and how these changes have affected the probability that small business loans default, using 1984–2001 data from the SBA's flagship lending program. On average, both distance and credit scoring are associated with higher default probabilities—the former suggests that distance interferes with information collection, while the latter suggests that production efficiencies encourage credit-scoring lenders to expand output by making riskier loans at the margin. The default-increasing effects of distance are substantially dampened at credit scoring banks, however, suggesting that hard-information lending approaches may outperform soft-information, relationship-based lending approaches in long-distance situations.

JEL classification: D81; G21; G28

☆ The views expressed here are those of the authors and do not necessarily reflect the views of the Office of the Comptroller of the Currency or the US Treasury Department.

* Corresponding author.
E-mail addresses: rdeyoung@ku.edu (R. DeYoung), dennis.glennon@occ.treas.gov (D. Glennon), pnigro@bryant.edu (P. Nigro).

1042-9573/$ – see front matter
doi:10.1016/j.jfi.2007.07.002

Keywords: Borrower–lender distance; Credit scoring; Small business loans

1. Introduction

Small business finance has traditionally been a local and close-knit affair. Small firms, whose informational opacity precludes them access to public capital markets, seek out local bank lenders whose geographic proximity allows them to observe and accumulate the "soft," non-quantitative information necessary to assess these firms' creditworthiness (Meyer, 1998; Stein, 2002; Scott, 2004). This relationship-based approach remains the method used by many, if not most, community banks to underwrite their small business loans today. But some exceptions to these tight location-based credit relationships began to emerge during the 1990s, when the geographic distance between small business borrowers and their commercial bank lenders began to increase (Cyrnak and Hannan, 2000; Degryse and Ongena, 2002; Petersen and Rajan, 2002; Wolken and Rohde, 2002; Brevoort and Hannan, 2004). At the same time, an increasing number of large banks began to evaluate small business loan applications using credit scoring models, which rely on "hard," quantitative information rather than soft, relationship-based information (Mester, 1997; Akhavein et al., 2005).

This apparent decline in the importance of borrower–lender proximity and relationships has potential implications for the supply and quality of small business credits; for the optimal trade-offs among information quality, customer service, loan production costs, and bank scale; and for the strategies banks use to offer small business credit.[1] All else equal, greater geographic distances between informationally opaque firms and their bank lenders should increase both the cost of traditional bank lending and the credit supply-reducing consequences of adverse selection—e.g., bankers will make fewer in-person visits because of high travel expenses, resulting in less accurate information, poorer credit assessments, higher default rates, and higher loan losses. Theoretically, the application of small business credit scoring models could either dampen or exacerbate these outcomes: On the one hand, credit scoring may improve the quality of banks' information about borrower creditworthiness; on the other hand, the automated nature of credit-scored lending processes may generate scale economies or other ancillary benefits that encourage banks to ramp-up loan output by lending to marginally less creditworthy borrowers.

At the same time, public policy in the US encourages banks to make loans to marginally less creditworthy small business borrowers. Chief among these policies is the Small Business Administration (SBA), which provides credit enhancements for small businesses unable to qualify for loans in regular credit markets. SBA-endorsed lenders (usually commercial banks) select the firms to receive loans, initiate the involvement of the SBA, and then underwrite the loans within SBA program guidelines. The SBA extends a partial guarantee in which the SBA shares a large portion of the loss on a *pro rata* basis with the lender. As a result, lenders have (perhaps reduced) incentives to screen applicants for creditworthiness, monitor borrowers on an ongoing basis, and set appropriate loan interest rates and contract terms. The data suggest that these partial government guarantees are instrumental in providing credit to small business borrowers. In 2001 the SBA reported a combined managed guaranteed loan portfolio of over $50 billion; in comparison,

[1] On the latter point, Boot and Thakor (2000) and DeYoung et al. (2004) offer analyses suggesting that advances in financial markets, instruments, and banking technology are driving the degree to which banks specialize in and or combine relationship lending with transactional lending.

the entire portfolio of small business loans (principle less than $250,000) held by US commercial banks in 2001 totaled about $120 billion.[2] However, SBA loan programs do little to solve the underlying, often severe information problems associated with these opaque small firms—not surprisingly, researchers have found positive correlations between the percentage of the loan guaranteed by the SBA and the probability of SBA loan default (Glennon and Nigro 2003, 2005).

This study investigates how increased distance between bank lenders and their small business borrowers affects the performance of SBA-guaranteed loans, and whether new lending technologies and/or the size of government loan guarantees mitigate or exacerbate those distance-related effects. We estimate the probability of loan default for a random sample of 29,577 loans made by US commercial banks between 1984 and 2001 under the Small Business Administration's flagship 7(a) loan program. Like the previous literature, we find that more generous SBA subsidies are associated with higher loan default rates. However, controlling for the level of SBA subsidy, we find two new and intriguing results. First, we find that greater borrower–lender distance is associated with higher default rates at non-credit-scoring banks, but on average distance has no impact on default rates at credit-scoring banks. This result suggests that credit-scoring models help mitigate the information problems associated with geographically distant borrowers. Second, we find that credit-scoring is on average associated with higher default rates. This perhaps unconventional result implies that scale-related cost reductions associated with automated small-business lending technology increase the profitability of the marginal loan application, which in turn creates incentives for lenders to expand their operations by approving lower-quality applicants.

We acknowledge that these findings—derived as they are from partially guaranteed SBA loans to firms that by definition do not have access to regular credit markets—will not generalize seamlessly to non-government-guaranteed small business lending. However, we argue that our empirical findings do have implications that stretch beyond the SBA program, and do provide insight into the design and performance of small business credit in general. The differences between subsidized and non-subsidized small business lending are not as fundamental as one may initially believe. First, the SBA guarantee allows lenders to put only a portion of their loan losses to the government—the guarantee represents a risk-sharing arrangement between the lender and the SBA. So long as SBA lenders incur some non-trivial share of potential loan default costs, they face the same qualitative tradeoffs between risk and return as do non-subsidized lenders. Second, the SBA put option is simply a hedge against credit risk, and is qualitatively similar to credit-insurance protection provided by third-party providers (e.g., mortgage insurers) or credit derivative products on non-guaranteed loans. Third, geographic distance confers *ex ante* information-gathering frictions on both subsidized and non-subsidized lenders independent of *ex post* risk allocation; both types of lenders face the prospect that increased borrower–lender distance will make loans more likely to fail, and that credit-scoring techniques could potentially mitigate or exacerbate these information problems. Finally, we stress that all of our empirical results regarding borrower–lender distance and credit scoring are derived holding constant the magnitude of the SBA loan guarantee, which as a first approximation statistically neutralizes the impact of this put option.

[2] This $120 billion total includes much of the $50 total of SBA loans. For example, in 2001 roughly 85 percent of the loans under SBA management were in amounts less than $250,000.

2. Background and relevant literature

Because it is difficult for investors in public capital markets to assess the financial condition and creditworthiness of small businesses, these firms depend disproportionately on private debt finance (Bitler et al., 2001; Petersen and Rajan, 1994; Berger and Udell, 1998).[3] Much of this funding is provided by nearby community bank lenders, located close enough to have personal knowledge of the firms' owners and managers (and often their suppliers and customers), to be a major participant in the local business environment, and to make frequent on-site visits. There is a growing body of academic work on small business lending and community banking; Berger and Udell (1998) and DeYoung et al. (2004) provide broad reviews of this literature. In this section we focus more closely on the recent increases in small borrower–lender distance and small business credit scoring alluded to above; the potential substitution of automated credit scored lending for high-touch relationship lending to small businesses; and the likely effects of these developments on the quantity and quality of bank loans to small businesses.

2.1. Borrower–lender distance

A number of recent studies demonstrate that the geographic distances between banks and their small business borrowers increased during the past two decades. For example, Petersen and Rajan (2002) estimated that the distances between US 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.[4] Consistent with this trend, Cyrnak and Hannan (2000) found that the share of small business loans made by out-of-market banks in the US approximately doubled between 1996 and 1998. 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.

Borrower–lender distance has been linked both theoretically and empirically to banking business strategies and banking industry structure. Using detailed spatial data from commercial loans made under the Community Reinvestment Act in 1997 through 2001, Brevoort and Hannan (2004) found that, within metropolitan areas, banks became less likely to lend as borrower–lender distance increased; moreover, this effect grew stronger over time and was strongest for smaller banks. Berger et al. (2005b) found that large banks are more likely than small banks to make long-distance loans, and are less likely to foster long-lasting relationships with those borrowers. These findings are consistent with recent theories that increased competition in lending markets 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, 2004).[5]

Improvements in communications technology (fax machines, the Internet), greater information availability (credit bureaus), and increased capacity to analyze information (personal com-

[3] According to Bitler et al.'s analysis of the 1998 Survey of Small Business Finances, 39 percent of small business respondents had a loan, a credit line, or a capital lease from a commercial bank. Finance companies, the second largest supplier, were used by only 13 percent of the respondents.

[4] 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.

[5] Ergungor (2005) concludes that increased competition has also reduced the risk-adjusted profitability of relationship lending at community banks in recent years.

puters, financial software, credit scoring) have facilitated faster and more accurate information flows from borrowers to lenders. On the one hand, these innovations have reduced the frequency of in-person visits by improving loan officers' ability to perform off-site screening and monitoring of small business borrowers. On the other hand, these innovations have made credit analysis more portable: the nexus of laptop computers, spreadsheet programs, and Internet connections has increased the productivity of loan officers' in-person visits to borrowers. By reducing the costs associated with borrower–lender distance, these developments have likely contributed to increases in these distances in recent years. Indeed, Hannan (2003) found that banks specializing in credit card lending (i.e., banks most experienced with credit scoring) accounted for the bulk of the increase in out-of-market small businesses lending in the US between 1996 and 2001.[6]

Although to our knowledge no previous study has tested whether borrower–lender distance affects loan performance, some studies have examined whether and how geographic distance impacts the performance of banks and other firms. Using data from US multi-bank holding companies during the 1990s, Berger and DeYoung (2001, 2006) concluded that geographic dispersion impedes the ability of senior headquarters managers to control the actions of local loan officers; they also found that this result dissipated over time, suggesting that technological progress allowed banks to partially mitigate distance-related control problems. Using division-level data from US corporations between 1991 and 2003, Landier et al. (2005) found that employees in divisions located geographically further from the corporate headquarters were more likely to be laid off; they also found that this relationship was weaker in "hard information intensive" industries, an indication that ease-of-information-transfer influences geographically distant business decisions. Using data on corporate loans made by a foreign bank in Argentina in 1998, Liberti (2005) found that soft information was more important for loans approved or denied at the local branch, while hard information was more important for loans approved or denied by loan officers located further away. These findings have obvious parallels for the costs associated with borrower–lender distance and the efficacy of credit-scoring lending technologies.[7]

2.2. *Small business credit scoring*

Credit scoring models use statistical techniques to transform quantifiable information—sometimes referred to as "hard" information—about a borrower (e.g., her income, debt load, financial assets, employment history, and credit history) or a small business (business credit reports, company financial ratios, sales figures, corporate structure, and industry identity) into a numerical "credit score" that ranks individual borrowers based on their likelihood of defaulting on a loan. Banks first used credit scores to screen applicants for consumer loans (credit cards, auto loans) and quickly became standard tools in mortgage lending.[8] The key innovation that spurred banks to apply credit-scoring models to small business lending was the discovery that a small business owner's *personal* credit information was a strong predictor of her *business* cred-

[6] For example, credit scoring enabled San Francisco-based Wells Fargo to make loans in virtually all 50 states prior to 1996, even though the bank had no branches outside California at the time.

[7] Other studies have found that large geographic distance between the home and host countries may impede the expansion of cross-border banking companies (e.g., Buch, 2003; Berger et al., 2004).

