Small Business Lending and Social Capital: Are Rural Relationships Different?

Robert DeYoung**
University of Kansas, Lawrence, KS USA

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

Peter Nigro
Bryant University, Smithfield, RI USA

Kenneth Spong*
Federal Reserve Bank of Kansas City, Kansas City, MO USA

June 2012

Abstract: Rural communities often are described as places where "everyone knows each other's business." Such intra-community information is likely to translate into a stock "social capital" that supports well-informed financial transactions (Guiso, Sapienza and Zingales 2004).

We investigate whether and how the "ruralness" of small banks and small business borrowers influences loan default rates, using data on over 18,000 U.S. Small Business Administration (SBA) loans originated and held by rural and urban community banks between 1984 and 2001. These data provide a good test of the value of soft information and lending relationships because (a) these borrowers tend to be smaller, younger, and more credit-challenged than other small businesses and (b) these loans were originated largely before the advent of small business credit scoring and securitization, and hence they were held in portfolio and put some bank capital directly at risk.

We have two main findings. First, loans originated by rural community banks and/or loans borrowed by rural businesses default substantially less often than loans made by urban banks and/or in urban areas. Second, loan default rates are significantly higher when borrowers are located outside the geographic market of their lenders, even after accounting for the physical distance between the bank and the small business. Thus, we conclude that loan defaults are lower in communities arguably expected to have large amounts of inexpensive soft information and at banks likely to have a high level of personal knowledge about their customers.

Our findings offer an explanation for why community banks—and in particular, rural banks—continue to exist despite operating at such small scale; why small local banks play a critical role in lending to small, informationally opaque borrowers; and why small rural banks are less likely to use small business credit scoring than their small urban counterparts. Moreover, our findings are consistent with the idea that a high stock of social capital is conducive to financial activity and development.

The authors are grateful for suggestions made by Tara Rice and by participants at a small business lending workshop the Federal Reserve Bank of Minneapolis. *The views and opinions expressed in this paper do not necessarily reflect the views and opinions of the Federal Reserve Bank of Kansas City, the Federal Reserve System, the Office of the Comptroller of the Currency, or the U.S. Treasury Department.
**Corresponding author: KU School of Business, 1300 Sunnyside Avenue, Lawrence, KS 66045, rdeyoung@ku.edu, 785-864-1806.

1. Introduction

Commercial banks play a central role in providing credit to opaque small businesses, chiefly because banks can foster borrower-lender relationships that transcend the information asymmetries that prevent these firms from accessing capital markets. These information problems are likely to be the greatest in the case of new businesses and other small businesses that have had blemished credit records. For such businesses, securing access to bank credit is an essential element that determines whether they have a chance to succeed and make their unique contributions to economic growth and job opportunities. At the same time, contracting with these businesses creates credit risk (and potentially insolvency risk) for lenders, so lenders must be expert in identifying the creditworthiness of these firms. A key research question is which lenders, and which lending technologies, perform best in this role of lending to new and struggling businesses.

It is an article of faith among many that community banks are better at this task than larger banks.[1] Community banks focus predominantly on local borrowers and local depositors, resulting in network economies in gathering information about the creditworthiness of local small businesses. Because community banks are in many ways small businesses themselves, they can empathize with small business people in ways that larger lenders cannot. And the geographically undiversified nature of community bank loan portfolios provides incentives for them to make efficient credit allocations. A trade association for the community banking industry has summarized these arguments as follows: "Community banks have only 12 percent of all bank assets but make 20 percent of all small business loans… Because most community banks are locally owned and operated, they have strong ties to their local communities. These banks have a close relationship with their customers and are quite familiar with their customers' financial condition and their history and ability to repay. The success of community banks is tied directly to the success or failure of their customers and their communities (ICBA 2011)." A

[1] Standing in potential contrast to this position are the aggregate lending data, which show that the lion's share of small business loans in the U.S. are extended by large banking companies, not by community banks. However, our study is about marginally qualified small business borrowers and the quality of their loans, not about average small business borrowers and the dollar value of credit they receive.

similar conclusion has been drawn by two Federal Reserve economists: "Community bankers typically know their customers better than bankers at larger organizations, and perhaps this knowledge of local people and local businesses offsets the exposure to local economic downturns. As a consequence, community banks seem to hang tight through the choppy waters of local economic downturns (Hall and Yeager 2002)."

These phenomena may be more pronounced in rural towns, where banks are even more closely tied to and dependent upon the local economy, where personal relationships are a more integral part of the social fabric, and where informal information about potential borrowers is likely to be more plentiful or easier to access. Despite two decades of continuous banking industry consolidation that has roughly halved the number of U.S. commercial banks, 59% of all remaining banks are located in rural counties, places that account for only 21% of the U.S. population. Based on conventional measures, nearly all of these banks are operating below minimum efficient scale, with undiversified loan portfolios exposing them to volatile local (agriculture-driven) economies. Do rural banks enjoy a competitive advantage in lending to small businesses so substantial that it offsets their size-based disadvantage?

Ascertaining the creditworthiness of rural small businesses can pose a number of challenges, not the least of which is that many rural small businesses are hard-information deficient. In rural local economies, the re-sale market for fixed investments and specialized assets is thin, which makes the value of seized collateral in the case of loan default uncertain. Rural small businesses are less likely than their urban peers to have audited financial statements, further reducing the amount and usefulness of hard information about their creditworthiness. The social and civic underpinnings of rural commerce are also fundamentally different from those in metropolitan markets. If one accepts the conventional wisdom that rural communities are closer knit than urban communities—so that "folks know each others' business"—then these personal informational nexuses no doubt extend into the rural business community as well. This gives rural banks a *costless* endowment of soft information about local businesses, which may serve to offset the production inefficiencies and risk diversification problems associated with small scale and the costly, labor-intensive nature of relationship lending and soft information collection.

One might characterize this hypothesized difference between rural and urban soft information environments as a difference in the level of "social capital." As described by Guiso, Sapienza and Zingales (2004), social capital impacts economic efficiency "by enhancing the prevailing level of trust. In high-social-capital communities, people may trust each other more because the networks in their community provide better opportunities to punish deviants. At the same time, in these communities people may rely more on others' keeping their promises because of the moral attitude imprinted with education." Guiso, Sapienza and Zingales (2004) identify exogenous differences in social capital across 95 Italian provinces, and find that financial transactions that involve trust (e.g., accepting personal checks as payment, extending credit to households) are more likely to occur in provinces with more social capital. In this study, we hypothesize that rural areas have higher levels of social capital than urban areas, and then test whether financial transactions that involve trust (relationship loans) are more likely to be successful in rural areas.

Using non-public data from the U.S. Small Business Administration (SBA) on over 18,000 loans made by community banks to small businesses between 1984 and 2001, we find evidence that small rural banks are especially good at underwriting and monitoring credit to small, informationally opaque firms. Loans made by small rural banks have a significantly lower likelihood of default than loans made by small urban banks. This performance advantage is positively related to the "ruralness" of the borrower-lender relationship: it intensifies as the size of the rural market declines and it weakens when the rural borrower and rural banker are located in different rural markets. Importantly, the data suggest that these advantages are not driven exclusively or even primarily by core competencies at rural banks, but more broadly, these advantages are made available to banks located in rural markets *via* the characteristics of rural businesses, rural economies and rural cultures and are indicative of the value of soft information and "knowing your customer."

Our findings suggest "ruralness" is a primary reason that small rural banks continue to exist in disproportionate numbers and operate successfully at less than efficient scale, and they offer an explanation for why rural banks are more likely to be portfolio lenders and less likely to engage in small business credit-scoring and securitization. In short, the classical description of banks as "special"—i.e., repositories of private local information that allows them to outperform both non-banks and non-local banks in the analysis of local creditworthiness—is nowhere more accurate than in rural places.

The rest of the paper proceeds as follows. We review the relevant literature on small business lending in Section 2. Formal, testable hypotheses are introduced in Section 3. We describe our small business loan data in Section 4, present the econometric methodology in Section 5, and report the results of our tests in Section 6. Section 7 concludes.

