**Abstract:** The purpose of this paper is to (1) investigate how bank branches are spatially distributed across the United States; (2) how bank branch closures between 2012 and 2017 are spatially distributed across the United States; and (3) if a net negative change in bank branches in a county affects access to credit as measured by mortgage originations. More specifically, this paper analyzes if bank branches and bank branch closures cluster in specific geographic areas. To analyze the spatial distribution of bank branches and bank branch closures, I use Local Moran’s I, Local Moran’s I with Empirical Bayes (EB) Rates, and Kernel Density Estimation. Next, to analyze how bank branch closures affect access to credit, I use Ordinary Least Squares (OLS) regression to analyze how net negative losses in bank branch access leads to a decrease in mortgage originations, controlling for county-characteristics. I find that bank branches cluster in urban and coastal regions, but this relationship flips when adding in population and using EB rates. Additionally, I find a similar relationship for bank branch closures. Finally, my results for the effect on bank branch closures on mortgage originations is inconclusive. While the analysis suggests statistically significant clustering of bank branches and bank branch closures, additional analysis is necessary to more thoughtfully carry out this analysis. Finally, the regression analysis provides mixed results, suggesting a more thoughtful construction of variables may be necessary to investigate this relationship.

1. Introduction

This project analyzes the spatial distribution of bank branches across the United States in 2017, the spatial distribution of bank branch closures across the United States between 2012 and 2017, and if bank branch closures within a county affect access to credit. The total number of bank branches has steadily declined in recent years, with a total of 93,391 branches nationally in 2012 compared with 86,476 branches nationally in 2017. This decreasing trend in bank branches coupled with the rise of mobile banking has led to questions of where and why banks choose to have or close branches, and how important the bank branch is for consumers in accessing credit.

In conducting a geospatial analysis of bank branches and bank branch closures, this project allows for an analysis of areas that are bank branch hotspots and bank-branch-closure hotspots. Identifying these geographical hotspots allows for a further analysis of those locations. For example, are bank branch hotspots more likely to be urban or rural counties? Is there something unique about the county populations—do they have higher incomes? Do they vary by race/ethnicity? Do they experience population or demographic changes over the period? The same questions apply to counties that are hot spots for bank branch closures. While this analysis cannot directly identify reasons for bank branch or bank-branch-closure hot spots, it can provide additional information about these areas and inform thinking on underlying mechanisms. Additionally, using regression analysis to analyze how a bank branch closure affects access to credit can help provide insight on the importance of bank branches for credit access[[1]](#footnote-1). This can also help identify areas that may well targeted by policy in increasing access to credit.

My hypothesis for this analysis is that bank branches will cluster in urban centers—places that have high populations and often many job opportunities—and that bank branch closures will cluster in areas with lower or decreasing population counts[[2]](#footnote-2) or areas that are economically declining, such as certain rural areas of the country. The hypothesis here is that banks will choose to locate bank branches in places with a high number of potential customers, and they will choose to close branches in areas with fewer potential customers. Additionally, understanding where and why banks choose to close branches will provide useful tools for policymaking. Understanding why banks choose to leave specific areas, particularly if they are underserved, can help develop policy to encourage banks to meet the needs of various communities. This may especially relevant to recent discussion for modernizing the Community Reinvestment Act (Brainard, 2020).

1. Methods

This section contains a description of each of the methods used. To conduct hotspot analysis I use Local Moran’s I, Local Moran’s I with EB Rates, and Kernel Density Estimation. To conduct regression analysis, I use Ordinary Least Squares (OLS). A description of each is provided below.