[8] In addition to aiding in the approval and denial of loan applications, credit-scoring models can be used to price new loans, monitor existing loans, identify candidates for cross-selling opportunities, and target prospective loan customers. For a discussion, see Mays (2004, Chapter 1).

itworthiness Berger et al. (2005a). Because households are less economically diverse than the small businesses they own and operate—which, for example, can range from traditional corner groceries to innovative software developers—this realization in effect reduced the degree of heterogeneity across small business loans and made it possible to manage them on a pooled, not individual, basis. As a result, some lenders were able to securitize and sell their small business loans in the secondary market. Indeed, Dell'Ariccia and Marquez (2006) argue that a reduction in the heterogeneity of borrower information across lenders (perhaps through widespread adoption of credit scoring) has implications for lender competition, loan quality, and loan supply. Mester (1997) provides a basic primer on credit scoring models for both consumer and small business lending. Frame and White (2004) review the relatively small academic literature on technological innovation in banking, of which small business credit scoring studies comprise a substantial portion. Roszbach (2004) reveals some systematic financial inefficiencies in the implementation of credit scoring models, using data on over 13,000 loan applications processed by a large Swedish bank in the mid-1990s.

Used in isolation, credit scoring may not improve the amount or accuracy of banks' information about borrowers and/or banks' abilities to make correct lending decisions—however, credit scoring approaches deliver a well-defined information set at less expense to the bank, and permit banks to make faster decisions on loan applications. Thus, the potential benefit of credit scoring depends on the manner in which banks use it. Consider two different ways that banks can use small business credit scoring models. First, credit scoring might be used as a first-stage filter to inexpensively identify loan applications that should clearly be rejected or approved, followed by a more thorough second-stage analysis of the "gray area" loan applications based on a broader set of (qualitative and quantitative) information. If done correctly, this approach can reduce a bank's overall loan screening expenses without compromising loan quality. Second, credit scoring might be used to automate the loan application accept/reject decision. This approach reduces loan screening expenses through even greater reductions in expensive human inputs; moreover, if loan volumes are large enough, automated credit scoring can fundamentally alter the economics of the lending process by capturing scale-based reductions in unit costs, diversifying away idiosyncratic risk, generating fee income from loan origination and loan servicing, and recycling scarce bank capital by securitizing the loans and selling them to other financial institutions (Mester, 1997; Rossi, 1998). On the potential downside, because credit scores are imperfect predictors of loan default based on a relatively small set of quantifiable information about a borrower, a greater number of unqualified loan applications are likely to be approved, all else equal. These loan screening mistakes, or type II errors, are likely to be more frequent for heterogeneous pools of loan applications like small business loans, and less frequent for more homogeneous pools of loan applications like conforming home mortgages.

DeYoung et al. (2006) construct a theoretical model of decision-making under risk and uncertainty, in which lenders have both imperfect information about borrowers and imperfect ability to use that information in the credit-approval process. In this framework, credit-scored lending has two independent, though not mutually exclusive, traits: credit scoring creates borrower information and credit-scoring generates loan production scale economies. If the predominant effect of credit scoring is to improve the quality of banks' information about borrower creditworthiness, or provide banks with better decision-making frameworks based on quantifiable rather than qualitative financial information, then credit scoring can mitigate the adverse information costs associated with borrower–lender distance. If, however, the predominant effect of credit scoring is to reduce banks' production costs by eliminating expensive on-site visits, reducing loan analysis time, or generating scale economies associated with automated lending processes, then credit scoring can increase the profitability of the marginal loan application without any mitigation of distance-related information costs. (We derive similar predictions below, using a much simpler theoretical framework.)

The only systematic database describing whether, when, and how bank lenders use small business credit scoring is based on a 1998 Federal Reserve Bank of Atlanta survey of the lead banks at 190 of the largest 200 commercial bank holding companies in the US (see Frame et al., 2001). Surveying large banks made sense: even prior to the dissemination of small business credit scoring technology, large banks were already more likely than small banks to use quantitative, hard-information approaches to business lending (Cole et al., 2004). These data have been used in a number of recent studies (including this one) of small business credit scoring by US banks, and have generated a number of useful findings and interesting conclusions.[9]

Frame et al. (2001) found a substantial increase in small business lending at the large banks that adopted credit scoring techniques, and concluded that credit scoring lowered information costs and by doing so reduced the value of traditional small bank–small borrower lending relationships. Frame et al. (2004) found that small business credit scoring contributed equally to increasing the supply of small business loans in both high income and low-to-moderate income areas. Akhavein et al. (2005) found that banks with large numbers of branch locations are more likely to adopt small business credit scoring techniques, suggesting that senior managers in these organizations use the technology as a control mechanism over branch bank managers and loan officers. Berger et al. (2005a) found that banks that use small business credit scoring experience higher nonperforming loan ratios, especially when these banks use credit scores to automatically accept or reject loan applications.

3. A theoretical illustration

The literature described above provides a starting point for our examination of borrower–lender distance, credit scoring, and their impact on the performance of small business loans. Evidence that geographic distance makes information gathering and monitoring more costly and less effective (e.g., Berger and DeYoung, 2001, 2006) implies that loan default rates will be higher at greater borrower–lender distances, all else equal. Evidence that hard information-based lending approaches may be especially effective at mitigating distance-related inefficiencies (e.g., Landier et al., 2005; Liberti, 2005) implies that loan default rates should be lower at credit-scoring lenders, all else equal. And conversely, evidence that automated lending production models exhibit scale economies (e.g., Rossi, 1998; Flannery and Samolyk, 2006, as well as pure casual empiricism) implies that loan default rates could be *higher* at credit-scoring lenders in equilibrium, all else equal: if scale economies make the marginal loan application more profitable, then applicants of marginally higher risk become more acceptable to the bank, and this provides an incentive to expand loan output which reduces average credit quality (DeYoung et al., 2006). In this section, we construct a simple theoretical framework to illustrate more formally how these three phenomena, individually and collectively, are likely to affect loan default rates.

Imagine a profit-maximizing lender makes loans to new borrowers located at various distances, d, from the bank using traditional (non-credit scoring) loan production processes. Both

[9] Taking a different tack, Agarwal and Hauswald (2006) gained access to small business credit-scoring data from a large US banking company. Although data from a single lender may not generate systematic conclusions, this alternative approach yields a variety of insights into small business credit scoring practices and outcomes.

loan markets and deposit markets are competitive, resulting in a market-clearing net interest margin, m. Loans are repaid with frequency ρ $(0 < \rho < 1)$, which implies a loan default frequency of $(1 - \rho)$. Banks maximize (expected) profit, defined as:

$$\pi = m \cdot \rho^e \cdot q - C(q) \tag{1}$$

where q is the quantity of loans and $C(q)$ is a total cost function that is twice differentiable and convex in q (i.e., $C_q > 0$, $C_{qq} > 0$). Differentiating Eq. (1) with respect to q and rearranging yields a standard equilibrium condition, $m \cdot \rho_e = C_q$, i.e., banks increase loan output until expected marginal revenue is just equal to marginal cost.

To complete the model, we make two economically intuitive assumptions about borrower–lender distance, both of which are consistent with the empirical findings in the literature cited above. First, we include distance as an argument in the cost function $C(q, d)$, and assume that costs increase with borrower–lender distance (i.e., $C_d > 0$). For example, the cost of originating and monitoring a loan using traditional lending techniques will increase with distance due to increased time and travel expenses. Second, we assume that lender uncertainty about borrower credit quality increases with borrower–lender distance, due to increasingly imperfect information about the loan applicant and local market in which the applicant operates, coupled with the lender's imperfect decision-making skills. Increased uncertainty results in both more frequent type I errors (i.e., rejecting good loans) and more frequent type II errors (i.e., approving bad loans). Hence, at greater borrower–lender distances the bank will expect a lower loan repayment frequency (i.e., $\rho_d^e < 0$), and so long as the bank's expectations are reasonably accurate, the bank will experience a higher actual loan default rate $(1 - \rho)$.[10] Incorporating these distance characteristics into the model requires us to rewrite the equilibrium condition as $m \cdot \rho^e(d) = C_q(d)$.

Figure 1 illustrates an initial equilibrium at point A for a borrower–lender distance of d. An increase in distance to d' $(d' > d)$ has two effects: the variable costs of originating loans will increase, illustrated by the upward shift in the marginal cost curve from $C_q(d)$ to $C_q(d')$, and the expected repayment rate will fall from $\rho^e(d)$ to $\rho^e(d')$, illustrated by the downward shift in the marginal revenue (MR) curve. The new equilibrium is at point B, where the bank makes fewer loans $(q' < q)$ and experiences a higher default rate (i.e., $1 - \rho^e(d') > 1 - \rho^e(d)$).

We can now use the model to illustrate several potential effects of credit scoring. First, credit scoring might reduce lender uncertainty, by improving the lender's information set and/or improving the lender's ability to use information—in either case, the lender's signal-to-noise ratio improves. Now better able to differentiate between good and bad borrowers, the lender will make fewer type I errors (fewer good loans will be rejected) and fewer type II errors (fewer bad loans will be accepted), and as a result the expected repayment rate will increase. (For now, assume that this is a pure reduction in uncertainty and does not affect the marginal cost curve.) This reduction in uncertainty works in the opposite direction from an increase in borrower–lender distance, shifting the expected marginal revenue curve up rather than down in Fig. 1. Starting from an equilibrium at point B, the reduction in uncertainty is illustrated by the shift in expected marginal revenue from $m \cdot \rho^e(d')$ to $m \cdot \rho^e(d', cs)$, moving the bank to a new equilibrium at point C

[10] The lender makes lending decisions based on the expected repayment rate $\rho^e(d)$, which may or may not equal the *ex post* observed repayment rate. However, if the expected rate is an unbiased estimate of the actual rate (a reasonable presumption) then both actual and expected loan performance will adjust similarly to changes in borrower–lender distance.

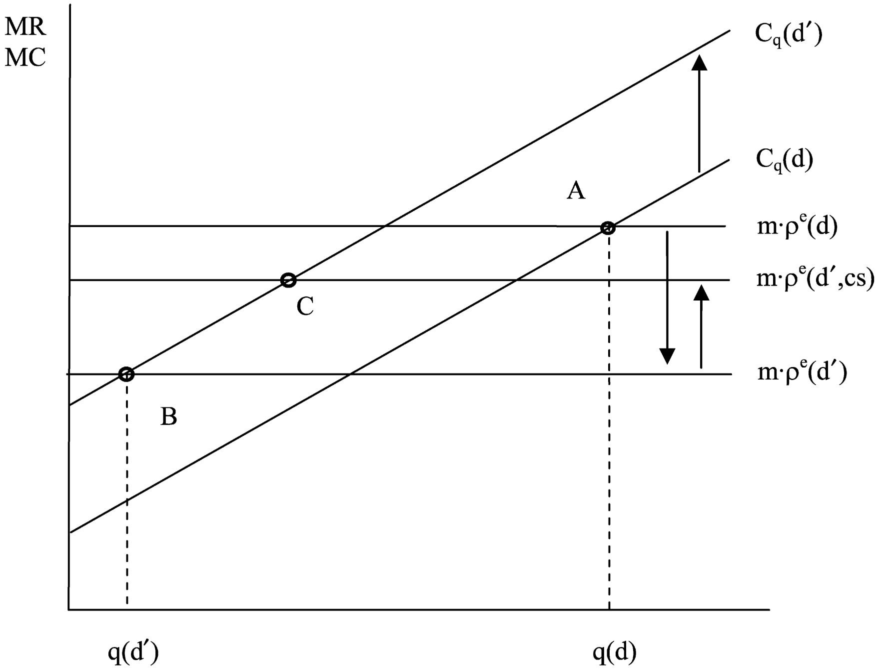

Fig. 1.

with a higher quantity of loans, better expected loan performance, and a lower actual loan default rate.[11]

Second, credit scoring might make the loan production process more efficient, for example, because acquiring and analyzing hard information is less expensive than acquiring and analyzing soft information.[12] The implications of these cost reductions are illustrated in Fig. 2. Recall that marginal costs are higher, and expected loan performance is worse, for more distant borrowers—this is illustrated, respectively, by the marginal cost functions $C_q(d)$ and $C_q(d')$ and the expected marginal revenue functions $m \cdot \rho^e(d)$ and $m \cdot \rho^e(d')$, where $d' > d$. Thus, in Fig. 2 the two distance-specific equilibria for a bank using traditional lending techniques are points A (shorter distance d) and B (longer distance d'). A switch to credit scoring techniques might reduce costs two different ways: by reducing loan production costs at all distances (hard information is cheaper than soft information) and by eliminating entirely the differential costs of distance (the expenses associated with credit scoring are invariant to distance). In the latter case, both of the marginal cost functions $C_q(d)$ and $C_q(d')$ would collapse to $C_q(cs)$, resulting in a new set of equilibria at points A^{cs} and B^{cs}. Note that moving from A to A^{cs} (and similarly, from B to B^{cs}) increases loan quantity but leaves loan default rates unchanged (i.e., expected marginal

[11] In Fig. 1, the upward shift in the expected marginal revenue curve (due to implementing credit scoring) is smaller than the previously discussed downward shift in expected marginal revenue (due to increased borrower–lender distance). This need not be the case. Even if the shift is greater, the results still hold.