2. Literature review

A large literature focuses on the importance of relationship lending and soft information in granting credit to small businesses—but very little of this literature examines whether there might be differences in lending technologies and credit performance across different types of community banks. The extant literature addresses a number of related questions, for example: how bank size influences small business borrower-lender relationships and soft information-based lending; how banking industry consolidation and the adoption of credit scoring have affected the supply of small business loans; and how relationship lending benefits small businesses (e.g., better access to credit, more favorable lending terms).

Much of the research on small business lending is based on the idea that small businesses are informationally opaque and that informational asymmetries between borrower and lender may limit the types of lenders and the lending technologies that may be used. Petersen (2004), for instance, discusses how the amount of information available on firms will influence their access to capital and the structure of financial markets and financial institutions. Petersen draws a distinction between hard and soft information, with hard information being more numerical and encompassing such items as financial statements, payment history, and output numbers while soft information is more qualitative and includes opinions, ideas, rumors, and statements of future plans. Much of what banks use in relationship lending and lending to more opaque small businesses would be characterized as soft information. According to

Petersen, such information is largely collected by the lending officer and is not easily transmitted to other parts of the banking organizations or readily available to other lenders or investors.

Boot (1999), Elysiani and Goldberg (2004), and Udell (2008) provide summaries of the literature that has developed on relationship lending, soft information, and other related topics. These surveys find that relationship banking has a distinct and important role to play in small business lending and financial intermediation. But these surveys also report that existing studies offer conflicting evidence regarding some dimensions of small business-bank lender relationships, and as such a number of interesting questions still remain unanswered.

Much of the relationship lending literature finds that smaller banks are likely to have an advantage in lending to informationally opaque small businesses and will direct more of their lending toward small businesses than larger banks will. Nakamura (1994) asserts that small banks have an informational advantage that makes them best able to lend to and closely monitor local small businesses, and the tight organizational structure of most small banks allows them to effectively exploit this advantage. Scott (2004) finds that small firms give higher satisfaction ratings to community banks (and lower ratings to large banks) regarding their performance in meeting credit needs and maintaining strong banking relationships. Using data on small firms in Argentina, Berger et al. (2001) conclude that large institutions and foreign-owned institutions have difficulty extending relationship loans to small firms. Berger et al. (2005) find suggestive evidence that small banks are better able to collect and act on soft information and tend to lend to more difficult credits, while large banks lend on a more impersonal basis and have shorter and less exclusive lending relationships.

The important role that small banks play in relationship lending would suggest that the banking consolidation of the past few decades might be expected to have an adverse effect on small business lending. But a number of studies (e.g., Avery and Samolyk 2004; Berger et al. 1998) find that any declines in small business lending at consolidated institutions are mostly offset by lending increases at other banks and by the entry of new banks. Other studies conjecture (Petersen and Rajan 2002) or document (DeYoung et al. 2010) that improvements in information processing, credit scoring, and communications are increasing the credit access of small firms by enabling them to contract with more distant sources. This extended geographic reach may be enabling larger, non-local banks to lend to small local businesses and thus eroding the relationship lending advantages of small banks (Berger and Rice 2010). In recent years, there has been an increased use of credit scores by small banks to evaluate small business loan applications; however, these banks tend to use the consumer credit score of the business owner, not the more encompassing small business credit score used by large banks (Berger, Cowan and Frame 2011). Hence, even when using hard information, small banks remain more likely to focus on the character and creditworthiness of the person applying for the loan.

Another line of research provides evidence that stronger relationships help to increase the availability of credit and improve lending terms and credit quality. For example, Petersen and Rajan (1994, 1995) and Cole (1998) generally find that relationships result in more available credit for small firms. The Petersen and Rajan studies, along with Berger and Udell (1995) and Garcia-Appendini (2011), also indicate that relationship lending and the availability of soft information can contribute to lower interest rates on small business loans. Only more recently have researchers (DeYoung et al. 2008; FDIC 2010; Garcia-Appendini 2011) looked at measures of credit quality in small business lending; these studies have typically found higher lending risks when banks go beyond local relationships and make little use of soft information, especially when lending outside of their own markets.

Only a few studies have looked at how lending relationships might differ across particular types of banks. Berger and Udell (2002) and Stein (2002) provide models of how a bank's organizational and management structure might influence its relationships with borrowers, and these models have implications for how lending relationships might vary across different community banks. Berger and Udell, for instance, assert that a loan officer's accumulation of soft information over time may create agency problems between the loan officer and a bank's management and stockholders, and these agency problems can be addressed best when the bank is a small, closely-held organization with few managerial layers. They characterize the loan officer as typically living in the local community and having the greatest access to soft information about a small business, its owner, and the community itself. These authors also contend that the resulting agency and contracting problems will be controlled better in small organizations where the president is actively engaged in making or reviewing most business lending decisions. As a result, their model would appear to imply that small banks in small rural markets might have the fewest agency problems since both the bank president and loan officer, if separate individuals, would each have extensive personal knowledge of most borrowers and might jointly be involved in many of the lending decisions.

A number of studies have compared and contrasted the management structure and lending operations at rural banks versus urban banks, although these studies do not systematically differentiate between large and small (i.e., community) urban banks. Kittiakarasakun (2010) compares banks in rural and urban areas and finds evidence that urban banks tend to rely more on verifiable (hard) information while rural banks tend to lend to nearby customers, about which they would have local, personal knowledge. These results mirror what Cole et al. (2004) found previously in comparisons of large and small banks. Cowan and Cowan (2006) further conclude that rural banks are less likely to use credit scoring for small business loans when compared to urban banks. In addition, Brickley et al. (2003) show that the office, ownership, and management structure of banks is likely to vary across different markets and customer bases, with rural areas primarily having locally controlled banks due to the advantages such banks may have in knowing customers and making decisions locally.

3. Hypotheses

The goal of this study is to generate evidence useful for addressing the following broad question: Are small rural borrowers more likely than small urban borrowers to have strong relationships with their banks? To answer this (and related) questions, we compare and contrast the default rates for loans made by rural and urban community banks to rural and urban small businesses.

It is natural to use default rates to draw inferences about the existence of lending relationships. A true borrower-lender relationship generates soft information about the borrower's creditworthiness that that bank can use to construct a sustainable lending strategy that reduces the likelihood of borrowing default. Just as important, lenders should be willing to incur short-term costs in order to develop and preserve a valuable long-term lending relationship, and hence will be more likely to restructure a troubled loan rather than calling the loan and forcing default.

We posit and test six inter-related hypotheses that the characteristics of agents (firms, banks), agent types (urban, rural) and/or agent locations (local, out-of-market) influence the probability of loan default. We state each of these hypotheses in a neutral fashion, so that each hypothesis supports a two-sided test that treats rural loans and urban loans symmetrically.

Let D(x) represent the probability that a loan will default, where x indicates locational information about the borrowing firm and the lending bank as follows:

x = RR indicates a rural firm borrowing from a rural bank

x = UU indicates an urban firm borrowing from an urban bank

x = RU indicates a rural firm borrowing from an urban bank

x = UR indicates an urban firm borrowing from a rural bank

This four-way taxonomy of borrower-lender pairs is exhaustive and mutually exclusive. We use this taxonomy to state our first two hypotheses.

H1. Ruralness Hypothesis: The efficiency of loan contracting and monitoring in rural markets will differ from that in urban markets, due to differences in informational, institutional and/or cultural conditions in those markets.

Using the notation introduced above, D(RR) < D(UU) is consistent with the existence of ruralness that makes loan contracting and monitoring more efficient—in other words, this finding would suggest that borrower-lender relationships are stronger in rural markets.[2] Similarly, D(RR) > D(UU) is consistent with the existence of ruralness that *reduces* the efficiency of loan contracting and monitoring. Should tests of this hypothesis find evidence consistent with ruralness, we will of course be interested in the source(s) of the ruralness. Ruralness might emanate from systematic differences in the practices of rural banks, from systematic qualities of rural small businesses, or from pairing rural banks with rural small businesses. Our second hypothesis examines the last of these three possibilities.

