Distant Lending, Specialization, and Access to Credit

Wenhua Di and Nathaniel Pattison*†

September 2019

Abstract

Small business lending has historically been very local, with most lenders serving nearby borrowers. Technological advances, however, have enabled lenders to reach a wider geographic market, thereby allowing them to specialize in other (non-geographic) firm characteristics, namely certain industries. This paper investigates the relationship between distant lending and industry specialization. We first document a recent increase in remote small business lending (loans to borrowers more than 100 miles away). Remote lenders tend to concentrate their loans within fewer industries and, consistent with industry expertise, have fewer defaults. We then examine the impact of specialized lending on credit access by exploiting the staggered entry of a large, remote, specialized lender. The results indicate that remote lending can complement local lending. We find that entry by the remote lender significantly increases total lending, with no evidence of substitution away from other lenders.

JEL G21, G23, L11
Keywords: Small business lending, Banking competition, Specialization, Distance, Credit access, Technology, Fintech

*Di: Federal Reserve Bank of Dallas, 2200 N. Pearl St. Dallas, TX 75201, wenhua.di@dal.frb.org. Pattison: Southern Methodist University, ULEE 301E, Dallas, TX 75275, npattison@smu.edu.

†Disclaimer: The views in this paper are those of the authors and do not necessarily represent those of the Federal Reserve Banks or the Federal Reserve System. An earlier version of this paper was circulated as “Remote Competition and Small Business Lending: Evidence from SBA Lending.” Acknowledgments: We thank SungJe Byun, Robert DeYoung, Janet Garufis, Daniel Millimet, Yichen Su, Gregory Udell, Mike Weiss, brown bag participants at the Federal Reserve Bank of Dallas and Southern Methodist University, and conference participants at the Federal Reserve System Applied Micro Conference, the Community Banking in the 21st Century research and policy conference at the Federal Reserve Bank of St. Louis, the Banking and Finance Workshop at the Federal Reserve Bank of Dallas, the Society for Government Economists at the 2018 ASSA Annual Meeting, and the Stata Texas Empirical Microeconomics Conference. We thank Benjamin Meier for his research assistance.

1 Introduction

Historically, distance has played an important role in lending. This is especially true for small business lending, where little public information is available about firms' risks and the information that does exist is "soft" or difficult to acquire and communicate at a distance. Physical proximity aids in the collection and transfer of such information, leading to better risk assessment and fewer defaults (Petersen and Rajan, 2002, DeYoung, Glennon and Nigro, 2008, Agarwal and Hauswald, 2010).[1] As a result, small business lending tends to be very local. The median distance between small businesses and their lenders is less than 10 miles, and the availability of credit depends on the presence of nearby bank branches (Nguyen, 2019, Granja, Leuz and Rajan, 2018).

Borrower-lender distances, however, have steadily increased over the past 30 years. The literature attributes this increase to technological advances that enable lenders to better collect, transmit, and process quantifiable or "hard" information.[2] Small business credit reports, credit scoring, the expansion of infomediaries, and improvements in information technologies have substantially increased the availability and use of hard information. The prevalence of hard information, in turn, decreases the reliance on locally collected "soft" information and allows for more distant lending. An extreme example of distant lending is online "Fintech" lenders, which use alternative data (e.g. social networks, rental history, transactions, and education data) and improved processing (e.g. machine learning, automation, and integration) to lend to very distant borrowers, often without any face-to-face interaction.

Our paper investigates a new lending "technology" that accompanies distant lending: specialization. Local lenders restrict their market to those that are nearby. If local lenders also specialized along non-geographic characteristics, there may be too few potential borrowers. However, as lenders expand their reach geographically, the larger potential market allows them to specialize along other characteristics (e.g. certain products, borrower types, or industries) and, perhaps, develop expertise that offset the disadvantages of lending at a distance. In small business lending, the borrowing firm's industry is a particularly important dimension. Small businesses operate in very different industries, and this heterogeneity makes underwriting less uniform than consumer loans and mortgages. Industry-specialized lenders may be able to focus on industries better suited for distant lending and also to take advantage of industry-specific expertise or economies of scale. Moreover, if specialized lenders use alternative information to screen borrowers, they may be able to focus on new segments of borrowers and expand access to credit.

In this paper, we examine the relationship between distant lending and specialization within the context of small business lending. We first document the presence of remote, industry-specialized lenders. We show a significant increase in remote small business lending (loans with a borrower-lender distance exceeding 100 miles), and then show that these remote lenders tend to concentrate their

[1] A related literature emphasizes the role of hierarchical distance and communication costs between loan officers and their superiors within an institution's organizational structure (Liberti and Mian, 2008, Qian, Strahan and Yang, 2015).

[2] Liberti and Petersen (2018) provide a recent review.

loans within fewer industries. Moreover, consistent with expertise, concentrated lenders experience better loan performance in the industries where they focus. Having established the presence of remote, industry-specialized lenders, we then ask whether remote, industry-specialized lending serves as a substitute or complement to traditional, geographically specialized lending? That is, do industry-specialized lenders compete for the same borrowers or are they able to expand credit access to a new segment of firms?

To examine the relationship between remote lending and industry specialization, we use loan-level data for the universe of Small Business Administration (SBA) 7(a) loans from 2001-2017. SBA 7(a) loans are common, relatively low-cost loans partially guaranteed by the SBA and given to credit-constrained small businesses.[3] The SBA 7(a) data are uniquely well-suited for our analysis, as they contain loan-level information on each borrower's location (address), industry (6-digit NAICS code), as well as the identity of the lender. We merge bank branch locations from the Federal Deposit Insurance Corporation's (FDIC) Summary of Deposits (SoD) data to compute the between borrower-lender distance for each loan.

We begin by documenting two new empirical facts about small business lending. First, in the past decade, the share of remote small business loans (those with a borrower-lender distance exceeding 100 miles) has increased. That is, the distribution of (log) borrower-lender distance has become increasingly bimodal.[4] We confirm that these changes in borrower-lender distances are not unique to the SBA program. We find a similar pattern using data from the Community Reinvestment Act (CRA), which contain information on most small business loans from larger banks.[5] Second, we find that lenders with a larger share of remote loans tend to be significantly more concentrated in the industries to which they lend. This correlation between borrower-lender distance and industry concentration holds across various measures of distance and concentration, as well as across lenders and within lenders over time.

We then investigate whether industry specialization is associated with industry-specific expertise in lending, perhaps offsetting some of the disadvantages of distant lending. To test this idea, we examine the relationship between industry specialization and loan performance. Specifically, we estimate how within-industry charge-off rates vary across lenders with different exposures to the industry. We first show, as in the prior literature, that the probability of default increases with borrower-lender distance. Consistent with industry-specific expertise, however, we find a correlation between greater industry exposure by a lender and lower charge-off rates relative to other lenders' loans to that same industry. We also find that, across lenders, industry concentration weakens the positive relationship between lending distance and charge-off rates, suggesting that greater industry

[3] In the 2017 Small Business Credit Survey Federal Reserve Banks (2017), 26% of employer small businesses seeking a loan or line of credit applied for an SBA loan and, among (nonapplicant) employer small businesses already holding a loan, 17% held an SBA loan. We discuss the size and importance of SBA 7(a) lending in Section 2.2.

[4] Similar to this change in the distribution, DeYoung et al. (2011) found that much of the increase in borrower-lender distances between 1993 and 2001 can be attributed to large increases in distances by banks that adopted credit scoring technology.

[5] The CRA requires business loans smaller than \$1 million to be reported by large banks or thrifts. The data do not contain information on industry, so we are unable to examine industry concentration using CRA data.

specialization helps offset the disadvantages of distance.

Having documented an increase in remote, industry-specialized lending, we then investigate how this new lending technology affects access to credit. The challenge in identifying the impact on credit access is that we do not observe the counterfactual number of loans that would have been originated absent remote lending. We address this challenge by examining the impact of entry by the largest remote SBA lender, Live Oak Bank, a branchless bank based in North Carolina. Live Oak is among the largest SBA lenders, originating more than 6% of all SBA 7(a) loans (dollar-weighted) and a significantly larger share in the industries in which it operates. Moreover, Live Oak exhibits the two key features of remote, industry-specialized lenders: (i) Live Oak gave 95% of its SBA loans to borrowers 100 or more miles from its single office in North Carolina, and (ii) more than 80% of its loans were to just six of the more than 800 industries receiving SBA loans. It describes industry-specific expertise as its primary advantage.[6] The combination of Live Oak's size and strategy for entering markets provides a unique setting for identifying the impact of remote lending. Specifically, we exploit Live Oak's staggered entry into specific industries. Upon entry, Live Oak generates a sharp increase in remote lending, providing 12-58% of post-entry SBA loans to these industries.

Our identification strategy compares changes in total lending in these "treated" industries (i.e. the industries that Live Oak enters) to changes in lending to a group of control industries that Live Oak did not enter. Instead of choosing comparison industries subjectively, we employ the Synthetic Control Method (SCM), developed in Abadie and Gardeazabal (2003) and Abadie, Diamond and Hainmueller (2010), to systematically construct a synthetic industry match for each treated industry and compare post-entry changes between the treated industry and this synthetic control. The key identification assumption is that the timing of entry by Live Oak into a specific industry does not coincide with other changes affecting lending to the treated industries. We support this identification assumption with information about the determinants of Live Oak's entry decisions. Additionally, to investigate whether these industries experience unusual growth independent of Live Oak, we test whether lending increased relative to the number of establishments and whether lending to these industries also increases in locations where Live Oak originated no loans.

Our results indicate that entry by this remote, industry-specialized lender significantly increased total lending to these industries. There are sharp increases in total SBA loans to the "treated" industries after Live Oak's entry, relative to the synthetic control. Moreover, we find no evidence of substitution away from other SBA lenders. Other institutions' SBA lending to these industries remains unchanged upon Live Oak's entry, suggesting that remote, industry-specialized lending complements local lending. One potential concern is that the increased SBA lending to these industries may reflect substitution away from non-SBA alternatives, which we do not observe in our main sample. This concern is partially mitigated by the SBA 7(a) program's "credit elsewhere" test, which requires lenders to certify that SBA borrowers would be unable to obtain a loan with

[6] "We are one of the nation's top originators of small business loans primarily because our expertise in specific industries enables us to lend to business owners who haven't had access to capital in the past" (Live Oak Bank, n.d.).

reasonable terms absent the SBA guarantee. Still, we further investigate non-SBA substitution using a proxy for total (SBA and non-SBA) lending within each industry: counts of financial statements collected by lenders as a part of the loan application and monitoring process. Again, we find no evidence of substitution away from other lenders.

This case study provides evidence that remote, specialized lending has the potential to complement local lenders and expand access to small business credit. Live Oak's size and staggered entry make it uniquely suited for our identification strategy, and it is the most prominent example of remote, industry-specialized lending. However, given that our analysis is of entry by a single lender, it raises questions about whether the effects of its entry would be representative of the effects of remote lending more broadly. We conclude with a discussion of the external validity of our results.

This research adds to several strands of the literature. The first studies banks' industry or sectoral specialization by banks. Winton (1999) and Stomper (2006) provide models of sectoral expertise and lending, demonstrating that sectoral specialization can be optimal for a bank (relative to diversification) if it facilitates industry expertise and improves monitoring. The related empirical literature generally finds that sectoral concentration by banks increases returns and reduces risk.[7] These papers use bank-level data on charge-offs and returns and measure sectoral specialization across fewer than 30 broad industry categories. An advantage of our data is that it contains *loan-level* information on the detailed industry code (NAICS code for more than 800 industries) and whether the loan was charged off. This information allows us to examine whether industry concentration lowers charge-off rates in the specific industries in which a lender specializes, rather than examining average bank-level charge-offs across all industries, as in the existing literature.[8]

Second, our paper connects this research on sectoral specialization to the literature on the role of physical distance in lending. Since it is difficult to assess the creditworthiness of small businesses, lenders have relied on personal relationships with borrowers and between loan officers to aid the transfer of "soft" information (Berger and Udell, 1995, Petersen and Rajan, 1994). A large theory literature examines the role of physical distance and information in banking competition (Sharpe, 1990, Rajan, 1992, Dell'Ariccia and Marquez, 2004, Von Thadden, 2004, Hauswald and

[7]Acharya, Hasan and Saunders (2006), using Italian bank-level data and exposure to 21 industry categories, finds that sectoral concentration increases returns and reduces risk, but only for high-risk banks (those with many doubtful or non-performing loans). Hayden, Porath and Westernhagen (2007), using German bank-level data and exposure across 23 sectors, also finds that concentration generally improves returns and loan performance. Similarly, Boeve, Duellmann and Pfingsten (2010) and Jahn, Memmel and Pfingsten (2013), using German bank-level data, find that sectoral specialization leads to better monitoring and fewer write-offs. Tabak, Fazio and Cajueiro (2011), using Brazilian bank-level data and exposure to 21 economic sectors, finds that sectoral concentration increases banks' returns and lowers default risk. Dincbas, Michalski and Ors (2017) use interstate banking deregulation to identify the impact of entry by banks more familiar with certain industries based on the industry compositions of the bank's original location. With state-level data on employment and output across 19 sectors, they find that, when a U.S. state that is highly exposed to an industry allows bank mergers with a state that is less exposed to that industry, there is subsequent growth of the industry in the less-exposed state. In contrast, using an international sample of large banks and inferring banks' concentration across 10 sectors, Beck, De Jonghe et al. (2013) find that concentration increases risk without raising returns.

[8]While not examining performance, Berger, Minnis and Sutherland (2017) are also able to examine within bank differences and find that concentrated banks are less likely to collected audited financial statements from firms in the industries and regions where they are concentrated.

Marquez, 2006, Frankel and Jin, 2015). Several empirical papers provide evidence that physical proximity facilitates information collection, lowers transaction and monitoring costs, and improves loan performance (Petersen and Rajan, 2002, Degryse and Ongena, 2005, DeYoung, Glennon and Nigro, 2008, Agarwal and Hauswald, 2010, DeYoung et al., 2011, Granja, Leuz and Rajan, 2018). Additionally, they emphasize that improvements in information technology and increasingly relying on quantifiable "hard" information allow lenders to expand geographically (Petersen and Rajan, 2002, Berger et al., 2005, DeYoung, Glennon and Nigro, 2008, Loutskina and Strahan, 2011). Our paper provides evidence of a complementary lending "technology" associated with distant lending: industry specialization.

Third, our paper relates to the literature examining entry and competition in lending and the effects on credit availability, particularly as it relates to local and distant firms. Much of this literature examines this question in the context of financial liberalization within developing countries. Detragiache, Tressel and Gupta (2008) and Gormley (2014) provide theoretical models examining lending competition by local and distant firms, and in particular, when foreign lenders compete with domestic banks. Similar to remote lenders in our setting, foreign lenders in these models have a lower ability to screen on local "soft" information but an offsetting comparative advantage in an improved ability or lower cost of screening along another dimension. Entry by these new lenders can expand access to new borrowers, provide a substitute for existing lenders, or induce a segmented credit market in which total lending falls. Empirically, in the context of countries' financial liberalization, papers have found mixed effects, with entry by foreign lenders sometimes reducing access to credit (Beck and Peria, 2010, Detragiache, Tressel and Gupta, 2008, Gormley, 2010) and sometimes increasing access to credit (Giannetti and Ongena, 2009, 2012, Bruno and Hauswald, 2013, Claessens and Van Horen, 2014).

Finally, a growing literature investigates the unique features of online FinTech lenders and their effects on access to credit. In mortgage lending, Buchak et al. (2018) examine how regulatory and technological advantages contribute to the rapid growth of mortgage originations by shadow banks and FinTech lenders. Fuster et al. (2019) find that FinTech mortgage lenders process applications faster than traditional lenders with no increase in default rates. In peer-to-peer (P2P) consumer lending, a growing literature examines the predictive content of the alternative information available to P2P lenders, such as pictures, borrower narratives, and social networks (see Liberti and Petersen (2018) for an overview). Another set of papers examines whether P2P lenders are substitutes or complements for traditional banks. Tang (2019), De Roure, Pelizzon and Thakor (2019), and Wolfe and Yoo (2018) all find evidence that P2P lenders and banks are substitutes, competing for an overlapping set of borrowers. Jagtiani and Lemieux (2017) find some evidence that P2P loans are more common in areas underserved by traditional banks.[9] Our paper adds to this literature by examining the unique characteristics of remote small business lenders and whether they substitute or complement local lenders.

[9]Outside of lending, Goodman, Melkers and Pallais (2019) find that an online college program complements traditional education and could satisfy unmet demand for computer science training.

The paper proceeds as follows. Section 2 provides background information on local and remote lenders, discusses the SBA 7(a) program, and describes the data. Section 3 examines the relationship between distance, loan performance, and industry concentration among SBA lenders. Section 4 examines a case study of entry by Live Oak, the largest remote, specialized lender, in order to assess the impact of industry specialization on credit availability. Section 5 concludes with a discussion of our results, external validity, and broader implications.

2 Background, Setting, and Data

This section first provides more information on the local nature of small business lending and discusses a framework for considering competition between local and remote lenders. We then discuss our setting of SBA 7(a) lending and the main data used in our analysis.

2.1 Background

Most small business lenders are geographic specialists. Both small and large banks typically define their markets as the area around their physical branches and the median borrower distance from the lender's branch is less than 10 miles (Agarwal and Hauswald, 2010, DeYoung, Glennon and Nigro, 2008, Granja, Leuz and Rajan, 2018).[10,11] Economic theory provides multiple reasons for this geographic proximity. Local lenders can use repeated interactions and relationships to collect and transfer "soft" information about firms, giving them an advantage over distant lenders (Berger and Udell, 1995, Petersen and Rajan, 1994, 2002). Even without information frictions, distance-related transaction costs associated with originating and monitoring a loan can lead to local lending (Degryse and Ongena, 2005). Additionally, local lenders may also be better informed about local economic conditions and their effect on a firm's profitability.

Alternatively, lenders may specialize along a non-geographic firm characteristic, namely industry. For most industries and geographic markets, the pool of potential borrowers would be too small for a lender to focus on both specific industries and within a local area (e.g. veterinarians within 20 miles), so we view geographic and industry specialization as two distinct alternatives. Industry specialization may offer two advantages. First, industry-specialized lenders to select industries with lower risks or less competitive markets. Second, industry specialization may facilitate expertise that

[10] The 2018 FDIC Small Business Lending Survey (Federal Deposit Insurance Corporation, 2018) surveyed approximately 1,200 banks, both small (assets less than $10 billion) and large (assets greater than $10 billion), about their geographic market relative to their physical branch locations. Among small banks, 73.8% have a geographic market within the city or county of their branches, and an additional 16.9% have a geographic market within their metropolitan statistical area (MSA) or state. Large banks have wider geographic areas, with 20.5% viewing their geographic market as the county of their branches, 18.3% at the MSA-level, and 42.8% at the state level. Only 18.4% of large banks report the geographic market as national (or other).