1. Local Moran’s I

The Local Moran’s I is a technique that identifies spatial clusters of features with high or low values (hot spots or cold spots) based off a set of weighted features. Additionally, Local Moran’s I can identify spatial outliers. Mathematically, Local Moran’s I calculates a statistic of spatial association for each feature by subtracting the mean from that feature’s value, dividing by the sample variance, and weighting by the spatial weight between the two features. The Local Moran’s I then outputs (1) a Local Moran’s Index; (2) Local Moran’s Z-score; (3) Local Moran’s P-value; and (4) Cluster-Outlier Type. The Local Moran’s I index provides information about whether the feature is part of a cluster (either hot or cold spot) by outputting a positive value. It outputs a negative value if the feature has neighbors with dissimilar values (outliers). If this number is statistically significant based off the p-value, then the Cluster-Outlier Type will provide a variable if it is a hotspot (high-high clustering), cold spot (low-low clustering), or outlier (either low-high or high-low).

To implement this for my analysis, I use the Local Moran’s I function in ArcGis on two different features: (1) the number of bank branches per county and (2) the number of bank branch closures per county. For both of these, I choose the inverse distance method of conceptualization of spatial relationships and let ArcGis use a default threshold, which assumes that every feature has at least one neighbor. I apply the False Discovery Rate (FDR) correction, which is a correction to account for multiple testing spatial dependency. Finally, I run this with 499 permutations—this means that for each permutation, ArcGis will randomly rearrange the features, calculate the Local Moran’s I values, and then compare this distribution to the actual observed Local Moran’s I values to determine the probability of that value being found in a random distribution.

1. Local Moran’s I with EB Rates

The Local Moran’s I with Empirical Bayesian (EB) Rates is an update to the Local Moran’s I to provide more accuracy when using Local Moran’s I with rates or proportions. Rates and proportions suffer from inherent variance instability, for example, in the case of counties with small sample populations and may have larger standard errors. This may lead to the misidentification of spurious outliers and can be corrected for using the Empirical Bayes technique. The smoothed rate is expressed as a weighted average of the observed rate and a reference rate, and it works by computing a weighted average between the given rate for each county and the reference weight, with weights proportional to the underlying population. Generally, small counties will have their rates adjusted much more than larger counties. This strategy is generally only useful when the populations vary substantially by geography (as in the case of counties), but less so when the populations are more or less equal across geographies (as in the case of census tracts). Because this analysis is focused at the county-level and I am using rates, I use the Local Moran’s I with EB Rates.

To implement this in my analysis, I use GeoDa and create a spatial weights matrix using the Queen’s Case contiguity option. This option defines the neighborhood for each feature such that every contiguous feature is within that neighborhood. It then implements the Local Moran’s I with EB rates analysis.

1. Kernel Density Estimation

Kernel Density works by calculating the density of point features around each output raster cell. In practice, a smoothly curved surface will be fit over each point, usually from a quartic or Gaussian function, where the highest point on the curve corresponds to the specified point at that raster cell. The height of the curve will then decrease as it gets further out from the point until it reaches zero at the search radius from the point. Kernel Density will do this for each point in the sample, and will then calculate the density for each raster cell by adding the values of all the overlapping kernel surfaces. Because I have point-level data of bank branches, I can create a kernel density map showing the densities of locations where bank branches were closed.

1. Ordinary Least Squares (OLS)

Regression is a technique to quantify relationships between variables, whereby a change in one variable can explain a change in other variables. These models provide information about the direction of the relationship (positive or negative), the strength of the relationship (the magnitude of the coefficient), and the statistical significance of the relationship (p-value). Ordinary Least Squares regression technique used for finding the best-fitting curve to a given set of points by minimizing the sum of the squares of residuals (differences between observed and predicted values).

I use OLS to investigate the relationship between bank branch closures, county characteristics and mortgage originations per capita. I choose two specifications. In the first, my dependent variable is the number of mortgages per capita and the explanatory variables are a binary indicator for whether the county had net negative branch closures, the share of population that is White, Black (non-Hispanic), and Hispanic, the share of population with at least a college degree, the share of population in poverty, and available housing units per population. In the second, my dependent variable is the number of mortgages per capita and the explanatory variables are the number of bank branch closures per population, the share of population that is White, Black (non-Hispanic), and Hispanic, the share of population with at least a college degree, the share of population in poverty, and available housing units per population.

1. Results

Below are the results of the analysis and a brief discussion of each.