[12] While informational economies are the most straightforward way to think about the production efficiencies associated with the credit-scored lending process, we note that credit-scored lending approaches may also yield scale efficiencies, diversification gains, and revenue synergies—all of which are conceptually consistent with the efficiencies we model here.

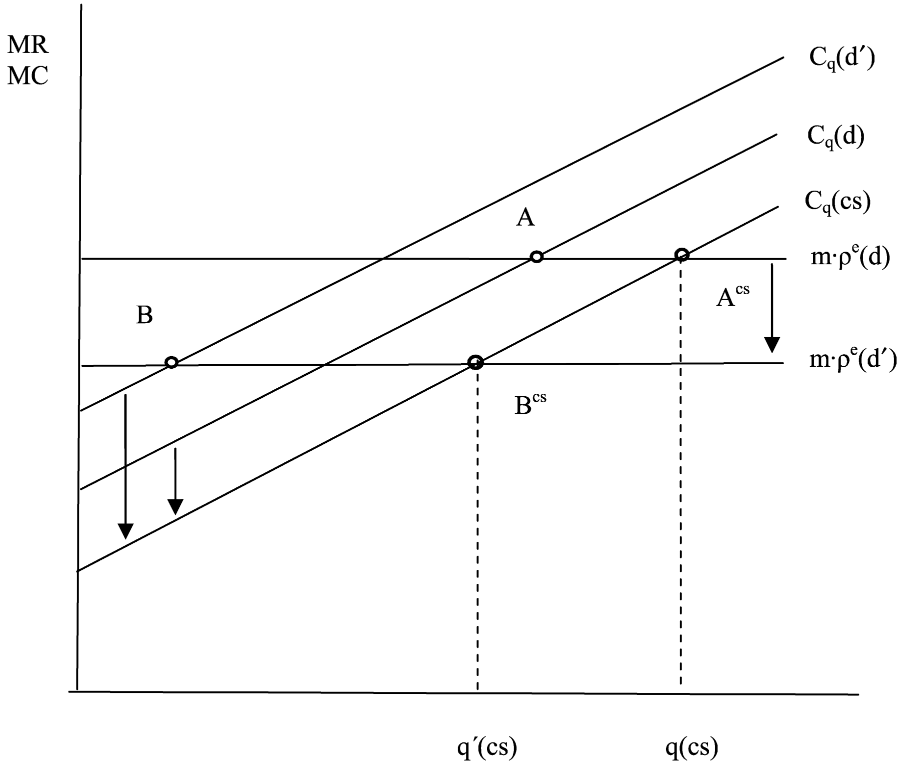

Fig. 2.

revenues $m \cdot \rho^e(d)$ and $m \cdot \rho^e(d')$ do not change). However, because the composition of the loan portfolio now includes a greater proportion of loans with longer borrower–lender distances (i.e., because $BB^{cs} > AA^{cs}$), the overall loan default frequency increases.[13]

Finally, because automated credit-scoring models rely on a limited set of quantifiable variables, it is *possible* that this lending approach is informationally inferior to traditional lending regimes, and as a result generates less complete information about borrower creditworthiness. (In this scenario, for a fixed quantity of loan applications, both type I and type II errors are higher under credit scoring.) If so, lenders will choose the credit scoring technology only if the production efficiencies (reduction in marginal cost) associated with credit scoring more than offset the reduction in expected loan performance (reduction in expected marginal revenue). This tradeoff is illustrated in Fig. 3, where point A shows the traditional, soft-information equilibrium and point B shows the credit-scoring equilibrium, holding borrower–lender distance constant. The figure is drawn so that increased loan output due to a reduction in marginal costs from $C_q(d)$ to $C_q(d, cs)$ is exactly offset by reduced loan output due to a reduction in expected marginal revenue from $m \cdot \rho^e(d)$ to $m \cdot \rho^e(d, cs)$, in this case leaving the contributions to profits (the integral

[13] In this highly stylized model, banks can increase their credit risk only by making loans to increasingly distant borrowers. But the general result illustrated here holds for any type (or cause) of risk: If the marginal cost of making loans declines, then banks will make additional loans at the margin, and this requires them to accept riskier loan applications that they would have previously rejected.

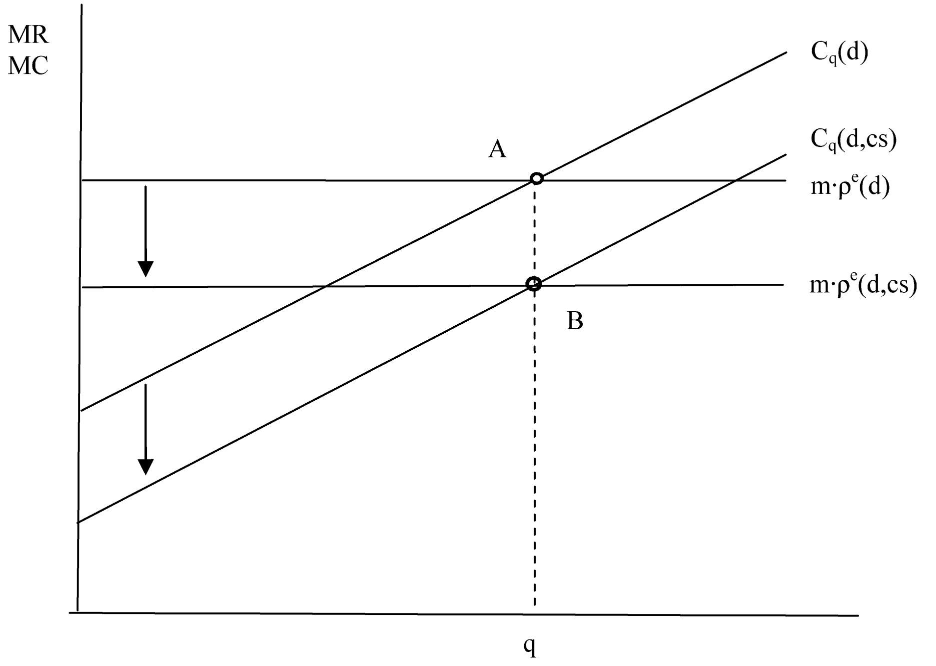

Fig. 3.

of $C_q - m \cdot \rho^e$ over q) unchanged. However, regardless of the net change in loan output, the loan default rate increases unambiguously at B relative to A—under this scenario, banks make both type I and type II errors more frequently under credit scoring, which reduces the credit quality of the accepted loan applicants.[14]

Thus, in this stylized theoretical framework we have illustrated the three main hypotheses that we test in our empirical analysis: (1) Borrower–lender distance is associated with higher loan default rates. (2) By reducing informational uncertainty, or by improving banks' decision-making abilities, credit scoring can mitigate the costs of distance and reduce loan default rates. (3) By increasing the profitability of the marginal loan, credit scoring creates incentives to make more (riskier) loans, increasing loan default rates as a result.

4. Data

Our data set is a random sample of 29,577 Small Business Administration 7(a) loans originated by 5535 qualified SBA program lenders between January 1984 and April 2001. The SBA 7(a) loan program provides loan guarantees for small business firms that are otherwise unable to access credit through conventional means, and as such, SBA loans are critical to the flow of

[14] This is a well-understood result in the industry. Lenders that adopt scoring models routinely evaluate the differences in credit quality across decision technologies (i.e., either the difference between judgmental and scoring models, or the difference between alternative scoring models) using a conventional "swap-in/swap-out" validation test. This test is designed to compare the expected performance of loans approved by, say, a judgmental decision process that would have been denied by a scoring model (i.e. swap-in), relative to the expected performance of loans approved by the scoring model but would have been denied by the judgmental process (i.e. swap-out).

credit to financially marginal small businesses. As indicated earlier, a substantial portion of dollar value of credit to small US businesses flows through this program. Since small businesses tend to create a disproportionate number of new jobs in market economies, these issues have special importance for economic policy.[15]

We use an event–history sample design. Under this design, each loan is observed multiple times, beginning with the quarter in which it was originated and continuing on through the quarter in which the loan either matured, paid off early (i.e., prepaid), or defaulted. Our sample of 29,577 loans generates 491,512 loan-quarters of observations. Table 1 provides some annual descriptive statistics for the loans in our data sample.

The SBA provides loan guarantees 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 of both principal and interest at the time of default and the contractual guarantee percentage stipulated by the SBA at the time of loan origination. Because lenders share in the losses, they have (perhaps reduced) incentives to screen for creditworthiness, monitor on an ongoing basis, or set appropriate loan interest rates and contract terms.[16] 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 SBA loans.[17] 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 remaining outstanding principal plus delinquent interest.

We cannot identify whether or not the individual loans in our data were originated using a credit scoring tool. Instead, we distinguish between credit-scoring and non-credit-scoring lenders based on the findings of a survey of the 200 largest US bank holding companies taken in 1998. (See Frame et al., 2001 for a description of this survey.) While this survey provides the best extant source on the dissemination of small-business credit-scoring techniques at US banking companies, these data have some obvious limitations. First, we do not know whether these lenders credit scored all, or just a portion, of their small business loan applications. Second, we cannot identify lenders that adopted credit-scoring technology after 1997. Third, we cannot identify credit-scoring lenders affiliated with banking companies too small to be included in the survey. While these limitations are not desirable, they are not especially problematic. The first limitation

[15] According to the SBA, small businesses "... provide 75 percent of the net new jobs added to the economy" (http://www.sba.gov). Some researchers argue that SBA estimates suffer from a variety of conceptual, methodological and measurement issues and as a result somewhat overstate the creation of jobs by small businesses (e.g., Davis et al., 1994, 1996). However, at least one recent study finds a positive (albeit economically small) relationship between the level of SBA lending in a local market and future personal income growth in that market (Craig et al., 2005). New job creation and macroeconomic impact aside, SBA figures indicate that small businesses employ 52 percent of the private work force, contribute 51 percent of private sector output, and produce 55 percent of innovations (US Small Business Administration 2000).

[16] The credit risk and intertemporal default patterns associated with long-term SBA loans have been shown to be quantitatively similar to those of speculative grade debt. See Glennon and Nigro (2005, Table 1).