[11] Agarwal and Hauswald (2010), using application-level data from a leading small business lender, finds a median distance from the firm to the bank branch of 2.62 miles for originated loans. Granja, Leuz and Rajan (2018) uses data from the Community Reinvestment Act to calculate the median distance between borrower's county and the county of the closest lender's branch. The median distance in 2016 was 6 miles. Using SBA 7(a) data to DeYoung, Glennon and Nigro (2008) calculates the median distance between a borrower and the lending office (rather than the closest branch) of the lender. The median distance to the lending office increased from 5.89 miles in 1984 to 21.28 miles in 2001.

help offset the informational disadvantages of lending at a distance. More experience in the industry may improve a lender's ability to screen borrowers (e.g. through industry-specific underwriting). For example, United Community Bank, an SBA lender with substantial online lending, reports that it mitigates the risk of "working with more borrowers it doesn't know well" by "originating SBA loans only within specific industries it has decided to cultivate after studying them carefully" (Schneider, 2016). Additionally, there may be industry-specific investments or economies of scale. For example, a lender could hire industry experts to screen applicants, lower borrower-acquisition costs through industry-specific advertising, or even provide consultancy for business development.

We view local lenders as having an advantage in locally communicated or "soft" information, whereas remote, industry-specialized lenders have an advantage in industry-specific information. Our dichotomy into local and remote lenders differs from the groupings of banks into large and small (or geographically concentrated and unconcentrated) that is common in the literature (e.g. Loutskina and Strahan (2011)). In particular, we split lenders based on distances between their branches and borrowers. As a result, in addition to community banks, we also define large banks that primarily lend to borrowers located around their physical branches as "local," i.e. geographically specialized. While these large lenders likely rely less on soft or local information than community banks, the majority view their market as the areas around their physical branches (Federal Deposit Insurance Corporation, 2018) and make use of some soft information in lending. For example, using data from the third-largest small business lender in the U.S., Agarwal and Hauswald (2010) found that the median borrower-lender distance was a few miles and that physical proximity facilitated the collection of soft information, leading to lower charge-off rates and more accurate risk assessment. Additionally, larger traditional lenders may have an advantage in general hard information, i.e. information that is not industry-specific. These features of local specialists contrast with remote lenders, which give the majority of their loans to borrowers more than 100 miles from the closest branch.

The implications of competition between industry and local specialists for the availability of credit is uncertain. Industry-specialists may use their comparative advantage to identify and extend credit to profitable but underfinanced borrowers, which would increase the total supply of credit. For example, some borrowers may be risky based on the information available to local lenders, but profitable based on the information available to industry-specific lenders. Alternatively, industry-specific and local lenders may compete for the same borrowers, resulting in little change in overall lending. Moreover, it is possible that cream-skimming by industry-specialists induces a segmented credit market, as in the financial liberalization models of Detragiache, Tressel and Gupta (2008) and Gormley (2014), ultimately reducing the availability of credit.[12]

[12] In the model of Detragiache, Tressel and Gupta (2008), foreign (in our case, remote) banks are better than domestic (local) banks at monitoring hard (industry-specific) information, but not soft information. Under some parametrizations, foreign bank entry causes "cream-skimming" of borrowers with good hard information. As a result, domestic banks are left with a worse pool of potential borrowers, causing them to raise interest rates and invest more in monitoring. Ultimately, the higher interest rates may reduce the equilibrium quantity of credit and, depending on the parameters, can result in an overall decline in welfare.

2.2 Setting

Our setting for examining distance, industry specialization, and competition is the market for Small Business Administration's 7(a) loans.[13] The primary advantage of this setting, as we discuss in detail in Section 2.3, is that we observe detailed information about each borrowers' industry and location for the universe of 7(a) loans from 2001-2017. The Small Business Administration is a federal agency tasked with helping to start, build, and grow small businesses. Through the 7(a) program, the SBA seeks to increase the supply of credit by providing guarantees for loans to credit-constrained small businesses. To qualify for an SBA 7(a) loan, the borrower must run a for-profit business that meets SBA industry-specific size standards for what constitutes a small business.[14] Additionally, to qualify, the borrower must be unable to obtain a loan elsewhere on "reasonable terms."[15] Lenders must document why the borrower could not obtain a loan on reasonable terms without the SBA guarantee and must review the personal resources of any applicants owning more than 20 percent of the small business. The loans can be used for working capital, expansions, to purchase a business or franchise, to buy commercial real estate, or to refinance debt.

The capital for loans in the SBA 7(a) program is provided by private lenders, which are mostly commercial banks, though there are also credit unions and other non-bank lenders. Lenders make most decisions regarding the SBA loans (subject to underwriting rules of the SBA such as a maximum interest rate and borrower requirements). Depending on a lender's experience, the SBA either re-analyzes the lender's underwriting decisions or delegates them to the lender. The Preferred Lender Program (PLP) status, given to the most experienced SBA lenders, allows a lender to make all underwriting and eligibility decisions. These PLP lenders make more than 80% of SBA 7(a) loans. The SBA provides a partial guarantee for the loan that, in the event of default, reimburses the lender for a share of the amount charged off. The maximum guarantee is 85% for loans up to \$150,000 and 75% for loans exceeding \$150,000 (with a maximum guarantee of \$3.75 million for a standard 7(a) loan).[16] In exchange, the SBA charges lenders a guarantee fee which depends on the features of the loan and the amount guaranteed.

SBA 7(a) lending is an important source of financing for small businesses, particularly for larger loans and small businesses with employees. In 2017, SBA 7(a) originated more than 60,000 loans totaling \$25.45 billion. Determining the share of total small business lending accounted for by these SBA 7(a) loans is difficult because the definition of a small business loan and the set of institutions

[13]The SBA also has a 504 loan program. We focus on 7(a) loans because they are the SBA's flagship program and also because the specific bank we focus on in the case study provides almost no loans in the SBA 504 loan program.

[14]`https://www.sba.gov/sites/default/files/2019-08/SBA`

[15]Temkin (2008) surveyed 23 banks that originate SBA loans about their application of the "credit elsewhere" requirement, and the surveys suggest that "the lenders are aware of the credit elsewhere requirement and adhere to the requirement." Lender representatives report that most SBA applicants are referred to the program if (i) the business shows insufficient net operating income to obtain a conventional loan, (ii) the collateral is limited, or (iii) the borrower does not have sufficient equity for the down payment.

[16]There have also been a few policy changes in SBA lending during the period we study. In particular, after the Great Recession dramatically reduced the supply of small business loans, Congress passed the Recovery Act in 2009 and raised the SBA loan guarantee to 90 percent and removed the guarantee fee, which revived the SBA loan program. Since these changes affect all industries similarly, they will be captured by the time controls in our empirical strategy.

that report varies across available data sources. One widely used measure of small business lending is the amount reported pursuant to the Community Reinvestment Act (CRA), which requires all commercial and savings banks with total assets over $1 billion to report small business loan originations.[17] These larger institutions represent 70% of all outstanding small business loans made by banks (Haynes and Williams, 2018). In 2017, total CRA reported small business lending was $242 billion, so SBA 7(a) lending amounted to 10% of total CRA lending.[18]

SBA 7(a) lending is more prominent among larger loans.[19] SBA 7(a) loans amount to less than 1% of the loan for less than $100,000. However, for the larger loan size categories ($100,000-$1 million), SBA 7(a) loans amount to 5-7% of the number of loans and 4-6% of the dollar volume. Moreover, these statistics do not include SBA 7(a) loans that were for more than $1 million, and these larger loans have accounted for more than 50% of SBA 7(a) lending (in $) each year since 2011.

Additionally, SBA 7(a) lending is a more significant source of financing for small businesses with employees. Of the 30 million small businesses in the U.S., only 20% have one or more employees (Mills and McCarthy, 2016). SBA 7(a) loans are often made to these employer businesses, reflecting a primary goal of SBA lending, job creation. In the 1990-2009 matched sample of SBA 7(a) borrowers in Brown and Earle (2017), the median number of employees among SBA 7(a) borrowers is seven and the mean is 14. In the 2017 Small Business Credit Survey Federal Reserve Banks (2017), a survey of over 8,000 small businesses with 1-499 employees, 26% of employer small businesses seeking a loan or line of credit applied for an SBA loan. Of those that already held loans and did not apply in the last year, 17% held an SBA loan or line of credit. Thus, our analysis of SBA lending accounts for a non-trivial share of small business financing, particularly for larger loans and employer small businesses.

2.3 Data

Our main analysis uses data from the SBA 7(a) Loan Data Report.[20] As mentioned, the SBA data are uniquely well-suited for our analysis, as they contain detailed information on the two key variables: industry and location. The data contain information on the loans (approval date, amount, term, repayment status), borrowers (address, NAICS industry code), and lenders (name, headquarter location). Interest rate information is available beginning in 2008. Our first analysis in Section 3 focuses on the relationship between the industry concentration and borrower-lender distance. Our sample for this analysis consists of all SBA 7(a) loans from 2007-2017, the period

[17] In the CRA, small business loans are defined as those with original amounts of $1 million or less and were reported on the institution's Call Report or Thrift Financial Report as either "Loans secured by nonfarm or nonresidential real estate" or "Commercial and industrial loans."

[18] These numbers are not directly comparable, as CRA data do not include loans for more than $1 million while SBA 7(a) statistics do.

[19] Appendix Figure A.1 shows the ratio of SBA 7(a) to CRA lending across the three loan size categories available in the CRA: loans less than $100,000, loans between $100,000 and $250,000, and loans between $250,000 and $1 million. Between 2004 and 2005, the asset threshold for CRA reporting increased from $250 million to $1 billion, which changed the set of institutions reporting. After 2005, the threshold continued to be adjusted for inflation.

[20] We drop loans that were approved but canceled before origination.

when remote lending became increasingly common. Later in Section 4, we expand the period to 2001-2017 to examine the effects of entry by remote lenders. This sample begins in 2001 because the borrower's industry code is frequently missing in earlier years.

For each loan, we calculate the distance between the SBA borrower and the closest branch of the institution making the loan. The SBA data contain the location (address) of the borrower and the name and headquarter address of the institution currently assigned the loan in 2017. For branch locations, we use the Federal Deposit Insurance Corporation's (FDIC) Summary of Deposits (SoD), which contains branch location data for all FDIC-insured institutions from 2001-2017. We then link the SBA 7(a) lending institutions to these branch networks using fuzzy matching, since lender names in the two datasets often do not exactly match.[21] We are able to match 92% of the borrowers to institutions, and the majority of the unmatched institutions are credit unions or non-bank lenders, which are not in the FDIC branch data. Next, to calculate the distance from a borrower to the branch, we need to convert borrower address to latitude and longitude coordinates. For the borrower-institution matches, we use the Census Geocoder to geocode the borrower's listed address and are able to generate a latitude and longitude coordinates for 72%, with slight increases in the match rate in more recent years. Our results, however, are also robust to using a lending distance measure based on the county centroid of the borrower's project (firm), which is available for all borrowers with a matched lending institution (Table A.2 Column 5). Finally, we calculate the (Haversine) distance between each matched borrower and the closest branch of the institution originating the loan. Appendix B provides more details on the matching procedure and how distances are calculated.

Panel A of Table 1 reports the summary statistics of the sample of 2007-2017 SBA 7(a) loans used in Section 3. SBA 7(a) loans had a median size of $100,000 (mean $310,000), median term of 84 months (mean 111 months), and median interest rate of 6% (mean of 6.3%). The median borrower-lender distance was 2.1 miles, although the mean distance was 91.5 miles, indicating that distances are skewed to the right. The next row shows that these distance measures are similar if we calculate distances using the centroid of the borrower's project's county, which is available for all borrowers matched to an institution. Finally, the average three-year charge-off rate (i.e. the share charged off within three years of origination) is 6%.

To investigate the representativeness of the SBA data, we supplement it with data from the Community Reinvestment Act (CRA). The CRA data reflect the broader small business lending market, reporting the volume and borrower location (county) of small business lending for all commercial and savings banks with total assets above $1 billion.[22] However, unlike the SBA data,

[21] To do this, we standardize lender names and addresses (following the procedure in Wasi, Flaaen et al. (2015)), and probabilistically match SBA lenders to 2017 bank headquarter locations in the FDIC SoD using bank name, address, city, state, and zip code. The institution assigned to the loan in 2017 may differ from the institution that originated the loan due to mergers and acquisitions or loan transfers. To account for changes in loan assignment from mergers and acquisitions, we combine the SoD with information on mergers and acquisitions from the Federal Deposit Insurance Corporation's Report of Structure Changes to determine, for each branch in each year, what institution currently holds the branch. We also verify that, apart from these structure changes, transfers of loan assignments are rare (see Appendix B and Appendix Figure B.1).

[22] Prior to 2005, the threshold for reporting was assets above $250 million. In 2005, it was increased to $1 billion

the CRA data do not contain information on the industries of small business borrowers, so our main analysis of industry concentration relies on data from the SBA. We replicate our distance measure in the CRA data by calculating the distance between the center of the borrower's county and the closest branch of the bank originating the loan. Since SBA 7(a) loans are most comparable to the larger CRA reported loan categories, as shown in Figure A.1, we calculate distance statistics using CRA loans above $100,000.

3 Lending Distance and Industry-Specialization

This section examines the relationship between remote lending and industry specialization. We first provide evidence of the growth in remote small business lending (defined as loans with a borrower-lender distance more than 100 miles). We then document two empirical facts about remote lending and industry specialization. First, there is a robust negative relationship between a lender's geographic diversification and its industry diversification, i.e. lenders with more remote loans tend to concentrate lending within fewer industries. Second, greater industry concentration is correlated with better loan performance in that industry, consistent with industry expertise.

The purpose of this section is to document the presence of remote, industry-specialized lenders. We show the correlation between distant lending and industry specialization, but do not aim to estimate a causal effect of distant lending or industry specialization. In practice, institutions jointly determine their degree of remote lending, industry specialization, and investments in developing industry-specific expertise, and causality among these choices is intertwined. For example, industry-concentrated lending may lead to industry-specific expertise, but pre-existing industry expertise may also cause a lender to concentrate in certain industries.

3.1 Changes in Borrower-Lender Distance

We begin our analysis by showing changes in distances between borrowers and lenders over the last twenty years. Figure 1 plots the average distance between the borrower and lender from 2001 to 2017 for both SBA 7(a) loans and loans reported in the CRA data. Both sources show that the average lending distance increased from less than 50 miles in 2001 to more than 150 miles in 2017. The steady increase in average (and median) lending distance over the last three decades has been documented in several papers (Petersen and Rajan, 2002, DeYoung et al., 2011, Granja, Leuz and Rajan, 2018).[23]

Rather than the increasing mean or median lending distance, our focus is on the rise of remote or very distant lending. Figure 2 plots the distribution of (log) borrower-lender distances for SBA 7(a) loans (panel a) and CRA loans above $100,000 (panel b) for three years: 2001, 2009, 2017. The figure reveals two striking features. First, much of the change in borrower-lender distances is from

and has been inflation-adjusted since that time.

[23]Granja, Leuz and Rajan (2018) focuses on the cyclicality of lending distance; loan distances increase during boom periods and decline during busts. The sample for our Figure 1 excludes CRA loans for less than $100,000. If we include these loans, our figure matches the cyclical fluctuations reported in Granja, Leuz and Rajan (2018).

an increased number of remote loans, i.e. those with more than 100 miles between the borrower and lender.[24] Between 2001 and 2017, the share of SBA loans with a distance of more than 100 miles, which we refer to as "remote" loans, increased from 4.6% to 16.4%. Second, most lending is still largely local. Even in 2017, 71% of loans had a borrower-lender distance of less than 10 miles. In summary, while local lending still dominates, the increasingly bimodal nature of the distance distribution shows a significant growth in remote lending.

3.2 Borrower-Lender Distance and Industry Concentration

Our primary measure of remote lending is the share of an institution's loans with a borrower-lender distance greater than 100 miles. We also construct two measures of industry concentration, with industry measured by the 5-digit NAICS code. Businesses from more than 800 distinct 5-digit NAICS codes received SBA loans in our sample. The first concentration measure is a Herfindahl-Hirschman Index (HHI). For lender b, the industry HHI is defined as $HHI_b = \sum_i S_{ib}^2$, where S_{ib} is the share of lender b's loans given to industry i. This is increasing in industry concentration and takes a value from close to 0 (least concentrated) to 1 (all loans to a single industry). As a second concentration measure, we calculate the top-five share, defined as the share of the institution's loans extended to its five most common industries.[25]

Figure 3 plots the relationship between institutions' shares of distant loans and their industry concentration. Lenders are grouped into 20 bins based on their share of loans from 2007-2017 with a borrower-lender distance over 100 miles (e.g. 0-5%, 5-10%, etc.). The figure plots the average industry HHI (Panel a) and average top-five share (Panel b) among institutions in each of the 20 distance bins. Both panels reveal a positive relationship between an institution's share of distant lending and its industry concentration. Institutions with few distant loans tend to diversify their lending across many industries, while institutions with more distant loans concentrate their lending within fewer industries. The figure also plots Live Oak Bank, a prominent example of a remote, specialized lender that we examine in Section 4.

We more formally examine the relationship between remote lending and industry concentration by estimating the following regression for institution b in year t:

$$HHI_{bt} = \alpha + \beta ShareDistant_{bt} + Controls_{bt} + \tau_t + \epsilon_{bt} \quad (1)$$

The dependent variable HHI_{bt} is the industry HHI for the loans originated by institution b in year t. The variable $ShareDistant_{bt}$ is the share of those loans originated to borrowers 100 or more

[24]This rise of remote lending can also be seen by looking at the largest lenders. For the fiscal year 2016, four of the top ten national SBA lenders (by total loan amount) had branches in two or fewer states, three of which (Live Oak Banking Company, Newtek Small Business Finance, and Celtic Bank Corporation) have only a single location. Additionally, some remote lenders are older community banks that have adopted a large online presence, e.g Carver State Bank founded in 1927, Evolve Bank & Trust founded in 1925, The Bankcorp Bank founded in 1876 each gave more than 85% of its loans to remote borrowers.

[25]Specifically, we index lender b's industry shares S_{ib} in decreasing order from the largest share S_{1b} to the smallest S_{Ib}. The top-five share for lender b is $\sum_{i=1}^{5} S_{ib}$.

miles away and β captures the relationship between remote lending and industry concentration. The specification also includes year fixed effects (τ_t), which capture shocks that are common to all lenders, such as changes in market-level industry composition or common economic shocks affecting lending. In some specifications, we also add $Controls_{bt}$, a set of lender-specific controls (lender volume ventiles or lender fixed effects). To account for serial correlation within a bank over time, standard errors are clustered at the bank level.

We restrict the sample of institution-year observations to those that made at least ten loans. The included observations originated 91% of SBA 7(a) loans during the period. Panel B of Table 1 reports the summary statistics for this sample. The median number of loans per institution-year is 24 (mean 95.8). The median share of loans given to borrowers located 100+ miles from the closest branch is 0 and the mean is 9%. For measures of concentration, the median industry HHI is 0.08 (mean 0.09) and the median share of loans given to the institutions' top five industries is 0.40 (mean 0.42).

