Remote Competition and Small Business Loans: Evidence from SBA Lending

Wenhua Di and Nathaniel Pattison*†

September 21, 2018

PRELIMINARY, DO NOT CITE

Abstract

While traditionally small business loans were largely made by local banks through relationship lending, distances between borrowers and their lenders continue to increase. Much of this increase in distance is from online or "remote" lending, as technological advances allow small or even branchless banks to reach a national market. This paper examines the impact of competition from remote lenders and, in particular, their impact on the market for Small Business Administration (SBA) guaranteed loans. Using data on all SBA loans from 2001-2017, we document increases in remote lending activity and also show that many remote lenders concentrate lending within a few industries. We then investigate the impact of remote competition on SBA lending, exploiting the staggered entry of a large remote lender into specific industries. We compare post-entry

*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. Acknowledgments: We thank Benjamin Meier for his research assistance.

loan volumes in the entered industries to loan volumes in a synthetic combination of similar industries that the remote lender did not enter. The results suggest that entry generates significant growth in SBA lending, with little evidence of a reduction in loans made by incumbent SBA lenders. We then explore the characteristics of those who borrow from remote lenders. Geography plays some role, as the borrowers of remote banks are more likely to live farther from a brick-and-mortar branch of an SBA lender. Additionally, we find that remote lenders have a greater market share in counties where SBA lending has previously been low.

1 Introduction

Local lenders have traditionally dominated small business lending. However, as innovations in information technology and credit scoring reduce the benefits of proximity, the distance between small business borrowers and lenders has grown (DeYoung, Glennon, and Nigro, 2008; Petersen and Rajan, 2002). At an extreme, some banks operate largely online and make loans to a national pool of borrowers. The impact that these new remote lenders will have on credit markets and total credit availability is uncertain. This paper examines the effect of entry by remote lenders on small business lending, and in particular, their impact on the market for Small Business Administration (SBA) 7(a) loans.

SBA loans are relatively low-cost small business loans originated by approved lenders and partially guaranteed by the SBA. We first document two facts about the prevalence and characteristics of remote lending in the market for SBA loans. First, distance in SBA lending has increased, with a significant increase in the share of loans with a borrower-lender distance of more than 100 miles. Second, while the loan portfolio of traditional local banks tends to be concentrated in certain locations, many remote banks concentrate their lending within a few industries. Given this, we view local lenders as having an advantage in assessing soft information and local risk, while the focus of remote lenders on specific industries helps them develop expertise and an ability to assess industry-specific risk.

We then examine the impact of entry by a remote lender specializing in certain industries. A major concern is the degree to which these new lenders are taking market share from incumbents (Mills and McCarthy, 2016). Moreover, the expected impact of entry when lenders have different informational advantages (e.g., local vs. industry) is uncertain. On the one hand, better risk assessment may allow the new entrant to identify profitable but under-financed firms and extend them credit, thereby increasing total credit and output. On the other hand, if new entrants "cream-skim" the most profitable firms, it may harm the local banks and the firms that rely on them, ultimately reducing total credit and output. For example, Detragiache, Tressel, and Gupta (2008) and Gormley (2014) provide models where "cream-skimming" by new entrants can induce a segmented credit market that forces existing banks out, causing some profitable investment opportunities to go unfunded. These conflicting predictions lead to the central question of this paper: Does entry by a remote lender with industry-specific expertise increase or decrease the total volume of lending to that industry?

To examine the impact of remote competition, we exploit the staggered entry a large remote lender into specific industries. Live Oak Bank is currently the largest SBA lender (by the dollar amount of loans), but the majority of its loans go to only six industries. Between 2007 and 2014, Live Oak gradually entered these industries and gained substantial market share (12-58%) of SBA lending in each. Our empirical strategy compares changes in total lending to these "treated" industries to changes in lending to a group of control industries that Live Oak has not entered. For several reasons, including the impact of the Great Recession on small business lending, changes in industry composition, and the fact that Live Oak endogenously selects which industries to enter, it can be difficult to select an appropriate group of control industries. Instead of choosing control industries in an ad hoc way, we employ the Synthetic Control Method (SCM), an econometric technique developed in Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010), to systematically construct a synthetic match. For each treated industry, the synthetic match is a weighted combination of control industries, where the weights are chosen so that this combination closely matches the outcomes of the treated industry in the years prior to Live Oak's entry. Then, similar to a difference-in-difference specification, we compare changes between the treated industry and this synthetic control.

Our data consists of loan-level observations of all SBA 7(a) loans, which we aggregate by year and industry (5-digit NAICS code) to evaluate the impact of Live Oak's entry. One caveat is that we only observe SBA loans. If Live Oak's entry causes borrowers to substitute from non-SBA loans to SBA loans, we will not be able to detect the decline in credit for non-SBA loans. However, substitution from non-SBA loans is limited by the "credit elsewhere test" of the SBA 7(a) loan program. It requires that 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.

Our results indicate that the entry of Live Oak significantly contributed to the growth in SBA loans to these industries. There are sharp increases in lending to these industries in the years after Live Oak entered, relative to the comparison industries. To provide a sense of the magnitude, only 0.6-1.6% of the control industries experienced larger increases in lending than those that Live Oak entered. We then examine the extent to which the additional remote loans caused substitution away from existing lenders. We find little evidence that Live Oak's entry resulted in a decline in SBA lending to these industries from existing lenders. Relative to the synthetic control, total lending in the treated industries increases by roughly the amount as the number of new remote loans, implying little to no substitution from other SBA lenders.

Finally, we examine the locations of borrowers to determine whether remote lenders offer loans in areas local loans are less available. Relative to loans by traditional banks, remote borrowers are located in counties with fewer pre-entry SBA loans per capita and fewer branches of traditional banks. Brown and Earle (2017) documents that SBA lending tends to be concentrated around the physical branch locations of lenders who develop expertise in SBA lending, and the geographic distribution of these lenders is uneven and changes over time. One implication of our analysis is that remote lenders are expanding SBA's guaranteed loan program to new borrowers, some of whom are in areas with have not had as much SBA lending in the past.

This paper provides a case study of the effects of entry by a large, remote bank into specific markets. Given that our results are derived from a particular lender in the SBA 7(a) market, they may not easily generalize to broader settings. However, as we discuss in Section 2, there is an increasing number of banks adopting remote, industry-specific models similar to that of Live Oak Bank. Additionally, as argued in DeYoung et al. (2008), the operation of the SBA 7(a) market can shed some light on the operation of small business lending more generally. SBA lenders still face default risk, though it is partially offset by the government guarantee, and must screen borrower and set rates. DeYoung et al. (2008) and DeYoung, Frame, Glennon, and Nigro (2011) show that information asymmetries, borrower-lender distances, and credit scoring technologies play a role in SBA lending.

This research adds to the literature examining the role of distance in lending. Into the late 1990s, the median distance between a small business and its creditor was less than 10 miles (DeYoung et al., 2008; Petersen and Rajan, 2002). The prevailing explanation for the close distance between borrowers and lenders is that lenders are better able to assess the quality of firms that are physically closer to the bank branch. A set of theory papers examine the role of physical distance and information acquisition in banking competition (Dell'Ariccia and Marquez, 2004; Frankel and Jin, 2015; Gormley, 2014; Hauswald and Marquez, 2006; Rajan, 1992; Sharpe, 1990; Von Thadden, 2004). In these models, banks use private information about borrowers to create a threat of adverse-selection and limit competition from more distant lenders. Consistent with the link between physical distance and information, DeYoung et al. (2008) show that more distant SBA loans were more likely to default, and Agarwal and Hauswald (2010) provide evidence consistent with physical distance promoting informational capture. In their paper, they find that that physical proximity to a branch is related to increased loan approval rates, higher interest rates, and more predictive power from banks' subjective borrower assessments. While physical proximity continues to play a large role in small business lending (Nguyen, 2017), in the last two decades, distances between borrowers and lenders have grown rapidly, with the increases attributed to advances in information technology (DeYoung et al., 2011; Petersen and Rajan, 2002). Jagtiani, Lemieux, et al. (2016) show that large banks have increased small business lending in areas where they do not have branches between 1997 and 2014, consistent with technology facilitating distant lending. Our paper focuses on the extreme of distance in lending: online or remote banks. We contribute to this literature by examining, through a case study that allows for clean identification, the impact of competition from a remote bank.

Second, our paper is related to the literature examining competition by lenders with a cost or information advantage in a setting with information asymmetry. Much of this literature focuses on the entry of foreign lenders into developing countries. Like our setting, the foreign lenders are thought to have a lower ability to screen on local "soft" information, but an offsetting comparative advantage, perhaps an improved ability to process information along another dimension (in our case, this is industry-specific information). In this case, the entry of a competitor with a different advantage can either increase or decrease total lending and output. A new lender with an informational advantage can either deepen the credit market by identifying profitable but under-financed firms, or they can induce a segmented credit market in which some worthwhile investments go unfunded (Detragiache et al., 2008; Gormley, 2014). Empirically, papers have found evidence of both effects. In some cases, cream-skimming by foreign lenders results in reduced access to credit, particularly in less-developed countries (Beck and Peria, 2010; Detragiache et al., 2008; Gormley, 2010), while others find that entry causes credit to be cheaper and more widely available (Bruno and Hauswald, 2013; Claessens and Van Horen, 2014; Giannetti and Ongena, 2009,1). Our paper examines the impact of entry by a lender with a disadvantage in gathering local information, but an advantage in assessing firms' industry-specific risk.

And third, this paper complements the existing research on alternative or fintech lending. Fuster, Plosser, Schnabl, and Vickery (2018) looked at the mortgage market and found that non-bank lenders use technological innovation to improve the efficiency in loan processing and refinances and respond more elastically to demand shocks. Jagtiani and Lemieux (2017) examined loan account data from the Lending Club found that the lender originates loans in areas with less access to traditional credit, such as those that lost bank branches or have a more concentrated banking sector. The growth of fintech lenders may be attributed to alternative underwriting, more efficient data processing, regulatory arbitrage, and marketing. New technology and customer-based services may attract borrowers who cannot get loans or need special advice elsewhere and help them access to credit. Buchak, Matvos, Piskorski, and Seru (2017) shows that 30% of shadow bank growth in mortgage lending can be attributed to the convenience and efficiency enabled by on-line lending technology. By exploiting the staggered entry of a remote lender into specific industries, our paper contributes to this literature by estimating the causal effect of competition from a new alternative lender on credit availability.

