**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. To analyze the spatial distribution of bank branches I use Local Moran’s I, Getis-Ord (gi\*) Hotspot Analysis, and Local Moran’s I with Empirical Bayes (EB) Rates. Similarly, to analyze the spatial distribution of bank branch closures, I use Local Moran’s I, Getis-Ord (Gi\*) Hotspot Analysis, and Local Moran’s I with Empirical Bayes (EB) Rates. Next, to analyze how bank branch closures affect access to credit, I use Ordinary Least Squares (OLS) regression to analyze if a net negative loss in bank branch access leads to a decrease in mortgage originations, controlling for county-characteristics. I find [results]. The analysis suggests [conclusion].

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. 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 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.

1. Methods

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[[1]](#footnote-1). Using these three datasets, I aggregate and combine them to have a unique county-level dataset to conduct this analysis.

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). [Add in about permutations and false discovery rate].

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.

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 where closed.

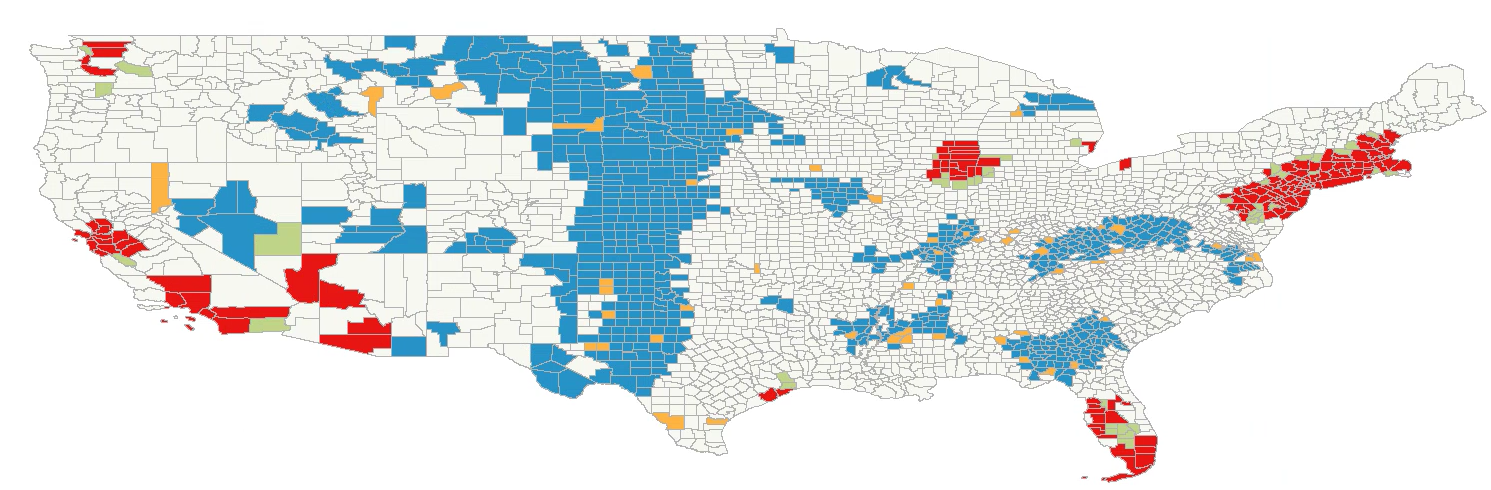
1. Ordinary Least Squares (OLS)

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 and mortgage originations per capita.