1. Data Description

In order to conduct this analysis, I create two datasets: a county-level dataset of bank branch counts; and a point-level dataset of branches that close between the two time periods. For the first, I combine data from three sources. First, I use the Census Bureau’s American Community Survey 5-year 2012-2017 estimates to obtain population and demographic information at the county level. Next, I use FDIC’s Summary of Deposits data and aggregate the number of full service brick and mortar offices or full service retail offices to the county-level. Finally, I use mortgage data from the Home Mortgage Disclosure Act (HMDA) from 2012 and 2017 and aggregate the count of mortgages that are approved for a home purchase loan[[3]](#footnote-3). Using these three datasets, I aggregate and combine them to have a unique county-level dataset to conduct this analysis. Table 1 shows summary statistics of this dataset.

Table 1: Summary Statistics

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Minimum** | **25th Percentile** | **Median** | **Mean** | **75th Percentile** | **Maximum** |
| **Bank Branches** | 1 | 5 | 10 | 27.81 | 21 | 1744 |
| **Bank Branch Closures (2012-2017)** | -23 | 0 | 1 | 2.20 | 2 | 189 |
| **Total Population** | 74 | 11,168 | 25,812 | 102,588 | 67,583 | 12,105,722 |
| **Total Mortgages** | 1 | 119 | 337 | 1,617 | 1,073 | 95,661 |

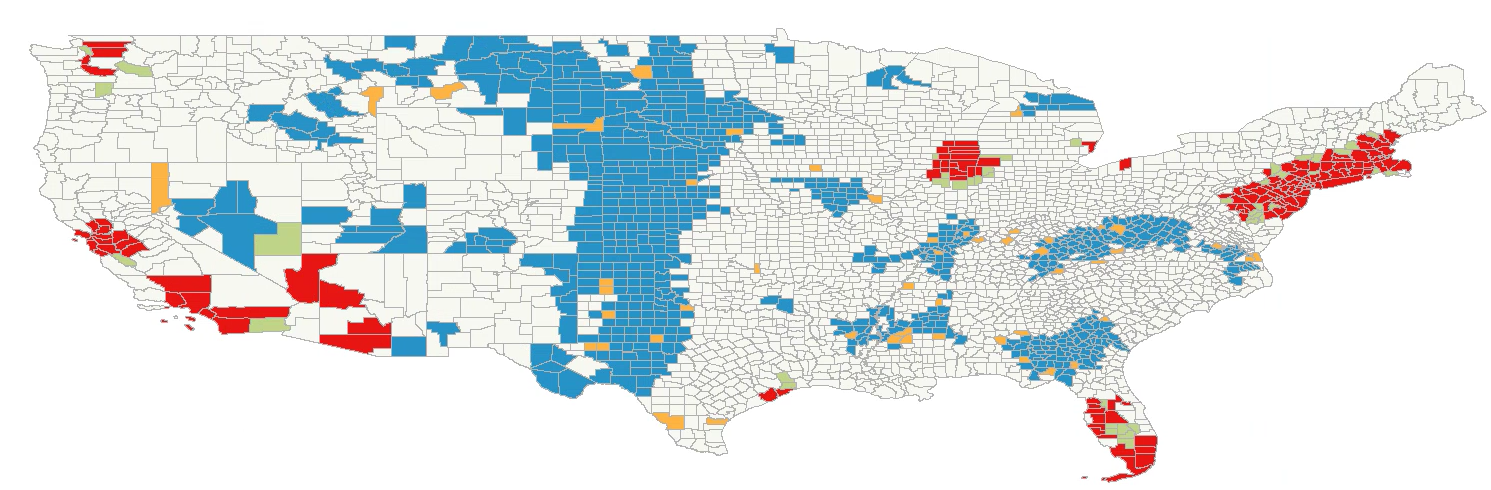
Note: n=3108

Table 1 shows summary statistics for total bank branches, bank branch closures between 2012 and 2017, total population, and total mortgages. The summary statistics show wide spread for all of these variables, especially population.