2 Background Information

2.1 SBA Lending

Our setting for examining the impact of remote lending competition is the Small Business Administration's 7(a) loan program. The general-purpose loans can be used for small businesses that need working capital, with a maximum loan of $5 million. Through the 7(a) program, the government provides loan guarantees for credit-constrained small businesses that cannot obtain credit elsewhere on reasonable terms.[1] In addition to meeting the "credit elsewhere" requirement, SBA 7(a) borrowers must run a for-profit business that meets the SBA's industry-specific size standard. The SBA provides lenders with guarantees of up to 85 percent of the loan amount when borrowers default on the loan, and the exact guarantee amount depends on the loan balance and terms. The maximum guarantee is $4.5 million. The 7(a) program is the SBA's largest (65% of all SBA loans in 2017), and it is partly funded by guarantee fees paid by lenders, with a higher fee for larger loans. The SBA 7(a) loan approved loans increased from $3.83 billion in 2008 to $12.41 billion in 2009 and reached $25.45 billion in 2017.[2]

The capital for the loan is provided by SBA lenders, which are mostly commercial banks, though there are also credit unions and other financial intermediaries. 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 the level of authority

[1] 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.

[2] 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 the SBA grants the lender, the SBA either re-analyzes the lender's decisions or delegates those decisions to the lender. The Preferred Lender Program (PLP) status, which is used by the most experienced SBA lenders, allow the lender to make all underwriting and eligibility decisions. PLP lenders make over 80% of SBA 7(a) loans.

SBA lenders still face default risk, despite the government guarantee. The average guarantee is 64% in our 2001-2017 sample, so the guarantee is partial, and many SBA lenders sell the guaranteed portion and only retain the unguaranteed part. Additionally, the SBA reviews lenders' decisions and can increase monitoring if portfolio performance is weak. Finally, DeYoung et al. (2008) and DeYoung et al. (2011) provide empirical evidence of the importance of credit-screening, default, and information asymmetries in lending through the SBA program.

2.2 Measuring Borrower-Lender Distance

To analyze the geographic distribution of lending activity, we construct a measure of the distance between each SBA borrower and the closest branch of the institution from which (s)he borrowed. The SBA data contain the address of the borrower, but for the lender it lists the name and address of the institution currently assigned the loan (as of 2017). In order to link these institutions to branch networks, we first standardize bank 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 Summary of Deposits (SoD) using bank name, address, city, state, and zip code. Overall, we match 75% of the institutions, and these institutions provide 91.8% of SBA loans. Many of the unmatched institutions are credit unions or non-bank lenders, which are not in the FDIC data.

Then, using the FDIC SoD from 2001-2017, we construct historical branch networks for each SBA lender. The SBA 7(a) loan data contain the institution that is currently assigned the loan, so in cases of bank closures, mergers, and acquisitions, the bank currently assigned the loan could differ from the bank that originated the loan.[3] For example, BankBoston merged with Bank of America in 2004, and all of its branches were converted to Bank of America. An SBA loan originated by BankBoston in 2001 may appear in the SBA data as currently assigned to 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 FDIC's Report of Structure Changes to determine the bank that holds that branch as of 2017. 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.

We then convert borrowers' addresses to longitude and latitude coordinates using the Census Geocoder. We are able to geocode 71% of borrower addresses. Finally, we calculate the distance between each borrower and the closest branch of the institution from which he borrowed (using our constructed branch network for that institution during the year that loan was given).[4] Overall, we are able to construct a measure of borrower-lender distance for 65% of SBA borrowers, with slight increases in the match rate in more recent years.[5] Appendix B provides more details on the matching procedure. These matched data are used to construct distance measures reported in certain figures, but we use the full SBA data (matched and unmatched) in our main analysis of Section 3.

[3] In Appendix Figure B.1, we show that for banks that were not involved in a merger or acquisition, there were very few differences between institutions' loan counts at the time of origination in 2012 and the counts of institutions assigned the loan in 2017. This indicates that the errors between the institutions that originate loans and those that are currently assigned the loans will come from changes in bank structures, rather than transfers of assignments across banks with no changes in structure.

[4] We calculate the Haversine distance, which is the shortest distance over the earth's surface. The FDIC SoD data contains longitude and latitude coordinates for the large majority of branches over this period, so we did not need to geocode branch addresses.

[5] The 2001-2005 match rate is 63.5%, while the 2011-2015 match rate is 68%.

2.3 Distance in Lending

Small business lending, including SBA lending, has historically been conducted through local banking relationships. Information about small businesses can be hard to communicate and may be best gathered through relationship banking and repeated interactions. Consequently, small business lending has been dominated by community banks lending to nearby borrowers. Into the late 1990s, the median distance between a small business and its creditor was less than 10 miles (DeYoung et al., 2008; Petersen and Rajan, 2002). However, in the last two decades, distances between borrowers and lenders have grown rapidly, with the increases attributed to advances in information technology (Petersen and Rajan, 2002). DeYoung et al. (2011) show that the borrower-lender distances accelerated rapidly in the mid-1990s, and a significant share of the acceleration is related to banks' adoption of credit scoring technologies. Although the Internet and technological innovations have made it easier to lend to distant borrowers, distance continues to be important (Agarwal and Hauswald, 2010; Nguyen, 2017). DeYoung et al. (2008), using SBA data, show that more distant loans were more likely to default.

Using loan-level observations of all 7(a) loans guaranteed by the SBA from 2001-2017, we provide some descriptive statistics on recent changes in borrower-lender distance for SBA loans, where distance is determined using the procedure described in Section 2.2. Figure 1 plots the distribution of borrower-lender distances for SBA loans for three years: 2001, 2008, 2017. The figure reveals two striking features. First, changes in the right tail of the distributions show that much of the change in borrower-lender distances is from an increased number of loans with 100 or more miles between the borrower and lender.[6] This rise of remote lending can also be seen by looking at the largest lenders. For fiscal year 2016, four of the

[6] By using the distance between the borrower and the closest branch of the lender, we may underestimate increases in distance. For example, borrowers may not borrow from the closest branch, or banks with large branch networks may make lending decisions out of a centralized location.

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. Second, Figure 1 also shows that there is still a large local component to lending. Even in 2017, for 71% of loans, the distance between the borrower and the closest branch of the lender was less than 10 miles. Overall, this is consistent with what DeYoung et al. (2011) found emerging in the late 1990s; there were large increases in borrower-lender distance among certain banks (those that adopt credit scoring technologies), while there was relatively little change for the majority of banks. The goal of this paper is to examine the impact that these changes in the right tail of the distribution, i.e., the rise of distant, remote lenders, have had on credit markets.

2.4 Industry Specialization and Discussion of Live Oak Bank

While traditional local banks specialize in lending within a certain geographic area, many remote lenders tend to specialize in lending to particular industries. Figure 2 shows the relationship between the number of branches of a bank and a measure of the industry concentration of the lender's loans. Each circle corresponds to an SBA lending institution, with the size of the circle reflecting the total amount of loans between 2014 and 2017. The industry concentration is measured with a Herfindahl-Hirschman Index (HHI), which for institution j is defined as $HHI_j = \sum_i S_{ij}^2$, where S_{ij} be the percentage (0-100) of institution j's loan volume given to industry i, where industries are measured at the level of the 5-digit NAICS code. This measure is increasing in industry concentration and takes a value from close to 100 (least concentrated) to 10,000 (most concentrated). The figure shows that lenders with the most branches tend to diversify their SBA lending across many industries, while lenders with few branches are more concentrated. Many of the most concentrated lenders have significant remote lending activity; the red circles show institutions for which at least twenty percent of their loans have a borrower-lender distance of more than 100 miles.

Industry-specific lenders have existed in SBA lending for decades. For example, Vision One Credit Union, founded in 1951, provides loans for private optometry practices and several banks almost exclusively fund agricultural loans. However, many of the most concentrated lenders are relatively new. 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 physicians offices.

The largest of these specialty lenders is Live Oak Bank. This national bank has no branches, but is currently the largest SBA lender by volume. Our empirical strategy examines the impact of entry by Live Oak into specific industries in order to examine the impact of competition by remote lenders on lending markets. Beginning in 2007 with veterinarians, Live Oak has gradually entered 20 industries after strategically identifying each industry's demand for SBA 7(a) financing, the competitive environment, and acquiring industry loan experts and industry expertise. Live Oak partners with third parties to provide efficient loan approval and comprehensive services and products enabled by digital banking platforms. The bank also goes beyond streamlining the underwriting and approval process with technologies and makes time to meet small business customers to help them succeed with the industry expertise of the bank's personnel.

3 Empirical Strategy

Our empirical analysis is motivated by the two figures in Section 2. Figure 1 suggests that an increasing number of loans are being made from banks located far from the borrower, i.e., remote lending is increasing. Using our distance measure, the number of SBA institutions with a median lending distance of at least 200 miles has increased from less than 10 in 2001 to more than 40 in 2017. Figure 2 shows that these remote lenders also tend to give loans to a more concentrated set of industries, which may allow them to better assess industry expertise. Indeed, the remote lender we focus on in our empirical analysis, Live Oak Bank, lists industry expertise as one of their key advantages.

Given the industry concentration of many remote lenders, remote competition can be viewed as the entry of a competitor with an informational advantage over local lenders in assessing firms' industry-specific profitability. The impact of entry by such lenders, most commonly examined in the context of entry by foreign banks into domestic markets, generates ambiguous predictions. On the one hand, an informational advantage may allow new entrants to identify profitable but under-financed firms and extend them credit, thereby increasing total credit and output. On the other hand, some worry that if new entrants are able to identify and lend to the most profitable firms, it may harm the local banks and the firms that rely on them, ultimately reducing total credit and output. Illustrating the latter case, Detragiache et al. (2008) and Gormley (2014) provide models where "cream-skimming" by new entrants can induce a segmented credit market that causes a reduction in total lending and output. Given these conflicting predictions, we examine the impact of entry by a remote lender empirically.

3.1 Entry of Live Oak Bank

The difficulty in estimating the impact of remote lending competition is that we do not observe the counterfactual number of loans that would have been extended without remote competition. Our empirical strategy attempts to overcome this challenge by examining a case study: the entry of Live Oak Bank into specific industries. The advantage of this approach, as we discuss below, is that we can exploit Live Oak's staggered entry into these specific industries, using the non-entered industries as a comparison group. Live Oak is currently the largest SBA lender by volume, yet has given the majority of its loans out to only a small set of industries. The bank describes its expertise in these industries as a key advantage that it has over other lenders. Beginning in 2007 with veterinarians, Live Oak has gradually added to the industries in which it operates. Table 1 presents the industries where Live Oak has given out at least 50 SBA loans as of 2017, along with the number of loans, Live Oak's post-entry share of SBA loans and share of loan volume 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 total loan amount is even greater, since they tend to give larger than average loans.