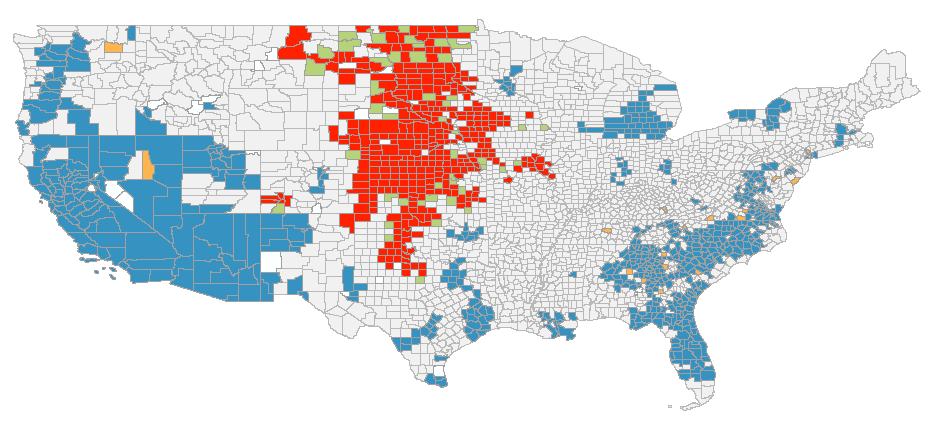
1. Results

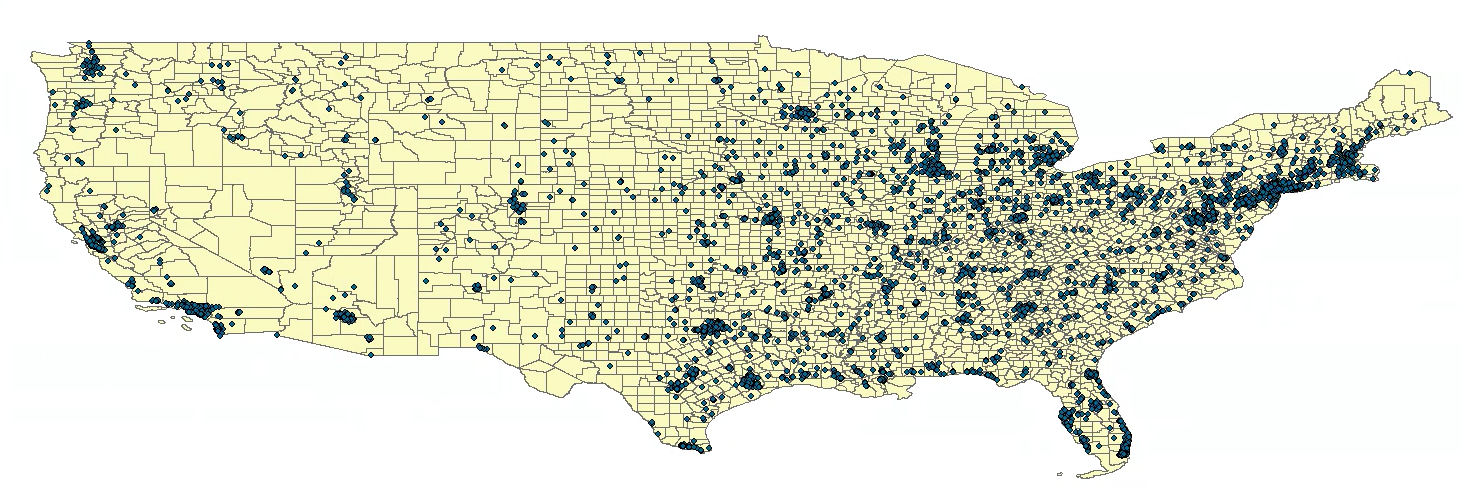
Local Moran’s I

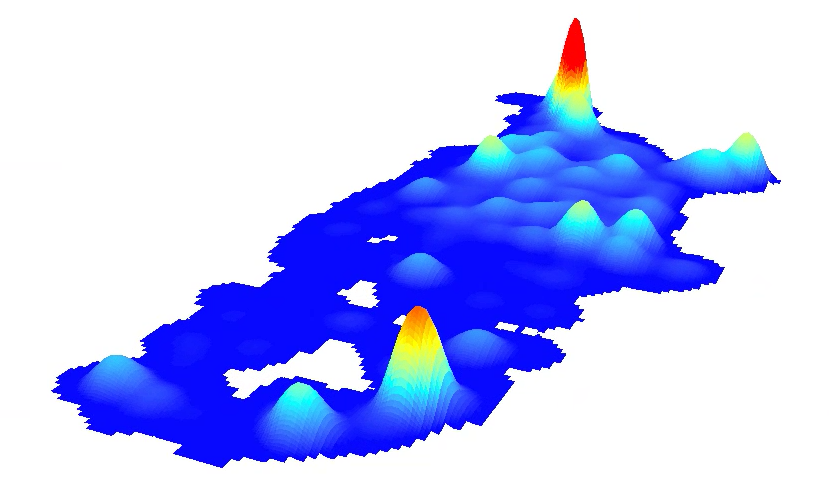
Map 1: Local Moran’s I of Bank Branch Distribution (2017)



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







1. I include mortgages resulting in both (1) approval but no origination; and (2) approval and origination. I include the former because I believe these are still indicators of access to credit—for example, these might be mortgages that are approved but the buyer pulls out of the home purchase last minute. In that case, the issue with the mortgage not originating isn’t related to access to credit (they are still able to get a loan) and should thus be included. [↑](#footnote-ref-1)