1. Bank Branch and Bank Branch per Capita Distribution

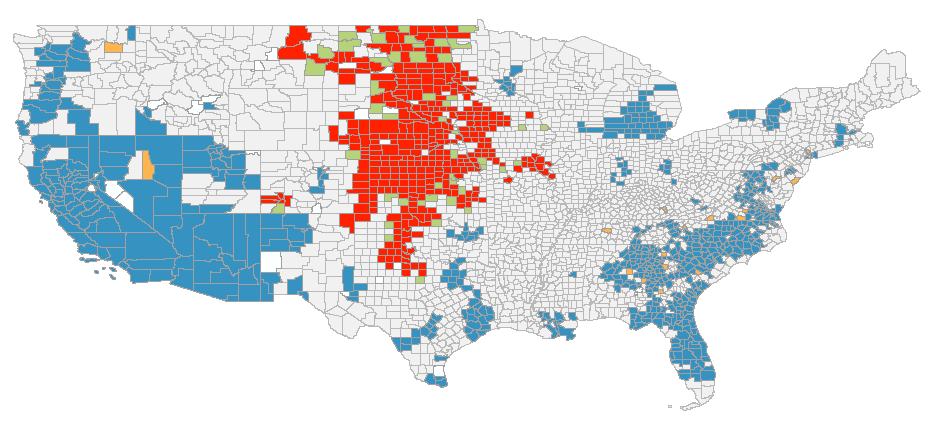
I start with using Local Moran’s I to analyze bank branch distribution in 2017, followed by Local Moran’s I with EB Rates to analyze bank branch closures per 100,000. Results and brief interpretation are below. I use ESRI’s ArcMap 10.7 to conduct Local Moran’s I and Regression analysis, and I use GeoDa to conduct Local Moran’s I with EB Rates.

Map 1: Local Moran’s I of Bank Branch Distribution (2017)



The results of this map are not too surprising—populous areas like the Northeast Corridor, Chicago, and parts of California are statistically significant hotspots per the analysis. This makes sense intuitively—where there are more people, there will likely be more bank branches. Additionally, many of these locations are home to affluent cities such as San Francisco or New York—prime areas for banks to do business. Cold spots are similarly not surprising—areas in the middle of the country with low population appear to be coldspots, with some high outlier spots nested within. These outliers may be smaller urban centers that may serve surrounding counties. Additionally, parts of the Mississippi Delta Region and Appalachia—two historically underserved areas—are cold spots on this map.

Map 2: Local Moran’s I with EB rates of Bank Branches\*100k/Population

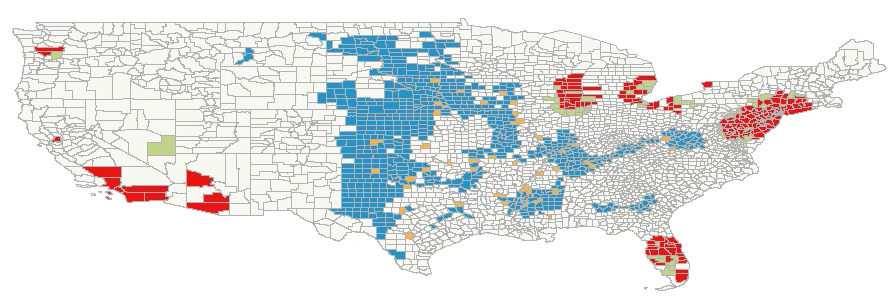


Interestingly, when adding in share of population, the map almost flips—coastal areas are now cold spots and inland, less populous areas are now hotspots. Despite using EB rates to correct, this may be driven by stark differences in population across counties and a non-linear relationship between bank branches and population.