[17] While is it permissible to securitize the unguaranteed portion of an SBA loan, most lenders retain this portion of the loan for its upside risk and to better maintain the borrower–lender relationship. Only 59 securitization transactions between 1994 and 2000 used either unguaranteed portions of SBA 7(a) loans or conventional small business loans as collateral (Board of Governors, 2000).

Table 1
Descriptive statistics for a selected set of factors

Year	# of loans	Cumulative (life-of-loan) default rate (%)	Prepaid (%)	Sold (%)	Percent interest rate premium over prime rate (%)	SBA guarantee percentage (%)	Median borrower–lender distance (miles) (%)	# of scoring banks
1984	722	26.59	55.54	16.48	8.21	86.75	5.89	0
1985	539	24.30	53.62	13.73	19.16	87.33	6.25	0
1986	820	19.39	55.12	12.20	26.78	85.26	5.65	0
1987	818	16.87	52.93	20.17	20.92	84.51	6.10	0
1988	748	18.85	50.27	19.25	16.55	84.10	5.92	0
1989	877	18.02	66.02	28.62	10.31	84.47	6.54	0
1990	958	20.98	63.26	32.78	14.47	84.81	6.26	0
1991	994	15.49	61.77	31.69	25.64	84.67	7.37	0
1992	1198	11.52	64.27	22.89	34.18	84.52	6.72	0
1993	1517	8.24	62.10	22.02	34.66	84.57	7.12	3
1994	2353	14.70	56.40	16.70	18.31	83.88	8.17	8
1995	4053	18.16	53.29	14.10	17.19	86.29	8.30	46
1996	2406	18.70	49.54	16.46	24.00	79.11	9.01	62
1997	2926	17.29	43.40	22.05	20.44	77.60	10.80	196
1998	2702	13.32	33.68	24.74	23.14	75.53	13.82	155*
1999	2576	8.27	28.96	30.33	22.62	70.34	18.02	101*
2000	2545	8.49	17.80	54.91	16.85	70.13	16.61	95*
2001	825	4.85	9.33	45.35	30.27	69.42	21.48	48*

Year	# of loans: Scoring banks	# of loans: Other banks	Default rate: Scoring banks	Default rate: Other banks	Prepaid: Scoring banks	Prepaid: Other banks	Sold: Scoring banks	Sold: Other banks	Premium: Scoring banks	Premium: Other banks	SBA guarantee: Scoring banks	SBA guarantee: Other banks	Distance: Scoring banks	Distance: Other banks	# of scoring banks
1995	198	3855	18.69	18.13	53.03	53.31	2.53	14.69	15.50	17.28	85.46	86.33	11.96	8.18	46
1996	240	2166	21.25	18.42	48.33	49.68	1.25	18.14	21.78	24.25	79.51	79.07	15.87	8.62	62
1997	870	2056	15.63	18.00	45.63	42.46	9.31	27.45	19.07	21.45	76.48	78.07	18.79	9.34	196
1998	949	1753	15.17	12.32	36.88	31.95	10.01	32.59	21.85	23.84	71.77	77.57	57.64	9.16	155*
1999	1134	1442	8.11	8.39	33.42	25.45	13.29	42.96	21.78	23.28	63.35	75.83	40.48	10.50	101*
2000	1031	1514	9.02	8.12	20.76	15.79	35.86	63.92	18.18	15.95	63.57	74.59	103.37	9.92	95*
2001	325	500	4.62	5.00	9.54	9.20	37.41	49.63	32.33	28.93	63.28	73.41	142.33	11.61	48*

Characteristics of a random sample of 29,577 SBA 7(a) loans to small businesses between 1984 and 2001. Mean data by cohort (year).
* Years for which we lack complete information for identifying banks that credit score small business loans.

simply constrains the form with which we state our credit-scoring hypothesis: We test whether *banks that use credit scoring models* have different default patterns, not whether *credit-scored loans* have different default patterns. We address the second limitation by estimating our regression models for a sub-sample of pre-1998 data. The third limitation is unlikely to be meaningful insofar as small business credit scoring was almost exclusively a large bank activity prior to 1998.

The data in Table 1 show that the number of scoring banks (by our definition) increases between 1993 and 1997 as this technology became more widely implemented, after which the number of scoring banks declines due to industry consolidation.[18] Although only a handful of the 5535 banks in our sample used credit scoring, these banks were generating approximately one out of every three loans during the final years of our sample. Cumulative default rates for SBA loans exhibit some cyclicality during our sample period, peaking at nearly 21% for loans originated just prior to, and falling to around 8% for loans originating just after, the 1990–1991 recession. The downward trend in loan defaults (and loan prepayments as well) toward the end of our sample period is due primarily to right-censoring in the data—that is, because it takes time for loans to "season," recently originated loans are less likely to default or prepay. (Note that the default percentages in Table 1 reflect the *ex post* probability of default over the full life of the loan. Table 2 displays statistics for the quarterly default rates, which better correspond to the spirit of our empirical model.)

The data suggest several changes in the SBA program over time. For example, the SBA guarantee percentage declined substantially during the late-1990s, from around 86% for loans originated in 1995 to less than 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 strengthen lenders' incentives to carefully screen and monitor loans, as well as more fully price risk into the loan interest rate. Indeed, the average interest rate premium (over the prime rate) on SBA loans increased over the 1995–2001 period as the average SBA guarantee percentage was on the decline—the linear correlation between these two annual averages is about −0.44 during these seven years. The percentage of loans sold off by the originating lenders also increased substantially at the end of the sample period, evidence of a more liquid secondary market for the guaranteed portion of SBA 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 5.65 miles and 7.37 miles, but began accelerating soon after that, reaching 10 miles by 1997 and 20 miles by 2001. Borrower–lender distance increased for *both* credit-scoring and non-credit-scoring banks, an indication that changes in industry conditions other than credit scoring (e.g., spatial structure of lenders relative to borrowers, new computer technology, remote Internet access) allowed or required lenders to reach further to make small business loans. But the most dramatic increase in borrower–lender distance is for the credit-scoring lenders: half of the loans originated by these banks in 2001 were to borrowers located 142 miles or more from the lending office.

Loans made by credit-scoring lenders carried lower SBA guarantee rates on average: for example, only 63 percent for loans originated by scoring lenders in 2001, compared to 73 percent for non-scoring lenders. Although we have no direct evidence to support our conjectures, this may indicate that credit-scoring banks were making riskier loans on average (and as such the

[18] Because we do not have data to identify lenders that adopted credit scoring for the first time after 1997, the final column of Table 1 understates the number of credit-scoring lenders in 1998 through 2001.

Table 2
Summary statistics

Variable name	Definition	Data source	Full sample (1984–2001) 491,512 loan-quarters 5212 lenders		Full sample (1984–2001) 29,577 loans unstacked	
			Mean	Std. Dev.	Mean	Std. Dev.
Default (time-varying)	Dependent variable = 1 if loan i defaulted at time t.	SBA (2000)	0.00896	0.09423	0.14889	0.35599
SBA%	= percentage of outstanding loan balance guaranteed by the SBA.	SBA (2000)	0.81426	0.09086	0.80002	0.10229
DISTANCE	= straight-line distance in miles between borrower and lending office of the lending institution.	SBA, Call Report	49.9667	237.619	65.7478	293.1521
ln DISTANCE	= natural log of DISTANCE.	SBA, Call Report	2.05697	1.90133	2.2134	1.94163
SCORER	= 1 if lender used credit scoring models to evaluate at least some of its small business loans at $t = 0$.	Akhavein et al. (2005)	0.11405	0.31787	0.16130	0.36782
MATURITY3	= 1 for 3-year loans.	SBA (2000)	0.07940	0.27037	0.13602	0.34281
MATURITY7	= 1 for 7-year loans.	SBA (2000)	0.62362	0.48448	0.64658	0.47804
NEWFIRM	= 1 if borrower is 3-year old or less.	SBA (2000)	0.30683	0.46118	0.32427	0.46811
FIRMSIZE	= number of full-time employees at borrowing firm.	SBA (2000)	12.4566	100.6588	12.3148	100.8332
HHI	= deposit-based Herfindahl index in local market of the borrower.	FDIC Summary of Deposits	0.19890	0.12171	0.19051	0.11603
URBAN	= 1 if borrower is located in a Metropolitan Statistical Area.	Call Report	0.79435	0.40417	0.82003	0.38417
CLP	= 1 if lender was a “certified” SBA lender.	SBA (2000)	0.19051	0.39270	0.16259	0.36900
PLP	= 1 if lender was a “preferred” SBA lender.	SBA (2000)	0.11599	0.32022	0.12790	0.33399
BANKSIZE	= lender assets (billions of 2001 $).	Call Report	12.8454	2.36233	13.1648	2.51285
CHARGEOFFS	= ratio of lender’s loan charge-offs to assets.	Call Report	1.65677	0.91779	1.64625	0.88016
RESERVES	= ratio of lender’s loan loss reserves to assets ($\times 100$).	Call Report	0.00393	0.00739	0.00369	0.00708
POLICY9401 (time-varying)	= 1 if loan was originated under liberal SBA credit policies in 1994–2001.		0.59504	0.49088	0.68925	0.46281
POLICYPOST89 (time-varying)	= 1 if loan was originated after the 1989 Federal Credit Reporting Act which required SBA to improve its risk management practices.		0.77987	0.41433	0.83832	0.36816
JOBGROWTH (time-varying)	= percent employment growth in borrower’s industry and home state in quarter t.	BEA/Haver	0.00526	0.02850	0.00337	0.02422
INCOMEGROWTH	= percent income growth in borrower’s industry and home state in loan origination quarter.	BEA/Haver	0.01488	0.01166	0.014939	0.011385
LOANAGE(x, y) (time-varying)	= 1 for loans between x and y quarters old in quarter t. This variable specifies a piecewise hazard function, enters regression multiple times, once each for the (x, y) values: (4, 5), (6, 7), (8, 9), (10, 12), (13, 15), (16, 20), and (21+).	SBA (2000)				

Variables used in estimation of Eq. (3). Random sample of 29,577 small business loans made in the SBA 7(a) loan program between 1984 and 2001 for which we have full information.

SBA provided less credit insurance) or that large credit-scoring lenders are less likely to be "certified" or "preferred" SBA lenders (and as such did not have access to higher levels of SBA credit insurance). Non-credit-scoring banks were substantially more likely to sell-off the guaranteed portions of these loans: for example, 50 percent of the loans originated by non-scoring lenders in 2001 were sold off, compared to only 37 percent for scoring lenders. This likely reflects the difference in the liquidity needs of the mostly small non-scoring banks (median assets of $231 million) and the mostly large scoring banks (median assets of $23 billion) that have much greater access to financial market funding (US Government Accounting Office, 1999).