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 extended loans because they either entered in mid-2015, so there is a relatively short post-period, or Live Oak makes up a such a small share of loans to that industry that is 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. In doing so, we assume that Live Oak's entry into the treated industries does not have spillover effects on lending to other industries. This would be violated if, for example, banks respond to Live Oak's entry in one industry by lending more to other industries or if growth in one industry spurs growth in another. Given that most lenders provide lending to many industries, we expect that Live Oak's entry into one of them is unlikely to have significant spillover effects on overall lending practices.

3.2 Data and Construction of 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.[7] 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, we drop the industries that Live Oak has given at least one loan to but are not in our set of six treated industries. So that we can compare loan originations over time, 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 310 control industries and the six treated industries that Live Oak has entered.

3.3 Synthetic Control Method

We examine the change in total SBA loans to firms in the industries that Live Oak enters, relative to the change in a group of control industries. 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 select industries that can serve as a suitable comparison group. Instead, we select comparison industries using the synthetic control method (SCM), developed by Abadie and Gardeazabal (2003) and Abadie et al. (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). The synthetic match is a weighted combination of the control industries (i.e., industries that Live Oak never enters), where the weights are chosen so that the pattern of loan volumes for the synthetic control closely matches that of the treated industry during the pre-treatment

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

period.

Formally, following the setup of Abadie et al. (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 enter 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 (entry), and $Y_{1t}(0)$ be the potential outcome without treatment. We want to know 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 $t > T_0$. Since we only observe $Y_{1t}(1)$ for the treated industry, the treatment effect requires an estimate of the counterfactual $Y_{1t}(0)$. Assume the potential outcomes for all industries i follow the factor model

$$Y_{it}(0) = \delta_t + \lambda_t \mu_i + \varepsilon_{it} \tag{1}$$

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 outcome of control industries equals the outcome of the treated industry for all pre-treatment periods:

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

As an estimator of the treatment effects τ_{1t} for $t > T_0$, Abadie et al. (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 (2) 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:

$$\{w_i^{j*}\}_{j\in\text{Control}} = \underset{\{w_i^j\}_{i\in\text{Control}}}{arg\,min} \sum_{t\leq T_0^j} [Y_{jt} - \sum_{i\in\text{Control}} w_i^j Y_{it}]^2$$

$$s.t. \sum\nolimits_{i\in\text{Control}} \quad w_i^j = 1$$

$$\text{and} \qquad w_i^j \geq 0 \quad \forall i,$$

where Y_{it} is the number of SBA loans given to industry i during year t. That is, we choose weights to minimize the mean square error of outcomes between the treated industry and the synthetic control during the pre-treatment period.[8] 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. For each treated industry j, we find the optimal weights then construct the corresponding synthetic controls 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.

The synthetic control method has several advantages over difference-in-differences estimators. It provides a data-driven, objective method of choosing control industries. By comparing pre-treatment fit, the method also provides a convenient way to assess the suitability of the comparison group. Moreover, the identification assumptions are weaker than those in

[8] 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 et al. (2010). As discussed in detail in Kaul, Klößner, Pfeifer, and Schieler (2015), this is equivalent to the minimization procedure above.

a standard difference-in-differences model. The model in equation (1) generalizes difference-in-differences models, which require λ to be constant over time (industry fixed effects) and impose specific time trends (e.g., year fixed effects). In addition to allowing these controls, equation (1) also allows industry-specific loadings to unobserved, time-varying factors ($\lambda_t \mu_i$).

While the identification assumptions are weaker than difference-in-differences, our empirical strategy still relies on the assumption that potential outcomes for all industries follow the factor model in equation (1). The key identification assumption is that the exact timing of entry by Live Oak into a specific industry does not coincide with other changes affecting the pattern of growth. For example, we assume that Live Oak does not enter specific industries because they anticipate abnormal future growth or a break from pre-existing trends. We support this assumption in four ways.

First, as mentioned, the synthetic control method allows for time trends and a fixed number of unobserved factors with loadings that can vary across industries. To the extent that the determinants of Live Oak's entry are reflected in these variables, we will be controlling for them. Second, Live Oak describes their entry decisions as depending on industry research, evaluation of payment levels, the current competition, and, most importantly, the ability to find a domain expert. The timing of entry depends on their ability to acquire the necessary expertise, and we have not found any evidence that they time entry based on anticipated unusual growth. Third, using the exact timing of Live Oak's entry, we argue, will limit bias due to unobserved factors affecting both entry and growth. Entry is a large and discrete change to the lending market in the industry, with Live Oak providing a significant share of the new loans. As long as the impact of this shock is large relative to the conditional variance of omitted factors that are correlated with entry and affect growth, the bias will be limited.[9] Fourth, as a falsification check, we examine changes in loans in the treated industries given to borrowers living in areas where Live Oak did not provide any loans. If our effects were driven

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

by changes to industry growth, rather than the entry of Live Oak, we would expect lending to these industries to increase, even where Live Oak did not operate. Alternatively, if our identification assumption is correct, the increased lending is due to Live Oak, so we would expect little change in lending where no Live Oak loans were given. Consistent with our identification assumption, we find small and insignificant changes to lending in the treated industries in locations where Live Oak gave no loans. A final concern is that other remote lenders may target the same industries as Live Oak. We address this in a robustness check by excluding loans from other remote lenders when constructing the sample.

4 Results

4.1 Main Results

The synthetic control results for the six industries that Live Oak enters are presented in Figure 3. We are unable to construct a good match for two of the industries. "Broilers" has a poor fit throughout the pre-period and the synthetic control for "Dentists" was already declining in 2008, prior to the entry of Live Oak. Consequently, we focus our analysis and discussion on the remaining four industries for which we are able to construct a good synthetic control match.[10] Appendix Table A.1 shows the industries that make up the synthetic controls. These industries are chosen to match the number of SBA loans given to the treated industries during the years prior to Live Oak's entry.

For the remaining four treated industries (Pharmacies, Investment Advice, Veterinarians, and Funeral Homes), the figure shows a good synthetic control match during the pre-treatment period. Relative to the synthetic control, all four industries show sharp and persistent increases in lending once Live Oak enters. We evaluate the statistical significance

[10] As discussed in Abadie et al. (2010), one should not use the synthetic control method when there is not a good pre-treatment fit for the treated unit.

of the increase in loans to industry i by estimating synthetic controls for each of the 310 control industries, assuming a placebo treatment in the same year that Live Oak entered industry i. Figure 4 plots the "gap" or difference between the number of loans for each treated industry and its synthetic control. We discard observations with poor pre-treatment fits, defined as having a pre-period mean squared prediction error (MSPE) of more than $\sqrt{3}$ times that of the treated industry.[11] In all four cases, the industry that Live Oak entered experienced increases relative to its synthetic match that were large relative to the distribution of placebo increases. The share of estimated placebo effects larger than the true treatment effect varied from 0.6-1.6% across the four treated industries.

We then evaluate the 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, Johnson, Kermani, Kwak, and Mitton (2016), we construct the test statistic

$$\widehat{\theta} = \sum_{j \in \text{Treat}} \left(\frac{\sum_{t=T_0^j+1}^{T} \frac{Y_{jt} - \widehat{Y}_{jt}(0)}{(T-T_0^j)} \Big/ Y_{jT_0^j} \widehat{\sigma}_j}{\sum_{j \in \text{Treat}} \frac{1}{\widehat{\sigma}_j}} \right) \tag{3}$$

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-

[11] The pre-treatment MSPE for industry j is defined as $\sum_{t \leq T_0^j} [Y_{jt} - \sum_{i \in \text{Control}} w_i^{j*} Y_{it}]^2$, where Live Oak entered the industry in year $T_0^j + 1$.

treatment year ($Y_{jT_0^j}$), and weighted by a measure of the quality of fit in the pre-treatment period ($\frac{1}{\hat{\sigma}_j}$). 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 (3). Figure (5) 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 placebo treatment effects are larger in absolute value than the actual average treatment effect, indicating that the magnitude of the loan increases to the treated industries is large relative to what would be expected from chance.

4.2 Substitution from Other SBA Lenders

The main results show that the entry of Live Oak caused an 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. By comparing the estimated increase in lending to the actual number of loans that Live Oak provided to industry i in each year, we can assess the degree to which Live Oak lends to new borrowers or simply diverts SBA borrowers who would have obtained loans from other lenders. If entry generates no substitution away from other lenders, then the total loan volume in the industry would be

$$\widehat{Y}_{it}^{NoSub} = \widehat{Y}_{it}(0) + LiveOak_{it}$$

where $\widehat{Y}_{it}(0)$ is the estimated counterfactual number of loans with no entry and $LiveOak_{it}$ is the number of loans Live Oak gave to industry i in year t. To evaluate the degree of substitution from existing lenders, we can compare $\widehat{Y}_{it}^{NoSub}$ to the actual number of loans Y_{it}. If $\widehat{Y}_{it}^{NoSub} \approx Y_{it}$, it would suggest that there was little substitution or business-stealing from existing SBA lenders and that remote entry only expanded the SBA market to new borrowers. On the other hand, if $\widehat{Y}_{it}^{NoSub} > Y_{it}$, it would suggest that entry caused a reduction in loans from existing lenders in addition to the expansion of the total number of loans.

Figure 6 shows the actual industry, synthetic control, and $\widehat{Y}_{it}^{NoSub}$ (labeled "Synth. + Live Oak") estimates for the four matched treatment industries. The actual number of loans is very similar to, or even above, the number of loans that would have been given out if there were no substitution away from existing lenders. This suggests that the large majority of Live Oak's loans were given to borrowers who would not have received an SBA loan otherwise. There is no indication that the entry of Live Oak caused a reduction in other SBA lending to these industries. Additionally, while the magnitudes are not statistically significant, the fact that $Y_{it} > \widehat{Y}_{it}^{NoSub}$ for three of the industries suggests the possibility of spillover effects. That is, remote entry increases total lending to that industry beyond simply the loans that the remote lender extends. In the next section, we examine one possible explanation for this: whether the increases can be explained by additional loans from other remote lenders.