Table 2 reports the estimates from specification 1. Column 1 confirms the positive relationship between an institution's share of distant loans and its industry concentration. The coefficient of 0.095 (significant at 1% level) indicates that a one standard deviation (20 pp) increase in the share of remote loans is associated with a 0.019 increase in the institution's industry concentration, which is a 21% increase over the mean HHI concentration of 0.09. Column 2 adds ventile indicators for the institution's size (the total number of loans originated by the institution during that year), Column 3 adds institution fixed effects, and Column 4 restricts the sample to a balanced panel of institutions who gave at least 10 SBA loans during each year from 2001-2017. Across all specification, the coefficient on the share of distant loans remains positive and statistically significant. Importantly, the institution fixed effects specifications in Columns 3 and 4 show that the positive relationship between distant loans and concentration also holds within institutions over time. Columns 5-6 replace $ShareDistant$ with the log of the median borrower-lender distance for that bank-year. The positive relationship between distance and concentration remains significant when using these alternative measures of distance and lender concentration. In Appendix Table A.1, we find a similar pattern when industry concentration is measured with the top-five industry share. Overall, the results of this section show a robust positive relationship between lending to distant borrowers and industry concentration.

3.3 Industry Concentration and Loan Performance

If industry concentration facilitates expertise in lending to these industries, concentrated lenders may experience better loan performance within the industries where they focus. To investigate this idea, we examine whether loans from concentrated lenders perform better than loans from less concentrated lenders. One possibility is that concentrated lenders simply choose to focus on industries with lower charge-offs, which would lead to better loan performance of concentrated lenders, even in the absence of expertise. So that our estimates will not be driven by such concentration in low-risk industries, our regressions will control for industry-specific charge-off rates with industry fixed effects. Thus, our strategy compares loan performance between more and less concentrated lenders that originate loans to the same industries.

Using the loan-level data, we estimate the following regression for a loan i from lender b to industry j originated in year t:

$$Chargeoff_{ibjt} = \alpha + \beta_0 log(dist_{ibjt}) + \beta_1 IndustryShare_{bjt} + X_{ibjt}\gamma + \delta_j + \tau_t + \epsilon_{ibjt} \qquad (2)$$

where $Chargeoff_{ibjt}$ is an indicator for whether loan i from lender b originated to industry j during year t was charged off within three years of origination. The variable $log(dist_{ibjt})$ is the log of the distance between the borrower and the closest branch of the institution originating the loan.[26] The main specification also includes loan-level controls for size and term length (X_{ibjt}) and industry (δ_j) and year (τ_t) fixed effects. Some specifications also include additional loan-level controls, state-by-year fixed effects, and lender-specific fixed effects.

Our measure of industry concentration, $IndustryShare_{bjt}$, is the share of total loans from lender b in year t that went to industry j.[27] A more common measure of the sectoral concentration of lenders is an HHI index (e.g. Acharya, Hasan and Saunders (2006) and Hayden, Porath and Westernhagen (2007)). The advantage of the industry-share, which was also used in Berger, Minnis and Sutherland (2017), is that it varies within a bank across industries.[28] The coefficient of interest, β_1, captures the correlation between the probability that a loan in industry j is charged off within three years and the lender's $IndustryShare_{bjt}$. If β_1 is negative, it would reflect that lenders giving a larger share of their loans to an industry experience lower charge-off rates relative to other lenders. Since the specification includes industry fixed effects, β_1 reflects how the probability of charge-offs varies among loans given to the same industry. In some specifications, we add the interaction of the share of loans to an industry and borrower-lender distance, to examine whether industry concentration can mitigate the disadvantages of lending at a distance.

Table 3 reports the results of specification (2). Consistent with the prior literature on distance and lending, the positive coefficient on the $log(dist)$ in Column 1 indicates that the probability of default increases with borrower-lender distance, controlling for loan characteristics (dummies for

[26] Table A.2 finds a similar pattern for small lenders, medium lenders, and large lenders, excluding Live Oak loans, using the county distance measure, and using the lagged industry share. Table A.3 finds similar results when excluding distance as a control.

[27] We focus on contemporaneous shares as our primary measure. If lenders build expertise, e.g. by hiring industry experts, then increase lending to the industry, current lending shares reflect expertise. However, if expertise is developed through past exposure to an industry, it may be more appropriate to use a lagged measure. In robustness checks, we find a similar effect using lagged shares. Moreover, contemporaneous and lagged shares are highly correlated; the coefficient of correlation is 0.92.

[28] An alternative measure concentration could be the *number* of loans a bank gave to the industry. This measure, however, would potentially conflate the effects of bank size and concentration. Instead, we adopt the common approach of using a measure that is comparable across banks of different sizes and then controlling directly for bank size in the regressions (Acharya, Hasan and Saunders, 2006, Hayden, Porath and Westernhagen, 2007, Berger, Minnis and Sutherland, 2017). However, to investigate the role of bank size, Columns 1-3 of Table A.2 estimate specification (2) separately for small, medium, and large lenders. Consistent with both the share and number of loans capturing industry expertise, the coefficient on *IndustryShare* increases in bank size, although the estimate for larger banks is very imprecisely estimated.

ventiles of loan size and term length). Column 2 adds the share of loans that a lender makes to the industry. The negative coefficient on the share in the industry indicates that having a greater share of loans to an industry is correlated with lower charge-off rates within that industry (relative to less concentrated lenders). To provide a sense of the magnitude, these estimates imply that an industry share of 52% would offset the additional risk of a 100-mile loan. The offsetting threshold is higher for more distant loans and lower for closer ones. This negative relationship between concentration and the probability of default remains similar when adding state-by-year fixed effects in Column 3. Column 4 includes the interaction of the "Share in industry" with the log of borrower-lender distance. The coefficient is negative and significant, which is consistent with concentration in lending mitigating the disadvantages of lending at a distance. Columns 5-8 repeat these specifications, but add lender fixed effects. The coefficients decrease in magnitude, but remain statistically significant. Thus, even within an institution, loan performance is better in the industries where the institution is more concentrated. However, in Column 8, the interaction of the industry share with $log(dist)$ is statistically insignificant and slightly positive, indicating that, within lenders, concentration does not weaken the role of distance.

4 Industry Specialization and Access to Credit

The last section documents an increase in remote small business lending over the last decade, and that remote lenders tend to concentrate their loans within fewer industries. That is, remote, industry-specialized lending has become increasingly common. What effect does this new lending technology have on access to credit?

This section investigates whether remote, industry-specialized lenders can expand access to small business loans. Industry-specialized lenders may use industry expertise to meet credit demand that would be unmet by traditional lenders, thereby increasing access to credit. Alternatively, industry-specialized lenders may compete for the same borrowers as traditional lenders, resulting in little change to the total quantity of small business credit. It is also possible that cream-skimming by new entrants with an informational advantage may induce a segmented credit market, as in the models of Detragiache, Tressel and Gupta (2008) and Gormley (2014), ultimately reducing the availability of credit.

The challenge in empirically examining the impact on credit is that remote lending has steadily and endogenously grown over time, and we do not observe the counterfactual number of loans would have been originated in the absence of this growth. To overcome this challenge, we examine the staggered entry by the largest remote, specialized SBA lender: Live Oak Bank. In recent years, Live Oak Bank has been the largest SBA lender (by dollar amount) and, as discussed in the next section, is a prominent example of a remote, industry-specialized lender. Moreover, their large and staggered entry into specific industries provides a unique surge in remote lending that allows us to estimate its impact on total credit.

4.1 Background Information: Live Oak Bank

As shown earlier in Figure 3, Live Oak Bank exhibits the two key features of remote, industry-specialized lenders. Live Oak gave 95% of its SBA loans to borrowers 100 or more miles from its single headquarters in North Carolina and 80% of its loans went to just six of the more than 800 industries receiving SBA loans. Moreover, Live Oak describes its expertise in these industries as its key advantage. "We are one of the nation's top originators of small business loans primarily because our expertise in specific industries enables us to lend to business owners who haven't had access to capital in the past" (Live Oak Bank, n.d.). Live Oak and other remote, industry-specialized lenders use industry experts and industry-specific underwriting criteria to assess firms. For example, concerning Live Oak's lending to Registered Investment Advisors (RIAs), "[O]ne of Live Oak's biggest advantages is that it understands the RIA industry and many banks don't ... A lot of lenders are uncomfortable with the RIA industry ... They don't understand this is a business without a lot of cash flow."[29] For their loans to healthcare professionals, Live Oak reports that "[m]ost lenders provide small business loans, along with other financial products and services, across all industries, never fully-understanding the needs and potential complications that come with lending to physicians... your loan application needn't be subjected to the same standards as an application from another type of business" (Voeller, 2018).

Table 4 presents the industries where Live Oak has given out at least 50 SBA loans as of 2017. This table also shows the number of loans, Live Oak's post-entry share of SBA loans (number and dollar amount) in that industry, and the month of entry. When Live Oak enters an industry, they provide a significant share of subsequent lending to that industry, ranging from 12% of SBA loans to offices of dentists to 58% of SBA loans to investment advice establishments. Live Oak's share of the total loan amount is even greater, since they tend to give larger than average loans. Relative to other SBA lenders in these industries, Live Oak tends to give out larger loans (unconditional mean of $1,161,378 vs. $472,794) for longer periods (unconditional mean of 209 months vs. 147 months) and charge lower interest rates (unconditional mean of 5.57% vs. 5.98%). Live Oak's loan performance within these industries is remarkably good. Their 3-year charge-off rate was 0.15%, compared to a charge-off rate of 1.1% for other lenders to these industries over the same period.

Finally, is among the largest SBA lenders and the largest by total dollar amount, originating more than 6% of all SBA 7(a) loans (dollar-weighted). Live Oak's combination of size, industry concentration, and staggered entry allows us to estimate their impact on total lending in the industries where they operate. We focus on entry into the six industries where Live Oak has given the most loans: veterinarians, dentists, investment advice establishments, pharmacies, broilers, and funeral homes. We exclude the remaining industries to which Live Oak has entered because they either entered in mid-2015, so there is a short post-period, or because they made only a small share of loans to that industry, and so are unlikely to have a noticeable impact. Our strategy will compare changes in loan volumes in the six "treated" industries that Live Oak enters to a group of control industries.

[29] Jamie Carvallo, co-founder of Park Sutton Advisors LLC, quoted in Shidler (2013).

4.2 Identification Strategy: Synthetic Control Method

Our strategy estimates the change in total annual SBA loans in the industries that Live Oak enters, relative to the change in a group of control industries.[30] Due to differences in industry-specific lending trends, changes in industry composition during the Great Recession, and the fact that Live Oak may choose to enter industries based on their past performance, it is challenging to manually select industries that serve as a suitable comparison group. Instead, we employ the synthetic control method, developed by Abadie and Gardeazabal (2003) and Abadie, Diamond and Hainmueller (2010), which provides a systematic way of constructing a synthetic match for each of the industries that Live Oak enters (i.e., the "treated" industries). For each treated industry, the synthetic match is a weighted combination of the control industries (i.e., those industries that Live Oak never enters), where the weights are chosen so that the pattern of annual loans for the synthetic control closely matches that of the treated industry during the period before Live Oak's entry.

Formally, following the setup of Abadie, Diamond and Hainmueller (2010), assume we observe a panel of I industries over T years and consider a single treated industry. Live Oak begins lending to industry 1 in year $T_0 + 1$, and does not lend to the other $I - 1$ control industries. Let Y_{it} be the observed number of loans to industry i at time t, $Y_{1t}(1)$ be the potential number of loans to industry 1 and time t with treatment (Live Oak's entry), and $Y_{1t}(0)$ be the potential number of loans without treatment. Our goal is to estimate the effect of the treatment on total lending to industry 1, $\tau_{1t} = Y_{1t}(1) - Y_{1t}(0) = Y_{1t} - Y_{1t}(0)$ for periods $t > T_0$. We only observe $Y_{1t}(1)$ for the treated industry during the post-treatment period, so estimating the treatment effect requires an estimate of the counterfactual number of loans, $Y_{1t}(0)$, that would have been given out if Live Oak had not entered.

Assume the potential outcomes for all industries i follow the factor model

$$Y_{it}(0) = \delta_t + \lambda_t \mu_i + \varepsilon_{it} \tag{3}$$

where δ_t is an unknown common factor (time fixed effect), λ_t is a $(1 \times F)$ vector of unobserved common factors, μ_i is a $(F \times 1)$ vector of unknown factor loadings, and ε_{it} is an unobserved, industry-level transitory shock with zero mean.

Suppose there are a set of weights $(w^*_{2t}, \ldots, w^*_{It})$, with $w^*_{it} \geq 0$ and $\sum_i w^*_{it} = 1$, such that a weighted combination of the outcomes of control industries equals the outcome of the treated industry for all pre-treatment periods:

$$\sum_{i=2}^{I} w^*_i Y_{i1} = Y_{11}, \quad \sum_{i=2}^{I} w^*_i Y_{i2} = Y_{12}, \quad \ldots, \quad \sum_{i=2}^{I} w^*_i Y_{iT_0} = Y_{1T_0}. \tag{4}$$

[30] With appropriate data, it would also be possible to apply our strategy to other outcomes, namely interest rates or charge-off rates of incumbent banks. This would allow us to assess other aspects of competition and cream-skimming. The SBA data, and Live Oak in particular, are not well-suited to this analysis. Interest rate data are available only after 2008, severely limiting the pre-treatment sample. For charge-offs, only 2 of the 2,511 Live Oak loans originated between 2007 and 2014 to our six treated industries were charged-off within three years of origination. While higher, the post-entry charge-off rates for other lenders within these industries were also low (0.3%).

As an estimator of the treatment effects τ_{1t} for $t > T_0$, Abadie, Diamond and Hainmueller (2010) suggests using

$$\hat{\tau}_{1t} = Y_{1t} - \sum_{i=2}^{I} w_i^* Y_{it},$$

which is asymptotically unbiased as the number of pre-treatment periods grows.

In practice, there is not a set of weights such that equations in (4) will hold exactly. Instead, we select weights such that the equation holds approximately. For each treated industry j, we construct a set of weights for the synthetic control by solving the following optimization problem:

$$\begin{aligned} \{w_i^{j*}\}_{j \in \text{Treated}} = \underset{\{w_i^j\}_{i \in \text{Control}}}{arg\,min} \sum_{t \leq T_0^j} [Y_{jt} - \sum_{i \in \text{Control}} w_i^j Y_{it}]^2 \\ s.t. \textstyle\sum_{i \in \text{Control}} w_i^j = 1 \\ \text{and} \quad w_i^j \geq 0 \quad \forall i. \end{aligned}$$

That is, we choose weights to minimize the mean squared error of annual lending between the treated industry and the synthetic control during the pre-treatment period. For each treated industry, the estimation window $1, \ldots, T_0^j$ covers the years 2001 to the year before Live Oak entered industry j.[31] We find the optimal weights then construct the synthetic control for treated industry j as $\hat{Y}_{jt}(0) = \sum_{i \in \text{Control}} w_i^{j*} Y_{it}$. The estimated impact of Live Oak entering on total loan volume is the difference between Y_{jt} and $\hat{Y}_{jt}(0)$ during the post-treatment period.

In this setting, the synthetic control method has several advantages over difference-in-differences estimators. While the difference-in-difference method restricts the weights on the control units to be equal, the synthetic control method varies the weights to capture the fact that some industries better match the treated unit during the pre-treatment period. Additionally, by examining pre-treatment fit, the method also provides a convenient way to assess the suitability of the comparison group. Additionally, the model in equation (3) also allows industry-specific loadings to common unobserved, time-varying factors ($\lambda_t \mu_i$). For example, the total number of loans across industries may respond differently to macroeconomic shocks.

Our empirical strategy still relies on the assumption that potential outcomes for all industries follow the factor model in equation (3). The key identification assumption is that the timing of entry by Live Oak into a specific industry does not coincide with other changes affecting the number of loans to an industry. For example, we assume that Live Oak does not enter specific industries because they anticipate abnormal future growth or a structural break in the factor model. We support this assumption in four ways.

First, as mentioned, the synthetic control method allows for time trends and a fixed number of

[31] Specifically, we include all pre-treatment outcomes as covariates in our baselines specification and use the default procedure of synth in Stata. By default, synth uses a regression-based approach to obtain variable weights in the V-matrix of Abadie, Diamond and Hainmueller (2010). As discussed in detail in Kaul et al. (2015), this is equivalent to the minimization procedure above.

unobserved factors with loadings that can vary across industries. To the extent that the determinants of Live Oak's entry are reflected in these past industry trends, we will be controlling for them. Second, Live Oak's description of their entry decisions does not suggest they enter industries that they predict will deviate from the trend. They describe their industry choices as depending on the evaluation of historical repayment performance, the current competition, and, most importantly, the ability to hire an industry expert.[32] The timing of entry depends on their ability to acquire or develop the necessary expertise. Third, using the exact timing of Live Oak's entry, we argue, will limit bias due to unobserved factors affecting both entry and growth. Given the number of loans that Live Oak provides, its entry is a large and discrete change to the lending market in the industry. As long as the impact of this shock is large relative to the omitted factors that are correlated with entry and affect growth, the bias will be limited.[33] Fourth, we show that the increases in lending to the treated industries are not driven by other remote lenders, that lending grows relative to the number of establishments in the industry, and that the growth in these industries occurs only in the location where Live Oak actually gave out loans.

4.3 Sample Construction: Treatment and Control Industries

We use data from the SBA 7(a) Loan Data Report to construct annual counts of approved SBA 7(a) loans by industry (5-digit NAICS code) from 2001-2017.[34] We begin in 2001 because in earlier years many of the observations of 7(a) loans are missing the industry code. Of the initial 835 5-digit NAICS industries receiving SBA loans, we drop the industries where Live Oak has given a small number of loans (i.e. those not among the six primary Live Oak industries). Thus, the control industries face no competition from Live Oak. To ensure consistency in industry definitions, we also drop industries which have had a change in the 5-digit NAICS code between 1997 and 2012, leaving 461 industries. Finally, we retain only the industries that have at least one SBA 7(a) loan approved for each year between 2001 and 2017. The final sample consists of a balanced panel from 2001-2017 of annual loan originations for 310 control industries and the six treated industries that Live Oak has entered. This forms the main sample for our analysis.

Our main sample only allows us to examine changes in SBA lending, so we supplement it with data from the Risk Management Association's (RMA) eStatement Studies.[35] Financial institutions provide the RMA with financial statements collected from commercial borrowers or applicants. Although participation is voluntary, hundreds of financial institutions including 9 of the 10 largest banks provide these statements to the RMA (Lisowsky, Minnis and Sutherland, 2017). The RMA's eStatement Studies publishes counts of the number of financial statements collected by industry

[32] "Our Emerging Markets group identifies new verticals by methodically analyzing payment records, level of competition, and most importantly, conducts a relentless search for a Domain Expert that not only understands the industry but also is a fit with our unique culture. We will strive to create at least four new verticals per year" (Bancshares, 2017).

[33] See Gentzkow, Shapiro and Sinkinson (2011) for a formal version of this argument.

[34] We drop canceled loans and loans given to borrowers in the U.S. territories.