5. Modeling loan performance using a panel-data design

We use a stacked-logit model to estimate the impact of borrower–lender distance and credit scoring on the probability of loan default, controlling for the effect of government loan subsidies and other variables. This approach takes good advantage of the panel design of our data. Each loan is represented by a series of binary variables $D_i(1), \dots, D_i(T_i)$, where $D_i(t) = 1$ if loan i ($i = 1, 2, \dots, N$) defaults during time period t ($t = 1, 2, \dots, T_i$), and $D_i(t) = 0$ otherwise, over the life of the loan.[19] 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 $\sum_{i=1}^{N} T_i$ rows. 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:

$$D_{it}^* = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{W}_{it} \boldsymbol{\gamma} + \varepsilon_{it} \tag{2}$$

where $\mathbf{X}$ is a vector of time-invariant covariates, $\mathbf{W}$ is a vector of time-varying covariates, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are the corresponding vectors of parameters to be estimated, and ε is an error term assumed to be distributed as standard logistic. We further define:

$$D_{it} = \begin{cases} 0 & \text{if } D_{it}^* \leqslant 0, \\ 1 & \text{if } D_{it}^* > 0. \end{cases}$$

Substituting the more compact notation $\mathbf{Z} = [\mathbf{X}, \mathbf{W}]$ and $\boldsymbol{\phi} = [\begin{smallmatrix} \beta \\ \gamma \end{smallmatrix}]$, the probability that $D_{it} = 1$ is given by:

$$\begin{aligned} &\operatorname{prob}(D_{it}^* > 0) = \operatorname{prob}(\mathbf{Z}\boldsymbol{\phi} + \varepsilon > 0), \\ &\operatorname{prob}(D_{it}^* > 0) = \operatorname{prob}(\varepsilon > -\mathbf{Z}\boldsymbol{\phi}), \\ &\operatorname{prob}(D_{it} = 1) = \Lambda(\mathbf{Z}\boldsymbol{\phi}) \end{aligned}$$

where $\Lambda(\cdot)$ is the logistic cumulative distribution function. We estimate Eq. (2) using standard binomial logit techniques.[20]

[19] Measuring time in quarters, the event history $D_i(1), \dots, D_i(t), \dots, D_i(T)$ 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.

[20] This approach is an empirical analog to the semi-parametric Cox proportional hazard model (Allison, 1995; Shumway, 2001; Brown and Goetzmann, 1995; Deng, 1997). It is sometimes referred to as a "discrete-time hazard model," because the event-history design of the data permits a hazard model to be estimated using discrete, qualitative dependent variable techniques. Indeed, $\operatorname{prob}(D_{it} = 1)$ in our model is the conceptual equivalent of a hazard rate, i.e., the probability that loan i defaults during period t conditional on having survived until period t. Compared to most other

We specify the stacked-logit model as follows:

$$\begin{aligned}\Pr[D_{it} = 1 \mid \mathbf{Z}] = \Lambda[&\mathrm{SBA\%}_i, \ln \mathrm{DISTANCE}_i, \mathrm{SCORER}_{ij}, \ln \mathrm{DISTANCE}_i * \mathrm{SCORER}_{ij},\\ &\mathrm{MATURITY}_i, \mathrm{FIRMSIZE}_i, \mathrm{NEWFIRM}_i, \mathrm{HHI}_i, \mathrm{URBAN}_i, \mathrm{CLP}_{ij}, \mathrm{PLP}_{ij},\\ &\mathrm{BANKSIZE}_{ij}, \mathrm{RESERVES}_{ij}, \mathrm{CHARGEOFFS}_{ij}, \mathrm{JOBGROWTH}_{it},\\ &\mathrm{INCOMEGROWTH}_i, \mathrm{POLICY9401}_i, \mathrm{POLICYPOST89}_i, \mathrm{LOANAGE}_{it}; \boldsymbol{\phi}] \quad (3)\end{aligned}$$

where i indexes the loan and j indexes the lender. The binary dependent variable D_{it} equals one if loan i defaulted in quarter t, and equals zero in all other quarters during the life of the loan. With the exception of the time-varying covariates (JOBGROWTH, INCOMEGROWTH and LOANAGE, defined below), all other variables are measured in the quarter in which the loan was originated. Table 2 shows definitions, summary statistics, and data sources for each of the variables specified in (3). Our main statistical tests are provided by the coefficient estimates on the variables SBA%, $\ln$ DISTANCE, SCORER, and $\ln$ DISTANCE $*$ SCORER. The remainder of the variables used to specify the model were selected largely based on the specification used in Glennon and Nigro (2005), and are discussed in detail in the Results section that follows.[21]

SBA% equals the percentage of the outstanding loan balance guaranteed by the SBA. SBA% can be thought of as an option held by the lender: it is the portion of a defaulted loan that can be put to the government in case of loan default. By reducing the cost of default to the lender, this option provides increased incentive for the lender to take credit risk; moreover, this incentive will be stronger as the size of the guarantee increases. Thus, we expect a positive estimated coefficient on SBA%, consistent with Glennon and Nigro (2003, 2005) and common economic sense. Holding the magnitude of SBA% constant, the put option should become more valuable—and hence be more likely to affect lender risk-taking—as the *a priori* chance of default increases. We control for this by including a variety of right-hand side variables that characterize the *a priori* chance of loan default, such as poor local economic conditions, the age of the borrower, the performance of loans previously underwritten by the lender, and the level of SBA-certification held by the lender.

DISTANCE equals the mile distance "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. 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.[22] We think of $\ln$ DISTANCE as the potential deterioration in the quality of lender information about borrower creditworthiness due to the increased costs of gathering soft information at longer distances. Stated in terms of the related theoretical literature, $\ln$ DISTANCE is our proxy for information imperfection.[23] Because less perfect information about borrower creditworthiness

multivariate hazard function models, this is a very flexible approach, because it allows for time-varying covariates on the right-hand side of the model, and it does not require the imposition of any parametric restrictions (e.g., a Weibull distribution) on the loan default distribution.

[21] We do not include a loan price or loan spread variable in our model. Theory suggests that the interest rate spread is a function of the probability of default, and not vice versa. Nevertheless, loan spreads *could* be used to *proxy* for credit-quality factors known to the lender but unobservable to the researcher. Because we already include some crude controls for the underlying credit quality of these small business borrowers (e.g., firm size, firm age), we do not also include loan spread in the regressions reported here, thus avoiding potential endogeneity issues.

[22] We also estimate Eq. (3) using discrete measures of borrower–lender distance. See Table 4 below.

[23] The degree of information available in the market about a borrowing firm is difficult to observe and measure. Studies in the finance literature typically use the size of a borrowing firm to proxy for information availability. (For a recent

makes it more likely that banks will lend to bad credit risks, ceteris paribus, we expect a positive estimated coefficient on ln DISTANCE.

SCORER is a binary variable equal to one if the lender is affiliated with a banking organization that used credit-scoring to screen at least some of its small business loan applications.[24] We think of SCORER as the potential for improved lender assessment of borrower creditworthiness made possible by credit-scoring models. Stated in terms of the related theoretical literature, SCORER is our proxy for more complete lender information. If this is the predominant effect of credit scoring, then we expect a negative estimated coefficient on SCORER, because more complete information about borrower creditworthiness makes it less likely that banks will lend to bad credit risks. However, as discussed above, it is also possible that credit scoring approaches result in higher likelihoods of loan default: First, because automated credit-scoring models rely on a limited set of quantifiable variables, they may be informationally inferior to traditional lending regimes and result in less complete lender information about borrower creditworthiness. Second, because automated credit-scoring approaches are likely to generate production efficiencies (e.g., scale economies, revenue synergies, risk diversification), the marginal loan will generate higher expected returns, creating a risk-return incentive for banks to expand production by making loans that are riskier at the margin. As an empirical question, the estimated marginal effect of SCORER on loan default will reflect the *net* impact of these decision-making, information imperfection, and financial performance effects.[25]

Given our information-based explanations of the expected coefficients on ln DISTANCE and SCORING—that is, borrower–lender distance reduces lender information, while credit-scoring models may improve lender information—we include the interaction term ln DISTANCE $*$ SCORER. If improved information generated by credit-scoring techniques either partially or fully mitigates the informational imperfection associated with borrower–lender distance, we expect a negative estimated coefficient on ln DISTANCE $*$ SCORER.

6. Results

We begin by estimating Eq. (3) as a standard logit model using the full data set, the results of which are displayed in Table 3. We estimated the standard errors in Table 3 under a variety of assumptions. Column [a] displays p-values generated using Huber–White robust standard errors, which control for heteroskedasticity but assume independence across observations. In the rest of the columns we reestimated the regression standard errors under a variety of assumptions that require the observations to be independent only across clusters defined by lender, borrower, and/or time. For example, because individual lenders may practice idiosyncratic underwriting policies and account management strategies, one might expect dependence within clusters of loans origi-

critique of this practice, see Holod and Peek, 2005, who use publicly-traded versus privately-held status to proxy for information availability.) In this study, we do not use firm size as a proxy for information because the borrowing firms in our sample are very small (mean number of employees = 12) and none are publicly traded; even the largest of the firms in our sample are informationally opaque to lenders other than those with whom they already have lending relationships.

24 We also estimate Eq. (3) using an alternative version of SCORER equal to one only if lenders used a fully automated credit scoring approach that prohibited loan officers from overruling their models' accept/reject decisions.

25 The theoretical ambiguity between the information-enhancing and cost-reducing effects of lending information technology is also present in Hauswald and Marquez (2003), although their model focuses primarily on the impact of these effects on competitive loan pricing.

Table 3
Results from stacked-logit estimation of Eq. (3)

Variable	Coefficient	[a] Huber–White standard errors $p >$ Chi sq.	[b] observations clustered by borrower $p >$ Chi sq.	[c] observations clustered by lender $p >$ Chi sq.	[d] observations clustered by origination quarter $p >$ Chi sq.	[e] observations clustered by lender and origination quarter $p >$ Chi sq.
Intercept	−6.7044	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
SBA%	0.4859	0.0150	0.0140	0.0290	0.0180	0.0170
ln DISTANCE	0.0416	< 0.0001	< 0.0001	< 0.0001	0.0010	0.0000
SCORER	0.3407	0.0010	0.0000	0.0010	0.0030	0.0010
SCORER ∗ ln DISTANCE	−0.0513	0.0370	0.0330	0.0440	0.1010	0.0500
MATURITY3	0.6244	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
MATURITY7	0.6120	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
FIRMSIZE	0.0003	< 0.0001	0.0060	0.0060	0.0050	0.0070
NEWFIRM	0.1994	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
HHI	0.0395	0.8610	0.8630	0.8640	0.8400	0.8640
HHI ∗ URBAN	−0.8422	0.0050	0.0060	0.0070	0.0050	0.0070
URBAN	0.1594	0.0560	0.0580	0.0700	0.0480	0.0610
CLP	−0.2016	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
PLP	−0.3276	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
BANKSIZE	−0.0198	0.0290	0.0290	0.0590	0.0280	0.0320
RESERVES	−0.0599	0.0030	0.0020	0.0050	0.0030	0.0030
CHARGEOFFS	6.4443	0.0010	< 0.0001	0.0010	0.0020	0.0000
INCOMEGROWTH	−1.0227	0.0580	0.0580	0.0570	0.0640	0.0590
JOBGROWTH	−4.1241	0.0010	0.0010	0.0010	0.0060	0.0010
POLICY9401	0.3230	< 0.0001	< 0.0001	< 0.0001	0.0030	< 0.0001
POLICYPOST89	−0.3805	< 0.0001	< 0.0001	< 0.0001	0.0010	< 0.0001
LOANAGE(4, 5)	1.4915	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
LOANAGE(6, 7)	1.7156	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
LOANAGE(8, 9)	1.8180	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
LOANAGE(10, 12)	1.7663	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
LOANAGE(13, 15)	1.7206	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
LOANAGE(16, 20)	1.6439	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001
LOANAGE(21+)	1.4128	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Dependent variable is loan by default. Random sample of 29,577 SBA 7(a) loans between 1984 and 2001 for which we have full information, of which 4044 defaulted during sample period. $N = 491{,}512$ loan-quarter observations. All variable definitions are displayed in Table 2.

nated by individual lenders.[26] Alternatively, because the economic, technological, regulatory, and political conditions surrounding the SBA program have changed over the course of our 1984–2001 data period, one might expect dependence within clusters of loans originated at the same point in time.[27] At the extreme, one might expect the data to display dependence for each individual loan over time, reflecting the idiosyncratic characteristics of each SBA borrower. The derived p-values using the cluster-specific standard errors are reported in columns [b], [c], [d], and [e], respectively, by borrower, lender, loan origination quarter, and both lender and loan origination quarter. As can be seen in Table 3, statistical inference is relatively robust across these alternative assumptions. Given this invariance, we simply use the Huber–White robust standard errors in the remainder of our tests.[28]

6.1. *Results from full-sample regressions*

Table 4 displays selected results from regressions using the full data sample and alternative specifications of the key test variables SCORER and DISTANCE. The top portion of the table shows the estimated logit regression coefficients along with Chi-square tests (p-values) of statistical significance. The bottom portion of the table shows the true marginal effects which reflect the economic significance of the test variables.[29] The results are consistent with each of our three main hypotheses, after conditioning on the effects of the SBA loan guarantees.