4.3 Robustness

The results above indicate that the entry of Live Oak resulted in an increase in total SBA lending, with little substitution away from existing SBA lenders. In this section, we examine two possible concerns with this interpretation. First, it is possible that other lenders targeted the same industries as Live Oak, so our estimates are picking up the effect of both Live Oak's entry and the subsequent entry of other lenders. Since other remote lenders would be most easily able to increase their lending to specific industries, we investigate this concern by dropping other remote lenders' loans from the sample. We define remote lenders as those whose median lending distance in the year was more than 100 miles.[12] Figure 7 shows the results. Increases in total lending still occur across all four industries. Moreover, the size of the increase more closely tracks with the amount expected if there were no substitution ("Synth. + Live Oak"). This provides additional evidence that entry by Live Oak increased total SBA lending to these industries without causing substitution from existing SBA lenders.

A second concern is that Live Oak enters industries which are about to break from trend and deviate from the model proposed in equation (1). To test this, we examine whether loans to the treated industries increased even in areas where Live Oak gave no loans. If the increase in lending activity is a result of Live Oak loans, we should see not see an increase in areas where Live Oak gave no loans. Alternatively, if our estimates are driven by overall growth in these industries, we would expect to see increases in lending to these industries even in areas where Live Oak gave no loans. In this exercise, 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 8 shows the results. Although the actual number of loans is above the synthetic control in some of the industries, the magnitude of the increase is much smaller than the main estimates in Figure 3. Using equation (3) to calculate the average treatment effect in areas with no Live Oak loans, and comparing it to the placebo distribution in Figure 5, 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. This is strong evidence that we are capturing the impact of the entry of Live Oak, rather

[12] To compute distance, we use the measures discussed in Section 2.2. We allow a bank to be a remote lender for some but not all years if there are years where their median lending distance is both above and below 100. In this case, we only drop loans from the bank during the years where the median distance is above 100. We explored several other definitions, and the results of this section are not sensitive to exactly how we define remote lending.

than heterogeneous trends in the industries that Live Oak chose to enter.

5 The Locations of Remote Borrowers

Section 3 shows that entry by a large remote lender into specific industries increased the total number of SBA loans. It is possible that remote lenders increase total lending because they expand access to the program geographically. Brown and Earle (2017) shows that access to the SBA lending program depends in part on physical proximity to a lender that offers SBA loans. In this section, we examine whether remote lenders serve borrowers that are located farther from or have less access to the SBA program through traditional banks.

Using our measures of borrower and branch locations described in Section 2.2, Figure 9 shows the distribution of the distance between the borrower and the closest branch of an SBA lender (not necessarily the lender from which the borrower obtained a loan). It shows these distributions for borrowers who ultimately obtained loans from Live Oak, some other remote lender, or a traditional bank. Here, we classify other remote lenders as banks with a median borrower-lending distance for their loans of at least 100 miles. The figure shows that both Live Oak and other remote lenders are more likely to lend to borrowers located farther a brick-and-mortar SBA lender, since more of the mass of their distributions are in the right end. However, for all three types of lenders, many loans go to borrowers located within a few miles of the nearest branch of an SBA lender. Thus, while remote lenders are more likely to lend to people located farther from a physical SBA branch, they also lend to many people living close to one. Moreover, the figure reveals that almost all borrowers live very close to an SBA lender. More than 95% of borrowers are within 5 miles of the closest branch of a bank that has granted SBA loans.

Physical distance does not fully capture the availability of SBA lending. We construct three additional measures of county-level geographic variation in lending. First, we compute the annual average number of SBA loans per capita and SBA loan volume per capita from 2000-2007 (prior to the entry of Live Oak and many remote lenders). This provides a measure of SBA lending in an area prior to the entry of many remote lenders. Second, we similarly construct per capita loans and volumes, but exclude any loans given by remote lenders. As above, we define remote lenders as banks where the median borrower-lender distance for their loans is at least 100 miles. This serves as a measure of lending by local banks. Finally, using the FDIC data, we construct county-level measures of bank branches per capita, using 2016 branch locations. During this period, the average county-level market share of remote lenders was 16.6% of loans and 22.5% of loan amounts. There are substantial differences in the market share of remote lenders. There were no remote loans 37% of counties between 2014 and 2017, while in 10% of counties remote lenders originated more than half of all SBA 7(a) loans.

Using the various measures of county-level access to lending, we estimate the following specification:

$$S_c = \alpha + \beta X_c + \varepsilon_c, \tag{4}$$

where S_c is the 2014-2017 market share of remote lenders to borrowers in county c, X_c is the county-level measure off access to non-remote SBA lending, and ε_c is the error term. This specification is similar to that in Buchak et al. (2017), which examines the geographical determinants of mortgage lending by shadow banks and fintech lenders.

Table 2 Panel A shows the results. Across all specifications, remote lenders have a higher market share in counties with less access to non-remote SBA lending. The coefficient in Column 1 indicates that, in counties where past (2000-2007) SBA loans per capita are 10% lower, the 2014-2017 market share of remote lenders increases by 0.65%. The coefficient is similar in Column 2, where the past loans per capita measure excludes loans made by remote lenders. Similarly, Column 3 shows that remote lenders have a higher market share in areas with fewer bank branches per capita. Columns 4-6 show that the results also hold when measure market share using loan amount, rather than the number of loans. Panel B of Table 2 shows a similar pattern for Live Oak, though not for the branches per capita measures. These results indicate remote lenders have the greatest market share in counties that in the past have had fewer SBA loans originated. This suggests that at least part of the growth in SBA lending is to areas that, in the past, have had less SBA lending.

6 Conclusion

The geographic distance between borrowers and lenders within the market for SBA 7(a) loans continues to increase, and much of the increase is due to remote lending activity from online banks. Many remote SBA lenders also specialize in lending to certain industries, perhaps acquiring industry expertise in assessing risk and assisting borrowers. This paper examines the competitive impact of these 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 SBA lenders. Consistent with this, we provide some evidence that remote lenders expand the SBA program geographically; they are more likely to lend to borrowers farther from brick-and-mortar branches, and the remote lending market share is highest in areas where SBA lending was less common. Live Oak Bank and other remote lenders have expanded the SBA 7(a) program to new borrowers and altered the industry composition of lending.

One question we cannot directly investigate is whether the entry of Live Oak caused a substitution away from non-SBA lenders. We do not observe whether non-SBA lending to these industries declined during this period. However, the ability of borrowers to switch between SBA and non-SBA lending is limited. The "credit elsewhere test" of the 7(a) program requires banks to certify that they would be unwilling to make loans outside of the SBA program and that they believe the borrower could not get other loans on reasonable terms. Additionally, the bank we focus on views other SBA lenders as its closest substitutes. 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." Since we observe no substitution from SBA lenders, and substitution from non-SBA lenders is limited, the SBA market expansion likely reflects an increase in total credit. Over time, it is possible that local and remote lenders, as well as non-SBA lenders, can use improved underwriting, technology, and industry expertise to increase the supply of credit to small-business owners.

References

Alberto Abadie and Javier Gardeazabal. The economic costs of conflict: A case study of the Basque country. *American Economic Review*, 93(1):113–132, 2003.

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. 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, 2010.

Daron Acemoglu, Simon Johnson, Amir Kermani, James Kwak, and Todd Mitton. The value of connections in turbulent times: Evidence from the united states. *Journal of Financial Economics*, 121(2):368–391, 2016.

Sumit Agarwal and Robert Hauswald. Distance and private information in lending. *The Review of Financial Studies*, 23(7):2757–2788, 2010.

Thorsten Beck and Maria Soledad Martinez Peria. Foreign bank participation and outreach: Evidence from mexico. *Journal of Financial Intermediation*, 19(1):52–73, 2010.

J David Brown and John S Earle. Finance and growth at the firm level: Evidence from sba loans. *The Journal of Finance*, 72(3):1039–1080, 2017.

Valentina Bruno and Robert Hauswald. The real effect of foreign banks. *Review of Finance*, 18(5):1683–1716, 2013.

Greg Buchak, Gregor Matvos, Tomasz Piskorski, and Amit Seru. Fintech, regulatory arbitrage, and the rise of shadow banks. Technical report, National Bureau of Economic Research, 2017.

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

Giovanni Dell'Ariccia and Robert Marquez. Information and bank credit allocation. *Journal of Financial Economics*, 72(1):185–214, 2004.

Enrica Detragiache, Thierry Tressel, and Poonam Gupta. Foreign banks in poor countries: theory and evidence. *The Journal of Finance*, 63(5):2123–2160, 2008.

Robert DeYoung, Dennis Glennon, and Peter Nigro. Borrower-lender distance, credit scoring, and loan performance: Evidence from informational-opaque small business borrowers. *Journal of Financial Intermediation*, 17(1):113–143, 2008.

Robert DeYoung, W Scott Frame, Dennis Glennon, and Peter Nigro. The information revolution and small business lending: The missing evidence. *Journal of Financial Services Research*, 39(1-2):19–33, 2011.

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

Andreas Fuster, Matthew Plosser, Philipp Schnabl, and James Vickery. The role of technology in mortgage lending. Technical report, National Bureau of Economic Research, 2018.

Matthew Gentzkow, Jesse M Shapiro, and Michael Sinkinson. The effect of newspaper entry and exit on electoral politics. *American Economic Review*, 101(7):2980–3018, 2011.

Mariassunta Giannetti and Steven Ongena. Financial integration and firm performance:

Evidence from foreign bank entry in emerging markets. *Review of Finance*, 13(2): 181–223, 2009.

Mariassunta Giannetti and Steven Ongena. “lending by example”: Direct and indirect effects of foreign banks in emerging markets. *Journal of International Economics*, 86(1):167–180, 2012.

Todd A Gormley. The impact of foreign bank entry in emerging markets: Evidence from india. *Journal of Financial Intermediation*, 19(1):26–51, 2010.

Todd A Gormley. Costly information, entry, and credit access. *Journal of Economic Theory*, 154:633–667, 2014.

Robert Hauswald and Robert Marquez. Competition and strategic information acquisition in credit markets. *The Review of Financial Studies*, 19(3):967–1000, 2006.

Julapa Jagtiani and Catharine Lemieux. Fintech lending: Financial inclusion, risk pricing, and alternative information. 2017.

Julapa Jagtiani, Catharine Lemieux, et al. Small business lending after the financial crisis: A new competitive landscape for community banks. *Economic perspectives*, 3:1–30, 2016.

Ashok Kaul, Stefan Klößner, Gregor Pfeifer, and Manuel Schieler. Synthetic control methods: Never use all pre-intervention outcomes together with covariates. 2015.