1. Bank Branch Closures and Bank Branch Closures per Capita Distribution

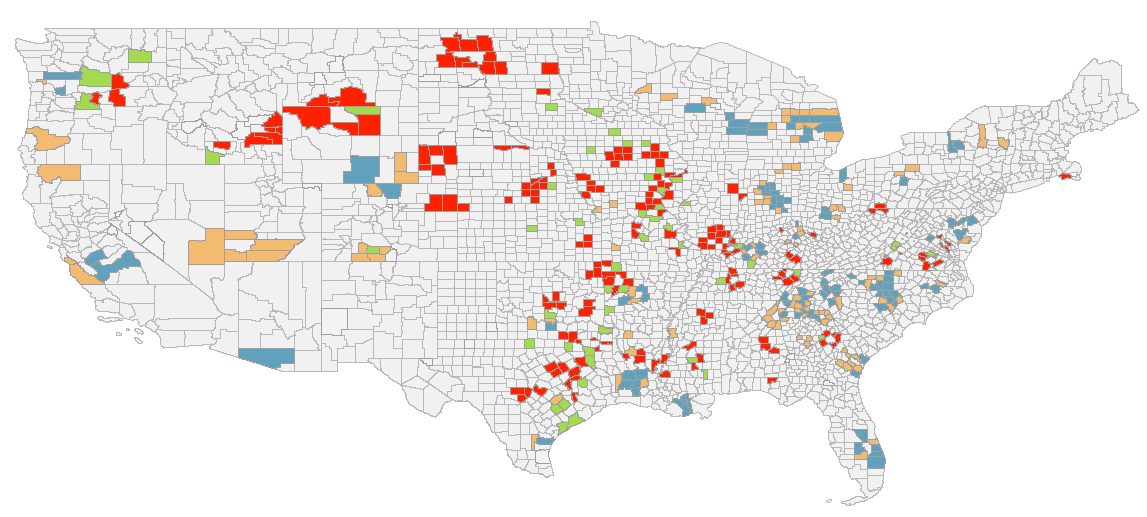
Next, I use Local Moran’s I and Local Moran’s I with EB Rates to analyze bank branch closure hotspots across the country. Additionally, I use Kernel Density Estimation to analyze point data of bank branches that have closed. Results are provided and discussed below.

Map 3: Local Moran’s I of Bank Branch Closures (2012-2017)



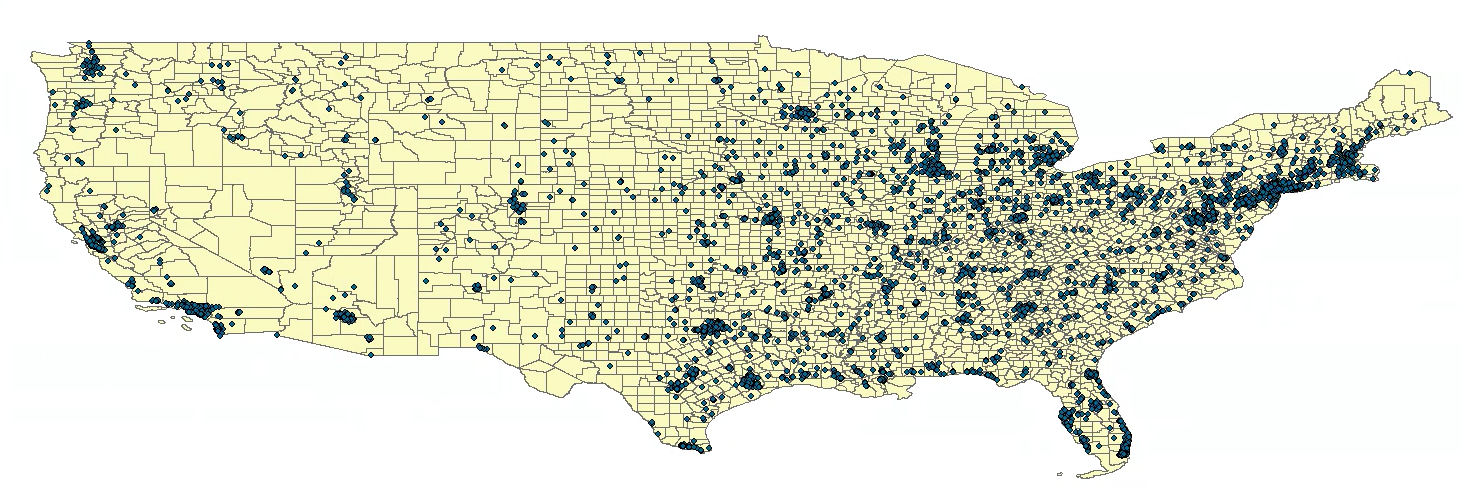
Note that this map has somewhat similar patterns to Map 1 in that populous cities or coastal regions are statistically significant hotpots and parts of the middle of the country are cold spots. Intuitively, I would expect that parts of the middle of the country would actually be hotspots for bank branch closures with more branches closing. However, it may still be the case that this is largely driven by population—rural counties with small population may actually see only slight decreases in bank branches, but have a much lower number of bank branches to start with. That is, they may be experiencing a significantly higher percent decrease in bank branches over the time period.