[35] For more detailed information on the participants and coverage of RMA's eStatement Studies, see Berger, Minnis and Sutherland (2017) and Lisowsky, Minnis and Sutherland (2017).

(6-digit NAICS). These counts of financial statements provide a measure of total (SBA and non-SBA) lending activity within an industry. Berger, Minnis and Sutherland (2017) find that there is a strong correlation (0.74) between the cumulative borrower sales reported by a bank in the RMA's financial statements and the size of the bank's commercial and industrial lending portfolio. Using the RMA industry-specific statement reports, we form annual counts of financial statements by industry. Because we manually code the data from RMA, we selected a subset of industries from the SBA sample.[36] The final RMA sample includes a balanced panel of annual financial statement counts for 63 industries from 2001-2017. Live Oak is not a participant in the RMA survey during our sample period, so the RMA data provide a proxy for total industry lending excluding Live Oak.[37]

4.4 Results

4.4.1 Total Credit

Figure 4 compares the actual number of loans in each of the six industries that Live Oak entered to the counterfactual number of loans predicted by the synthetic control.[38] For each industry, the vertical line represents the year before Live Oak's entry. The fit during the pre-treatment period, i.e. the average pre-treatment mean squared prediction error (MSPE), can be used to asses the quality of the synthetic control. The average pre-treatment MSPE for industry j is defined as $\frac{1}{T_0^j}\sum_{t\leq T_0^j}[Y_{jt} - \sum_{i\in \text{Control}} w_i^{j*} Y_{it}]^2$, where Live Oak entered the industry in year $T_0^j + 1$. Based on this measure, we are unable to construct a well-fitted synthetic control for "Broilers" and "Dentists." The average mean squared prediction error (MSPE) for these two industries is over 8,000. However, for the remaining industries, the synthetic control fits well as the average pre-treatment MSPE ranges from 12.4 to 137. As discussed in Abadie, Diamond and Hainmueller (2010), one should not use the synthetic control method when there is not a good pre-treatment fit for the treated unit.

Consequently, we focus our analysis and discussion on the remaining four industries for which we are able to construct a well-fitting synthetic control match. For these four treated industries (Pharmacies, Investment Advice, Veterinarians, and Funeral Homes), Figure 4 shows sharp and persistent increases in the number of loans (relative to the synthetic control) once Live Oak enters. This indicates that Live Oak's entry generated an increase in SBA lending to these industries.

To evaluate the statistical significance of the increases in loans to treated industry j, we estimating placebo synthetic controls for each of the 310 control industries, assuming a "treatment" in the

[36]To create the RMA sample, we begin with the 310 industries in the final SBA sample and keep those industries that have at least 20 SBA loans per year. The minimum number of loans in a year for any of the industries that Live Oak entered is 36, so these larger industries are more similar to those that Live Oak entered. There are 140 industries with at least 20 SBA loans per year. Next, our SBA sample is at the 5-digit NAICS level, while the RMA data are available at the 6-digit NAICS level. Therefore, we keep industries with a one-to-one mapping between the 5- and 6-digit NAICS (as measured in the SBA data). Of these 92 industries, we have complete data in the RMA from 2001-2017 for 63.

[37]Live Oak is not in the List of Participants published for the 2015-2018 eStatement Studies, and we confirmed they did not participate with the RMA in earlier years.

[38]Appendix Table A.4 shows the industries that make up the synthetic controls.

same year that Live Oak entered industry j. For both the treated and control industries, Figure 5 plots the "gap" or difference between the number of loans for each industry and its synthetic control. We discard observations with poor pre-treatment fits, defined as having an average pre-treatment MSPE of more than $\sqrt{3}$ times that of the treated industry.[39] In all four treated industries, the gap for the actually treated industry is large relative to the placebo gaps for the control industries. Across the four treated industries, the share of estimated placebo effects with average post-period treatment effects larger (in absolute value) than the true treatment average varies from 0-2% across the four treated industries.[40]

We then evaluate the joint significance of the four treatment effects by examining the size of the average increase relative to a placebo distribution. Specifically, using a formula similar to that in Acemoglu et al. (2016), we construct the test statistic

$$\widehat{\theta} = \sum_{j \in \text{Treat}} \left(\frac{\sum\limits_{t=T_0^j+1}^{T} \frac{Y_{jt} - \widehat{Y}_{jt}(0)}{(T - T_0^j)} \Big/ \left(Y_{jT_0^j} \widehat{\sigma}_j \right)}{\sum\limits_{j \in \text{Treat}} \frac{1}{\widehat{\sigma}_j}} \right) \tag{5}$$

where

$$\widehat{\sigma}_j = \sqrt{\sum_{t=1}^{T_0^j} \left(Y_{jt} - \widehat{Y}_{jt}(0) \right)^2 / T_0^j}.$$

In the formula, $T_0^j + 1$ is the treatment year for industry j, and T is the total number of periods. The test statistic is $\widehat{\theta}$ is the average annual effect across the treated industries, where the effect is normalized by the number of loans to that industry in the last pre-treatment year ($Y_{jT_0^j}$), and weighted by a measure of the quality of fit in the pre-treatment period ($\frac{1}{\widehat{\sigma}_j}$). Normalizing converts the measure into the percentage change relative to the last pre-treatment year, so the magnitudes are comparable across industries of different size. We then construct a placebo distribution of average effect sizes for control industries. To do this, we randomly select 5,000 sets of four control industries. We assign each of the four a placebo treatment year corresponding to an actual treatment year (i.e., 2007, 2009, 2011, and 2013), then estimate a placebo treatment effect for each using the synthetic control method. Finally, for this placebo group of four, we construct the corresponding average effect $\widehat{\theta}^{PL}$ as in formula (5). Figure 6 shows the distribution of all 5,000 placebo estimates $\widehat{\theta}^{PL}$ compared to the actual treatment effect $\widehat{\theta}$. Only 4.74% of the 5,000 normalized placebo treatment effects are larger in absolute value than the actual normalized treatment effect, indicating that the magnitude of the loan increases to the treated industries is large relative to what would be expected from random variation.

[39] All significance results are similar if we use larger (5 times or 20 times the treatment pre-period MSPE) or include all control industries.

[40] If we include placebo industries with a poor pre-treatment fit, i.e. pre-period MSPE more than $\sqrt{3}$ times that of the treated industry, the share remains under 3.9%.

4.4.2 Substitution from Other Lenders

The results in the last section show that the entry of Live Oak caused a significant increase in total SBA lending to certain industries. It is not clear, however, the extent to which entry also caused substitution away from other lenders. To examine this, we drop Live Oak loans from the sample and repeat the synthetic control analysis. Since Live Oak loans are excluded, the synthetic control now reflects the counterfactual number of loans other SBA lenders would have given to the treated industry if Live Oak did not enter. If Live Oak causes substitution away from existing SBA lenders, the actual number of loans will be lower than the synthetic control. Alternatively, if Live Oak complements existing lenders by serving a different segment of borrowers, the actual and counterfactual number of loans would be roughly equal.

Figure 7 presents the results of the synthetic control excluding loans from Live Oak. In all four industries, the actual number of loans given by SBA lenders is similar to (or slightly greater than) the synthetic control.[41] That is, other SBA lenders continued lending similar amounts to these industries, and there is no evidence that Live Oak's entry generated substitution away from other SBA lenders. This suggests that Live Oak's loans were given to borrowers who would not have otherwise received an SBA loan.

A remaining question is whether the Live Oak caused substitution from non-SBA lending. Substitution to non-SBA lending is limited by the "credit elsewhere test" of the SBA 7(a) loan program, which requires the bank to certify that they would be unwilling to make the loan outside of the SBA program and that they believe the borrower could not get other loans on reasonable terms, although banks do have discretion in interpreting this language. Additionally, other SBA lenders are likely the closest substitutes for loans from Live Oak.[42] Given that we found no substitution within the SBA program, it is likely that substitution outside of SBA lending is also limited. Still, to investigate possible substitution, we examine the impact of Live Oak's entry on counts of financial statements collected by lending institutions from the RMA data. As discussed in Section 4.3, these counts provide a measure of both SBA and non-SBA lending by industry for a broad set of financial institutions. Importantly, Live Oak is not included in these counts. Figure 8 presents the results. In all four industries, the actual number of financial statements collected is similar to the synthetic control, implying little change in total (SBA and non-SBA lending) to these industries by other lenders.

4.4.3 Threats to Identification

This section investigates two potential threats to our interpretation of the increases in lending as the causal effect of Live Oak's entry. First, perhaps some of the increase is due to other remote lenders targeting the same industries. If so, our estimates are picking up the effect of both Live

[41] In unreported results, we find that the similarity of the actual number of loans and the synthetic control remains if we separately analyze small or large banks, defined as banks with less or more than \$10 billion in assets.

[42] Live Oak's 2017 Annual Report states that "[i]f we lose our status as a Preferred Lender, we may lose some or all of our customers to lenders who are SBA Preferred Lenders."

Oak's entry and the subsequent entry of other remote lenders. To investigate this, we repeat the synthetic control analysis but exclude loans from other remote lenders, defined as lenders as those whose median lending distance in the year was more than 100 miles.[43] Figure 9 reports the results. Similar increases in total lending still occur across all four industries. Moreover, the size of the increase closely tracks with the actual number of loans Live Oak gave out, as seen by the line "Synth. + Live Oak", which adds the actual number of Live Oak loans to amount predicted by the synthetic control.

A second concern is that Live Oak targets industries that will experience rapid growth. If there is overall growth in these industries, independent of Live Oak, we would expect to see increases in lending to these industries nationwide. We think this explanation is unlikely, as Live Oak's description of its industry selection emphasizes past performance and the ability to hire industry experts, rather than anticipation of abnormal growth. Moreover, the increases occur immediately upon Live Oak's entry and the magnitudes of increases match closes with the number of loans Live Oak originates (Figure 9).

To empirically test whether these industries experienced abnormal growth, we estimate the synthetic control but replace the dependent variable with the annual number of SBA loans to each industry divided by the number of national establishments in that industry-year from the Quarterly Census of Employment and Wages. If these industries were experiencing unusual growth, then establishments would increase at a similar rate to lending, resulting in little change in the ratio of loans to establishments. Instead, Figure 10 shows similar increases in SBA lending to these industries relative to the number of establishments. As an alternative test of whether lending to these industries increased independently of Live Oak, we estimate a synthetic control, but exclude from the sample any loans given to borrowers in zip codes where Live Oak ever provided a loan to any industry. Figure 11 shows the results. The actual number of loans in these zip codes is close to the synthetic control. Using equation (5) to calculate the average treatment effect in areas with no Live Oak loans, and comparing it to the placebo distribution in Figure 6, the corresponding two-sided p-value is 0.483. That is, there is no significant increase in lending to treated industries in areas where Live Oak gave no loans; the treatment effect these areas is smaller than almost 50% of the placebo treatment effects. Overall, the results from these tests indicate that lending to these industries increased relative to number of businesses in these industries, and that the increases occurred only where Live Oak gave out loans.

4.5 Mechanism: Industry Selection and Industry Expertise

The results above indicate that Live Oak is able to extend loans to borrowers that would not have obtained a loan otherwise. Additionally, in our sample, Live Oak maintained very low charge-off rates (only two charge-offs within three years of origination among loans in the six treated industries).

[43] We allow a bank to be a remote lender for some but not all years if there are years some years when their median lending distance is more than 100 miles and other years when it is less than 100 miles. In this case, we only drop loans from the bank during the years where the median distance is greater than 100 miles. We explored several other definitions, and the results of this section are not dependent on the choice of a specific threshold for remote lending.

How is Live Oak able to identify low-risk but underfinanced borrowers? This paper emphasizes the role of industry specialization, and this section provides additional evidence about two potential aspects of industry specialization - industry selection and industry expertise.

One advantage of industry specialization is that it allows specialized lenders to select industries that are lower risk, less competitive, or better suited for distant lending. To investigate this in the case of Live Oak, we estimate a regression to compare the characteristics of the six industries that Live Oak enters to other industries receiving SBA loans. To focus on industry characteristics, rather than the performance of Live Oak, we exclude loans from Live Oak and estimate the following regression for loan i originated to industry j in year t:

$$y_{ijt} = \alpha + \beta_0 LO\_Industry_j + \beta_1 log(dist_{ijt}) + \beta_2 LO\_Industry_j \times log(dist_{ijt}) + X_{ijt} + \varepsilon_{ijt} \quad (6)$$

The dependent variable is an indicator for whether the loan was charged-off within three years of origination or, in separate regressions, the interest rate on the loan.[44] $LO\_Industry$ is an indicator for the six industries that Live Oak entered, and β_0 captures the difference in average charge-off rates (or interest rates) between the industries that Live Oak enters and the other SBA industries. If $\beta_0 < 0$, it would indicate that Live Oak enters industries with lower average charge-off rates. As mentioned, we exclude Live Oak loans from the sample, so the estimates do not reflect differences due to Live Oak's loan performance. We also control for the log of borrower lender distance, $log(dist)$, and the interaction of this term with the indicator for Live Oak industries. The coefficient β_2 reflects how the relationship between distance and charge-offs differs in Live Oak's industries and other SBA industries. We add loan-level controls, X_{ijt}, consisting of year fixed effects and dummy variables for the ventiles of loan size and term length, as well as industry fixed effects in some specifications.[45]

A second potential advantage of industry specialization is that it may facilitate industry expertise, either through experience or economies of scale. To investigate this, we examine differences between Live Oak's loans and other loans to the same industries. Restricting the sample to Live Oak's six industries, we estimate the following specification:

$$y_{ijt} = \alpha + \beta_0 LiveOak_{ijt} + \beta_1 log(dist_{ijt}) + \beta_2 LiveOak_{ijt} \times log(dist_{ijt}) + X_{ijt} + \varepsilon_{ijt} \quad (7)$$

Again, we estimate separate regressions with either the three-year charge-off indicator or the interest rate as the dependent variable. $LiveOak$ is an indicator for whether Live Oak originated the loan, and β_0 captures the difference in charge-off rates (or interest rates) between Live Oak loans and other loans. We also control for the log of borrower lender distance, $log(dist_{ijt})$, and the interaction of this term with the indicator for Live Oak industries. The coefficient β_2 reflects how the relationship between distance and charge-offs differs in Live Oak's industries and other SBA industries. All regressions include controls for loan characteristics, year fixed effects, and industry fixed effects.

[44]The charge-off sample consists of loans from 2007-2014. Interest rate data are available beginning in late 2008, resulting in a slightly smaller sample for regressions with the interest rate as the outcome.

[45]When industry fixed effects are included, we cannot separately identify the coefficient on $LO\_Industry$.

Table 5 reports the results from these regressions.[46] Columns 1-4 report estimates from specification 6. Column 1 shows that loans in Live Oak's industries are 1.35 percentage points less likely to be charged off, relative to SBA loans in other industries (controlling for the ventiles of loan size and term length and year fixed effects). As in the regressions of Section 3.3, charge-off rates are increasing in distance. In Column 2, we replace the Live Oak industry indicator with industry-specific fixed effects, and also include the interaction of the industry-indicator with lending distance. Live Oak's industries exhibit a weaker relationship between the distance and charge-offs than other industries; the coefficient on the interaction of the log of borrower-lender and indicator for a Live Oak industry is negative and statistically significant.

Columns 3 and 4 repeat these specifications with interest rates as the outcome. Column 3 shows no significant difference in interest rates relative to other industries. However, Column 4 reveals that interest rates increase as borrower-lender distance increases, and this is positive correlation is larger in the industries that Live Oak enters. Overall, these results suggest that industry selection offers an advantage. Live Oak entered industries with lower average charge-off rates overall and a weaker relationship between distance and charge-offs. Moreover, the interest rate results suggest that these industries may be less competitive or that other lenders are mispricing the lower credit risk. Despite lower charge-off rates, there are no significant differences in interest rates. Additionally, although charge-off rates rise at a slower rate than other industries (Column 2), interest rates rise more rapidly with distance (Column 4).

Table 5 Columns 5-8 report estimates from specification 7. Column 5 and 6 compare the probability that a Live Oak loan is charged-off, relative to other loans to the same industries. Column 5 reveals that Live Oak's charge-off rates are significantly lower than other lenders. For completeness, we include Column 6, which investigates how Live Oak's relative charge-off rate varies with distance. These estimates are imprecise, though this is not surprising. Live Oak experienced only two charge-offs in our sample, limiting our ability to examine heterogeneity in their charge-offs. Columns 7-8 repeat these regressions with the interest rate as the outcome. Live Oak charges lower interest rates than other lenders operating in the same industries. Overall, these results support industry expertise as a second advantage of industry specialization. Live Oak experiences lower charge-off rates and charges lower interest rates than other lenders in the same industries.

Industry specialization can expand access by allowing lenders to select industries that are better suited to distant lending and allowing lenders to develop expertise in lending to these industries. It is also possible that remote lenders expand access to borrowers in locations underserved by traditional banks. To investigate this, we examine whether remote borrowers are located far from physical branches of SBA lenders. The results suggest that physical distance to a branch is unlikely to limit the supply of credit. Indeed, our distance measures indicate that 99% of remote SBA borrowers are within 10 miles of a branch of a bank that has granted SBA loans. Moreover, Appendix Figure A.3, which plots the distribution of distances between remote and local SBA borrowers and the

[46] Table A.5 repeats the regressions using the county-based measure of borrower-lender distance, which is available for a greater number of SBA loans.

closest branch of any SBA lender, shows that remote borrowers are not located farther from physical branches than local borrowers.

5 Conclusion

While small business lending is largely local, distances between small business borrowers and lenders have increased over the past several decades. This paper documents that a significant portion of the increase is due to remote lending, i.e. loans where the borrower and lender are more than 100 miles apart. The lenders providing a significant number of remote loans tend to concentrate within fewer industries. That is, geographically diversified lenders are more likely to be industry-concentrated. Industry concentration may facilitate the development of industry expertise in lending, and consistent with this, we find that concentrated lenders have lower charge-offs.

We then examine the competitive impact of entry by the largest of these specialized, remote lenders: Live Oak Bank. We find that the entry of Live Oak Bank into specific industries resulted in sharp and persistent increases in the number of SBA loans granted to firms in these industries. Moreover, there was little to no resulting decline in lending from existing lenders. While we do not observe non-SBA lending directly, we find no evidence of substitution based on one measure of total lending: financial statements collected from firms as a part of the lending process. This case study shows that remote, industry-specific lending strategy has the potential to deepen credit markets. In particular, our paper shows an increase in SBA 7(a) lending, which is targeted to credit-constrained borrowers. Moreover, the default rate on Live Oak's loans is extremely low (only two charge-offs within three years of origination in our sample), suggesting that they expand access to the program without reducing loan performance or increasing the SBA 7(a) program's costs.

One question raised by this research involves the extent to which these results hold for other lenders and outside of the SBA program. While we cannot address this question directly, we show that remote, industry-specialized lending is not isolated nor unique. Within SBA 7(a) lending, lenders originating many remote loans generally tend to concentrate in fewer industries. Many of the most concentrated SBA lenders are relatively new and the practice is expanding.[47] Moreover, the rise of niche or specialty lending has received attention outside of SBA small business lending.[48] A recent article by Karen Mills, former Administrator of the Small Business Administration, emphasizes specialization of lenders in specific industries as a key innovation of emerging small business lenders Mills (2019).