Karen Mills and Brayden McCarthy. The state of small business lending: Innovation and technology and the implications for regulation. 2016.

Hoai-Luu Q Nguyen. Are credit markets still local? evidence from bank branch closings. *American Economic Journal: Applied Economics*, 2017.

Mitchell A Petersen and Raghuram G Rajan. Does distance still matter? the information revolution in small business lending. *The Journal of Finance*, 57(6):2533–2570, 2002.

Raghuram G Rajan. Insiders and outsiders: The choice between informed and arm’s-length debt. *The Journal of finance*, 47(4):1367–1400, 1992.

Steven A Sharpe. Asymmetric information, bank lending, and implicit contracts: A stylized model of customer relationships. *The journal of finance*, 45(4):1069–1087, 1990.

Kenneth Temkin. An analysis of the factors lenders use to ensure their sba borrowers meet the credit elsewhere requirement. 2008.

Ernst-Ludwig Von Thadden. Asymmetric information, bank lending and implicit contracts: the winner's curse. *Finance Research Letters*, 1(1):11–23, 2004.

Nada Wasi, Aaron Flaaen, et al. Record linkage using stata: Preprocessing, linking, and reviewing utilities. *Stata Journal*, 15(3):672–697, 2015.

Table 1: **Live Oak Industries**

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

This table shows the industries (5-digit NAICS codes) where Live Oak Bank has approved at least 50 loans. "Live Oak's Share of Loans" shows the number of Live Oak loans to that industry divided by the total number of SBA loans to that industry since the entry of Live Oak. Similarly, "Live Oak's Share of Volume" calculates Live Oak's share of total loan volume to that industry. "Enter Month" is the month that Live Oak first approved a loan to that industry.

Table 2: **Remote Lending and Geography**

Panel A: Remote Lending

	Market Share (# loans)			Market Share (amount)		
	(1)	(2)	(3)	(4)	(5)	(6)
log(per capita SBA loans)	-6.78***			-6.52***		
	(0.56)			(0.70)		
log(per capita non-remote SBA loans)		-7.08***			-6.68***	
		(0.52)			(0.65)	
log(branches per capita)			-3.19***			-4.36***
			(0.98)			(1.21)
Observations	2,422	2,422	2,419	2,422	2,422	2,419
Mean of Dep. Var.	16.6	16.6	16.6	22.5	22.5	22.5

Panel B: Live Oak Lending

	Market Share (# loans)			Market Share (amount)		
	(1)	(2)	(3)	(4)	(5)	(6)
log(per capita SBA loans)	-2.14***			-2.43***		
	(0.30)			(0.44)		
log(per capita non-remote SBA loans)		-1.97***			-2.33***	
		(0.28)			(0.42)	
log(branches per capita)			0.045			-0.57
			(0.51)			(0.76)
Observations	2,422	2,422	2,419	2,422	2,422	2,419
Mean of Dep. Var.	3.87	3.87	3.87	7.28	7.28	7.28

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This table reports the estimates from specification (4). Panel A regresses the remote lender market share of loans (or dollar amount) in a county from 2014-2017 on county-level measures of access to SBA lending from traditional banks. Remote lenders are defined as banks where the median borrower-lender distance for their loans is more 100 miles. Panel B replaces the dependent variable with Live Oak's market share of loans (or loan amount) in a county from 2014-2017. The county-level measures of access are (i) per capita loans, (ii) non-remote per capita loans, and (iii) county-level branches per capita. Per capita SBA loans are averages from 2001-2007, and branches are from the 2016 FDIC Summary of Deposits. Since we take the log, counties with zero loans or branches are dropped. Data: SBA 7(a) Loan Report.

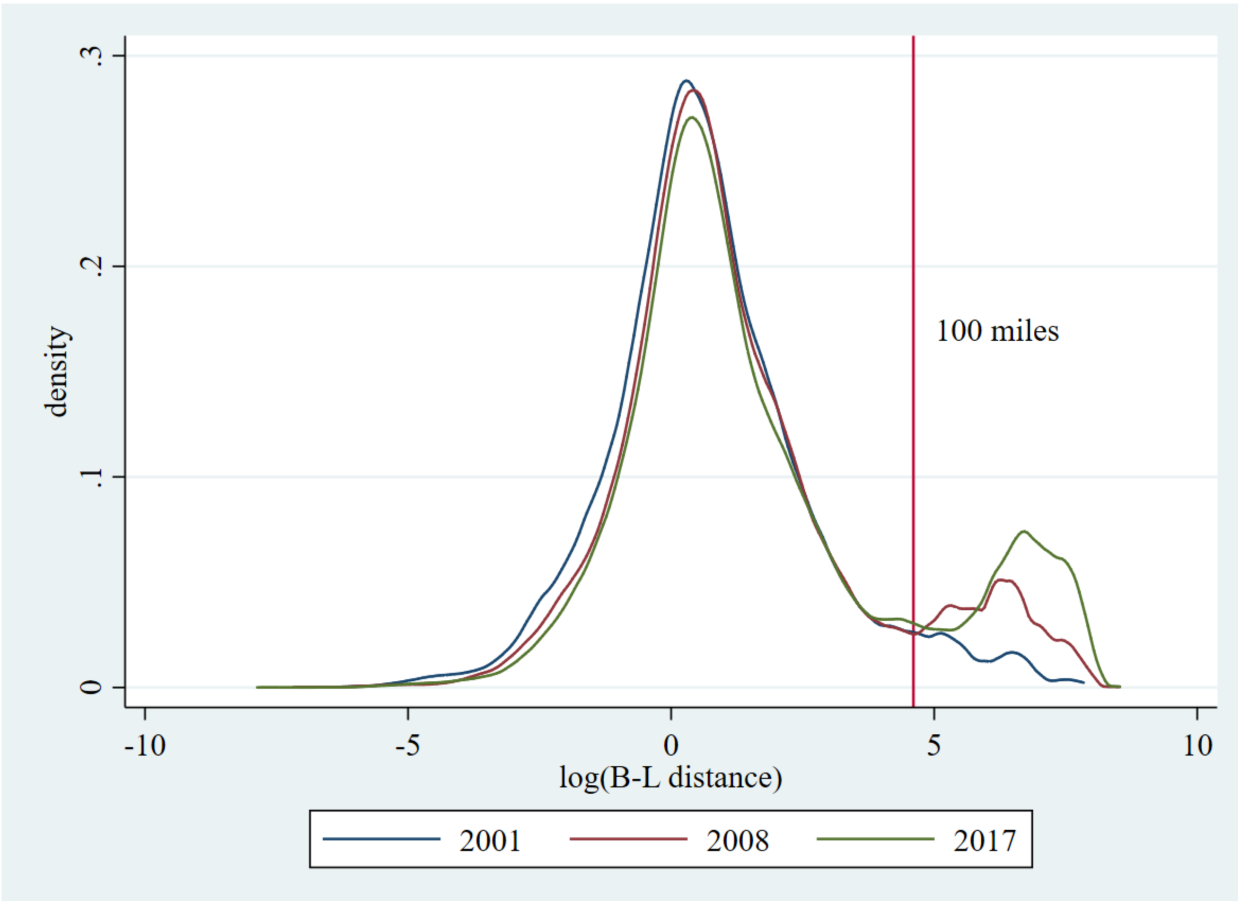


Figure 1: **Distribution of (log) Borrower-Lender Distance for SBA Loans** This graph shows the distribution of the distance between borrowers and the closest branch of the institution from which they borrowed. Borrower-lender distance is calculated according to the procedure described in Section 2.2.

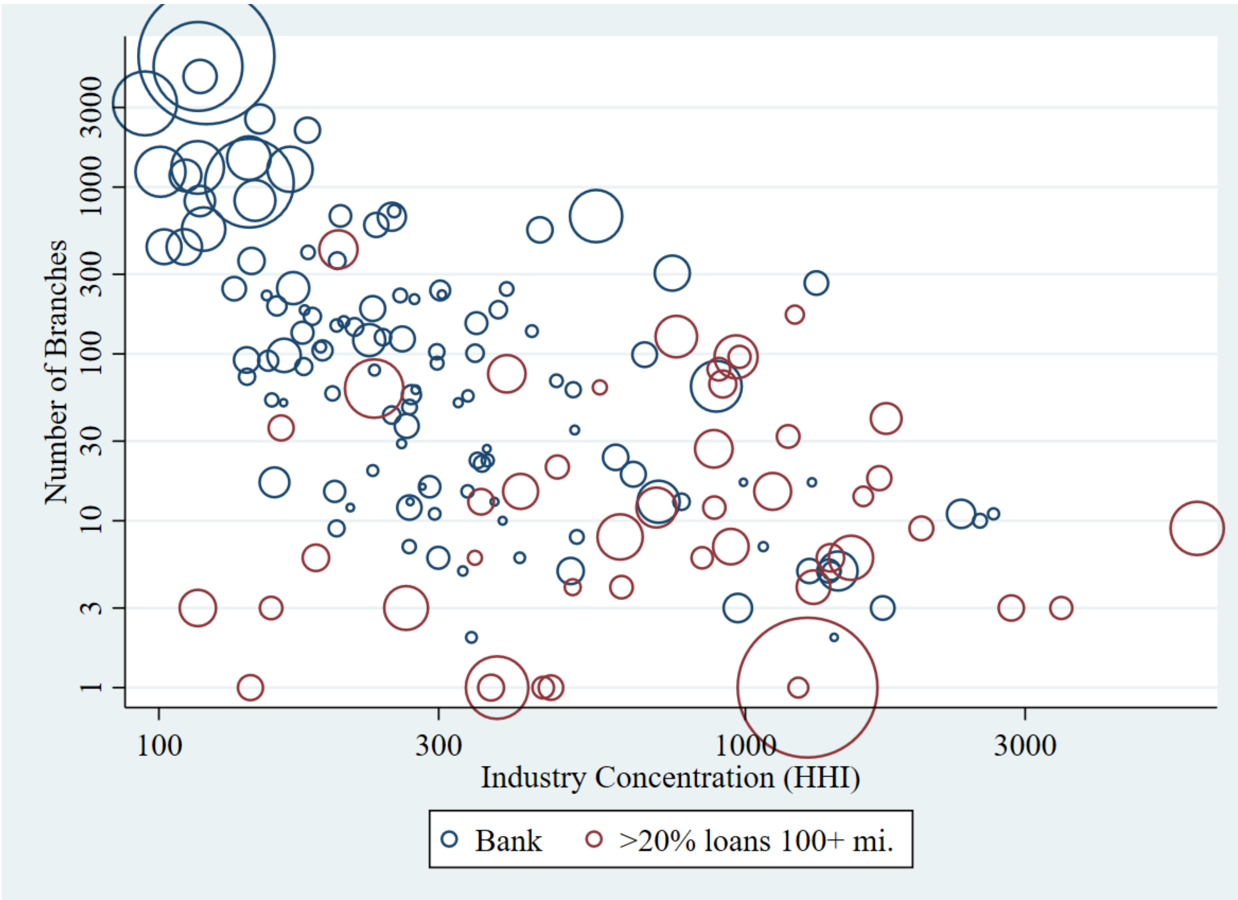


Figure 2: **Branches and Industry Concentration (2014-2017)** Each observation is an SBA lender, and the size of the circle reflects the total amount of SBA 7(a) lending from that lender between 2014-2017. The vertical axis shows the (log-scale) number of branches, and the horizontal axis is a (log-scale) measure of how concentrated the lender's loans are in a certain industry. The industry concentration (HHI) for lender j is $HHI_j = \sum_i S_{ij}^2$, where S_{ij} is the percentage (0-100) of lender j's loan volume given to industry j (5-digit NAICS code). This measure is increasing in industry concentration and takes a value between 100 (least concentrated) and 10,000 (most concentrated). The red circles are lenders with significant remote lending, defined as having at least 20% of their loans with a borrower-lender distance of more than 100 miles.