H2. Borrower-Lender Empathy Hypothesis: Agents of the same type—that is, urban banks paired with urban firms, or rural banks paired with rural firms—share informational, institutional and/or cultural similarities that make loan contracting and monitoring more efficient.

All else equal, D(RR) < D(RU) and D(RR) < D(UR) are consistent with borrower-lender empathy in rural markets. Similarly, D(UU) < D(RU) and D(UU) < D(UR) are consistent with borrower-lender empathy in urban markets. (While borrower-lender empathy may be an underlying cause for ruralness, it is important to note that borrower-lender empathy can exist independent of ruralness.)

Testing our remaining hypotheses requires the following more detailed taxonomy of the relative locations of borrower-lender pairs:

x = RRL indicates a rural firm borrowing from a local rural bank

x = RRNL indicates a rural firm borrowing from a non-local rural bank

x = UUL indicates an urban firm borrowing from a local urban bank

x = UUNL indicates an urban firm borrowing from a non-local urban bank

x = RU indicates a rural firm borrowing from an urban bank

x = UR indicates an urban firm borrowing from a rural bank

[2] Note that a lower default rate for relationship lending does not necessarily map into higher profitability. Numerous other facets of loan production—such as pricing, optimal scale, diversification effects, and ancillary revenues—differ across the relationship and non-relationship lending models and have important impacts on bank profitability.

where "non-local" refers to a lender located in a city, town or county different from the borrower (details below). This six-way taxonomy splits same-type loans (RR, UU) into local loans (RRL, UUL) and non-local loans (RRNL, UUNL), and as such continues to be exhaustive and mutually exclusive. The finer detail permits us to isolate the effects of "localness" apart from the effects of "ruralness."

H3. Local Lending Hypothesis: Pairs of agents in the same local market have relatively low costs of information sharing and/or information revelation, making loan contracting and monitoring more efficient.

All else equal, D(RRL) < D(RRNL) is consistent with local market lending efficiency effects in rural markets. Similarly, D(UUL) < D(UUNL) is consistent with local market lending efficiency effects in urban markets. While finding evidence of localness *per se* would not be surprising—for instance, DeYoung et al (2008) found localness, measured in terms of the geographic distance between borrowers and lenders, to be a strong determinant of small business loan performance—this test may determine whether localness is relatively more or less important in rural or urban settings.

Should we find evidence of ruralness in the data, the finer detail in the six-way taxonomy provides us with a way to identify the roots of ruralness: Does ruralness reside primarily in differences between rural and urban small businesses borrowers? Or does it arise due to differences between rural and urban small community bank lenders?

H4. Credit Analysis Hypothesis: Lender types (urban banks vs. rural banks) are different in ways that impact the efficiency of loan contracting and monitoring.

If rural banks are better at credit analysis than urban banks, then all else equal, we should find the joint result D(RRNL) < D(RU) <u>and</u> D(UUNL) > D(UR). This follows from holding constant the type of the borrower (rural or urban) on either side of each inequality, while allowing the type (rural or urban) of the *non-local lender* to change on either side of each inequality to change. By looking only at loans made by out-of-market banks, this test neutralizes the potential effects of localness, which may be different in purely rural local lending environments (RRL) and purely urban local lending environments (UUL). Alternatively, a joint finding that D(RRNL) > D(RU) and D(UUNL) < D(UR) would indicate that urban banks are better at credit analysis than rural banks.

H5. Credit Quality Hypothesis: Borrower types (urban firms vs. rural firms) are different in ways that impact their credit risk.

If rural firms are better credit risks than urban firms, then all else equal, we should find the joint result D(RRNL) < D(UR) and D(UUNL) > D(RU). This follows from allowing the type of the *local* borrower (rural or urban) to change on either side of each inequality, while holding constant the *non-localness* of the lender on either side of each inequality. Again, by looking only at loans made by out-of-market banks, this test neutralizes the potential effects of localness. Alternatively, a joint finding that D(RRNL) > D(UR) and D(UUNL) < D(RU) would indicate that urban firms are better credit risks than rural firms.

Unfortunately, evidence that is consistent with the Credit Quality Hypothesis that rural (or urban) firms have higher credit quality would also be consistent with the alternative hypothesis that credit supply restrictions force rural (urban) firms to look to urban (rural) banks for loans. When credit access is restricted for a given group of borrowers—due to market frictions that are exogenous to those borrowers, such as market power or discrimination in lending—then that group of borrowers will exhibit a relatively low rate of loan default, because the marginal accepted loan applicant in the credit restricted group will be more qualified than the marginal accepted loan applicant outside that group. Thus, our final hypothesis:

H6. Credit Access Hypothesis: Systematic credit restrictions in local markets cause local firms to seek credit in non-local markets. These ex-patriot credit-restricted firms will exhibit above-average credit quality in the non-local markets in which they find a loan.

In other words, the Credit Quality Hypothesis is pooled with the Credit Access Hypothesis in our tests.[3] In the latter case, low default rates are consistent with strong lending relationships and high quality credit information. In the former case, low default rates are consistent with reduced credit access in which only low-risk firms get loans.

We make efforts to disentangle these two pooled effects in our regression tests. Ideally, we would be able to absorb differences in the restrictiveness of credit across the local markets in our data by including control variables for aggregate local loan supply in our tests (e.g., small business loans per capita). However, we cannot construct such control variables from our database because these loans were drawn randomly across (not within) geographic markets. Moreover, systematic aggregate data on small business loans are not available from other sources. Instead, we attempt to adjust for local credit conditions using the second-best approach of including in our regression tests control variables that measure local lender concentration and state-level restrictions on bank entry.

We also appeal to the (scant) academic literature that attempts to reveal objectively whether rural borrowers face limited credit access. Using data from the 1993 National Survey of Small Business Finance—data that falls in the middle of our 1984-2001 data—Walraven (1999) finds evidence that (a) rural small businesses were significantly less likely than urban small businesses to apply for a loan, and (b) rural small business loan applications were significantly more likely to be accepted than urban small business loan applications. In addition, Briggeman and Akers (2010), using data from both the 2005 Agricultural Resource Management Survey and the 2003 National Survey of Small Business Finance,

[3] Loan defaults provide high-quality information about both credit risk and credit access. Because they are market outcomes, loan defaults do not suffer from some of the problems associated with other measures of credit access, such as (a) self-reported borrower data and surveys on the number of loan applications made and rejected or (b) restrictive loan contract terms which can pool information about credit risk with supply-side attempts to restrict access to credit.

conclude that rural small business owners reported having fewer problems in receiving credit than their urban counterparts. Taken together, the findings from these two studies are inconsistent with the notion that credit access in rural markets is restricted relative to urban markets.[4] Thus, should our tests of the Credit Quality Hypothesis generate evidence consistent with relatively high credit quality for rural small businesses, we could reasonably argue that Hypotheses H5 and H6 are not pooled in our data, based on the findings of these two studies.

4. Data and bivariate tests

We test our hypotheses using non-public data on 18,160 small business loans made by U.S. banks between January 1984 and April 2001 under the Small Business Administration (SBA) flagship 7(a) loan program. The 7(a) loan program provides loan guarantees for small business firms that are unable to access the same credit on reasonable terms from non-federally guaranteed sources; as such, SBA loans are critical to the flow of credit to small firms. Lenders 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. The SBA guarantee is a *pro rata* loss sharing arrangement between the SBA and the lender, so the lender puts some of its capital at risk for every loan. The percentage of the loan principle that is guaranteed varies across loans, and we control for this in our regression tests.