As industry-specific lending grows, it may affect broader labor market and banking outcomes. If more loans are extended to certain industries, it may alter the industrial composition of small

[47] For example, Bank of George, founded in 2007 with two offices in Nevada, and The Mint National Bank, opened in 2007 in Texas, give more than half of their SBA 7(a) loans to firms in the hotel and motel industry. Finwise Bank, started in 2000, gives more than half of its SBA loans to offices of lawyers. Affinity Bank, started in 2002 but renamed in 2010, gives 43% of SBA loans to dentists and another 20% to physician offices. Others give a large share (> 20%) to retirement communities, dry cleaners, restaurants, gas stations, lawyers, liquor stores. Some banks (e.g. Hanmi Bank) that specialize in certain ethnicities.

[48] See American Banker (2013) and American Banker (2012) for examples of other niche lenders.

businesses. Live Oak Bank and other remote lenders have altered the industry composition of SBA 7(a) lending. Additionally, remote, industry-specialized lenders face different risks than local banks. Since these lenders are not concentrated geographically, they are less exposed to regional economic downturns. However, industry concentration may increase their exposure to industry-specific risks.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association*, 105(490): 493–505.

Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review*, 93(1): 113–132.

Acemoglu, Daron, Simon Johnson, Amir Kermani, James Kwak, and Todd Mitton. 2016. "The Value of Connections in Turbulent Times: Evidence from the United States." *Journal of Financial Economics*, 121(2): 368–391.

Acharya, Viral V, Iftekhar Hasan, and Anthony Saunders. 2006. "Should Banks Be Diversified? Evidence from Individual Bank Loan Portfolios." *The Journal of Business*, 79(3): 1355–1412.

Agarwal, Sumit, and Robert Hauswald. 2010. "Distance and Private Information in Lending." *The Review of Financial Studies*, 23(7): 2757–2788.

American Banker. 2012. "Specialty Lending Takes Off." *https://www.americanbanker.com/slideshow/specialty-lending-takes-off*, Accessed: 2019-09-01.

American Banker. 2013. "Niche Lending Gains Allure." *https://www.americanbanker.com/slideshow/niche-lending-gains-allure*, Accessed: 2019-09-01.

Bancshares, Live Oak. 2017. "Live Oak Bancshares, 2017 Annual Report." *https://investor.liveoakbank.com/static-files/417a703e-1b87-4155-b723-0a5f24df2cc7*.

Beck, Thorsten, and Maria Soledad Martinez Peria. 2010. "Foreign Bank Participation and Outreach: Evidence from Mexico." *Journal of Financial Intermediation*, 19(1): 52–73.

Beck, Thorsten, Olivier De Jonghe, et al. 2013. "Lending Concentration, Bank Performance and Systemic Risk: Exploring Cross-country Variation." The World Bank.

Berger, Allen N, and Gregory F Udell. 1995. "Relationship Lending and Lines of Credit in Small Firm Finance." *Journal of Business*, 351–381.

Berger, Allen N, Nathan H Miller, Mitchell A Petersen, Raghuram G Rajan, and Jeremy C Stein. 2005. "Does Function Follow Organizational Form? Evidence from the Lending Practices of Large and Small Banks." *Journal of Financial Economics*, 76(2): 237–269.

Berger, Philip G, Michael Minnis, and Andrew Sutherland. 2017. "Commercial Lending Concentration and Bank Expertise: Evidence from Borrower Financial Statements." *Journal of Accounting and Economics*, 64(2-3): 253–277.

Boeve, Rolf, Klaus Duellmann, and Andreas Pfingsten. 2010. "Do Specialization Benefits Outweigh Concentration Risks in Credit Portfolios of German Banks?"

Brown, J David, and John S Earle. 2017. "Finance and Growth at the Firm Level: Evidence from Sba Loans." *The Journal of Finance*, 72(3): 1039–1080.

Bruno, Valentina, and Robert Hauswald. 2013. "The Real Effect of Foreign Banks." *Review of Finance*, 18(5): 1683–1716.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2018. "Fintech, Regulatory Arbitrage, and the Rise of Shadow Banks."

Claessens, Stijn, and Neeltje Van Horen. 2014. "Foreign Banks: Trends and Impact." *Journal of Money, Credit and Banking*, 46(s1): 295–326.

Degryse, Hans, and Steven Ongena. 2005. "Distance, Lending Relationships, and Competition." *The Journal of Finance*, 60(1): 231–266.

Dell'Ariccia, Giovanni, and Robert Marquez. 2004. "Information and Bank Credit Allocation." *Journal of Financial Economics*, 72(1): 185–214.

De Roure, Calebe, Loriana Pelizzon, and Anjan V Thakor. 2019. "P2p Lenders Versus Banks: Cream Skimming or Bottom Fishing?"

Detragiache, Enrica, Thierry Tressel, and Poonam Gupta. 2008. "Foreign Banks in Poor Countries: Theory and Evidence." *The Journal of Finance*, 63(5): 2123–2160.

DeYoung, Robert, Dennis Glennon, and Peter Nigro. 2008. "Borrower-lender Distance, Credit Scoring, and Loan Performance: Evidence from Informational-opaque Small Business Borrowers." *Journal of Financial Intermediation*, 17(1): 113–143.

DeYoung, Robert, W Scott Frame, Dennis Glennon, and Peter Nigro. 2011. "The Information Revolution and Small Business Lending: The Missing Evidence." *Journal of Financial Services Research*, 39(1-2): 19–33.

Dincbas, Neslihan, Tomasz Kamil Michalski, and Evren Ors. 2017. "Banking Integration and Growth: Role of Banks' Previous Industry Exposure."

Federal Deposit Insurance Corporation. 2018. "FDIC Small Business Lending Survey."

Federal Reserve Banks. 2017. "2017 Small Business Credit Survey: Report on Employer Firms."

Frankel, David M, and Yu Jin. 2015. "Securitization and Lending Competition." *The Review of Economic Studies*, 82(4): 1383–1408.

Fuster, Andreas, Matthew Plosser, Philipp Schnabl, and James Vickery. 2019. "The Role of Technology in Mortgage Lending."

Gentzkow, Matthew, Jesse M Shapiro, and Michael Sinkinson. 2011. "The Effect of Newspaper Entry and Exit on Electoral Politics." *American Economic Review*, 101(7): 2980–3018.

Giannetti, Mariassunta, and Steven Ongena. 2009. "Financial Integration and Firm Performance: Evidence from Foreign Bank Entry in Emerging Markets." *Review of Finance*, 13(2): 181–223.

Giannetti, Mariassunta, and Steven Ongena. 2012. "Lending by Example: Direct and Indirect Effects of Foreign Banks in Emerging Markets." *Journal of International Economics*, 86(1): 167–180.

Goodman, Joshua, Julia Melkers, and Amanda Pallais. 2019. "Can Online Delivery Increase Access to Education?" *Journal of Labor Economics*, 37(1): 1–34.

Gormley, Todd A. 2010. "The Impact of Foreign Bank Entry in Emerging Markets: Evidence from India." *Journal of Financial Intermediation*, 19(1): 26–51.

Gormley, Todd A. 2014. "Costly Information, Entry, and Credit Access." *Journal of Economic Theory*, 154: 633–667.

Granja, João, Christian Leuz, and Raghuram Rajan. 2018. "Going the Extra Mile: Distant Lending and Credit Cycles." National Bureau of Economic Research.

Hauswald, Robert, and Robert Marquez. 2006. "Competition and Strategic Information Acquisition in Credit Markets." *The Review of Financial Studies*, 19(3): 967–1000.

Hayden, Evelyn, Daniel Porath, and Natalja v Westernhagen. 2007. "Does Diversification Improve the Performance of German Banks? Evidence from Individual Bank Loan Portfolios." *Journal of Financial Services Research*, 32(3): 123–140.

Haynes, George, and Victoria Williams. 2018. "Small Business Lending in the United States, 2016." U.S. Small Business Administration Office of Advocacy.

Jagtiani, Julapa, and Catharine Lemieux. 2017. "Fintech Lending: Financial Inclusion, Risk Pricing, and Alternative Information."

Jahn, Nadya, Christoph Memmel, and Andreas Pfingsten. 2013. "Banks' Concentration Versus Diversification in the Loan Portfolio: New Evidence from Germany."

Kaul, Ashok, Stefan Klößner, Gregor Pfeifer, and Manuel Schieler. 2015. "Synthetic Control Methods: Never Use All Pre-intervention Outcomes Together with Covariates."

Liberti, Jose M, and Atif R Mian. 2008. "Estimating the Effect of Hierarchies on Information Use." *The Review of Financial Studies*, 22(10): 4057–4090.

Liberti, José María, and Mitchell A Petersen. 2018. "Information: Hard and Soft." *Review of Corporate Finance Studies*, 8(1): 1–41.

Lisowsky, Petro, Michael Minnis, and Andrew Sutherland. 2017. "Economic Growth and Financial Statement Verification." *Journal of Accounting Research*, 55(4): 745–794.

Live Oak Bank. n.d.. "Our Difference is Your Advantage." *https://www.liveoakbank.com/advantages/*, Accessed: 2019-09-01.

Loutskina, Elena, and Philip E Strahan. 2011. "Informed and Uninformed Investment in Housing: The Downside of Diversification." *The Review of Financial Studies*, 24(5): 1447–1480.

Mills, Karen. 2019. "Small-Business Banking is About to Get a Whole Lot Better." *American Banker*.

Mills, Karen, and Brayden McCarthy. 2016. "The State of Small Business Lending: Innovation and Technology and the Implications for Regulation."

Nguyen, Hoai-Luu Q. 2019. "Are Credit Markets Still Local? Evidence from Bank Branch Closings." *American Economic Journal: Applied Economics*.

Petersen, Mitchell A, and Raghuram G Rajan. 1994. "The Benefits of Lending Relationships: Evidence from Small Business Data." *The Journal of Finance*, 49(1): 3–37.

Petersen, Mitchell A, and Raghuram G Rajan. 2002. "Does Distance Still Matter? the Information Revolution in Small Business Lending." *The Journal of Finance*, 57(6): 2533–2570.

Qian, Jun, Philip E Strahan, and Zhishu Yang. 2015. "The Impact of Incentives and Communication Costs on Information Production and Use: Evidence From Bank Lending." *The Journal of Finance*, 70(4): 1457–1493.

Rajan, Raghuram G. 1992. "Insiders and Outsiders: The Choice between Informed and Arm's-length Debt." *The Journal of Finance*, 47(4): 1367–1400.

Schneider, Howard. 2016. "Flourishing Forte." *Independent Banker*.

Sharpe, Steven A. 1990. "Asymmetric Information, Bank Lending, and Implicit Contracts: A Stylized Model of Customer Relationships." *The Journal of Finance*, 45(4): 1069–1087.

Shidler, Lisa. 2013. "Fidelity's M&a Program Reload Looks 'game-changing' After It Partners with a Middleman to Get Uncle Sam to Guarantee Ria Deals." *Riabiz*.

Stomper, Alex. 2006. "A Theory of Banks' Industry Expertise, Market Power, and Credit Risk." *Management Science*, 52(10): 1618–1633.

Tabak, Benjamin M, Dimas M Fazio, and Daniel O Cajueiro. 2011. "The Effects of Loan Portfolio Concentration on Brazilian Banks' Return and Risk." *Journal of Banking & Finance*, 35(11): 3065–3076.

Tang, Huan. 2019. "Peer-to-Peer Lenders Versus Banks: Substitutes or Complements?" *The Review of Financial Studies*, 32(5): 1900–1938.

Temkin, Kenneth. 2008. "An Analysis of the Factors Lenders Use to Ensure Their Sba Borrowers Meet the Credit Elsewhere Requirement."

Voeller, Jon. 2018. "Looking for Financing? Find Your 'Niche'." *https://www.liveoakbank.com/medical-practice-resources/look-financing-find-niche/*, Accessed: 2018-11-28.

Von Thadden, Ernst-Ludwig. 2004. "Asymmetric Information, Bank Lending and Implicit Contracts: The Winner's Curse." *Finance Research Letters*, 1(1): 11–23.

Wasi, Nada, Aaron Flaaen, et al. 2015. "Record Linkage Using Stata: Preprocessing, Linking, and Reviewing Utilities." *Stata Journal*, 15(3): 672–697.

Winton, Andrew. 1999. "Don't Put All Your Eggs in One Basket? Diversification and Specialization in Lending."

Wolfe, Brian, and Woongsun Yoo. 2018. "Crowding Out Banks: Credit Substitution by Peer-to-Peer Lending."

Table 1: **Bank-Year Level Summary Statistics**

Variable	Median	Mean	Std. Dev.	Obs
Panel A: Loan-level summary statistics (2007-2017)				
Loan amount ($1,000s)	100	310.33	582.65	565,610
Loan term (in months)	84	111.34	75.55	565,610
Interest rate (%)	6	6.33	1.43	432,635
Borrower-lender distance (mi.)	2.09	91.5	327.97	374,651
B-L county dist. (mi.)	5.76	97.32	327.29	515,476
Charge-off rate, 3-year	0	.06	.24	389,552
Panel B: Institution-year summary statistics (2007-2017)				
Number of loans	24	95.84	414.46	4,930
Share of loans > 100 mi.	0	.09	.2	4,930
Industry HHI	.08	.09	.07	4,930
Share to top 5 industries	.4	.42	.19	4,930

Panel A reports summary statistics for (non-canceled) SBA 7(a) loans from 2007-2017. Interest rate is available beginning in 2008. Borrower-lender distance is the distance between the borrower and the closest branch of the institution from which she borrower. It is computed as discussed in 2.3. "B-L county dist." is the distance between the centroid of the borrower's project's county and the closest branch. The charge-off rate is an indicator for whether a loan was charged off within three years of origination (available for loans originated in 2014 or earlier). Panel B reports summary statistics for institution-year observations from 2007-2017. The sample of institutions is restricted to institution-years with at least 10 loans. These lenders originated 93% of all SBA loans during the period.

Table 2: **Share of Distant Loans and Lender Industry Concentration**

	Dependent variable: Bank's Industry Concentration (HHI)					
	(1)	(2)	(3)	(4)	(5)	(6)
Share 100+ mi.	0.0952***	0.125***	0.0694***	0.0499***		
	(0.0207)	(0.0198)	(0.0129)	(0.0153)		
log(med. distance)					0.0187***	0.00917***
					(0.00283)	(0.00184)
Observations	4,930	4,930	4,930	1,602	4,930	4,930
Mean Dep. Var.	0.0904	0.0904	0.0904	0.0634	0.0904	0.0904
Year FE	X	X	X	X	X	X
Bank size ventiles		X	X	X	X	X
Bank FE			X	X		X
Balanced panel				X		

This table examines the correlation between the share of loans made at 100+ miles and the lender's industry concentration. Observations are at the lender-year level from 2007-2017 and standard errors are clustered at the lender level. The sample is restricted to lender-year observations with at least 10 loans. Bank size ventiles are ventile indicators for the number of SBA loans each year.

Table 3: **Lender Industry Concentration and Loan Performance (within Industry)**

	Dependent variable: Indicator for Charge-off within 3 Years							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$log(dist)$	0.00476***	0.00500***	0.00437***	0.00521***	0.00208***	0.00210***	0.00252***	0.00208***
	(0.000362)	(0.000358)	(0.000332)	(0.000401)	(0.000388)	(0.000388)	(0.000369)	(0.000425)
Share in industry		-0.0441***	-0.0333***	-0.0391***		-0.0170***	-0.0174***	-0.0176***
		(0.00340)	(0.00284)	(0.00386)		(0.00428)	(0.00416)	(0.00507)
Share$\times log(dist)$				-0.00298**				0.000273
				(0.00144)				(0.00129)
Observations	255,871	255,871	255,871	255,871	255,871	255,871	255,871	255,871
Industry FE	X	X	X	X	X	X	X	X
Year FE	X	X	X	X	X	X	X	X
Loan char.	X	X	X	X	X	X	X	X
State-by-year FE			X				X	
Lender FE					X	X	X	X

This table estimates specification (2). Observations are at the loan level from 2007-2014 and standard errors are clustered at the industry (5-digit NAICS) level. Loan characteristics include dummies for ventiles of the size of the loan and the term length. The state in the state-by-year fixed effects is determined by the location of the borrower's business.

Table 4: **Live Oak Industries**

Industry	**Live Oak Loans**	**Share of Live Oak's Loans**	**Share of SBA Loans**	**Share of SBA Volume**	**Live Oak's Entry Month**
Veterinarians	1,455	0.25	0.33	0.49	06/2007
Offices of Dentists	1,038	0.18	0.12	0.27	03/2009
Investment Advice	814	0.14	0.58	0.75	02/2013
Pharmacies	799	0.14	0.30	0.56	11/2009
Broilers	520	0.09	0.37	0.60	03/2014
Funeral Homes	311	0.05	0.28	0.41	09/2011
Self-Storage	131	0.02	0.34	0.53	05/2015
Insurance Agencies	105	0.02	0.09	0.20	11/2015
Breweries	97	0.02	0.09	0.20	04/2015
Physicians	80	0.01	0.02	0.06	07/2012
Other	378	0.07	0.01	0.03	

This table shows the industries (5-digit NAICS codes) where Live Oak Bank has approved at least 50 loans. "Share of Live Oak's Loans" is the share of Live Oak's 2007-2017 loans going to that industry. "Share of SBA Loans" and "Share of SBA Volume" show Live Oak's post-entry share of SBA loans in each industry by number and dollar amount. "Entry Month" is the month that Live Oak first approved a loan to that industry.

Table 5: **Live Oak Bank: Industry Selection and Industry Expertise**

Sample:	Excluding Live Oak Loans				Loans to Six Industries Live Oak Entered			
Dependent variable:	Charge-off Indicator		Interest Rate (%)		Charge-off Indicator		Interest Rate (%)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LO industry	-0.0135***		-0.00371					
	(0.00230)		(0.0136)					
$log(dist)$	0.00465***	0.00516***		0.00630***	0.00118**	0.00122***	0.0129***	0.0134***
	(0.000201)	(0.000210)		(0.00147)	(0.000460)	(0.000467)	(0.00495)	(0.00504)
LO industry$\times log(dist)$		-0.00323***		0.0232***				
		(0.000987)		(0.00599)				
Live Oak loan					-0.00749**	0.000478	-0.383***	-0.294*
					(0.00379)	(0.0181)	(0.0383)	(0.177)
Live Oak loan $\times log(dist)$						-0.00122		-0.0138
						(0.00271)		(0.0265)
Observations	254,178	254,178	167,071	167,071	10,610	10,610	8,724	8,724
Year FE	X	X	X	X	X	X	X	X
Loan char.	X	X	X	X	X	X	X	X
Industry FE		X		X	X	X	X	X

The sample consists of loans originated between 2007-2014. Columns 1-4 exclude loans originated by Live Oak. Columns 5-8 restrict the sample to loans within the six Live Oak industries (including loans originated by Live Oak). The dependent variable is either an indicator for whether the loan was charged-off within three years of origination or the loan's interest rate (%). Interest rate data are available from 2008Q4. Loan characteristics include dummies for ventiles of the size of the loan and the term length. Industry fixed effects are indicators for the 5-digit NAICS code.