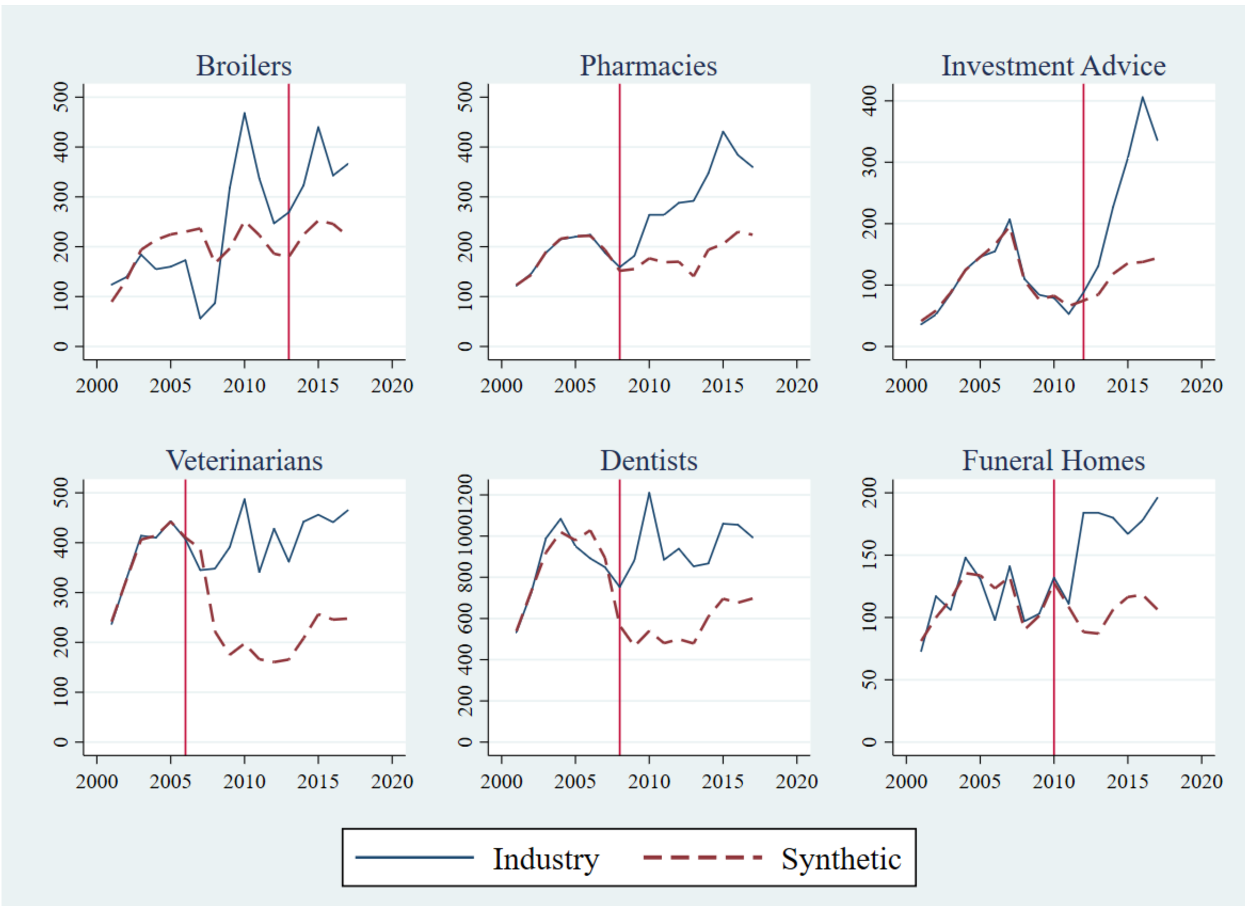
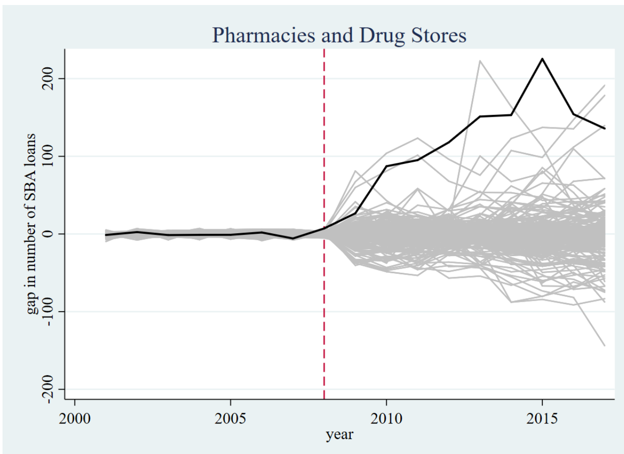
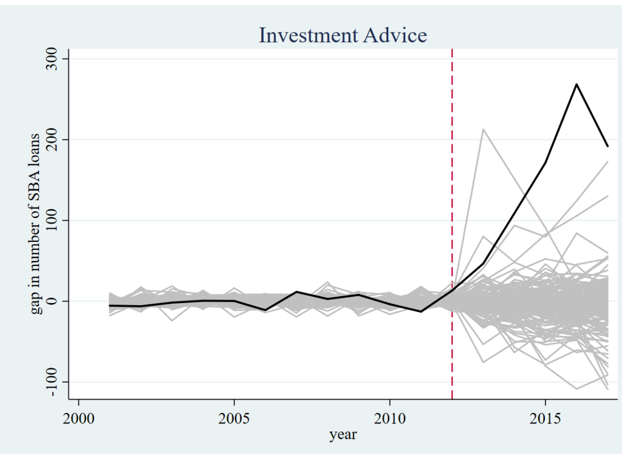


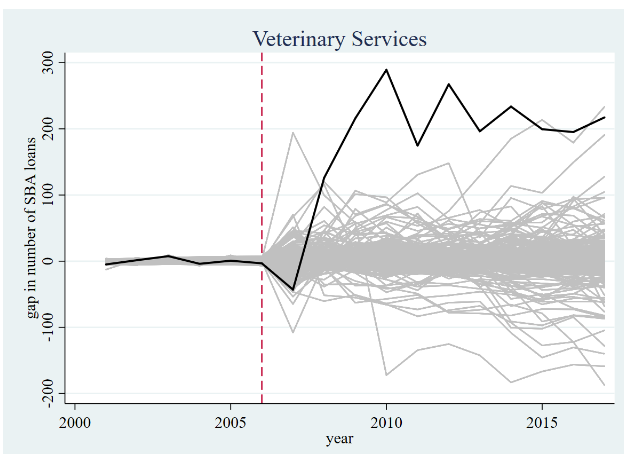
Figure 3: **Number of Loans - Treated Industry vs. Synthetic Control** This figure compares the number of loans in each industry that Live Oak enters to its synthetic control. The synthetic controls are formed by matching on all pre-treatment years beginning in 2001, with no additional covariates. The vertical line shows the year before Live Oak entered.



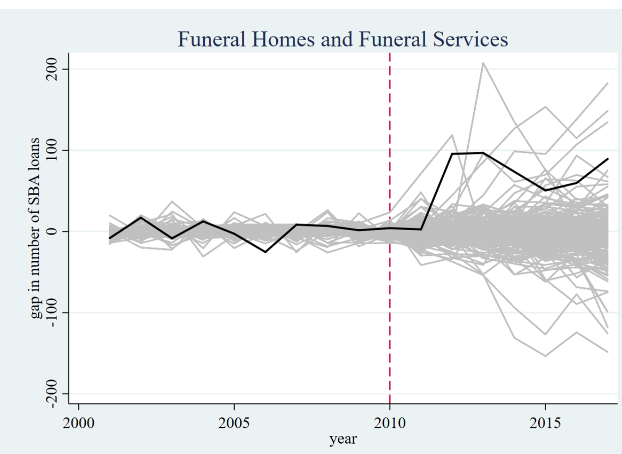
(a)



(b)



(c)



(d)

Figure 4: **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.