Map 4: Local Moran’s I with EB rates of Bank Branch Closures\*100k/Population (2012-2017)

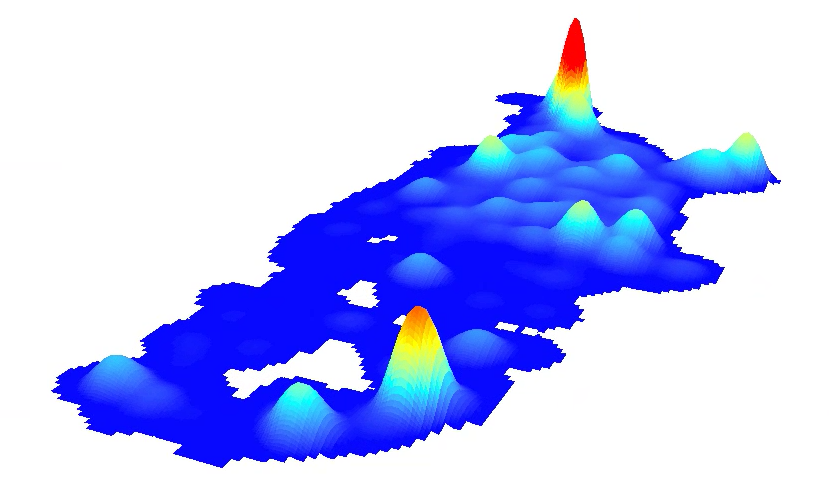


When accounting for population and using EB Rates, the map shows something more intuitive. Hotspots cluster in the middle of the country and rural areas, and obvious urban centers and coastal areas are no longer hotspots.

Map 5a: Bank Branch Closures Point Locations



Map 5b: Kernel Density Estimation of Bank Branch Closures



Map 5a plots the points of all locations where a bank branch closed between 2012 and 2017. Map 5b provides a kernel density estimation of areas that lost the most bank branches. The map shows that populous urban centers are still hotspots for losing bank branches. However, it also shows density for bank branch closures in the middle of the country. Still these results are likely to be at least partially driven by population.

1. Ordinary Least Squares Regression

Finally, while the above analyses provide good exploratory information, I use regression to attempt to quantify the relationship between bank branch closures, county characteristics and mortgage originations. The results are below.

Table 1a: Regression 1 Results

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable Names** | **Full names** | **Diagnostics:**  AICc: -23535.558  Adjusted R-Squared: .234230  Joint F-Statistic: .00000  Joint Wald Statistic: .00000  Koenker (BP Statistic): .00000  Jarque-Bera Statistic: .00000 | | | |
| **Dependent Variable** | **Label** |
| Ttl\_mrt | Share of mortgages per population |
| **Covariates** | **Label** | **Probability** | **Significant Predictor?** | **Positive or Negative relationship** | **Coefficient** |
| Branch\_L | Binary variable for whether county lost bank branches | .00000 | **Yes** | Positive | .00189 |
| White\_sh | Share of population that is White | .00171 | **Yes** | Positive | .00407 |
| Black\_sh | Share of population that is Black | .00185 | **Yes** | Positive | .00514 |
| Hspnc\_sh | Share of population the is Hispanic | .37576 | **No** | Positive | .00685 |
| Share\_b\_h | Share of population with at least a bachelors | .00000 | **Yes** | Positive | .02223 |
| Pov\_sh | Share of population in poverty | .00000 | **Yes** | Negative | -.01770 |
| Hunits\_sh | Share of housing units per population | .00000 | **Yes** | Positive | .00619 |