We find positive and statistically significant coefficients on SBA% in all of the regression specifications displayed in Table 4, in line with the prediction that higher loan guarantees will yield higher loan default rates. The economic effect is non-trivial. Based on the estimates in column [1], a ten percentage point increase in SBA% at the means of the data (from 80 to 90%, or approximately one standard deviation) is associated with about a 4.8 percent increase in the probability of loan default in a given quarter.[30] Coupled with the sub-sample results reported in later tables, these findings are consistent with the straightforward financial notion that insurance (e.g., in the form of partial loan guarantees) provides incentives for risk-taking.

Holding constant the effect of loan guarantees, we find positive and statistically significant coefficients on ln DISTANCE in both columns [1] and [2], consistent with the idea that greater information uncertainty (e.g., due to increased borrower–lender distance) will yield higher loan default rates. Based on the estimates in column [1], a doubling of borrower–lender distance at the means of the data (from 66 miles to 132 miles, well less than one standard deviation) is

[26] We note that clustering on lenders is theoretically problematic, given that the size and strategies of individual banks changed dramatically during our sample periodic due to changes in regulation, technology, non-bank competition, and growth via acquisition.

[27] Note that we cluster on loan origination time rather than on chronological time. The latter approach would cluster together observations at different points in their loan life-cycles, points at which the base probability of loan default varies dramatically. Clustering by loan origination year-quarter avoids this problem.

[28] We evaluated the statistical significance of the coefficients in Tables 4 through 8 using all of the alternative assumptions from Table 3, with equally robust results for statistical inference. Results are available upon request.

[29] The marginal effects are constructed as follows: We calculated the derivatives with respect to the three main test variables (ln DISTANCE, SCORER, and SBA%) based on the estimated logit coefficients; 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.

[30] The calculation is $(0.00430 * 0.10)/0.00897 = 0.04796$, where 0.00897 is the mean quarterly loan default rate from Table 2.

Table 4
Selected results from stacked-logit estimation of Eq. (3)

Variable	[1]		[2] SCORER = AUTOAPP		[3] Discrete DISTANCE	
	Coefficient	$p >$ Chi sq.	Coefficient	$p >$ Chi sq.	Coefficient	$p >$ Chi sq.
SBA%	0.4859	0.0150	0.4302	0.0260	0.4985	0.0120
ln DISTANCE	0.0416	< 0.0001	0.0398	< 0.0001		
DIST2550					0.1282	0.0210
DIST50UP					0.2627	< 0.0001
SCORER	0.3407	0.0010	0.2877	0.0340	0.2873	< 0.0001
SCORER ∗ ln DISTANCE	−0.0513	0.0370	−0.0548	0.0960		
SCORER ∗ DIST2550					−0.1423	0.3670
SCORER ∗ DIST50UP					−0.3060	0.0040
N	491,512		491,512		491,512	
$D = 1$	4044		4044		4044	
$D = 0$	487,468		487,468		487,468	
Marginal effects:						
SBA%	0.00430		0.00381		0.00441	
ln DISTANCE	0.00031		0.00032			
DIST2550					0.00097	
DIST50UP					0.00198	
SCORER	0.00204		0.00150		0.00202	
ln DISTANCE (SCORER = 0)	0.00036		0.00035			
ln DISTANCE (SCORER = 1)	−0.00011		−0.00016			
DIST2550 (SCORER = 0)					0.00110	
DIST2550 (SCORER = 1)					−0.00016	
DIST50UP (SCORER = 0)					0.00226	
DIST50UP (SCORER = 1)					−0.00049	

Dependent variable is loan by default. Huber–White standard errors.

associated with a 2.4 percent increase in the probability of loan default in a given quarter.[31] As discussed above, the costs associated with borrower–lender distance are unlikely to be linear, and simply specifying DISTANCE in natural logs (because travel expenses increase at a decreasing rate with distance) may not be enough. We us a more flexible functional form in column [3], where borrower–lender distance is specified in discrete rather than continuous terms. DIST2550 equals one when borrower–lender distance is between 25 and 50 miles, and DIST50UP equals one when borrower–lender distance is greater than 50 miles.[32] Relative to "local" loans made to borrowers less than 25 miles from the bank, loans between 25 and 50 miles from the bank were about 10.8 percent more likely to default, and loans more than 50 miles from the bank were about 22.1 percent more likely to default.[33] Thus, we find that borrower–lender distance is positively associated with loan default probability (holding the underwriting criteria constant) both at the means, and away from the means, of the data.

[31] The calculation is $(0.00031 * \ln 2)/0.00897 = 0.02395$.

[32] While 25 miles and 50 miles are *ad hoc* choices, note that both thresholds exceed the median borrower–lender distance in every year of our analysis (see Table 1). The loans in our sample were distributed as follows: 74% in the less-than-25-miles category, 10% in the 25-to-50-miles category, and 16% in the greater-than-50-miles category. This distribution is consistent with Wolken and Rohde (2002) who reported that 70 percent of all small business loans in 1998 were accessed from financial institutions within 30 miles of the business.

[33] The calculations are $0.00097/0.00897 = 0.10814$ and $0.00198/0.00897 = 0.2207$, respectively.

We also find positive and statistically significant coefficients on SCORER in all of the regression specifications reported in Table 4. These results are consistent with either or both of the following two scenarios: First, credit scoring-related reductions in loan defaults (due to improved lender information and decision-making ability) could be more than offset by increased risk-taking by lenders (due to incentives to expand loan supply to capture scale economies and other credit scoring-related improvements in bank profit functions). Second, credit-scoring may simply reduce the quantity and/or quality of information about small business borrower credit risk available to the lender, resulting in increased type II decision errors (i.e., approving bad loans). Regardless, the net economic effects are substantial. Based on the estimates in column [1], and holding borrower–lender distance constant, the quarterly probability of loan default was about 22.7 percent higher for loans made by banks that used credit scoring to underwrite at least some of their small business.[34] We find similar results in column [2] where we redefine SCORER to equal one only for lenders that used a fully automated, non-discretionary approach to small-business credit scoring in which loan officers were not allowed to overrule their models' accept/reject decisions. Holding borrower–lender distance constant, the quarterly probability of loan default was about 16.8 percent higher for loans made by banks that used an automated credit scoring approach.

While it is not possible to completely disentangle the various effects of credit scoring on loan default rates, our regressions yield some suggestive evidence. The coefficient on the interaction term $\ln \text{DISTANCE} * \text{SCORER}$ is negative and statistically significant, consistent with the idea that the improved decision-making ability associated with credit scoring techniques should partially or fully mitigate the informational uncertainty associated with borrower–lender distance. Based on the estimates in column [1], the marginal effect of $\ln \text{DISTANCE}$ conditional on $\text{SCORER} = 0$ equals 0.00036, which for a doubling of borrower–lender distance translates into a 2.8 percent increase in the quarterly probability of loan default. When $\text{SCORER} = 1$, however, the marginal effect equals -0.00011, which translates into a relatively small 0.8 percent reduction in the loan default rate for a doubling of distance. These results suggest that the deleterious impact of increased borrower–lender distance on loan default rates was, on average, largely neutralized at banks that used credit scoring techniques.

Distance may impact the transmission of information differently for local loans and longer-distance loans, and this differential may in turn determine the choice of a lending technology. For example, banks may choose a relationship lending approach for borrowers close enough to visit inexpensively, but choose a transactional credit-scored lending approach for far-away borrowers too costly to visit in-person. Indeed, the data displayed in Table 1 above imply a causal connection between the introduction of small business credit-scoring and average borrower–lender distance. Moreover, the regression results in column [3] of Table 4 suggest that applying alternative underwriting technologies in this way is optimal. For "local" loans made to borrowers less than 25 miles from the bank, credit scoring *increased* the quarterly probability of loan default by about 23 percent—but for loans made to borrowers more than 50 miles from the bank, credit scoring *decreased* the likelihood of loan default by about 5.4 percent.

6.2. Control variables

Returning to Table 3, we find that the coefficients on the remainder of the control variables have sensible signs and tend to be statistically significant. Small business loans with shorter

[34] The calculation is $(0.00204 * 1)/0.00897 = 0.2272$.

maturities (i.e., 3-year and 7-year maturities: MATURITY3 and MATURITY7) were more likely to default than loans with longer maturities (i.e., 15-year). 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 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.[35] Borrowers that are less than 3-year old at loan origination (NEWFIRM) were more likely to default on their loans than more mature small businesses. Holding these age effects constant, larger borrowers (FIRMSIZE) were also more likely to default.

Competitive rivalry among banks is associated with higher loan defaults. Although the coefficient on HHI is never statistically significant, the coefficient on the interaction term HHI $*$ URBAN is statistically negative, indicating that increased concentration (reduced competitive rivalry) was associated with lower default rates in urban markets. This finding is consistent with Petersen and Rajan (1994, 1995), who argue that lenders in concentrated markets are more likely to cultivate relationships with their small business clients (e.g., engage in careful monitoring that reduces loan default rates). We emphasize that the welfare implications of this result are ambiguous, because concentrated markets are likely to generate a smaller supply of loans.

Experienced SBA lenders with good lending records (i.e., certified lenders and preferred lenders: CLP and PLP) were less likely to make loans that defaulted. Large lenders (BANKSIZE) also had lower loan default rates, perhaps because these lenders could afford to attract and retain high-quality staff that specialize in underwriting and monitoring loans, or perhaps because small banks face more limited growth opportunities and face pressure to grow their loan portfolios by reaching further into the risk pool. All else equal, banks that wrote off large amounts of bad loans in the recent past (CHARGEOFFS) were more likely to originate small business loans that eventually defaulted (perhaps indicating an appetite for credit risk), and banks with high levels of loan loss reserves (RESERVES) were less likely to originate small business loans that eventually defaulted (perhaps indicating a low tolerance for insolvency risk).

Robust economic activity at the time of loan origination (INCOMEGROWTH) and during the life of the loan (JOBGROWTH) were both associated with reductions in loan defaults. As expected, default rates were higher for loans originated under the relatively liberal SBA credit policies between 1994 and 2001 (POLICY9401), and default rates were lower for loans originated after the passage of the Federal Credit Reporting Act of 1990 which required the SBA to improve its risk-management practices (POLICYPOST89), ceteris paribus. The piecewise hazard function (the LOANAGE terms), which we include to capture the time path of default associated with loan seasoning, has the familiar concave shape, with quarterly default rates peaking on average eight or nine quarters after loan origination.

6.3. *Results from sub-sample regressions*

Tables 5 through 8 display selected regression results for various sub-samples of the data. We reestimated Eq. (3) for sub-samples of loans based on the years in which the loans were originated (Table 5), the size of the loan (Table 6), the size of the lender (Table 7), and borrower–lender distance (Table 8). The main results from the full-sample regressions in Table 4 are relatively

[35] 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.

robust to these additional sub-sample tests. When these new results deviate from those in Table 4, they tend to be economically sensible and in some cases are instructive for public policy.