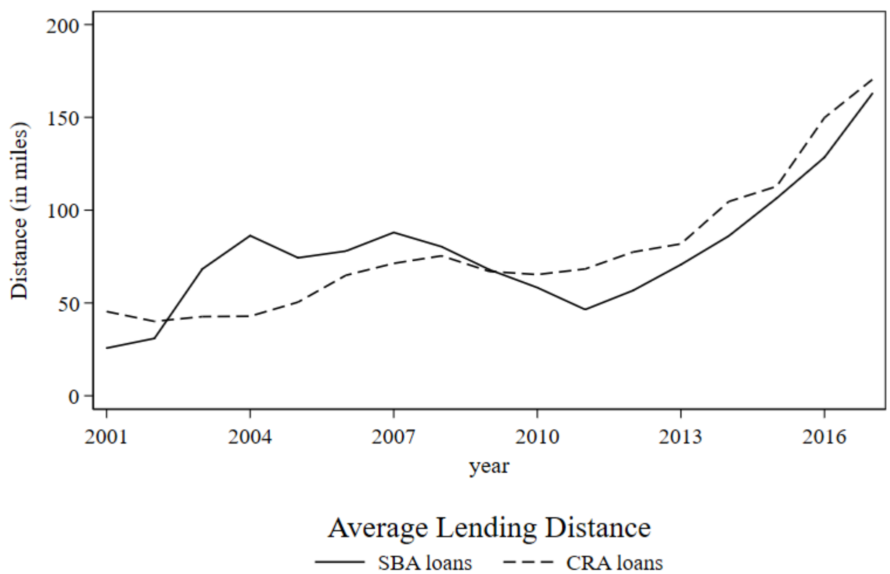
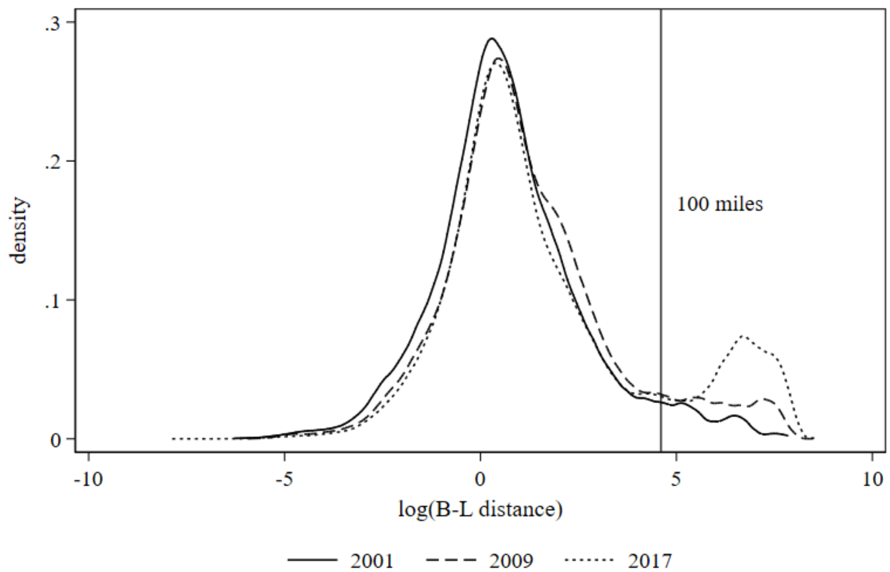
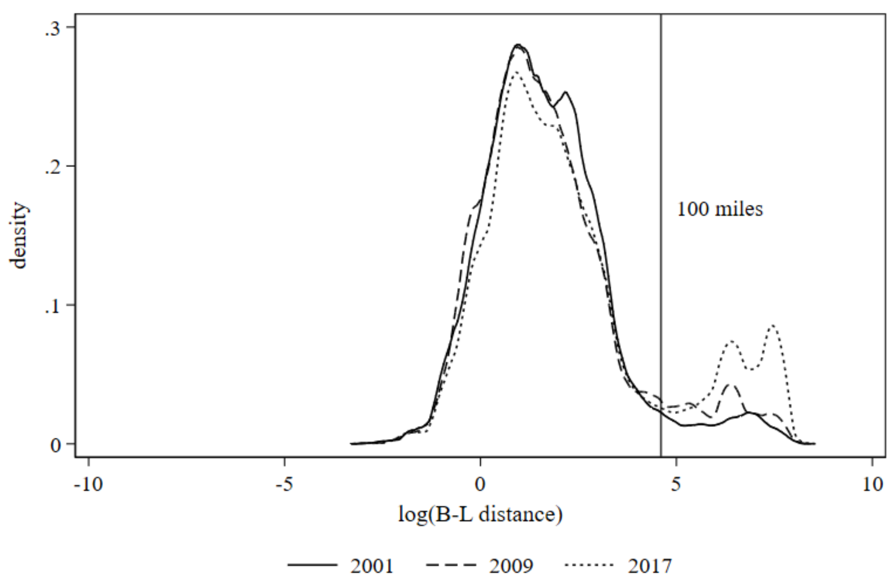


Figure 1: **Changes in Borrower-Lender Distances** This figure shows the average distance between the borrower and lender for SBA 7(a) loans and CRA loans for more than $100,000. Distance is calculated using the method discussed in Section 2.3.



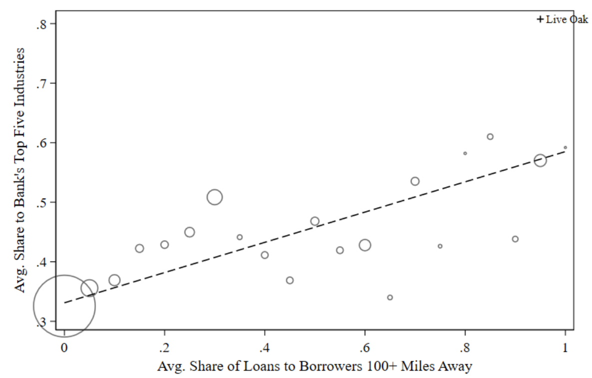
(a) SBA 7(a) Lending



(b) CRA Lending

Figure 2: **Distribution of (log) Borrower-Lender Distance** This figure shows the distribution of the distance between borrowers and the closest branch of the institution from which they borrowed. Borrower-lender distance is calculated according to the procedure described in Section 2.3.

(a)



(b)

Figure 3: **Distant Lending and Industry Concentration** This figure plots the relationship between institutions' share of distant loans and industry concentration. Each institution's share of 100+ mile loans and industry concentration are measured from 2007 to 2017 and the sample is restricted to institutions with at least 20 loans during this period. Institutions are grouped into 20 bins based on the share of loans with a borrower-lender distance over 100 miles (0-5%, 5-10%, etc.). Each institution's industry concentration is measured with the industry HHI or the share of the institution's loans given to its top five industries. The figure plots the average industry HHI (Panel a) and average top-five share (Panel b) among institutions within each of the 20 distance bins. The size of the circle reflects the number of loans by institutions in each bin. The figures also plots Live Oak separately.

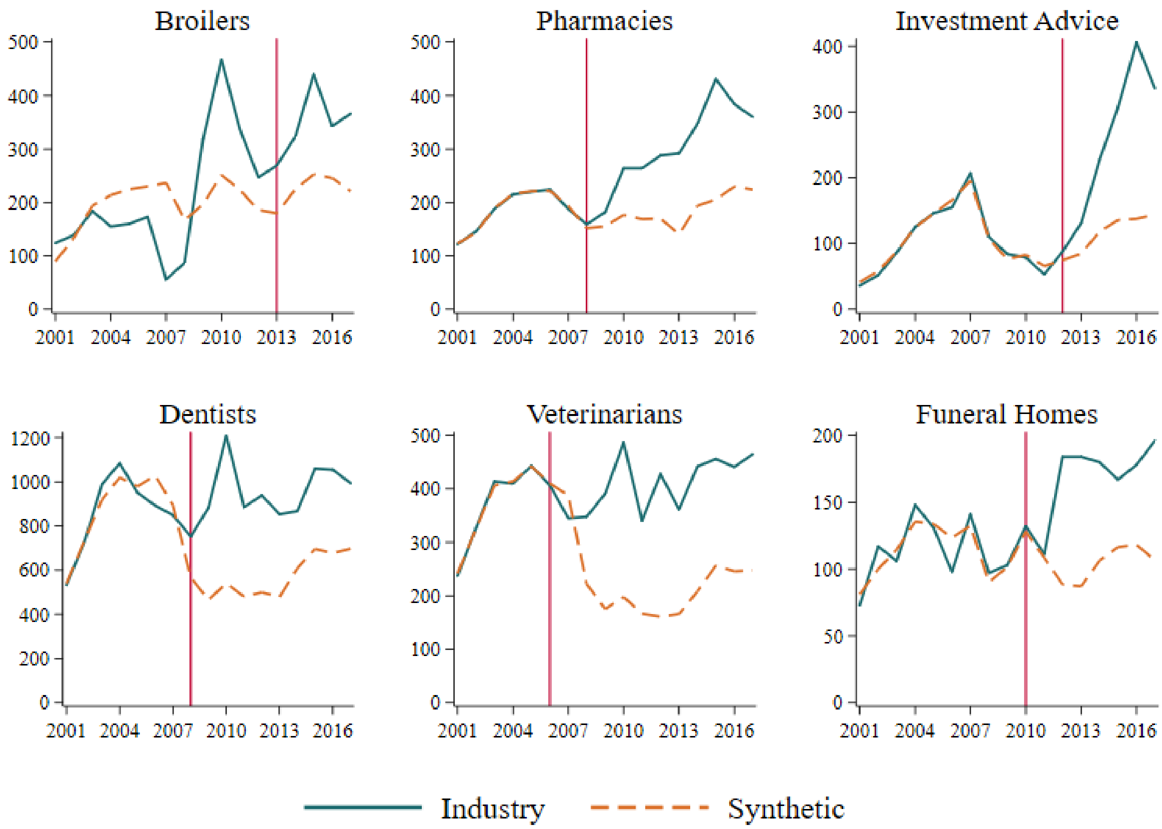


Figure 4: **Number of Loans - Treated Industry vs. Synthetic Control** This figure compares the number of loans in each industry that Live Oak enters to its synthetic control. The synthetic controls are formed by matching on all pre-treatment years beginning in 2001, with no additional covariates. The vertical line shows the year before Live Oak entered.

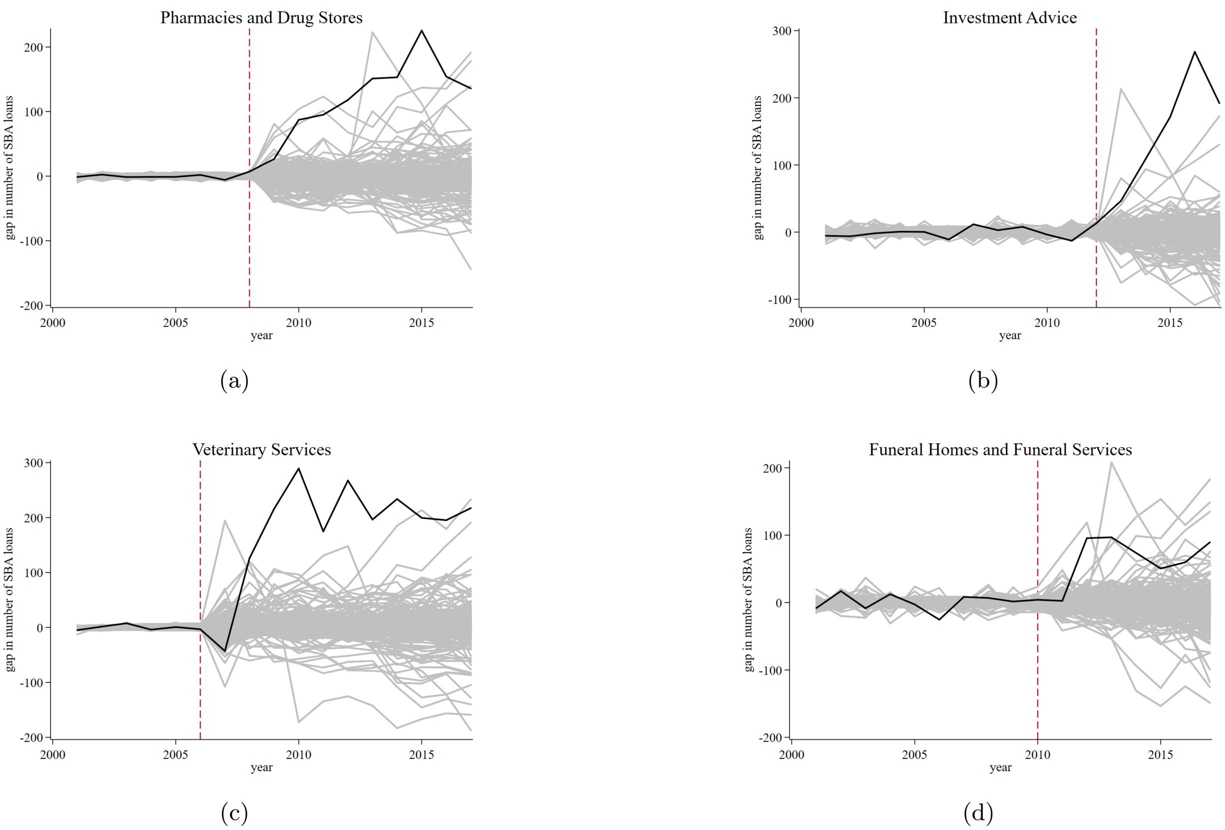


Figure 5: **Comparison of Treatment Effect and Simulated Placebo Effects** The vertical axis shows the "gap" or the difference between the number of loans in an industry and its synthetic control for each year from 2001-2017. The vertical line shows the year before Live Oak entered. The bold line shows the gap for the industry that live Oak entered, while the grey lines show the gap for the placebo industries. The figure discards industries with poor pre-period matches, defined as having pre-entry MSPE $\sqrt{3}$ times higher than that of the treated industry.

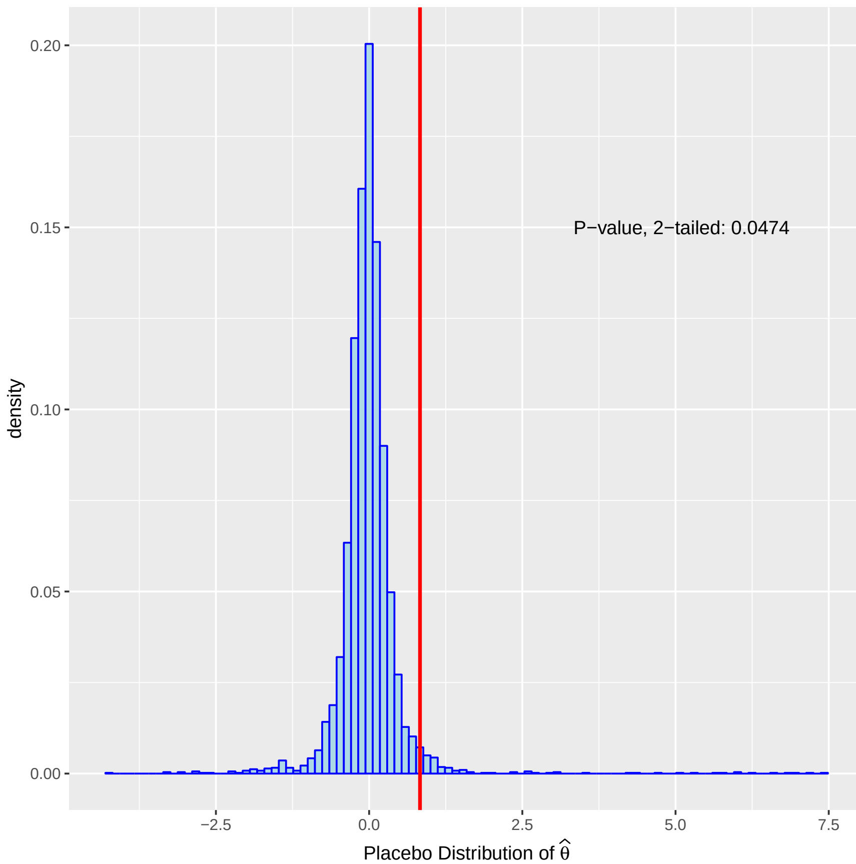


Figure 6: **Placebo Distribution of** $\widehat{\theta}^{PL}$ The vertical red line shows the magnitude of the average treatment effect $\widehat{\theta}$ for the treated industries, calculated from equation (5).

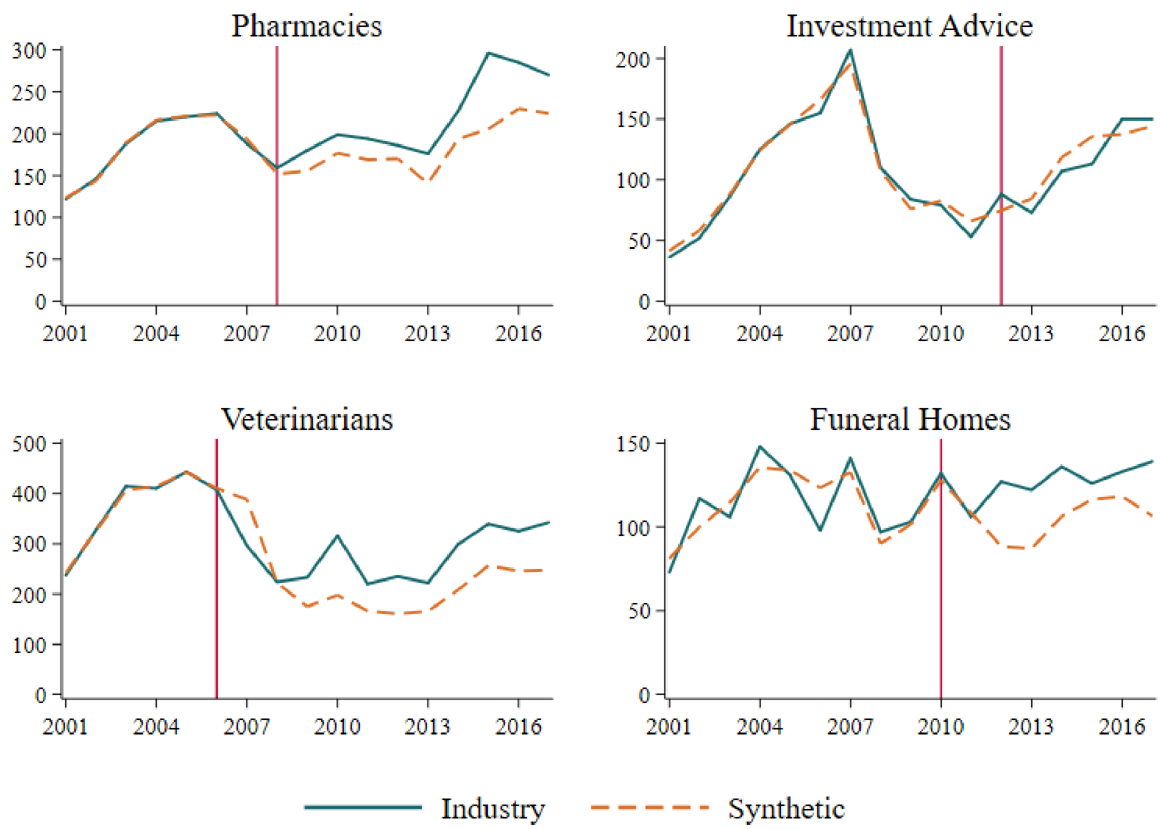


Figure 7: **Synthetic Control Excluding Loans from Live Oak** This figure compares the number of loans from other lenders (excluding Live Oak) to its synthetic control. The synthetic controls are formed by matching on all pre-treatment years beginning in 2001, with no additional covariates. The vertical line shows the year before Live Oak entered.