All of the loans in our data were made by relatively small, or "community," banks. For the purposes of this study, we define a community bank as a U.S. commercial bank with less than $1 billion in assets in inflation-adjusted 2000 dollars. The 18,160 loans made by these banks is a subsample of a larger initial sample of 31,880 SBA loans made by banks of all sizes during the sample period. The larger initial group of loans is a 20% random sample from the population of all 7(a) loans originated during the sample period, weighted by the number of loans in the population each year. As suggested by the data in Table 1, the number of SBA loans has varied substantially from year to year; these fluctuations mainly

[4] By itself, the Walraven (1999) finding could indicate that limited credit access discouraged rural firms from applying for loans. But the Briggeman and Akers (2010) finding makes this possibility unlikely.

reflect changes in the level of federal funding of SBA programs across time, and we control for these annual fluctuations in our main tests.[5]

The historical nature of these data is an advantage for the purposes of our study. These loans were originated between 1984 and 2001, before automated, hard information-based lending techniques (credit scoring, loan securitization) were widely applied to small business loan applications. Although large banks began credit scoring small business loan applications in the early 1990s, community banks did not adopt this new technology until some years later (Frame, et al. 2001). So during our sample period, community banks in both urban and rural markets were using very similar lending technologies for small business loan applicants—i.e., screening loan applicants based disproportionately on local, soft information and holding the resulting loans on their balance sheets. This allows us to test more cleanly for the pure effects of ruralness, without the potentially confounding effects that emerged in later years when urban banks became more reliant on hard-information lending than rural banks.[6]

Using data on SBA loans—borrowers that were unable to access regular market (non-federally guaranteed) lending channels—is also an advantage for the purposes of this study. Because these firms typically lack well-documented credit histories, the evaluation of their creditworthiness turns disproportionately on the ability of banks to gather and utilize soft information about the loan applicant. Much of this information arises from the bank's relationship with the applicant, but just as importantly information comes from the applicant's customers, suppliers and non-business acquaintances in the local market. While the government subsidy reduces the bank's exposure to credit risk, risk is not fully mitigated—the loan principle is only partially guaranteed by the SBA—so the lender has clear incentives to collect and analyze this information.

The SBA data include the Zip code of each borrower (the address of the business establishment) and each bank (the address of the lending office, usually but not always the bank's headquarters). We use

[5] For further details of how this initial 20% random sample was drawn, see Glennon and Nigro (2003).
[6] Cowan and Cowan (2006), for instance, find that the greater a bank's investment in farm loans relative to total loans (their measure of ruralness), the lower the likelihood that the bank will use credit scoring.

this information to determine whether borrowers and lenders are "urban" or "rural," which allows us to assign each loan to the four-way and six-way taxonomies defined above. Urban borrowers or lenders are located in Metropolitan Statistical Areas (MSAs) and rural borrowers are located in non-MSA counties.[7] Knowing the MSA or non-MSA county in which the borrower is located allows us to merge in a variety of information about the borrower's local market conditions, such as banking market structure calculated from the FDIC Summary of Deposits database. Knowing the identity the lender allows us to merge in detailed bank financial statement data from the FDIC Reports of Condition and Income (the call reports).

Overall, 15.4% of the loans in our data set defaulted. Table 2 displays the average default percentages for each of the borrower-lender pairs in the four-way (Panel A) and six-way (Panel B) taxonomies. The difference-in-means tests provide bivariate (uncontrolled) tests of our six hypotheses. Only a few of the hypotheses receive statistically significant support. There is no support for the Ruralness hypothesis, as average D(RR) is not statistically different from average D(UU). There is partial support for the Borrower-Lender Empathy hypothesis, both in rural markets where D(RR) < D(UR) and in urban markets where D(UU) < D(UR). In contrast, there is strong support for the Local Lending Hypothesis, both in rural markets where D(RRL) < D(RRNL) and in urban markets where D(UUL) < D(UUNL). There is no support for the Credit Analysis, Credit Quality, or Credit Access Hypotheses. Finally, when the differences-in-means are statistically significant, the economic magnitudes of these differences are non-trivial: the bivariate differences in loan default rates in Table 2 range in size from 2.7% to 3.2%, on the order of 20% of the average sample loan default rate of 15.4%.

5. Multivariate model

The preliminary difference-in-means tests in Table 2 do not control for a host of exogenous conditions shown elsewhere to affect the performance of SBA loans (e.g., DeYoung, Glennon and Nigro 2008). For our main tests, we re-examine our hypotheses using a "stacked-logit" model of loan default

[7] To make these determinations, we compared the Zip code of each borrower and lender with Zip code information in the 1990 and 2000 U.S. Census databases. We were unable to make confident location identifications for 670 loans, less than 4 percent of the final sample.

that allows us to control for these exogenous conditions. The model is constructed as follows: Assume that each loan i (i =1,2,…,N) is originated during period t=0 and enters the model T times as a series of binary variables $D_i(1),...D_i(T_i)$. $D_i(t)=1$ if loan i defaults during time period t and $D_i(t)=0$ otherwise, over the life of the loan. For example, measuring time in calendar quarters, the event history for a 3-year loan will be five zeros followed by a one (0,0,0,0,0,1) if the loan defaults in the sixth quarter after it was originated, but will be a string of twelve zeros (0,0,0,0,0,0,0,0,0,0,0,0) if the loan does not default.[8] The 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:

$$D^*_{it} = \mathbf{X_i} \boldsymbol{\beta} + \mathbf{W_{it}} \boldsymbol{\gamma} + \varepsilon_{it}$$

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:

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

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

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

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

[8] 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.

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

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

where $\Lambda(\cdot)$ is the logistic cumulative distribution function. We estimate equation (1) using standard binomial logit techniques.[9] We specify the model as follows:

$$\Pr[\, D_{it}{=}1|\mathbf{Z}\,] = \Lambda[\, \textit{BORROWER-LENDER}_i, \textit{loan controls}, \textit{lender controls}, \textit{borrower controls}, \textit{market controls}, \textit{YEAR}_t; \phi \,] \quad (1)$$

where 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. Our main statistical tests are provided by the coefficient estimates on the dummy variables in *BORROWER-LENDER*, which is a vector comprised either of (RR, UU, RU and UR) or (RRL, UUL, RRNL, UUNL, RU and UR) in different specifications. We include four vectors of control variables (*loan controls*, *lender controls*, *borrower controls*, *market controls*) to capture default-relevant variation in the terms of the loan contracts, the characteristics of bank lenders, the characteristics of borrowing firms, and the characteristics of the borrowers' local markets. The vector of *YEAR* dummies is included to control for annual variation in loan default rates. Table 3 displays descriptive statistics and definitions for all of the variables used in the stacked-logit tests. Unless otherwise indicated as time-varying, all of the variables are observed at the time of loan origination.

Some of the control variables bear special mention because their inclusion may be important in identifying some of our hypotheses tests. *Distance* is the "as the crow flies" mileage between the

[9] This approach 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, 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 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.

borrower's home market and the market in which the lending office is located.[10] We include this variable to soak up any pure effects of distance on loan default (i.e., procuring information is more costly, and monitoring is more difficult, at a distance), which leaves only the impact of ruralness on loan default in our cross-market RU, UR, RRNL and UUNL coefficients. *HHI* is the deposit-share Herfindahl index for banks and thrifts in the borrower's local market, and *Branching* is a dummy equal to one in states with restrictions of bank branches. We include HHI and Branching to soak up market power-related effects on credit access, although we have no *a priori* expectations regarding their coefficient signs. On the one hand, if market power results in less output, then default rates will be lower because marginal loan applicants will be denied loans. On the other hand, market power could expand output by making lenders are more willing to invest in a soft-information relationship; in this case, default rates will be lower (higher) if the investment in soft information improves credit screening by more (less) than the increased output reduces the creditworthiness of the marginal borrower.[11] Due to their small size, rural banking markets tend to be especially concentrated, so we also include the interaction term *HHI*Urban*, where *Urban* is a dummy variable equal to one in urban borrower markets. *%SBA* is the percentage of the loan principle that is guaranteed by the SBA. We include this variable to control for the severity of potential moral hazard incentives, in which banks originating and holding loans with higher guarantees may have reduced incentives to carefully screen and monitor loans.

6. Results

The stacked-logit estimates displayed in Tables 4-7 are generally more consistent with our hypotheses than are the bivariate difference-in-means tests in Table 2. To test our hypotheses we need to compare loan default probabilities across all combinations of borrower-lender pairs. This required us to estimate equation (1) multiple times (12 times for the four-way taxonomy, 30 times for the six-way

---

[10] The exact locations used to calculate *Distance* are the geographic Zip Code centroids for the borrower and the lender. In the regressions, we specify this variable as ln(*Distance* + 1).