Table 5 displays selected results from three sub-sample regressions: the portion of the sample during or before the Atlanta Fed credit-scoring survey (pre-1999), the first half of the sample period (1984–1992), and the second half of the sample period (1993–2001). The coefficient on SBA% remains positive throughout, but is statistically insignificant in regressions [4] and [5]. This is more likely due to the general lack of variation in SBA% during the early part of our sample period (see Table 1) than any differential response of lending banks to loan subsidies. The coefficient on $\ln$ DISTANCE remains positive and statistically significant throughout, but the magnitude of this effect declines substantially over time. This suggests that advances in information, financial, and/or communications technologies over time better equipped banks to deal with the uncertainties of lending at long geographic distances. Perhaps because of these improvements, the marginal impact of SCORER declined somewhat over time, although the coefficient on this variable remained positive and significant throughout.

Table 6 contains selected results from sub-sample regressions for loans in amounts less than \$100,000, for loans in amounts less than \$250,000, and for loans in amounts greater than \$100,000. The coefficient on SBA% remains positive and significant for the small loan sub-samples, but equals zero for the large loan ($>$ \$100,000) sub-sample. This result supports the hypothesis that, because smaller firms are more likely to have the highest level of information uncertainty, lenders require higher guarantees for this segment of the portfolio. The increased subsidy results in more loans to smaller firms; however, because of the increased uncertainty, a greater percentage of these loans ultimately perform poorly. Moreover, this result implies that the higher SBA subsidy encourages banks to approve loan applications from the smallest businesses, but does not appear to influence banks' accept/reject decisions for loans from larger businesses, which (assumedly) have better access to credit in any event. The coefficients on $\ln$ DISTANCE and SCORER are positive and significant throughout, and both tend to get larger as the loans in the sub-sample become larger. These results may indicate, respectively, that larger borrowers search longer and further for loans (resulting in a positive link between loan size, distance, and loan default) and that the informational inadequacies of credit-scored based lending (relative to relationship lending) are amplified at larger, more complex and likely more heterogeneous borrowers. The coefficient on the interaction term SCORER $*$ $\ln$ DISTANCE is negative as before, although it becomes statistically non-significant for the sub-sample of very small loans ($\leqslant$ \$100,000).

Table 7 contains selected results from sub-sample regressions for loans originated by small banks (less than \$1 billion in growth-adjusted 2001 dollars) and by large banks (all other banks).[36] The results for the small bank sub-sample are robust to the full-sample results with the exception of the coefficient on SBA%, which was statistically non-significant. Evidently, the underwriting practices of small, relationship-based lenders tend to be unaffected by loan guarantees, in contrast to larger banks who are more likely to treat small business loans as financial transactions and/or are willing to relax their underwriting standards in the presence of this credit insurance. As a result, larger lenders are more likely to put applicants that are harder to assess, or appear to be of lower quality, in the SBA program. The results for the large

[36] The size threshold was \$1 billion for banks in 2001. For previous years, an index (2001 = 100) was used to reduce the size threshold by an amount equal to the nominal annual growth rate of the median US commercial bank.

Table 5
Selected sub-sample results for stacked-logit estimation of Eq. (3)

Variable	[4] Survey years only 1984–1998		[5] First half 1984–1992		[6] Second half 1993–2001		[7] Second half 1993–2001	
	Coefficient	$p >$ Chi sq.	Coefficient	$p >$ Chi sq.	Coefficient	$p >$ Chi sq.	Coefficient	$p >$ Chi sq.
SBA%	0.3081	0.2000	0.1579	0.7020	0.3786	0.0840	0.4927	0.0310
ln DISTANCE	0.0408	< 0.0001	0.0583	0.0010	0.0240	0.0290	0.0303	0.0150
SCORER	0.3773	0.0010					0.3002	0.0040
SCORER $*$ ln DISTANCE	−0.0480	0.1170					−0.0383	0.1350
N	435,305		166,933		324,579		324,579	
$D = 1$	3935		1412		2992		2992	
$D = 0$	431,370		165,521		321,587		321,587	
Marginal effects:								
SBA%	0.00275		0.00132		0.00345		0.00449	
ln DISTANCE	0.00032		0.00049		0.00022		0.00021	
ln DISTANCE(SCORER = 0)	0.00036						0.00027	
ln DISTANCE(SCORER = 1)	−0.00008						−0.00009	
SCORER	0.00250						0.00191	

Data sub-samples based on year in which loan was originated. Dependent variable is loan by default. Huber–White standard errors.

Table 6
Selected sub-sample results for stacked-logit estimation of Eq. (3)

Variable	[8] Loan $\leqslant$ \$100,000		[9] Loans $\leqslant$ \$250,000		[10] Loan > \$100,000	
	Coefficient	$p >$ Chi sq.	Coefficient	$p >$ Chi sq.	Coefficient	$p >$ Chi sq.
SBA%	0.7033	0.0050	0.6641	0.0020	−0.0646	0.8577
ln DISTANCE	0.0349	0.0070	0.0391	0.0010	0.0509	0.0020
SCORER	0.2241	0.0660	0.3306	0.0020	0.5878	0.0010
SCORER ∗ ln DISTANCE	−0.0225	0.4550	−0.0440	0.0910	−0.1052	0.0160
N	271,307		409,390		220,205	
$D = 1$	2746		3775		1658	
$D = 0$	268,561		405,615		218,547	
Marginal effects:						
SBA%	0.00703		0.00605		−0.00048	
ln DISTANCE	0.00032		0.00030		0.00030	
ln DISTANCE(SCORER = 0)	0.00034		0.00035		0.00037	
ln DISTANCE(SCORER = 1)	0.00014		−0.00006		−0.00057	
SCORER	0.00176		0.00216		0.00266	

Data sub-samples based on size of loan. Dependent variable is loan by default. Huber–White standard errors.

Table 7
Selected sub-sample results for stacked-logit estimation of Eq. (3)

Variable	[11] Bank assets $\leqslant$ \$1 billion in growth-adjusted 2001 dollars		[12] Bank assets > \$1 billion in growth-adjusted 2001 dollars	
	Coefficient	$p >$ Chi sq.	Coefficient	$p >$ Chi sq.
SBA%	0.2390	0.4870	0.5764	0.0202
ln DISTANCE	0.0530	< 0.0001	0.0257	0.0840
SCORER	0.8734	0.0020	0.2535	0.0190
SCORER ∗ ln DISTANCE	−0.3447	0.0070	−0.0281	0.2990
N	224,492		267,019	
$D = 1$	2049		2355	
$D = 0$	222,444		264,664	
Marginal effects:				
SBA%	0.00216		0.00502	
ln DISTANCE	0.00045		0.00017	
ln DISTANCE(SCORER = 0)	0.00048		0.00022	
ln DISTANCE(SCORER = 1)	−0.00408		−0.00002	
SCORER	0.00244		0.00160	

Data sub-samples based on size of originating bank. Dependent variable is loan by default. Huber-White standard errors.

bank sub-sample are robust to the full-sample results with the exception of the coefficient on SCORER ∗ ln DISTANCE, which remains negative but becomes statistically non-significant.

Table 8 contains selected results from sub-sample regressions for loans made to borrowers at various geographic distances from the lending office: 0-to-10 miles, 10-to-25 miles, 25-to-50 miles, and greater than 50 miles. Given this method of sub-sampling (which in the first three sub-samples truncates the upper end of DISTANCE), it is not surprising that the coefficient on ln DISTANCE is statistically positive only for the greater-than-50-miles regression. The coefficient on SCORER is positive and significant only in the 0-to-10 mile regressions. This result is consistent with those reported earlier in Table 4, column [3]: loans to borrowers closest to the lender are exactly the loans for which transactional credit-scored lending should perform the

Table 8
Selected sub-sample results for stacked-logit estimation of Eq. (3)

Variable	[13] DISTANCE between 0 and 10 miles		[14] DISTANCE between 10 and 25 miles		[15] DISTANCE between 25 and 50 miles		[16] DISTANCE greater than 50 miles	
	Coefficient	$p >$ Chi sq.	Coefficient	$p >$ Chi sq.	Coefficient	$p >$ Chi sq.	Coefficient	$p >$ Chi sq.
SBA%	0.6167	0.0420	0.2424	0.5820	0.0148	0.9818	0.3990	0.3460
ln DISTANCE	0.0103	0.5360	0.0407	0.7530	0.3024	0.2490	0.0966	0.1020
SCORER	0.4067	0.0010	0.9067	0.4310	4.0637	0.0850	−0.2280	0.6330
SCORER ∗ ln DISTANCE	−0.0556	0.4560	−0.3134	0.4540	−1.0969	0.1030	0.0640	0.4660
N	269,057		110,674		46,507		65,274	
$D = 1$	2346		977		436		645	
$D = 0$	266,711		109,697		46,071		64,629	
% SCORER = 1	7.85%		12.36%		16.29%		48.71%	
Marginal effects:								
SBA%	0.00532		0.00211		0.00014		0.00389	
ln DISTANCE	0.00005		0.00010		0.00149		0.00118	
ln DISTANCE(SCORER = 0)	0.00009		0.00035		0.00273		0.00090	
ln DISTANCE(SCORER = 1)	−0.00055		−0.00250		−0.00900		0.00165	
SCORER	0.00311		0.00047		0.00201		0.00101	

Data sub-samples based on borrower–lender distance. Dependent variable is loan by default. Huber-White standard errors.

worst relative to relationship lending, as close geographic proximity should maximize the informational advantages of personal borrower–lender relationships. Similarly, the coefficient on SBA% is positive and significant only in the 0-to-10 mile group, suggesting that SBA loan guarantees enticed banks to make loans to opaque borrowers located close-by where monitoring is relatively easy, but not at longer distances where monitoring is more costly.

7. Conclusions and implications for policy

Over the past two decades the geographic distance between small business borrowers and their commercial bank lenders has increased dramatically, largely due to improvements in information, communications, and financial technologies that allow quicker and more efficient analysis of information about small borrower creditworthiness. Improved or otherwise altered information sets about borrower creditworthiness are most likely to make a difference for financially marginal small businesses. In this study, we examine the interrelated ways that increased borrower–lender distance and increased use of credit-scoring technologies have affected loan default probabilities for a large random sample of 29,577 small business loans made by US commercial banks between 1984 and 2001 under the SBA 7(a) loan program, holding constant the effects of the government subsidies associated with these loans. 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 find strong patterns in the data that are consistent with hypotheses derived from a simple theoretical model of lender profit-maximization, in which the lender has imperfect information about the creditworthiness of its borrowers. The model is grounded in two sensible maintained assumptions: the cost of making loans increases with borrower–lender distance (e.g., greater travel costs) and the expected performance of those loans declines with borrower–lender distance (e.g., information becomes more uncertain). Given these assumptions, we demonstrate that credit scoring may either mitigate or exacerbate distance-related loan default rates, depending on the net effects of credit scoring on information collection and/or the shape of the loan production function. Indeed, we find strong evidence consistent with our maintained assumption that loan performance declines with distance: the probability of loan default increased with borrower–lender distance during our sample period, both at the means of the data as well as for small business borrowers located further away. According to our estimates, this deleterious impact of distance declined over time, implying that changes in banking industry structure (e.g., a consolidated and thus more efficient industry) and/or information and communications technologies (e.g., portable computers, spreadsheet analysis, the Internet) during our sample period improved banks' abilities to lend to small businesses.

Importantly, we find that the default-increasing effects of borrower–lender distance were substantially mitigated at banks that used credit-scoring models to screen at least some of their loan applications. After controlling for these distance-related effects, however, we find that lenders that used credit-scoring models experienced *higher* default rates. This finding suggests that banks are willing to accommodate the costs of higher expected default rates in exchange for ancillary benefits associated with high-volume credit-scored and securitized lending strategies (e.g., scale economies, fee generation, diversification, recycling of equity capital). Alternatively, though perhaps less likely, this finding could also indicate that credit-scoring underwriting processes simply collect less information, and thus lead to more loan approval errors. In either case, these findings complement those of Berger et al. (2005a), who found that credit-scoring banks not only tend to make higher volumes of small business loans, but these loans tend to be riskier. Our results are robust to whether credit-scoring banks used a fully automated approach or allowed loan officers the discretion to overrule these models.