Table 1b: Regression 2 Results

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Variable Names** | **Full names** | **Diagnostics:**  AICc: -23450.82961  Adjusted R-Squared: .213067  Joint F-Statistic: .00000  Joint Wald Statistic: .00000  Koenker (BP Statistic): .00000  Jarque-Bera Statistic: .00000 | | | |
| **Dependent Variable** | **Label** |
| Ttl\_mrt | Share of mortgages per population |
| **Covariates** | **Label** | **Probability** | **Significant Predictor?** | **Positive or Negative relationship** | **Coefficient** |
| C12\_17\_per | Change in bank branches from 2012 to 2017 over population | .68298 | **No** | Negative | -.51291 |
| White\_sh | Share of population that is White | .00618 | **Yes** | Positive | .00452 |
| Black\_sh | Share of population that is Black | .00002 | **Yes** | Positive | .00598 |
| Hspnc\_sh | Share of population the is Hispanic | .41150 | **No** | Positive | .00644 |
| Pov\_sh | Share of population in poverty | .00000 | **Yes** | Negative | -.01716 |
| Share\_b\_h | Share of population with at least a bachelors | .00000 | **Yes** | Positive | .02440 |
| Hunits\_sh | Share of housing units per population | .00000 | **Yes** | Positive | .00496 |

In the first regression, the variable of interest branch\_L is slightly positive and statistically significant. Intuitively, this does not make much sense—this interpretation means that, on average, if a county loses at least one bank branch, it will see an increase in mortgage originations. However, in the second regression, even though it is not statistically significant, the coefficient on the C12\_17\_per variable is negative, which does make intuitive sense. This means that, on average, an additional increase in bank branch closures leads to a decrease in mortgage originations. The differences in sign between these two analyses (despite the lack of statistical significance for the latter) is interesting and suggests additional analysis may be necessary. For example, in the first regression that uses the binary indicator, this indicator variable makes no distinction between a county that loses one branch and a county that loses one-hundred branches. It may be the case that the relationship between bank branch loss and access to credit is not actually linear.

Additionally, while both models have low values for the adjusted R-squared, many of the variables have coefficients that make sense and are statistically significant. For both models, there is a negative relationship between the share of a county living in poverty and mortgage originations. Intuitively this makes sense—banks are not likely to originate mortgages to those who may not have the ability to repay such as those living poverty. Additionally, small dollar mortgages—mortgages that might be particularly useful for those living in high poverty areas—are costly for banks and are less likely to be originated. Additionally, there is a positive relationship between the share of housing units per population, the share of population with at least a bachelor’s degree, and the share of white and black populations.

1. Discussion/Conclusion

This paper set out to analyze the distribution of bank branches and bank branch closures across the United States. Additionally, this paper investigates how bank branch closures affect access to credit as measured by mortgage originations. The original hypothesis is that bank branches would cluster in high population, affluent areas such as coastal regions and cities, and that bank branch closures would cluster in rural areas that may have higher poverty or lower or decreasing population counts. Additionally, I hypothesized that bank branch closures in a county would decrease access to credit. The results are mixed.

First, we can see that when not including population, bank branches do indeed show statistically significant clusters in coastal and urban areas. However, when adding in population and using EB rates, this relationship flips. While this may simply be the nature of this relationship, I believe this analysis could be improved in a few ways. First, I think that the varying population sizes of counties may be problematic for this analysis even with the use of EB rates. To account for this, one might use census tracts as opposed to counties as the base unit of geography since they are more homogenous in terms of population distribution. Additionally, an investigation of the relationship between bank branches and total population should be pursued. For example, there may be diminishing returns for banks to add more branches—a bank branch may go further for the first 1,000 people of the population than the last 1,000. This may lead to a county with a population of 10,000 having 10 bank branches, but a population with a county of 100,000 having 50 bank branches (as opposed to 100, which would make the relationship consistent).