[11] Petersen and Rajan (1994, 1995) argue that a bank will be more willing to invest in costly information collection as its market power increases because the bank is less likely to lose the borrower to rivals.

taxonomy), each time excluding a different borrower-lender dummy. To conserve space, we report the estimated logit coefficients in Table 4 (four-way taxonomy) *only for* specifications in which RR was the excluded borrower-lender dummy; then in Table 5 we report the marginal default probabilities (expressed as odds ratios) associated with *all* borrower-lender pairs. Similarly, we report logit coefficients in Table 6 (six-way taxonomy) only for specifications in which RRL was the excluded borrower-lender pair dummy, followed by a full set of marginal default probabilities in Table 7.

In addition to the full sample tests, we also estimate the model for three subsamples of loans. The first subsample contains only those loans written by the very small banks having less than $100 million of assets. If ruralness exists, it may be more likely to reside at smaller banks that exhibit conditions typically associated with relationship lending, such as flat management structures, local focus, and easy borrower access to bank officials. The estimates from this subsample may indicate whether ruralness is an important and separate factor among very small banks that are most likely to have strong relationships with their customers. The second subsample excludes loans to firms located in rural markets with populations greater than 50,000 people. If ruralness exists, it may be more likely to be found in smaller places where information, institutions and culture are more conducive to relationship lending. The third subsample excludes loans in which the borrowing firm and the lending bank were located in different geo-political markets but were still less than 25 miles apart in terms of *Distance*. Because close proximity between two out-of-market agents may facilitate borrower-lender relationships similar to those between in-market agents, removing these observations may make improve identification for some of our hypothesis tests.

6.1. Testing the main hypotheses. We find strong evidence consistent with the Ruralness hypothesis. Based on the data in Table 5, Panel A, loans between rural firms and rural banks (RR) are only 70% as likely to default as loans between urban firms and urban banks (UU), all else held equal. This result is somewhat stronger among smaller community banks (66% in Panel B) and in smaller rural markets (63% in Panel C). Evidence consistent with the last result can be found in Table 4, where the estimated coefficient on *Rural(<10k)* is negative and significant in regression (2). This indicates a kind of "hyper-ruralness" in which loans originated by banks in very small rural markets with populations less than 10,000 were statistically less likely to default than the typical RR loan.

We also find strong evidence consistent with the Borrower-Lender Empathy Hypothesis—but crucially, only for rural borrower pairs. Based on the data in Table 5, Panel A, local rural loans (RR) were only 78% as likely to default as loans in which rural firms borrowed from urban banks (RU) and only 55% as likely to default as loans in which urban firms borrowed from rural banks (UR). This rural borrower-lender empathy result is robust for all three of the subsamples (see Panels B, C and D). However, there is no similarly robust evidence consistent with special empathy between urban banks and urban borrowers—in all four Panels of Table 5, the odds ratios in the cells intersected by the UU row and the RU and UR columns are seldom statistically significant. This suggests that our general finding of Ruralness (i.e., the strong D(RR) < D(UU) results discussed in the previous paragraph) is driven by some (as yet undefined) inter-agent empathy among rural borrowers and lenders.

The more detailed six-way taxonomy permits more discriminating hypothesis tests, and permits us to better identify the root sources of ruralness. We find clear support the Local Lending Hypothesis for both rural and urban lending. Based on the data in Table 7, Panel A, rural firms that borrow locally (RRL) were only 90% as likely to default as rural firms that borrow from non-local rural banks (RRNL). Similarly, urban firms borrowing locally (UUL) were only 78% as likely to default as urban firms borrowing from non-local urban banks (UUNL). These results are relatively robust across the subsamples—although the relatively weaker (statistical and economic) results for the rural loans suggest that the advantages of "ruralness" are to some degree portable across rural markets.

We find no evidence to support the Credit Analysis Hypothesis for either rural banks or urban banks—hence, our general finding of Ruralness is not driven by superior credit screening or monitoring by rural banks. However, we do find evidence consistent with the Credit Quality Hypothesis for rural firms. Based on the data in Table 7, Panel A, out-of-market rural loans (RRNL) were only 65% as likely to default as urban firms borrowing in rural markets (UR), while at the same time, out-of-market urban loans (UUNL) were 31% more likely to default than rural firms borrowing in urban markets (RU). In other words, when we hold constant the Local Lending effect by looking only at out-of-market loans, we find that rural firms were better credit risks than urban firms. We also find support for the Credit Quality Hypothesis for both the small lender subsample (Panel B) and the small rural market subsample (Panel C), but the result weakens to statistical non-significance in the Panel D subsample.

As discussed above, it is difficult to separate evidence consistent with the Credit Quality Hypothesis that rural borrowers are better credit risks from the alternative Credit Access argument that scarce credit supply in rural markets could result in relatively higher quality rural borrowers seeking loans out-of-market. Despite including variables in the regressions that control for *supply* side conditions in local lending markets (*HHI*, *HHI*Urban*, *Branching*), and despite the results of previous research suggesting the rural small businesses are not more credit constrained (and if anything, are less credit constrained) than urban small businesses, we cannot definitely reject the Credit Access Hypothesis as a potential explanation for our findings.

6.2. Control variables. As mentioned above, the control variables *Distance*, *HHI*, *Branch* and *SBA%* are perhaps more important than the other control variables for identifying our main hypotheses. As expected, default rates are positively related to *Distance*, an indication that including this control variable helps separate pure borrower-lender distance effects from non-local lending effects, both of which increase the probability of loan default. The probability of default is lower in highly concentrated urban markets (*HHI*Urban*) but not in highly concentrated rural markets (*HHI*), an indication that including these variables helps control for reduced credit access due to market-power supply side constraints. The *Branch* variable, included to control for the supply side impact of potential market entry on competition, is never statistically significant. The probability of default is unrelated to *SBA%* for the typical loan in the data, an indication that no serious moral hazard incentives were associated with the typical SBA loan during our sample period. However, the probability of default increases with the size of the loan guarantee for so-called “low-doc” loans (for which the borrower provides only limited amounts of personal financial information in the loan application).

7. Conclusions

The disproportionate share of small business loans in the portfolios of community banks is *prima facie* evidence that these banks have a comparative and/or competitive advantage at extending credit to small, informationally opaque borrowers. The traits and conditions so often cited for these advantages of community banks—local focus, customer relationships, and flat organizational structures—should be even more in evidence at rural community banks. Rural banks are relatively small even for community banks, and this should reinforce the advantages of flat structure and local focus. Rural social mores and institutions and the small size of most rural markets emphasize the value of personal relationships, and help generate plentiful and cheap soft information, the mother's milk of lending to opaque firms.

While there are scores if not hundreds of research studies on community banks, there is very little evidence comparing the small business lending effectiveness of rural and urban community banks. This study aims to rectify that shortage. We examine the incidence of loan default among 18,000 Small Business Administration (SBA) loans made by rural and urban community banks between 1984 and 2001. Because SBA borrowers tend to be smaller, younger, and more credit-challenged than other small businesses, these loans provide a good test of the *ex ante* efficiency with which small banks screen small business loan applications and the *ex post* efficiency with which they monitor small business loans. In particular, SBA lending should provide a basis for judging the value of soft, personal information in limiting the risk of extending such credit.

Our primary finding is that loans originated by rural community banks default substantially less frequently than loans originated by urban community banks. This performance advantage increases for relatively smaller rural banks, increases in relatively smaller rural markets, and diminishes when the borrower and banker are located in different rural markets. These results are consistent with the existence of what we refer to here as "ruralness," a difficult-to-measure quality that may be rooted in higher levels of "social capital" in rural society (Guiso, Sapienza and Zingales 2004).