We find that government loan guarantees and competition in local lending markets are both associated with higher probabilities of loan default. These findings illustrate the inherent tradeoff between public policies aimed at increasing the quantity of small business credit (e.g., providing government subsidies, encouraging market competition) and the resulting information problems that may encourage banks to allocate funds to uncreditworthy borrowers. On average, our results indicate that SBA guarantees have little effect on loan default probabilities at small banks, but are associated with substantially higher default probabilities for small loans—hence, these government loan subsidies tend to influence the behavior of large transactions-based lenders that treat small business loans like consumer lines of credit, but tend not to influence the underwriting practices of small relationship-based lenders. The link in the data between market competition and high default rates is consistent with Petersen and Rajan (1994, 1995), who argue that lenders in concentrated (noncompetitive) markets are more likely to cultivate relationships with their small business clients, e.g., careful monitoring that reduces loan default rates, but in doing so may limit overall loan supply by denying loans to other good applicants.

Our results also highlight some interesting questions for SBA policy. Starting in 2005, the SBA 7(a) program became self-funded; general tax revenues (which have helped fund the program since its creation in the early 1950s) are no longer available to subsidize losses on defaulted loans.[37] Under the new budgeting, lender fees in 2005 increased to 0.54% from 0.25% of the total loan amount, and borrower fees doubled from 0.125 to 0.25% of the loan amount. To remain a financially viable program under these tighter fiscal constraints, it is imperative that the SBA have the tools and expertise needed to measure and manage risk at least as well as the lenders that use the program. Our results suggest that credit-scoring has improved the ability of (predominantly large) banks to measure small business credit risk; thus, if the loan subsidies provided by the SBA are not adequately risk-priced, it will be a straightforward proposition for these banks to shift credit risk to the SBA. The SBA faces a difficult problem: while its financial viability may depend on linking its fee structure to credit scores, doing so will reduce the availability of loan and loan guarantees to truly opaque small businesses for which credit scoring is less efficient—the exact market failure that SBA was created to address.

We acknowledge that these findings, derived as they are from partially subsidized SBA loans to firms that by definition do not have access to regular credit markets, will not generalize seamlessly to non-government-guaranteed small business lending. However, we argue that our empirical findings do have implications that stretch beyond the SBA program, and do provide some important learning for the creation and performance of small business credit in general. So long as SBA lenders incur some non-trivial share of potential loan default costs, they face the same qualitative tradeoffs faced by a non-subsidized lender. For both subsidized and non-subsidized lenders, credit scoring might (a) improve bank performance by generating better information on borrower creditworthiness, and hence yield reduced loan default rates, or (b) improve bank performance by generating production scale economies, revenue synergies, or diversification effects that encourage banks to expand volume by approving riskier loans at the margin, and hence incur higher loan default rates. Similarly, geographic distance confers information-gathering frictions

[37] See John Roesti, "Effort to Revive Subsidy for 7(a) Fails," *American Banker*, February 11, 2005, and Rob Blackwell and Hannah Bergman, "US Budget: What's In And What's Out For Industry," *American Banker*, February 08, 2005. The subsidy to the SBA 7(a) program amounted to approximately $80 million in fiscal 2004. We note the possibility that, under the newly Democratic-controlled Congress, this change to a self-funded program may be rolled back.

on all banks, regardless of whether they hold options to put loan losses to the government. For both subsidized lenders and non-subsidized lenders alike, increased borrower–lender distance is likely to reduce the quality of borrower information, and credit-scoring techniques could either mitigate or exacerbate these information problems as discussed above.

Acknowledgments

The authors are especially grateful to Dan McMillen for sharing his time and geographic mapping expertise, Scott Frame for graciously providing data and expertise on credit-scoring, and an anonymous referee for excellent guidance on the final version of the manuscript. We also thank Allen Berger, Elena Carletti, Ed Kane, Stefano Lovo, Robert Marquez, Richard Nelson, Evren Örs, Tony Saunders, and Chiwon Yom for comments that have improved our work.

References

Akhavein, J.D., Frame, W.S., White, L.J., 2005. The diffusion of financial innovations: An examination of the adoption of small business credit scoring by large banking organizations. J. Bus. 78, 577–596.

Agarwal, S., Hauswald, R., 2006. Distance and information asymmetries in lending decisions. Unpublished manuscript.

Allison, P., 1995. Survival analysis using the SAS system: A practical guide. SAS Institute, Cary, NC.

Berger, A.N., DeYoung, R., 2001. The effects of geographic expansion on bank efficiency. J. Finan. Services Res. 19, 163–184.

Berger, A.N., DeYoung, R., 2006. Technological progress and the geographic expansion of the banking industry. J. Money, Credit, Banking 38, 1483–1513.

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

Berger, A.N., Buch, C., DeLong, G., DeYoung, R., 2004. The comparative advantages of nations at exporting financial institutions management via M&As. J. Int. Money Finance 23 (3), 333–366.

Berger, A.N., Frame, W.S., Miller, N., 2005a. Credit scoring and the availability, price, and risk of small business credit. J. Money, Credit, Banking 37, 191–222.

Berger, A.N., Miller, N.H., Petersen, M.A., Rajan, R.G., Stein, J.C., 2005b. Does function follow organizational form? Evidence from the lending practices of large and small banks. J. Finan. Econ. 76, 237–269.

Bitler, M.P., Robb, A.M., Wolken, J.D., 2001. Financial services used by small businesses: Evidence from the 1998 survey of small business finances. Fed. Reserve Bull. 87, 183–205.

Board of Governors, 2000. Report to Congress on Markets for Small-Business and Commercial-Mortgage-Related Securities. Federal Reserve System, Washington, DC.

Boot, A.W.A., Thakor, A.V., 2000. Can relationship banking survive competition? J. Finance 55, 679–713.

Brevoort, K.P., Hannan, T.H., 2004. Commercial lending and distance: Evidence from community reinvestment act data. Working paper 2004-24. Federal Reserve System Board of Governors, FEDS.

Brown, S., Goetzmann, W., 1995. Performance persistence. J. Finance 50 (2), 679–698.

Buch, C.M., 2003. Information or regulation: What is driving the international activities of commercial banks? J. Money, Credit, Banking 35 (6), 851–870.

Cole, R.A., Goldberg, L.G., White, L.J., 2004. Cookie-cutter versus character: The micro structure of small-business lending by large and small banks. J. Finan. Quant. Anal. 39.

Craig, B., Jackson III, W.E., Thomson, J.B., 2005. SBA-loan guarantees and local economic growth. Unpublished manuscript.

Cyrnak A., Hannan, T., 2000. Non-local lending to small business. Working paper, Board of Governors of the Federal Reserve System.

Davis, S., Haltiwanger, J., Schuh, S., 1994. Small business and job creation: Dissecting the myth and reassessing the facts. In: Solomon, L., Leoinson, A. (Eds.), Labor Markets, Employment, and Job Creation. Westview Press, Boulder, CO, pp. 13–32.

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

Degryse, H., and Ongena, S., 2002. Distance, lending relationships, and competition. Working paper, University of Tilburg.

Dell'Ariccia, G., Marquez, R., 2004. Information and Bank Credit Allocation. J. Finan. Econ. 72 (1), 185–214.
Dell'Ariccia, G., Marquez, R., 2006. Lending booms and lending standards. J. Finance 61, 2511–2546.
Deng, Y., 1997. Mortgage termination: An empirical hazard model with stochastic term structure. J. Real Estate Finance Econ. 14 (3), 309–331.
DeYoung, R., Hunter, W.C., Udell, G.F., 2004. The past, present, and probable future for community banks. J. Finan. Services Res. 25 (2/3), 85–133.
DeYoung, R. Glennon, D., Nigro, P., 2006. Borrower–lender distance, credit scoring, and the performance of small business loans. Working paper 2006-04. Federal Deposit Insurance Corporation, Center for Financial Research.
Ergungor, O.E., 2005. The profitability of bank-borrower relationships. J. Finan. Intermediation 14, 485–512.
Flannery, M., Samolyk, K., 2006. Payday lending: Assessing the importance of scale for store performance. Working paper. Federal Deposit Insurance Corporation.
Frame, W.S., White, L.J., 2004. Empirical studies of financial innovations: Lots of talk, little action? J. Econ. Lit. 42, 116–144.
Frame, S., Srinivasan, A., Woosley, L., 2001. The effect of credit scoring on small business lending. J. Money, Credit, Banking 34 (August), 813–825.
Frame, S.M., Padhi, M., Woosley, L., 2004. Credit scoring and the availability of small business credit in low- and moderate-income areas. Finan. Rev. 39 (February), 35–54.
Glennon, D., Nigro, P., 2003. Measuring the default risk of small business loans: A survival analysis approach. J. Money, Credit, Banking 37, 923–948.
Glennon, D., Nigro, P., 2005. An analysis of SBA loan defaults by maturity structure. J. Finan. Services Res. 28, 77–111.
Greene, W.H., 1997. Econometric Analysis, third ed. Prentice Hall, Upper Saddle River.
Hannan, T.H., 2003. Changes in non-local lending to small business. J. Finan. Services Res. 24, 31–46.
Hauswald, R., Marquez, R., 2003. Information technology and financial services competition. Rev. Finan. Stud. 16, 921–948.
Holod, D., Peek, J., 2005. Asymmetric Information and Liquidity Constraints: A More Complete Test. In: Proceedings from a Conference on Bank Structure and Competition. Federal Reserve Bank of Chicago.
Landier, A., Nair, V.B., Wulf, J., 2005. "Geographic dispersion and corporate decision-making. Unpublished manuscript.
Liberti, J.M., 2005. How does organizational form matter? Communication, distance, and soft information. Unpublished manuscript.
Mays, E., 2004. Credit Scoring for Risk Managers. Thomson, South Western, Mason, OH.
Mester, L.J., 1997. What's the point of credit scoring? Fed. Reserve Bank Philadelphia Bus. Rev. 5 (September/October), 3–16.
Meyer, L., 1998. The present and future roles of banks in small business finance. J. Banking Finance 22, 1109–1116.
Petersen, M.A., Rajan, R.G., 1994. The benefits of lending relationships: Evidence from small business data. J. Finance 49, 3–37.
Petersen, M.A., Rajan, R.G., 1995. The effect of credit market competition on lending relationships. Quart. J. Econ. 110, 407–443.
Petersen, M.A., Rajan, R.G., 2002. The information revolution and small business lending: Does distance still matter? J. Finance 57, 2553–2570.
Rossi, C.V., 1998. Mortgage banking cost structure: Resolving an enigma. J. Econ. Bus. 50 (2), 219–234.
Roszbach, K., 2004. Bank lending policy, credit scoring, and the survival of loans. Rev. Econ. Statist. 86, 946–958.
Scott, J.A., 2004. Small business and the value of community financial institutions. J. Finan. Services Res. 25, 207–230.
Shumway, T., 2001. Forecasting bankruptcy more accurately: A simple hazard approach. J. Bus. (January), 101–124.
Stein, J.C., 2002. Information production and capital allocation: Decentralized vs. hierarchical firms. J. Finance 57, 1891–1921.
US Government Accounting Office (GAO), 1999. Small business administration size of the SBA 7(a) secondary markets is driven by benefits provided. May.
US Small Business Administration (SBA), 2000. Small business lending in the United States. June.
Wolken, J.D., Rohde, D., 2002. Changes in the location of small businesses' financial service suppliers between 1993 and 1998. Internal memorandum, Federal Reserve Board.