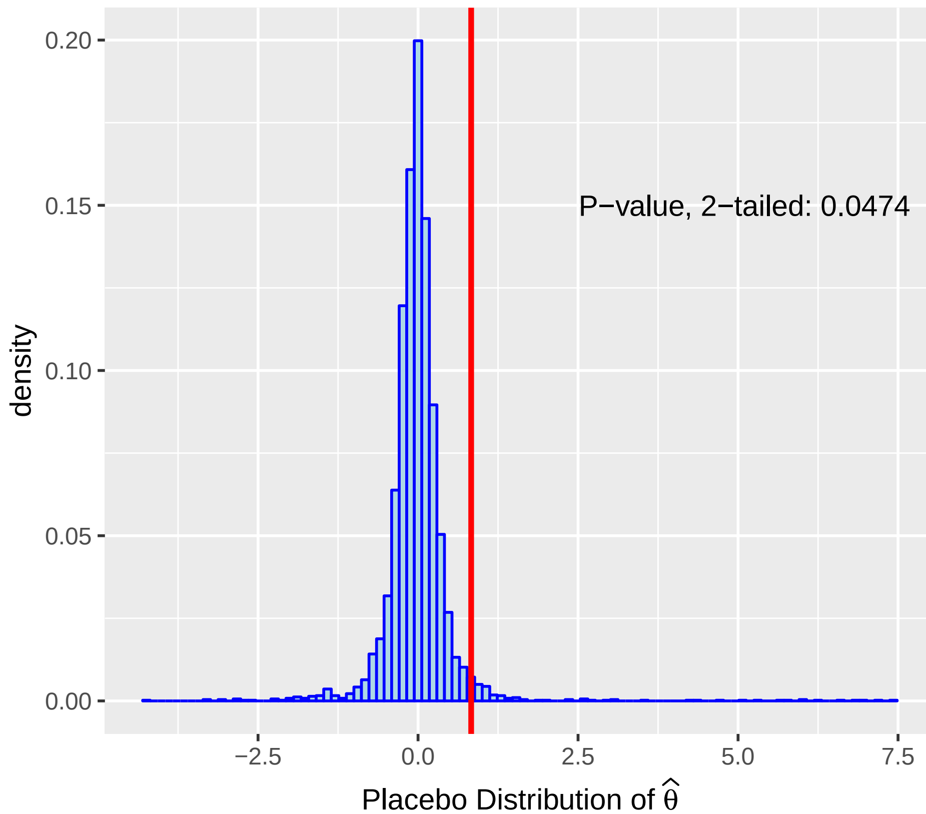


Figure 5: **Placebo Distribution of $\widehat{\theta}^{PL}$** The vertical red line shows the magnitude of the average treatment effect $\widehat{\theta}$ for the treated industries, calculated from equation (3).

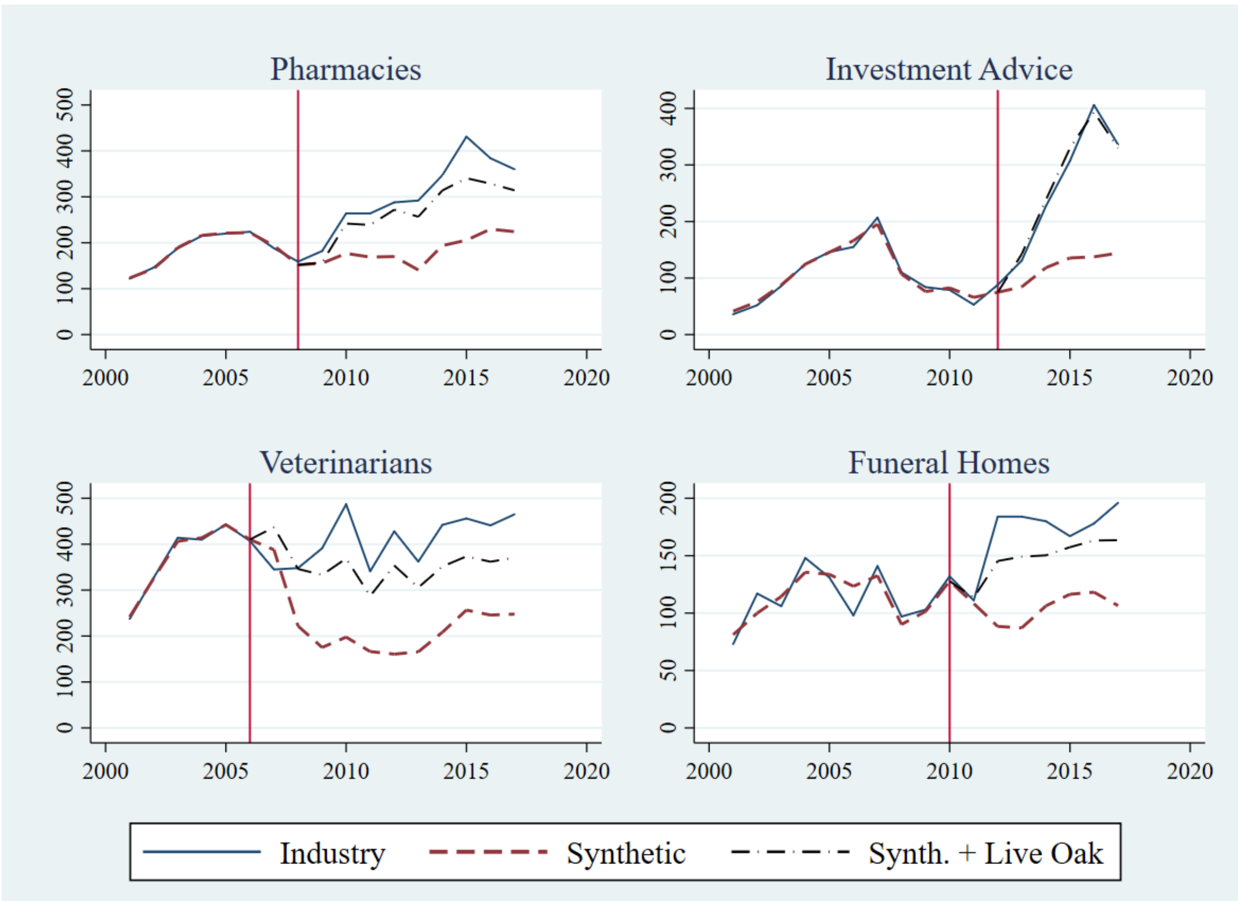


Figure 6: **Examining Substitution from Existing Lenders** This figure compares the number of loans in each industry that Live Oak enters to its synthetic control. The black dotted line "Synth. + Live Oak" adds the number of Live Oak loans to the outcome for the synthetic control, which reflects the number of loans that would be expected with no substitution from existing lenders. 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.

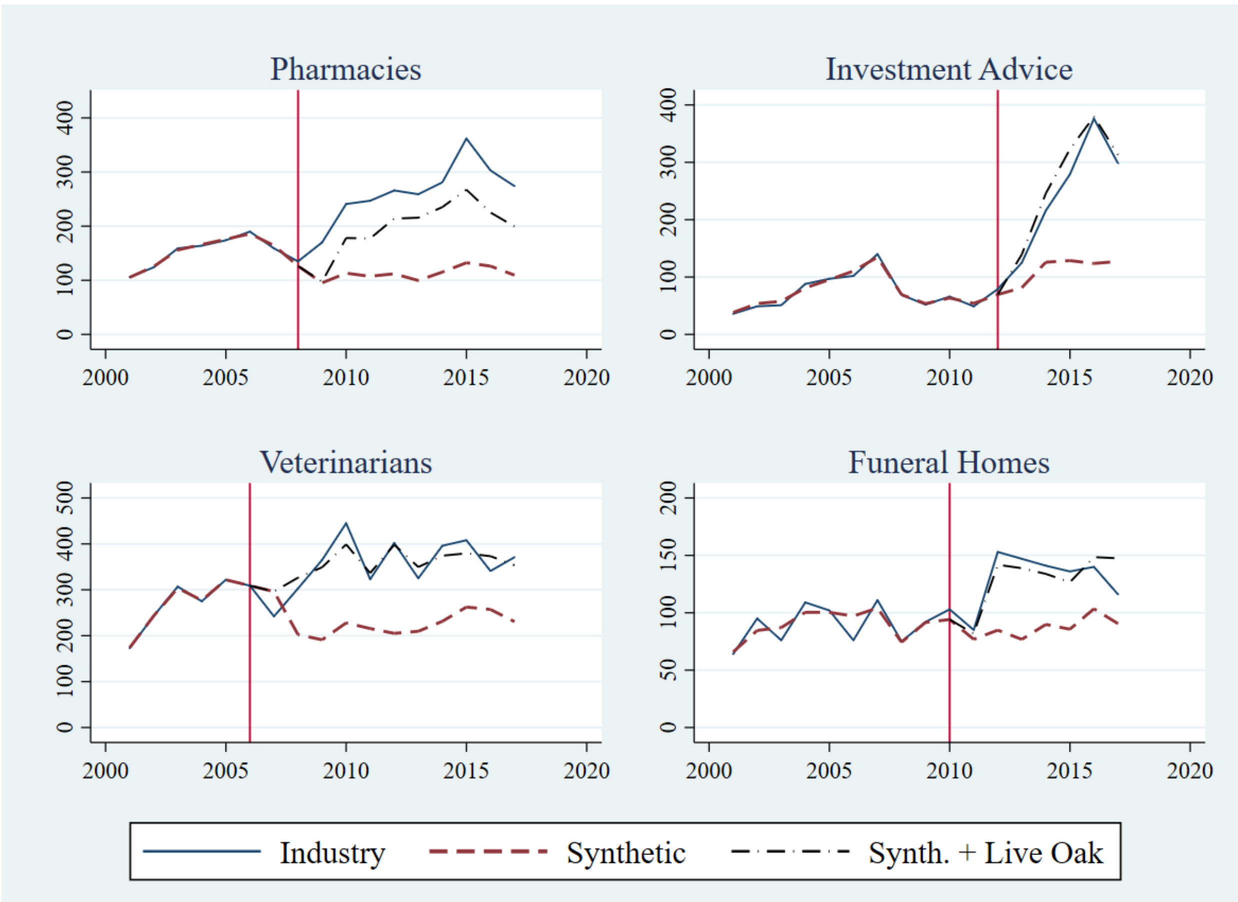


Figure 7: **Number of Loans - Treated Industry vs. Synthetic Control (excluding remote loans)** We exclude any loans from other remote lenders, defined as an institution-year observation with a median lending distance of more than 100 miles. This figure compares the number of loans in each industry that 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.