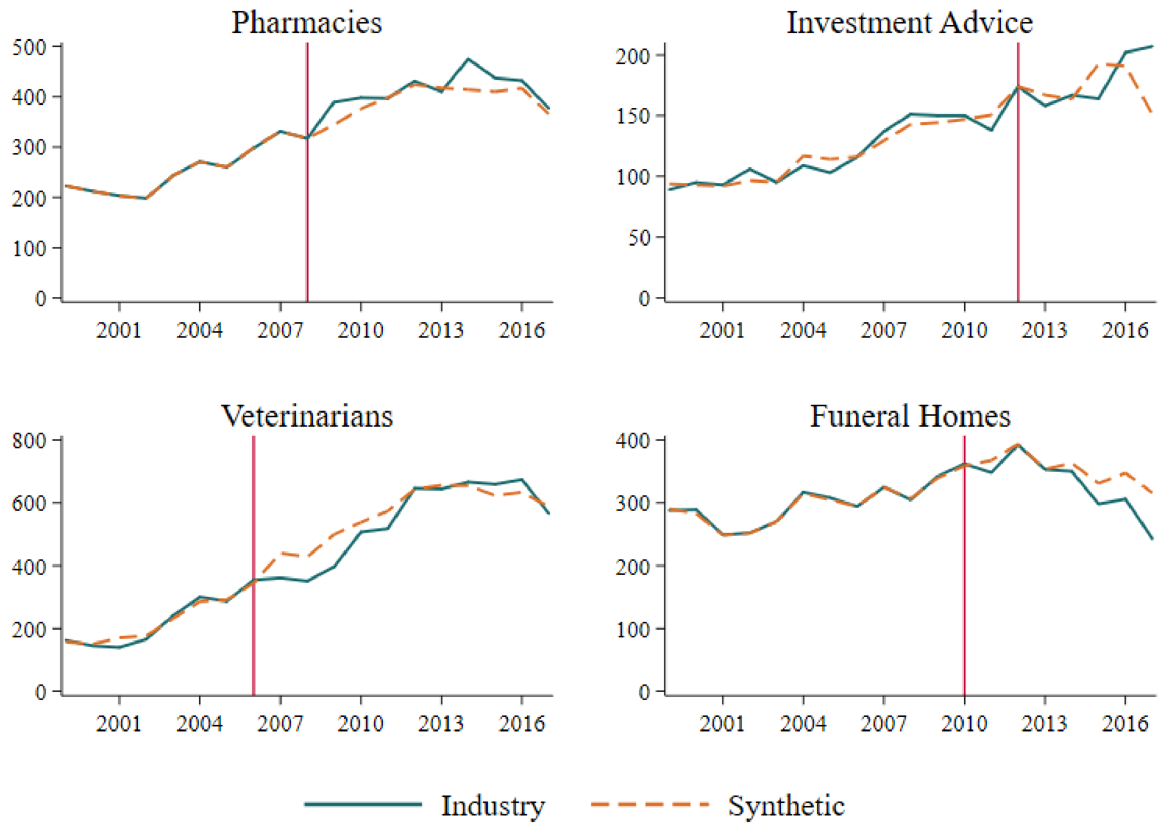


Figure 8: **Synthetic Control using RMA Counts of Financial Statements** This figure shows the change in counts of borrowers' financial statements collected by other lenders upon Live Oak's entry. The figure compares the number of statements collected in each industry that Live Oak enters to its synthetic control. The synthetic controls are formed by matching on all pre-treatment years beginning in 2001, with no additional covariates. The vertical line shows the year before Live Oak entered.

Source: Risk Management Association's Annual eStatement Studies

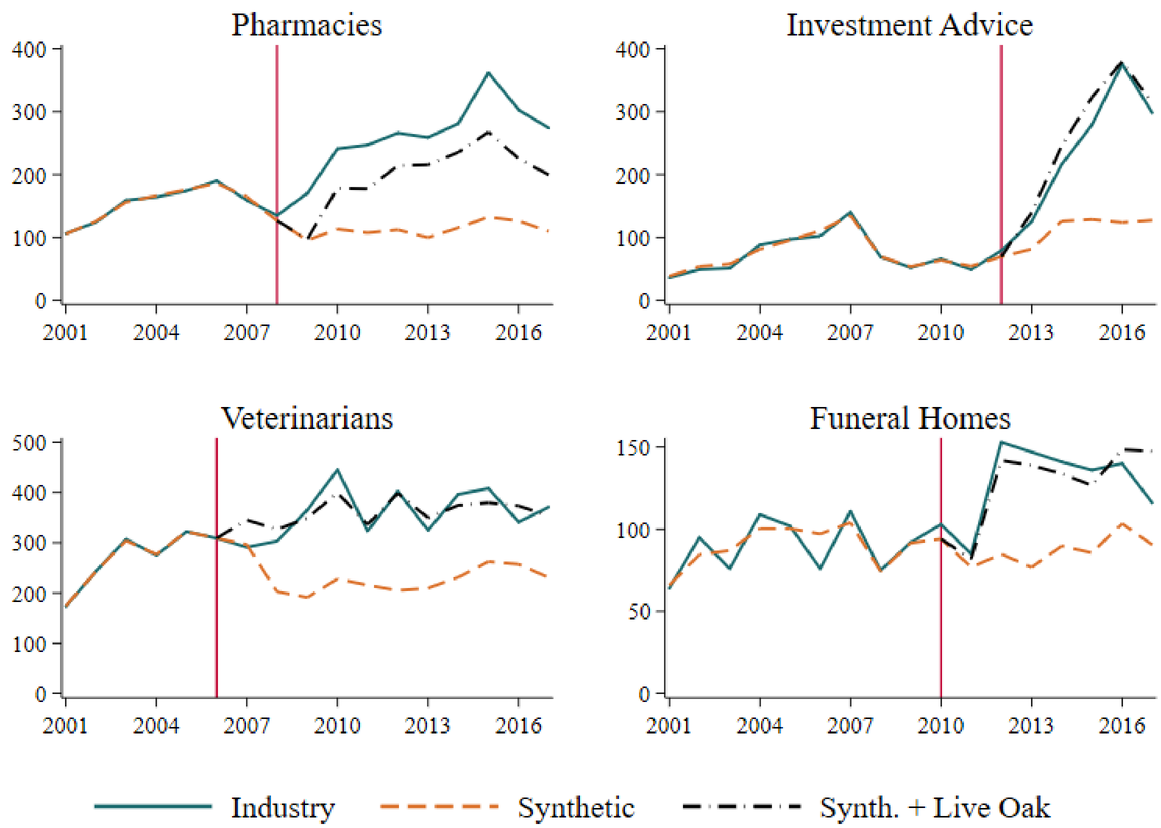


Figure 9: **Number of Loans - Treated Industry vs. Synthetic Control (excluding remote loans)** We exclude any loans from other remote lenders, defined as an institution-year observation with a median lending distance of more than 100 miles. This figure compares the number of loans in each industry that Live Oak enters to its synthetic control. The synthetic controls are formed by matching on all pre-treatment years beginning in 2001, with no additional covariates. The vertical line shows the year before Live Oak entered. The black dotted line "Synth. + Live Oak" adds the number of Live Oak loans to the outcome for the synthetic control.

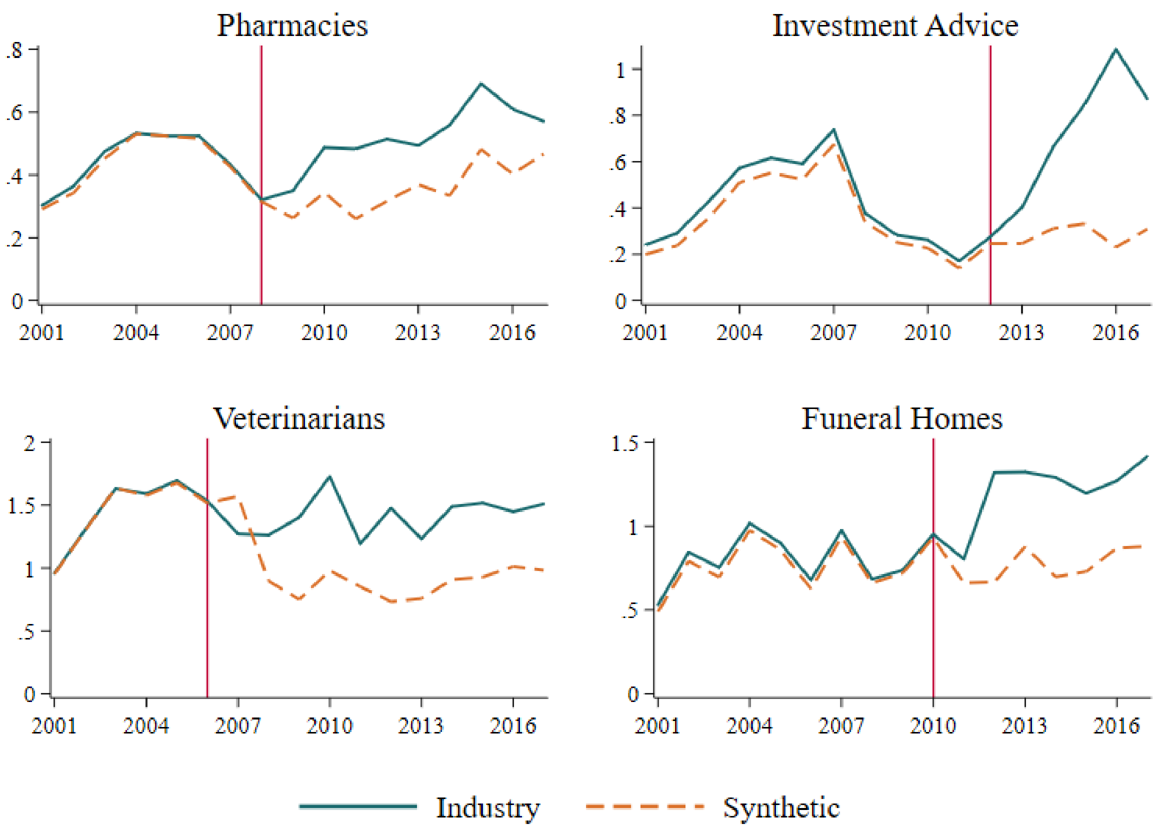


Figure 10: **Treated Industry vs. Synthetic Control: Loans per 100 Establishments** This figure estimates the synthetic control with loans per 100 establishments as the outcome. Establishment data are from the Quarterly Census of Employment and Wages.

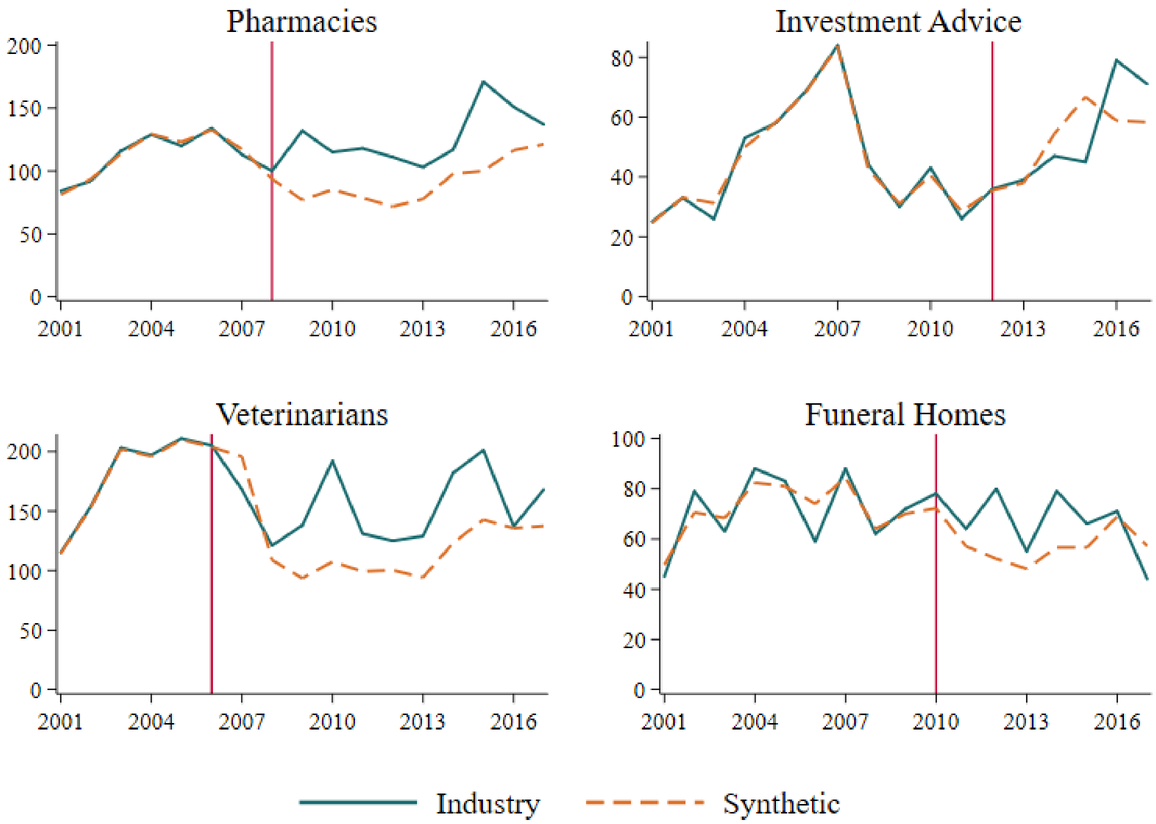
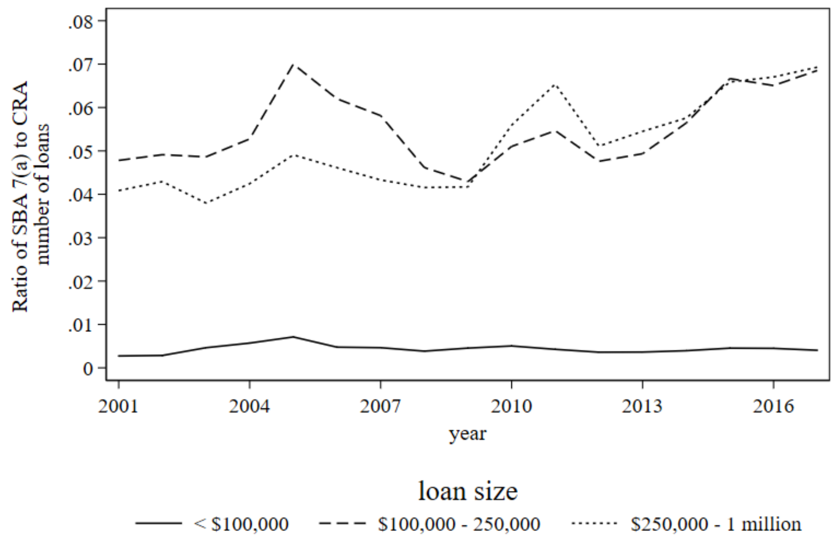
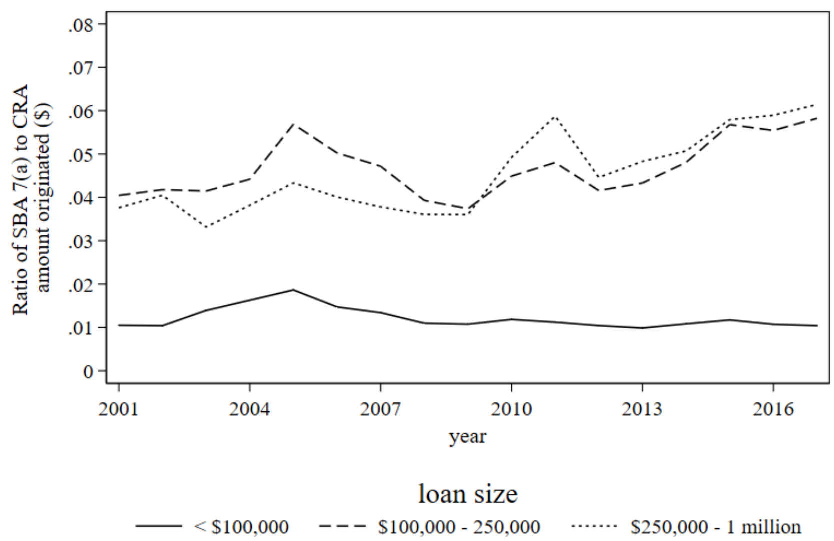


Figure 11: **Treated Industry vs. Synthetic Control in Zip Codes with Zero Live Oak Loans** This figure provides a falsification check by showing growth in loans to the treated industries in zip codes where Live Oak gave no loans. The two-sided p-value of the average effect on these four groups, computed using equation (5), is 0.483.

A Appendix Tables and Figures



(a) Number of Loans



(b) Loans in Dollars

Figure A.1: **Comparison of SBA 7(a) and CRA Loan Volume** These figures show the ratio of SBA 7(a) lending to CRA lending for the three loan size categories available in the CRA.