These findings may also help explain why small banks, and small rural banks in particular, continue to exist in the U.S. in disproportionate numbers and are able to operate successfully at less than efficient scale. The small business loans in our data are more likely to perform as (a) banks and borrowers grow smaller, (b) banks and borrowers share the same geographic market, and (c) banks and borrowers share a common set of social practices and institutions. The evidence we find thus provides support for the value of soft information and personal knowledge in lending to small and informationally opaque businesses and the value of community banks in performing this function. Given that the data used in this study come from small business startups and other small businesses that are especially credit-constrained, our findings may provide some lessons for economic policymakers aiming to more efficiently channel credit to small businesses.

References

Avery, R.B., and Samolyk, K.A., 2004. Bank consolidation and small business lending: The role of community banks. Journal of Financial Services Research 25, 291-325.

Berger, A.N., Cowan A.M., and Frame, W.S., 2011. The Surprising Use of Credit Scoring in Small Business Lending by Community Banks and the Attendant Effects on Credit Availability, Risk and Profitability. Journal of Financial Services Research 39, 1-17.

Berger, A.N., Klapper, L.F., and Udell, G.F., 2001. The ability of banks to lend to informationally opaque small businesses. Journal of Banking and Finance 25, 2127-2167.

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

Berger, A.N., and Rice, T., 2010. Do small businesses still prefer community banks. Working paper.

Berger, A.N., Saunders, A., Scalise, J.M., and Udell, G.F., 1998. The effects of bank mergers and acquisitions on small business lending. Journal of Financial Economics 50, 187-229.

Berger, A.N., and Udell, G.F., 1995. Relationship lending and lines of credit in small firm finance. Journal of Business 68, 351-382.

Berger, A.N., and Udell, G.F. 2002. Small business credit availability and relationship lending: The importance of bank organizational structure. The Economic Journal 112, F32-F53.

Boot, A.W.A., 1999. Relationship banking: What do we know? Journal of Financial Intermediation 9, 7-25.

Brickley, J.A., Linck, J.S., and Smith, C.W. Jr., 2003. Boundaries of the firm: evidence from the banking industry. Journal of Financial Economics 70, 351-383.

Briggeman, B.C., and Akers, M.M., 2010. The credit advantage of farm and rural small business ownership. Agricultural Finance Review 70, 353-364.

Cole, R.A., 1998. The importance of relationships to the availability of credit. Journal of Banking and Finance 22, 959-977.

Cole, R.A., Goldberg, L.G., and White, L.J., 2004. Cookie cutter vs. character: The micro structure of small business lending by large and small banks. Journal of Financial and Quantitative Analysis 39, 227-251.

Cowan, C.D., and Cowan, A.M., 2006. A survey based assessment of financial institution use of credit scoring for small business lending. Small Business Research Summary No. 283. Office of Advocacy. Small Business Administration.

DeYoung, R., Glennon, D., and Nigro, P., 2008. Borrower-lender distance, credit scoring, and loan performance: Evidence from informational-opaque small business borrowers. Journal of Financial Intermediation 17, 113-143.

DeYoung, R., Frame, W.S., Glennon, D., and Nigro, P., 2010. The Information Revolution and Small Business Lending: The Missing Evidence. Journal of Financial Services Research 39, 19-33.

Elyasiani, E., Goldberg, L.G., 2004. Relationship lending: a survey of the literature. Journal of Economics and Business 56, 315-330.

Federal Deposit Insurance Corporation, 2010. Insights from the FDIC's credit and consumer products/services survey. Supervisory Insights (Winter).

Frame, W. Scott, Aruna Srinivasan, and Lynn Woosley, 2001. The Effect of Credit Scoring on Small Business Lending. Journal of Money, Credit, and Banking, 33(3), 813-825.

Garcia-Appendini, E., 2011. Lending to small businesses: the value of soft information. *ssrn.com/abstract=1750056*.

Guiso, L., Spaienza, P., and Zingales, L. 2004. The Role of Social Capital in Financial Development. American Economic Review 94, 526-556.

Hall, J.R. and T.J. Yeager, 2002. Community Ties: Does Relationship Lending Protect Small Banks When the Local Economy Stumbles? Federal Reserve Bank of St. Louis, The Regional Economist, April.

Independent Community Bankers of America, 2001. Statement for the Record at hearing on "To review credit conditions in rural America," U.S. House of Representatives, Subcommittee on Operations, Oversight, and Credit, April 14, 2011.

Kittiakarasakun, J., 2010. Does location impact how banks make their lending decisions? A paper presented at the Southwestern Finance Association 2010 Annual Meeting (March 4).

Nakamura, L.I., 1994. Small borrowers and the survival of the small bank: Is mouse bank mighty or mickey? Business Review (November/December), Federal Reserve Bank of Philadelphia, 3-15.

Petersen, M.A., 2004. Information: hard and soft. Working paper. Northwestern University.

Petersen, M.A., and Rajan, R.G., 1994. The benefits of lending relationships: Evidence from small business data. Journal of Finance 49, 3-37.

Petersen, M.A., and Rajan, R.G., 1995. The effect of credit market competition on lending relationships. Quarterly Journal of Economics 110, 406-443.

Petersen, M.A., and Rajan, R.G., 2002. Does distance still matter? The information revolution in small business lending. Journal of Finance 57, 2533-2570.

Scott, J.A., 2004. Small business and the value of community financial institutions. Journal of Financial Services Research 25, 207-230.

Stein, J.C., 2002. Information production and capital allocation: Decentralized versus hierarchical firms. Journal of Finance 57, 1891-1921.

Udell, G.F., 2008. What's in a relationship? The case of commercial lending. Business Horizons 51, 93-103.

Walraven, N.A., 1999. Lending by Rural Banks Involved in Mergers. Journal of Agricultural and Applied Economics 31, 201-214.

Table 1
Random sample of 18,160 small business loans made under the Small Business Administration 7(a) loan guarantee program between 1984 and 2001.

Year loan originated	Number of loans	Percent of sample
1984	574	3.2%
1985	538	3.0%
1986	821	4.5%
1987	865	4.8%
1988	754	4.2%
1989	751	4.1%
1990	811	4.5%
1991	825	4.5%
1992	1,026	5.6%
1993	1,283	7.1%
1994	1,726	9.5%
1995	2,656	14.6%
1996	1,532	8.4%
1997	1,526	8.4%
1998	1,165	6.4%
1999	1,128	6.2%
2000	556	3.1%
2001	112	0.6%
total	18,160	100.0%

Table 2

Loan default rates by borrower-lender pair types. Sample of 18,160 small business loans made in the U.S. between 1984 and 2001. All lenders are banks with assets less than $1 billion measured in real 2000 dollars. All loans were made under the Small Business Administration 7(a) loan guarantee program. ***, ** and * indicate differences from zero at the 1%, 5% and 10% levels, respectively.

Panel A: Four-way borrower-lender taxonomy

variable name	*borrower*	*lender*	*# of loans*	*loan default rate* mean	*loan default rate* std dev	*difference-in-means, relative to the omitted row*			
RR	rural	rural	5,362	0.153	0.390	--	0.000	-0.026	-0.029*
UU	urban	urban	11,690	0.152	0.369	-0.000	--	-0.025	-0.030**
RU	rural	urban	548	0.178	0.382	0.026	0.025	--	-0.005
UR	urban	rural	560	0.182	0.386	0.029*	0.030**	0.005	--

Panel B: Six-way borrower-lender taxonomy

variable name	*borrower*	*lender*	*# of loans*	*loan default rate* mean	*loan default rate* std dev	*difference-in-means, relative to the omitted row*					
RRL	local rural	local rural	4,491	0.150	0.357	--	-0.000	-0.015	-0.029**	-0.027*	-0.032**
UUL	local urban	local urban	10,969	0.151	0.358	0.000	--	-0.015	-0.028**	-0.027*	-0.032**
RRNL	local rural	non-local rural	871	0.165	0.372	0.015	0.015	--	-0.014	-0.012	-0.017
UUNL	local urban	non-local urban	721	0.179	0.384	0.029**	0.028**	0.014	--	0.002	-0.003
RU	local rural	non-local urban	548	0.177	0.382	0.027*	0.027*	0.012	-0.002	--	-0.005
UR	local urban	non-local rural	560	0.182	0.386	0.032**	0.032**	0.017	0.003	0.005	--

Table 3

Descriptive Statistics. Sample of 18,160 small business loans made in the U.S. between 1984 and 2001 and observed quarterly during the life of each loan. All lenders are banks with assets less than $1 billion measured in real 2000 dollars. All loans were made under the Small Business Administration 7(a) loan guarantee program. Unless otherwise indicated as time-varying, all variables are observed at time of loan origination.