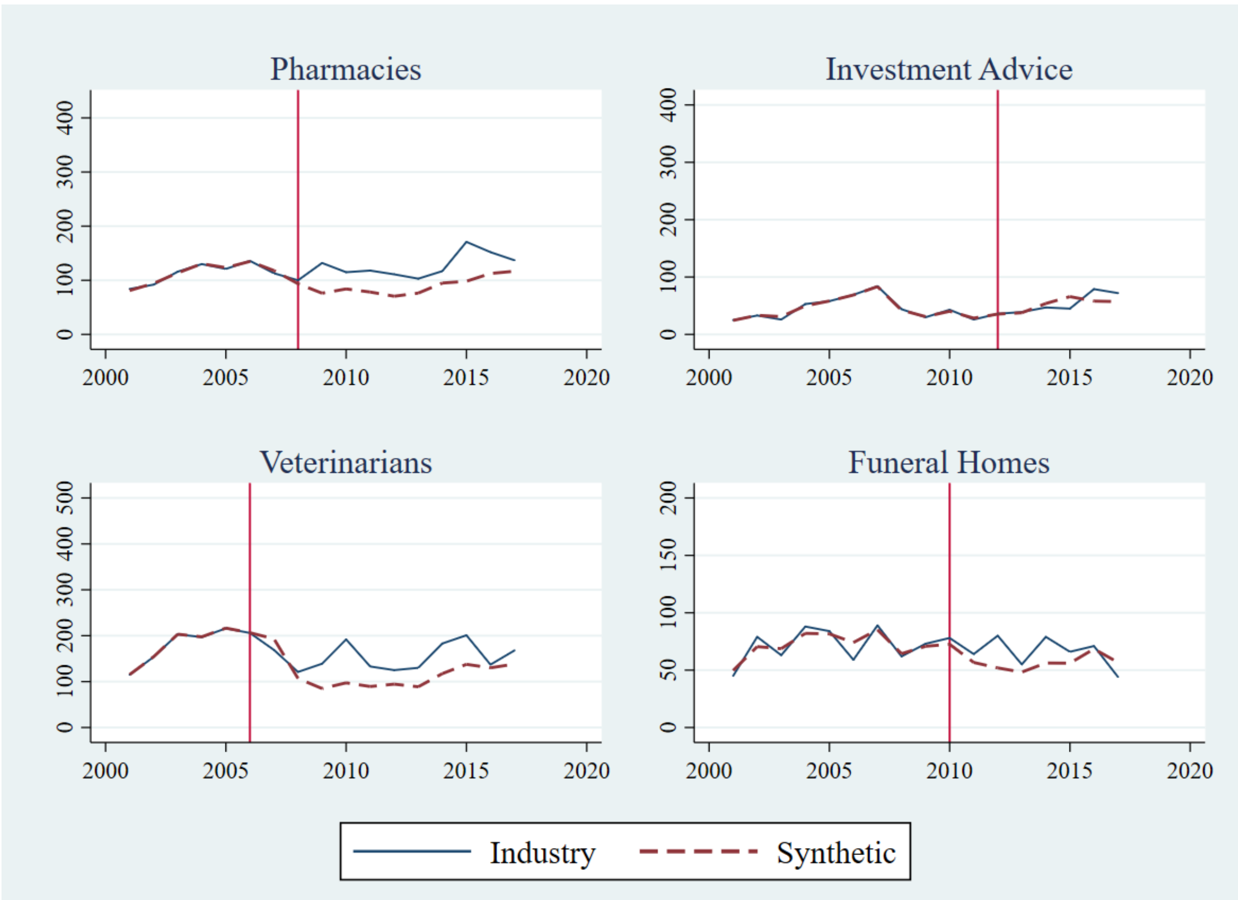


Figure 8: **Treated Industry vs. Synthetic Control in Zip Codes with Zero Live Oak Loans** This figure provides a falsification check by showing growth in loans to the treated industries in zip codes where Live Oak gave no loans. The two-sided p-value of the average effect on these four groups, computed using equation (3), is 0.483.

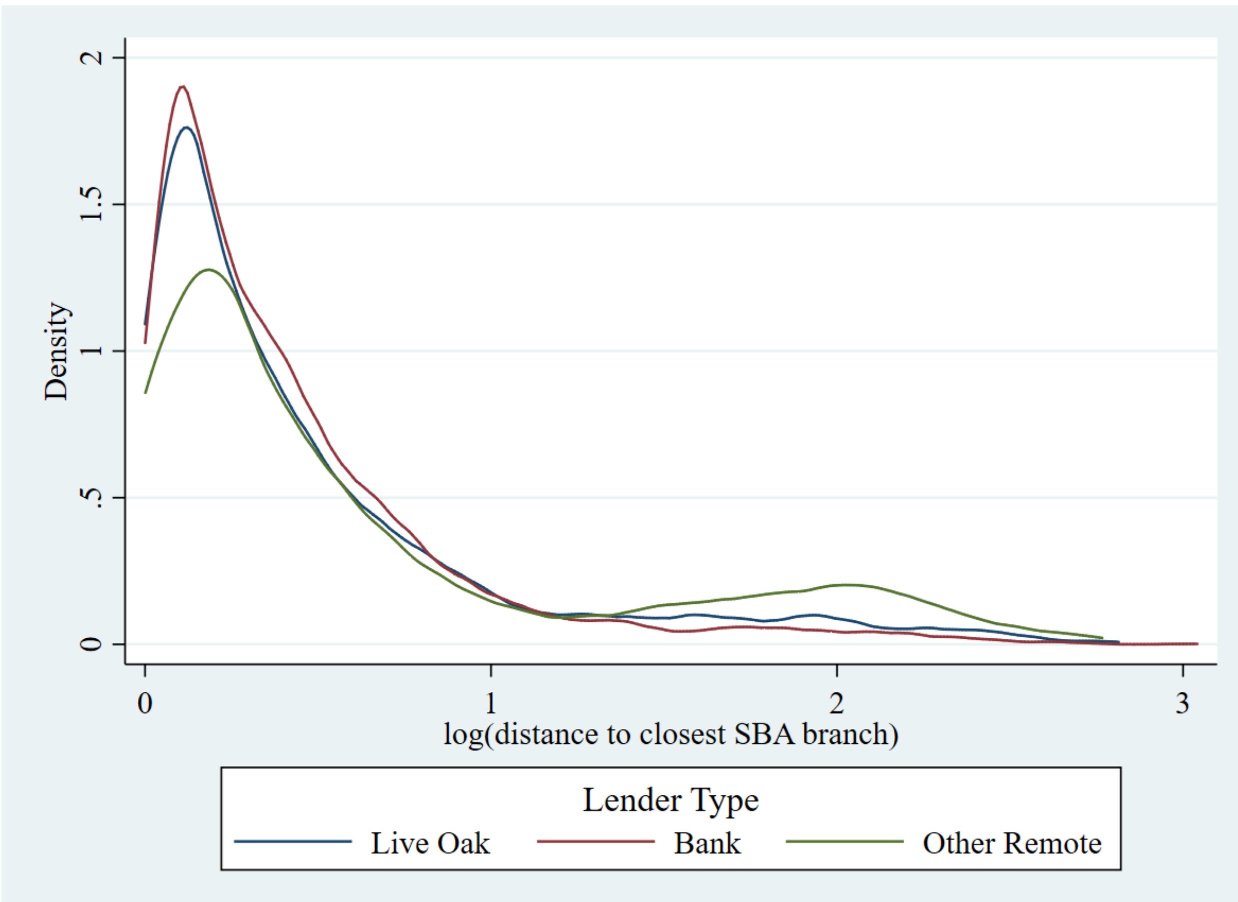


Figure 9: **Distance to Closest SBA Branch** This graph shows the distribution of the distance between borrowers and the closest branch of any institution that grants SBA loans. Distance is calculated according to the procedure described in Section 2.2, except it is the distance to the closest branch of any SBA lender.

A Appendix Tables and Figures

Table A.1: **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
	Continuing Care Retirement Communities	0.25
	Machine Shops	0.30
	Offices of Physical, Occupational and Speech Therapists, and Audiologi	0.28
	Other Direct Selling Establishments	0.00
	Photography Studios, Portrait	0.05
	Solid Waste Collection	0.04
	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 Property and Casualty 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
	Commercial Lithographic Printing	0.02
	General Automotive Repair	0.06
	Motion Picture and Video Production	0.42
	Offices of Lawyers	0.03
	Other Business Service Centers (including Copy Shops)	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 Business Service Centers (including Copy Shops)	0.05
	Shellfish Fishing	0.17

B Appendix: Matching Procedure

In this section we describe the procedure used to construct a measure of borrower-lender distance. The measure we use is the distance between the borrower and the closest branch of the institution from which they borrowed.

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.[13] 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,

[13] 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.

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 match 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, we use the FDIC's Report's 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.

One possibility is that banks 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,

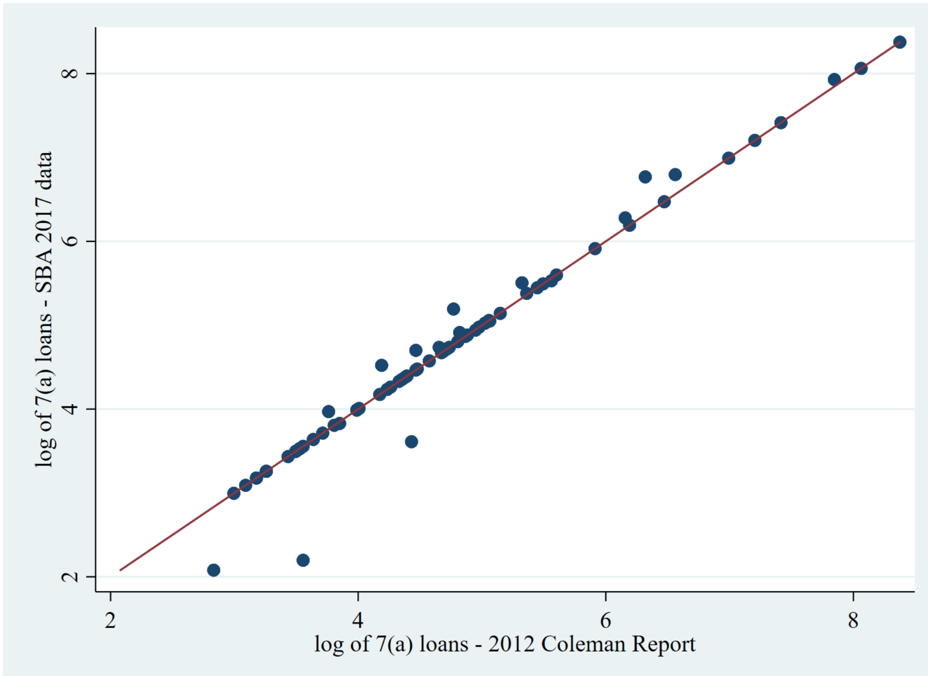


Figure B.1: Differences between current and contemporaneous counts.

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.

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.[14] Finally, we calculate the (haversine)

[14] We drop the 1.5% of branches that are missing longitude and latitude data.

distance between the borrower and (i) the closest branch of the institution that originated the loan and (ii) the closest branch of a competing SBA lender.[15]

[15]The Haversine distance, which is the shortest distance over the earth's surface.