Next, a similar pattern holds for bank branch closures. When not including population, bank branch closures also cluster in coastal and urban regions. This is opposite to my hypothesis—I believed that urban areas would see very few bank branch closures and rural areas would see more. My reasoning is the documented migration patterns to big cities as well as their relative affluence in the country making them enticing locations for banks to put branches. My results lead me to believe that further investigation of mechanisms is necessary. For example, banks might be leaving rural areas because of the reasons I hypothesized—changes in population or higher rates of poverty. The data show that banks are also closing branches in urban areas too, but the mechanisms here may be different. For example, it may be the case that in cities—places with higher shares of young, college-educated people and widespread availability of high-speed internet—the branch is just not as important for being able to reach consumers. Further investigation of this mechanism is necessary. It is also worth noting that when adding in population and using EB rates, bank branch closure per population hotspots do appear to concentrate in rural areas.

Finally, the regression results give some telling information but ultimately suggest further analysis is needed. As discussed earlier, the relationship between the share of those living in poverty and access to credit is negative. This makes sense—banks generally tend to prefer lending to those with higher incomes. Additionally, a higher ratio of housing units to population has a positive coefficient which also makes sense intuitively. An increase in housing units has a positive association with mortgage originations. The regressions however give mixed results for the explanatory variables of interest—bank branch closures. The indicator variable for losing at least one bank branch is positive (and statistically significant), while the variable for share of bank branch closures per population is negative (though not statistically significant). This might suggest that this variable could be constructed more thoughtfully to reflect a non-linear relationship between bank branch closures and population. Similarly to the mechanism discussed before, it could be the case that closing one bank branch just simply isn’t enough to see a noticeable effect on access to credit. I also think that this model might be improved by incorporating some measure of internet access to account for different mechanisms of lending in different geographies. As mentioned earlier, it could simply be the case that banks can reach consumers much easier through mobile banking in cities (so they will close branches there without having an effect on loan originations) and may be the opposite for rural areas. While this analysis serves as a useful starting point, additional study is needed to assess trends in bank branch closures, mechanisms for closing, and effect on access to credit.

1. Citations

Brainard, Lael. 2020. “Strengthening the CRA to Meet the Challenges of Our Time.” Board of Governors

of the Federal Reserve System. Accessed here

<https://www.federalreserve.gov/newsevents/speech/brainard20200921a.htm>

Census Bureau. American Community Survey (ACS). 2013-2017. Retrieved using tidycensus in R

(Walker).

Consumer Financial Protection Bureau (CFPB). 2012-2017. Home Mortgage Disclosure Act Data.

Retreived from <https://www.consumerfinance.gov/data-research/hmda/>.

Federal Depository Insurance Corporation (FDIC). 2012-2017. Summary of Deposits Data. Retrieved from

<https://www7.fdic.gov/sod/sodInstBranch.asp?barItem=1> .

Kyle Walker (2020). tidycensus: Load US Census Boundary and AttributeData as 'tidyverse' and 'sf'-Ready

Data Frames. R package version 0.9.5. <https://CRAN.R-project.org/package=tidycensus>

Nguyen, Hoai-Luu. 2014. Department of Economics, Massachussets Institute of Technology. Job Market

Paper.

Tabit, PJ and Josh Winters. January 2019. “‘Rural Brain Drain’: Examining Millennial Migration Patterns

and Student Loan Debt.” *Consumer and Community Context,* Board of Governors of the

Federal Reserve System.

1. See Nguyen, 2014 for an analysis of the effect of closings on local economic outcomes. [↑](#footnote-ref-1)
2. See Tabit and Winters, 2019 for a discussion of the effect of migration patterns in rural areas on student loans. Out migration patterns from rural areas may have similar effects on mortgage markets. [↑](#footnote-ref-2)
3. I include all mortgages except for those that are incomplete or withdrawn by the applicant. [↑](#footnote-ref-3)