			Stacked-logit data (319,112 loan-quarters)		**Raw loan data (18,160 loans)**	
	definition	***time-varying variable?***	***mean***	***std dev***	***mean***	***std dev***
Dependent variable						
Default	= 1 if loan defaulted in current period	yes	0.009	0.093	0.154	0.361
Borrower-Lender						
RR	= 1 if Rural borrower, Rural lender	no	0.295	0.456	0.319	0.457
RRL	= 1 if Rural borrower, Rural lender, in same market	no	0.270	0.444	0.247	0.431
RRNL	= 1 if Rural borrower, Rural lender, in different markets	no	0.049	0.216	0.048	0.214
UU	= 1 if Urban borrower, Urban lender	no	0.644	0.479	0.621	0.485
UUL	= 1 if Urban borrower, Urban lender, in same market	no	0.585	0.493	0.604	0.489
UUNL	= 1 if Urban borrower, Urban lender, in different markets	no	0.036	0.187	0.040	0.195
UR	= 1 if Urban borrower, Rural lender	no	0.030	0.169	0.031	0.173
RU	= 1 if Rural borrower, Urban lender	no	0.031	0.172	0.030	0.171
Loan controls						
Loan amount	Loan amount (2000 dollars).	no	144,917	162,146	155,558	166,676
Distance	Straight-line distance in miles between business and lending office.	no	16.268	35.387	16.651	35.806
LowDoc	= 1 if loan is a "low documentation" loan.	no	0.269	0.444	0.323	0.468
SBA%	Percent of the outstanding loan principle guaranteed by the SBA.	no	0.829	0.072	0.824	0.073

Continued on next page.

Table 3
Continued from previous page.

	definition	time-varying variable?	Stacked-logit data (319,112 loan-quarters) mean	Stacked-logit data (319,112 loan-quarters) std dev	Raw loan data (18,160 loans) mean	Raw loan data (18,160 loans) std dev
Lender controls						
Assets	Lending bank assets (thousands of 2000 dollars).	no	227,596	284,237	250,016	308,386
PLP	= 1 is lender is an SBA ”preferred loan provider.”	no	0.086	0.280	0.096	0.294
CLP	= 1 if lender is an SBA “certified loan provider.”	no	0.176	0.381	0.158	0.365
Borrower controls						
Corporation	= 1 if borrower is organized as a corporation.	no	0.502	0.500	0.523	0.499
Partnership	= 1 if borrower is organized as a partnership.	no	0.086	0.280	0.079	0.270
New Business	= 1 if borrower is a new business start-up.	no	0.299	0.458	0.319	0.466
SIC(I)	= 1 if borrower is in SIC code “I”.	no	0.307	0.461	0.295	0.456
Market controls						
Rural(<10k)	= 1 if borrower is in rural county with population less than 10,000.	no	0.032	0.175	0.029	0.169
Branching	= 1 if borrower is in a State with bank branching restrictions.	no	0.352	0.477	0.388	0.487
NE	= 1 if borrower is in a Northeast state.	no	0.140	0.347	0.140	0.347
MW	= 1 if borrower is in a Midwest state.	no	0.122	0.327	0.119	0.324
CEN	= 1 if borrower is in a Central state.	no	0.211	0.408	0.204	0.403
SW	= 1 if borrower is in a Southwest state.	no	0.155	0.362	0.161	0.367
WST	= 1 if borrower is in a Western state.	no	0.256	0.436	0.252	0.434
Urban	= 1 if borrower is in an urban (MSA) area.	no	0.674	0.468	0.651	0.476
HHI	Deposit-based Herfindahl index for borrower market.	no	0.203	0.128	0.196	0.123
ΔIncome	Percent change in state-specific, industry-specific income relative to level at the time of loan origination.	yes	0.196	0.182	0.267	0.204
ΔEmployment	Percent growth in state-specific, industry-specific employment relative to level at the time of loan origination.	yes	0.005	0.029	0.004	0.025

Table 4

Estimated coefficients for stacked logit loan default model (1) using 319,112 loan-quarter observations. Specification uses the four-way borrower-lender taxonomy. All data and variables are described above in Table 2. ***Coefficients on main test variables are highlighted; the omitted test variable is RR.*** Superscripts ***, ** and * indicate a statistically significant difference from zero at the 1%, 5% and 10% levels, respectively.

	[1]	[2]	[3]	[4]	[5]
Sample:	Full sample	Full sample	Banks with assets over $100 million excluded	Rural markets with population over 50,000 excluded.	Out-of-market loans with distances less than 25 miles excluded
Intercept	-6.0713***	-6.0560***	-4.5420***	-6.5388***	-6.1169***
UU	***0.3566****	***0.3421****	***0.4223****	***0.4627****	***0.3623****
RU	***0.2551***	***0.2339***	***0.4158**	***0.2769***	***0.2648***
UR	***0.5016****	***0.4931****	***0.5125***	***0.6402****	***0.4347****
Rural(<10k)		-0.2371*			
age1	0.6836***	0.6835***	0.6921***	0.8066***	0.7034***
age2	-0.0643***	-0.0643***	-0.0651***	-0.0793***	-0.0668***
age3	0.2641***	0.2640***	0.2604***	0.3459***	0.2770***
age4	-0.0049***	-0.0049***	-0.0047***	-0.0068***	-0.0052***
age5	0.0000***	0.0000***	0.0000***	0.0000***	0.0000***
ln(Distance)	0.0304**	0.0304**	0.0399**	0.0359**	0.0334***
ln(Assets)	-0.0392**	-0.0405**	-0.0802	-0.0479**	-0.0439**
HHI*Urban	-1.5552***	-1.5759***	-1.8761***	-1.9480***	-1.6200***
HHI	-0.1432	-0.1091	-0.3376	0.0402	-0.0668
Branching	-0.0319	-0.0308	-0.1314	0.0207	-0.0410
Corporation	0.0136	0.0132	0.0813	0.0211	0.0126
Partnership	-0.1187	-0.1187	-0.0323	-0.0582	-0.1097
New Business	0.1749***	0.1750***	0.1865***	0.1884***	0.1609***
SIC(I)	-0.3972***	-0.3974***	-0.4324***	-0.3878***	-0.3974***
LowDoc*SBA%	2.2108**	2.2042**	5.6434***	1.9513*	2.3094**
NE	-0.3251***	-0.3225***	-0.7491***	-0.3792***	-0.3359***
MW	-0.3255***	-0.3214***	-0.5039***	-0.3719***	-0.3355***
CEN	-0.413***	-0.4074***	-0.4772***	-0.4402***	-0.4162***
SW	-0.1405**	-0.1360**	-0.1344	-0.1611**	-0.1424**
WST	-0.3379***	-0.3363***	-0.4577***	-0.2977***	-0.3573***
LowDoc	-1.8083**	-1.8029**	-4.6743***	-1.6325*	-1.8861**
SBA%	0.6663	0.6672	-0.4181	0.9175*	0.7093*
PLP	-0.3273***	-0.3269***	-0.2645	-0.3490***	-0.3174***
CLP	-0.2162***	-0.2149***	-0.1237	-0.2243***	-0.2035***
Loan Amount	-0.0000*	0.0000*	0.0000	0.0000	0.0000
ΔIncome	-1.2977***	-1.3011***	-0.9810	-1.6023***	-1.3283***
ΔEmployment	-0.1681	-0.1694	0.7562	-0.7834	-0.2762
year dummies	Yes	yes	yes	yes	yes
observations	319,112	319,112	114,415	249,931	306,373

Table 5

Odds ratios for main test variables in stacked logit loan default model (1) specified using the four-way borrower-lender taxonomy. Each column is based on a different omitted test variable. For example, the odds ratios in the first columns of each panel correspond with the coefficients in Table 4, where RR was omitted test variable. Superscripts ***, ** and * indicate a statistically significant difference from one at the 1%, 5% and 10% levels, respectively.