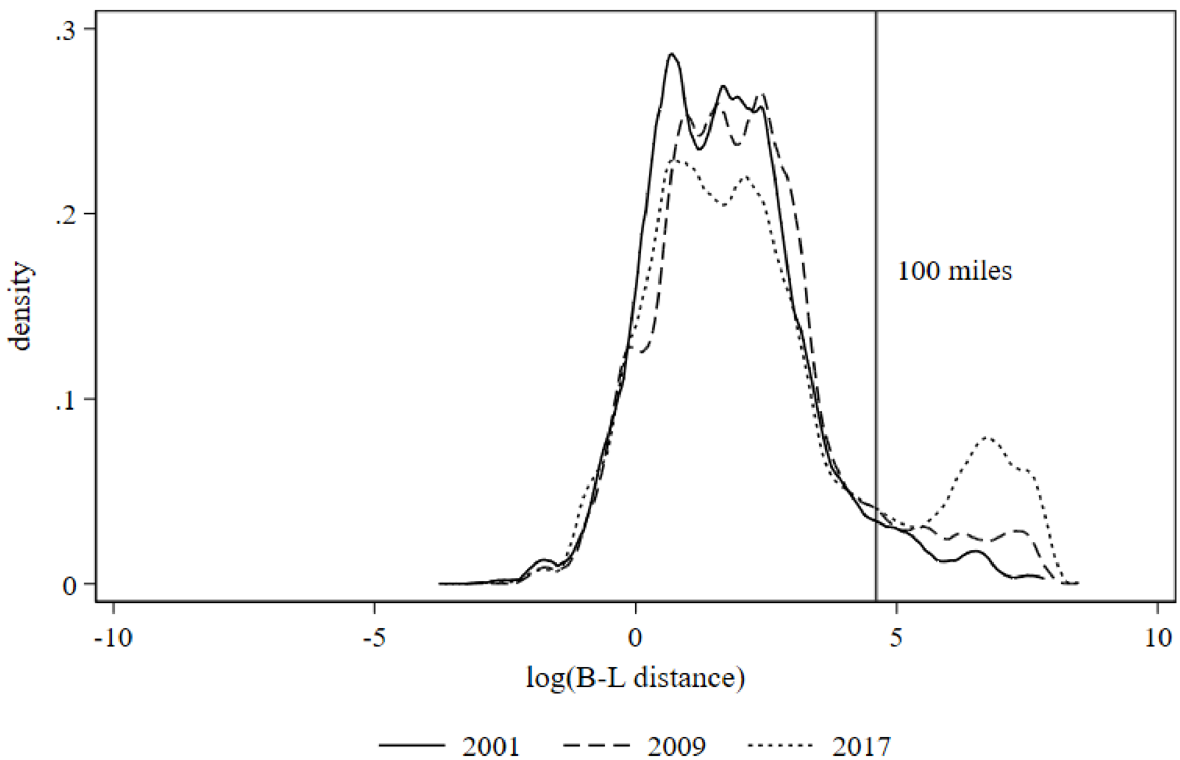
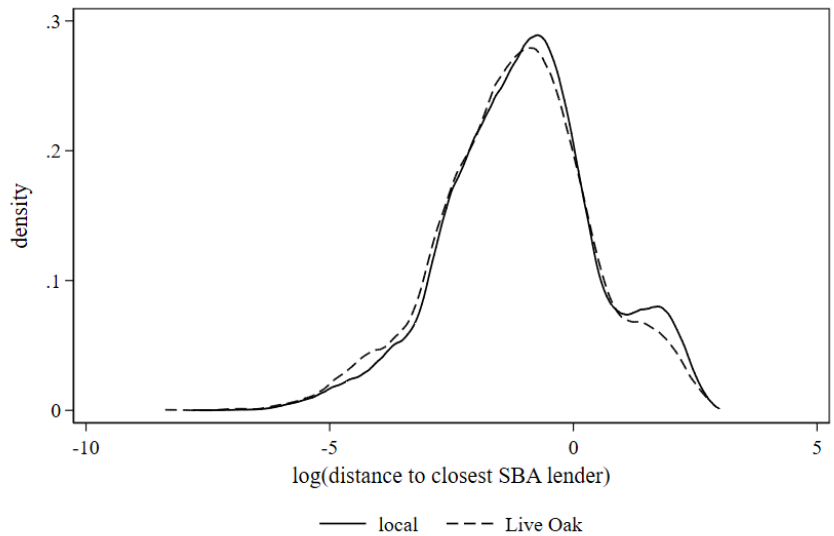
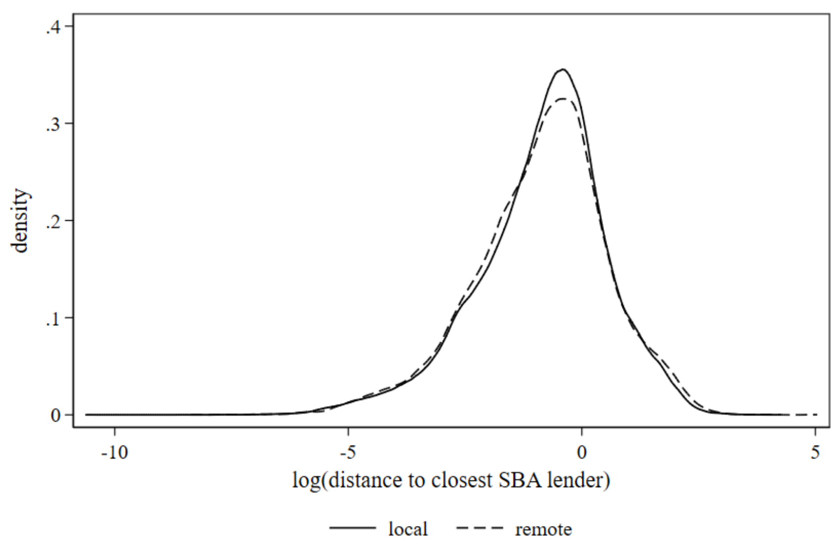


Figure A.2: **Distribution of (log) Borrower-Lender Distance for SBA Loans (County Measure)** This figure shows the distribution of the distance between borrowers and the closest branch of the institution from which they borrowed. Borrower-lender distance is calculated between the centerpoint of the project county and the closest branch according to the procedure described in Section 3.1.



(a) Comparison of local loans and Live Oak loans (in Live Oak industries)



(b) Comparison of local loans and remote loans

Figure A.3: **Distance to Closest SBA Branch** This graph shows the distribution of the distance between borrowers and the closest branch of any institution that grants SBA loans for SBA borrowers between 2007 and 2017. The first figure compares local loans (from a lender within 100 miles) to Live Oak loans for borrowers in the six treated industries. The second figure compares local loans to remote loans (from a lender more than 100 miles away). Distance is calculated according to the procedure described in Section 3.1, except it is the distance to the closest branch of any SBA lender.

Table A.1: **Lender Industry Concentration and Share of Distant Loans**

	Dependent variable: Bank's Top Five Share					
	(1)	(2)	(3)	(4)	(5)	(6)
Share 100+ mi.	0.185***	0.297***	0.155***	0.185***		
	(0.0357)	(0.0341)	(0.0294)	(0.0572)		
log(med. distance)					0.0457***	0.0207***
					(0.00493)	(0.00419)
Observations	4,930	4,930	4,930	1,602	4,930	4,930
Mean Dep. Var.	0.416	0.416	0.416	0.312	0.416	0.416
Year FE	X	X	X	X	X	X
Bank size ventiles		X	X	X	X	X
Bank FE			X	X		X
Balanced panel				X		

This table examines the correlation between the share of loans made at 100+ miles and the lender's industry concentration. Concentration is measured as the share of loans a lender makes its to its top five industries. Observations are at the lender-year level from 2007-2017 and standard errors are clustered at the lender level. The sample is restricted to lender-year observations with at least 10 loans. Bank size ventiles are ventile indicators for the number of SBA loans each year.

Table A.2: **Robustness: Lender Industry Concentration and Loan Performance**

	Dependent variable: Indicator for Charge-off within 3 Years					
	Small Lenders (1)	Medium Lenders (2)	Large Lenders (3)	Excluding Live Oak (4)	County Distance (5)	Lagged Industry Share (6)
$log(dist)$	0.00136***	0.00804***	0.00374***	0.00500***		0.00461***
	(0.000250)	(0.000603)	(0.000670)	(0.000365)		(0.000525)
Share in industry	-0.00509**	-0.0652***	-0.107	-0.0446***	-0.0382***	
	(0.00228)	(0.0119)	(0.233)	(0.00348)	(0.00308)	
$log(dist)$ (county measure)					0.00453***	
					(0.000370)	
Lag share in industry						-0.0398*
						(0.0205)
Observations	81,865	68,587	105,419	254,178	351,429	148,411
Industry FE	X	X	X	X	X	X
Year FE	X	X	X	X	X	X
Loan char.	X	X	X	X	X	X

This table examines the correlation between the lender's share of loans given to an industry (5-digit NAICS) and the share of loans charged off within three years. Observations are at the loan level from 2007-2014 and standard errors are clustered at the industry (5-digit NAICS) level. Small lenders are those that gave less than 100 loans in the year, medium lenders gave 100 to 1,000 loans in the year, and large lenders gave more than 1,000 loans in the year. The county measure of $log(dist)$ calculates the distance between the midpoint of the borrower's project's county and the closest branch. Lag share in industry is the lender's share of loans to the industry over years $t-2$ to $t-1$.

Table A.3: **Lender Industry Concentration and Loan Performance Excluding Distance**

	Dependent variable: Indicator for Charge-off within 3 Years			
	(1)	(2)	(3)	(4)
Share in industry	-0.0331***	-0.0239***	-0.00863**	-0.00914**
	(0.00320)	(0.00254)	(0.00391)	(0.00378)
Observations	389,548	389,548	389,548	389,548
Industry FE	X	X	X	X
Year FE	X	X	X	X
Loan char.	X	X	X	X
State-by-year FE		X		X
Lender FE			X	X

This table estimates specification (2) excluding $log(dist)$ as a control. Observations are at the loan level from 2007-2014 and standard errors are clustered at the industry (5-digit NAICS) level. Loan characteristics include dummies for ventiles of the size of the loan and the term length. The state in the state-by-year fixed effects is determined by the location of the borrower's business.

Table A.4: **Industries Comprising Synthetic Controls**.

Industry	**Synthetic Makeup**	**Weight**
Broilers and Other Meat Type		
	Chicken Egg Production	0.67
	Offices of Lawyers	0.33
Pharmacies and Drug Stores		
	All Other Miscellaneous Schools and Instruction	0.07
	Hazardous Waste Collection	0.04
	Homes for the Elderly	0.25
	Machine Shops	0.30
	Offices of Physical, Occupational and Speech Therapists, and Audiologists	0.28
	Other Direct Selling Establishments	0.00
	Photography Studios, Portrait	0.05
	Specialized Freight (except Used Goods) Trucking, Local	0.00
Investment Advice		
	All Other Miscellaneous Schools and Instruction	0.17
	Clothing Accessories Stores	0.08
	Cosmetics, Beauty Supplies, and Perfume Stores	0.05
	Direct Title Insurance Carriers	0.37
	General Freight Trucking, Long Distance, Truckload	0.04
	Offices of Mental Health Practitioners (except Physicians)	0.28
	Offices of Real Estate Agents and Brokers	0.01
Veterinary Services		
	Automotive Body, Paint, and Interior Repair and Maintenance	0.31
	Digital Printing	0.02
	General Automotive Repair	0.06
	Motion Picture and Video Production	0.42
	Offices of Lawyers	0.03
	Private Mail Centers	0.16
Offices of Dentists		
	Car Washes	0.25
	General Automotive Repair	0.33
	Offices of Lawyers	0.42
Funeral Homes and Funeral Services		
	Art Dealers	0.11
	Chicken Egg Production	0.46
	Cosmetics, Beauty Supplies, and Perfume Stores	0.03
	Hobby, Toy, and Game Stores	0.06
	Offices of Lawyers	0.12
	Other Marine Fishing	0.17
	Private Mail Centers	0.05

Table A.5: **County-Based Distance Measure: Industry Selection and Industry Expertise**

Sample:	Excluding Live Oak Loans				Loans to Six Industries Live Oak Entered			
Dependent variable:	Charge-off Indicator		Interest Rate (%)		Charge-off Indicator		Interest Rate (%)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
LO industry	-0.0134***		-0.00705					
	(0.00190)		(0.0110)					
$log(dist)$	0.00444***	0.00484***		0.0128***	0.00142***	0.00145***	0.0217***	0.0231***
	(0.000202)	(0.000209)		(0.00144)	(0.000444)	(0.000453)	(0.00474)	(0.00486)
LO industry$\times log(dist)$		-0.00262***		0.0287***				
		(0.000981)		(0.00582)				
Live Oak loan					-0.00675**	-0.00176	-0.399***	-0.213
					(0.00306)	(0.0151)	(0.0309)	(0.147)
Live Oak loan $\times log(dist)$						-0.000767		-0.0286
						(0.00227)		(0.0222)
Observations	348,905	348,905	229,850	229,850	15,562	15,562	12,796	12,796
Year FE	X	X	X	X	X	X	X	X
Loan char.	X	X	X	X	X	X	X	X
Industry FE			X	X	X	X	X	X

This Table repeats Table 5, but uses the distance measure constructed from the centerpoint of the borrower's county. The sample consists of loans originated between 2007-2014. Columns 1-4 exclude loans originated by Live Oak. Columns 5-8 restrict the sample to loans within the six Live Oak industries (including loans originated by Live Oak). The dependent variable is either an indicator for whether the loan was charged-off within three years of origination or the loan's interest rate (%). Interest rate data are available from 2008Q4. The sample is restricted to loans to the six treated industries described in Section 4.1. Loan characteristics include dummies for ventiles of the size of the loan and the term length. Industry fixed effects are indicators for the 5-digit NAICS code.

B Appendix: Matching Procedure

In this appendix, we describe the procedure used to construct a measure of borrower-lender distance.

B.1 Matching SBA Lenders to FDIC Summary of Deposits

The SBA 7(a) loan data contain the name and address of the institution that is currently assigned the loan. There are 5,815 institutions that gave out SBA loans between 2001 and 2017. For these institutions, we conduct a series of probabilistic matches using bank name, address, city, state, and zip code to link the SBA lending institutions to institutions in the 2017 FDIC Summary of Deposits. First, the matching procedure produces a match score between 0 and 1 based on the similarity of the text in the variables listed above, with more weight given to the bank name and address, since they are more likely to uniquely identify banks.[49] Of the 5,815 unique institutions, we find an exact match for 3,041. After checking for accuracy, we also count the roughly 800 institutions with a bigram match score greater than 0.98 as a match. For those with a score less than 0.98, we conduct a clerical review to determine whether the best match is accurate. After this first round of matching, we conduct a second round of matching and clerical review using different weights for the variables. We then manually match any unmatched institution that gave more than 100 SBA loans between 2001 and 2017 (provided that the institution is a bank and is not closed). Overall, we match 75% of the 5,815 institutions and these institutions provide 91.8% of SBA loans from 2001-2017. The majority of unmatched SBA institutions are credit unions or non-bank lenders, for which we do not have bank branch locations in the FDIC Summary of Deposit data, or they closed banks whose assets were transferred.

B.2 SBA Lenders' Branch Locations

Having matched banks in the SBA data to banks in the FDIC Summary of Deposits, we now construct historical branch networks. The FDIC Summary of Deposits contains annual counts and locations for bank branches from 1994-2017. For each matched SBA lender, we can therefore determine its branch locations at the time the loan was originated. The matches are imperfect, however, since the SBA 7(a) data contain the institution currently assigned the loan, rather than the institution that originated the loan. Bank closures, mergers, and acquisitions will generate differences between the banks currently assigned the loan and the bank that originated the loan. For example, BankBoston merged with Bank of America in 2004, and all of its branches were converted to Bank of America. Consequently, an SBA loan originated by BankBoston in 2001 may appear in the SBA data as currently held by Bank of America. To construct historical branch networks in light of these changes in bank structure, for each branch in each year from 2001-2017, we use the

[49]Specifically, we first standardize the bank names and addresses, then use reclink command in Stata. To assess similarity, reclink uses bigram comparison to score two strings based on the number of common 2-4 consecutive letter combinations. The first probabilistic match uses relative weights of 14 (out of 20) given to the name, 8 given to the address, 4 given to city, and 4 given to the zip code. The second match uses the same variables, but weights of 16,4,4, and 4. In both, we require state to match exactly.

FDIC's Reports of Structure Changes to determine the bank that holds that branch as of 2017. For example, we consider a branch to be a part of Bank of America's network if that branch is a Bank of America branch or would later become a Bank of America branch. That is, for a given year t, we consider a branch to be a part of an institution j's network in year t if that branch either (i) belongs to institution j in year t or (ii) would become a branch of institution j by 2017.

Another possible source of error is that banks may transfer loan assignments, even if there were no changes in bank structure. In order to gauge the error introduced by transfers of assignments, we compare the top 100 lenders in FY2012 from the 2012 Coleman Report to the top 100 lenders in FY2012 based on who is currently assigned the loan. These top 100 lenders provided 59% of all SBA loans and 60% of SBA volume in FY2012. Of the top 100 lenders, we are able to match 70 in our 2017 data. The unmatched banks are due to name changes, closures, mergers, and acquisitions between 2012 and 2017. Of the matched banks, the number of loans attributed to them in our data is very similar to the loans attributed to them in the 2012 Coleman Report (see Figure B.1), suggesting that absent changes in bank structure, banks rarely transfer the assignment of SBA loans.

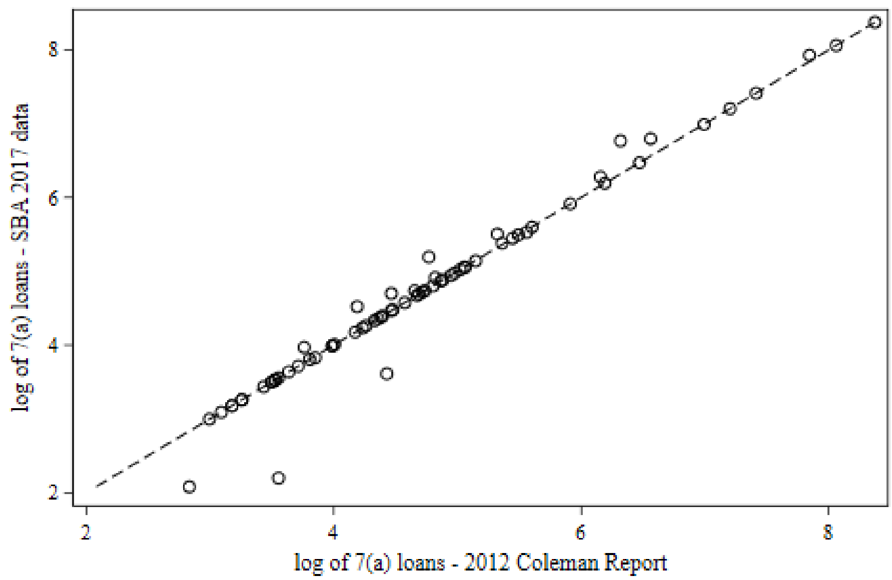


Figure B.1: Difference between counts at origination in 2012 and counts recorded in 2017

B.3 Borrower-Lender Distance

Starting with the 962,527 non-canceled SBA loans from 2001-2017 (and dropping the 179 that are missing industry info), we are able to match 885,166 to a lending institution in the FDIC Summary of Deposits. We then run these loans through the Census Geocoder, using the borrower's listed address, and are able to match 629,946 of the addresses to a latitude and longitude. Then, based on the borrower's institution and year, we match each borrower to the historical branch network for that institution.[50] Finally, we calculate the (Haversine) distance between the borrower and (i) the closest branch of the institution that originated the loan and (ii) the closest branch of any SBA lender.[51]

.

[50]We drop the 1.5% of branches that are missing longitude and latitude data.
[51]The Haversine distance, which is the shortest distance over the earth's surface.