Panel A: Full sample

	RR omitted	UU omitted	RU omitted	UR omitted
RR	--	0.70***	0.78***	0.55***
UU	1.43***	--	1.11	0.81**
RU	1.29**	0.90	--	0.77
UR	1.65***	1.16	1.29	--

Panel B: Banks with assets over $100 million excluded

	RR omitted	UU omitted	RU omitted	UR omitted
RR	--	0.66***	0.66*	0.53***
UU	1.53***	--	1.01	0.82
RU	1.52*	0.99	--	0.89
UR	1.67**	1.09	1.10	--

Panel C: Rural markets with population over 50,000 excluded.

	RR omitted	UU omitted	RU omitted	UR omitted
RR	--	0.63***	0.76**	0.48***
UU	1.59***	--	1.20	0.77
RU	1.32**	0.83	--	0.69*
UR	1.96***	1.19	1.44*	--

Panel D: Out-of-market loans with distances less than 25 miles excluded

	RR omitted	UU omitted	RU omitted	UR omitted
RR	--	0.70***	0.77**	0.58***
UU	1.44***	--	1.10	0.86
RU	1.30**	0.91	--	0.83
UR	1.55***	1.08	1.19	--

Table 6

Estimated coefficients for stacked logit loan default model (1) using 319,112 loan-quarter observations. Specification uses the six-way borrower-lender taxonomy. All data and variables are described above in Table 2. ***Coefficients on main test variables are highlighted; the omitted test variable is RR.*** Superscripts ***, ** and * indicate a statistically significant difference from zero at the 1%, 5% and 10% levels, respectively.

	[1]	[2]	[3]	[4]	[5]
Sample:	Full sample	Full sample	Banks with assets over $100 million excluded	Rural markets with population over 50,000 excluded.	Out-of-market loans with distances less than 25 miles excluded
Intercept	-6.0702***	-6.0552***	-4.5414***	-6.5534***	-6.0953
RRNL	***0.1092***	***0.1192***	***0.1540***	***0.2134***	***0.2660***
UUL	***0.3599****	***0.3462****	***0.4305****	***0.4808****	***0.3722***
UUNL	***0.5778****	***0.5647****	***0.7317****	***0.6901****	***0.5692***
RU	***0.3087****	***0.2898*****	***0.4766*****	***0.3605****	***0.3363***
UR	***0.5477****	***0.5412****	***0.5723*****	***0.7103****	***0.4920***
Rural(<10k)		***-0.2435****			
age1	0.6835***	0.6834***	0.6923***	0.8066***	0.7032
age2	-0.0643***	-0.0642***	-0.0651***	-0.0793***	-0.0668
age3	0.2640***	0.2639***	0.2604***	0.3460***	0.2769
age4	-0.0049***	-0.0049***	-0.0047***	-0.0068***	-0.0052
age5	0.0000***	0.0000***	0.0000***	0.0000***	0.0000
ln(Distance)	0.0182	0.0177	0.0268	0.0198	0.0187
ln(Assets)	-0.0393**	-0.0406**	-0.0798	-0.0476**	-0.0444
HHI*Urban	-1.5297***	-1.5487***	-1.8632***	-1.8993***	-1.5660
HHI	-0.1727	-0.1405	-0.3754	-0.0175	-0.1371
Branching	-0.0327	-0.0316	-0.1312	0.0210	-0.0427
Corporation	0.0147	0.0143	0.0828	0.0223	0.0144
Partnership	-0.1181	-0.1183	-0.0324	-0.0585	-0.1100
New Business	0.1749***	0.1751***	0.1862***	0.1889***	0.1597
SIC(I)	-0.3967***	-0.3968***	-0.4322***	-0.3870***	-0.3974
LowDoc*SBA%	2.1787**	2.1704**	5.6258***	1.9258*	2.2791
NE	-0.3254***	-0.3227***	-0.7476***	-0.3801***	-0.3348
MW	-0.3223***	-0.3180***	-0.4983***	-0.3685***	-0.3360
CEN	-0.4064***	-0.4005***	-0.4710***	-0.4332***	-0.4134
SW	-0.1321*	-0.1273*	-0.1284	-0.1519**	-0.1393
WST	-0.3318***	-0.3298***	-0.4483***	-0.2913***	-0.3546
LowDoc	-1.7830**	-1.7763**	-4.6579***	-1.6133*	-1.8622
SBA%	0.6696	0.6710	-0.4244	0.9231*	0.7059
PLP	-0.3324***	-0.3319***	-0.2755	-0.3530***	-0.3199
CLP	-0.2192***	-0.2179***	-0.1304	-0.2278***	-0.2066
Loan Amount	0.0000*	0.0000**	0.0000	0.0000	0.0000
ΔIncome	-1.2940***	-1.2974***	-0.9726	-1.5919***	-1.3145***
ΔEmployment	-0.1677	-0.1689	0.7536	-0.7796	-0.2784
year dummies	yes	Yes	yes	yes	yes
observations	319,112	319,112	114,415	249,931	306,373

Table 7

Odds ratios for main test variables in stacked logit loan default model (1) specified using the six-way borrower-lender taxonomy. Each column is based on a different omitted test variable. For example, the odds ratios in the first column of each panel correspond with the coefficients in Table 6, where RRL was omitted test variable. Superscripts ***, ** and * indicate a statistically significant difference from one at the 1%, 5% and 10% levels, respectively.

Panel A: Full sample

	RRL omitted	UUL omitted	RRNL omitted	UUNL omitted	RU omitted	UR omitted
RRL	--	0.78***	0.90	0.56***	0.73***	0.58***
UUL	1.43***	--	1.29**	0.80**	1.05	0.83*
RRNL	1.11	0.70**	--	0.63***	0.82	0.65***
UUNL	1.78***	1.24**	1.60***	--	1.31*	1.03
RU	1.36***	0.95	1.22	0.76*	--	0.79
UR	1.73***	1.21*	1.55***	0.97	1.27	--

Panel B: Banks with assets over $100 million excluded

	RRL omitted	UUL omitted	RRNL omitted	UUNL omitted	RU omitted	UR omitted
RRL	--	0.65***	0.86	0.48***	0.62**	0.56***
UUL	1.54***	--	1.32	0.74	0.96	0.87
RRNL	1.17	0.76	--	0.56**	0.72	0.66*
UUNL	2.08***	1.35	1.78**	--	1.29	1.17
RU	1.61**	1.05	1.38	0.78		0.91
UR	1.77**	1.15	1.52*	0.85	1.10	--

Panel C: Rural markets with population over 50,000 excluded.

	RRL omitted	UUL omitted	RRNL omitted	UUNL omitted	RU omitted	UR omitted
RRL	--	0.61***	0.81	0.50***	0.70***	0.49***
UUL	1.62***	--	1.31	0.81**	1.13	0.80
RRNL	1.24	0.77	--	0.62**	0.86	0.61**
UUNL	1.99***	1.23**	1.61**	--	1.39**	0.98
RU	1.43***	0.89	1.16	0.72**	--	0.70
UR	2.04***	1.26	1.64**	1.02	1.42	--

Panel D: Out-of-market loans with distances less than 25 miles excluded

	RRL omitted	UUL omitted	RRNL omitted	UUNL omitted	RU omitted	UR omitted
RRL	--	0.69***	0.77**	0.57***	0.71***	0.61***
UUL	1.45***	--	1.11	0.82*	1.04	0.89
RRNL	1.31**	0.90	--	0.74*	0.93	0.80
UUNL	1.77***	1.22*	1.35*	--	1.26	1.08
RU	1.40***	0.97	1.07	0.79	--	0.86
UR	1.64***	1.13	1.25	0.93	1.17	--