Banking Fallout: Spatial Patterns of the 2008 Financial Crisis

Madison Poore

Spatial Data Analysis, Fall 2025

Introduction

For almost a decade, The United States set the framework for one of the largest housing market bubbles in history. Average home prices more than doubled between 1998 and 2006, which is the sharpest increase in US history[1]. Subsequently, with an increased demand for mortgages, lenders created mortgage-backed securities (MBS), which bundled mortgages together to sell as a security with "minimal risk" (since they were backed by credit default swaps, CDS)[2]. However, poor regulations and minimal underwriting led to the inevitable plummet of the stock market and a stark increase in unemployment rates worldwide. While all areas of the United States suffered from the recession, not all regions were impacted equally, which prompts our research question[3].

How well do changes in housing prices and unemployment rates explain the spatial pattern of bank closures across the US throughout the 2008 financial crisis?

To answer this question, I collected data from the FDIC[4], Census Bureau[5], Bureau of Labor Statistics (BLS)[6], and Zillow[7]. The specific use of each source will be discussed in a following section.

[1] Federal Reserve
[2] Corporate Financial Institute
[3] TIME Magazine
[4] FDIC Data
[5] Census Data
[6] BLS
[7] Zillow Data

Data Exploration

The data we will analyze includes an observation for each of the mainland 48 states over four years in question - 2007, 2009, 2011, and 2013. The data contains the following variables: state abbreviation (AL, CO, etc.), year of observation, total number of branches within the state (FDIC), total number of bank closures since the last year recording period (FDIC), the closure rate of banks within the state (my calculation), state-wide unemployment rate (BLS), and a smoothed housing price index (Zillow). There is more information regarding the housing price index (HPI) in the appendix. Our response variable is the closure rate, which is calculated by the total number of closures in the state divided by the total number of branches in the state. Since the first reporting period tracked in this report is 2007, we will start observing bank closures in 2011. That is, 2007 is the basis for bank closure calculations, so there is no interesting rate to be observed there.

```
> summary(df)
  state_abbr             year      total_branches   total_closures      closure_rate       unemployment_rate      HPI
 Length:364         Min.   :2007   Min.   :  22.0   Min.   :-341.000   Min.   :-0.098039   Min.   : 2.400    Min.   : 94260
 Class :character   1st Qu.:2008   1st Qu.: 138.0   1st Qu.:  -2.000   1st Qu.:-0.009269   1st Qu.: 5.400    1st Qu.:132401
 Mode  :character   Median :2010   Median : 402.0   Median :   0.000   Median : 0.000000   Median : 7.650    Median :166575
                    Mean   :2010   Mean   : 849.3   Mean   :   0.978   Mean   : 0.001681   Mean   : 7.338    Mean   :189193
                    3rd Qu.:2012   3rd Qu.:1056.5   3rd Qu.:   5.000   3rd Qu.: 0.015311   3rd Qu.: 9.000    3rd Qu.:226657
                    Max.   :2013   Max.   :7162.0   Max.   : 194.000   Max.   : 0.189815   Max.   :13.700    Max.   :477112
```

Figure 1: Numerical Summaries of Variables

As seen in Figure 1, the bank closure rate may be positive or negative. A positive bank closure rate indicates that banks are closing within the state, while a negative closure rate indicates banks are *opening* within the state. We can visually understand the closure rate variable in Figure 2. Because the values of our response are not all positive, we cannot model this data with a typical Poisson process. Therefore, we will implement a SAR and

CAR model to study the spatial dependence of residuals. Furthermore, since this paper discusses implementation of a SAR and CAR model, we may choose covariates to describe the bank closure rate throughout the chosen years. To summarize the covariates, unemployment rate is a percentage as provided by the Bureau of Labor Statistics for each state. HPI is a housing price index determined by Zillow to calculate the average home price within the state for the given year. Visual representations of each variable follow.

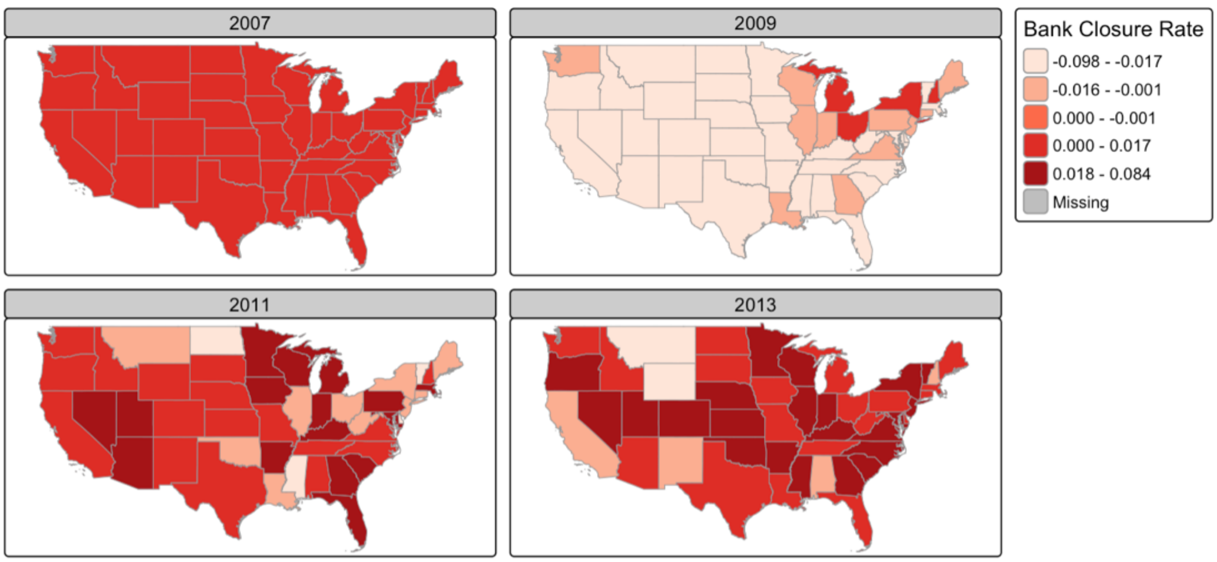

Figure 2: Bank Closure Rate - Calculated by number of banks closed / total number of banks

Again, As seen in the graphic, 2007 is the "reference level" for the data, so all bank closure rates are constant. However, in 2009, we can calculate bank closures since 2007. In 2011, we calculate closures since 2009, and so on. A pattern to take note of before analysis is the drastic increase in the number of banks between 2007 and 2009, which is representative of the housing bubble that existed in 2008. However, the recession that occurred after the housing bubble appears in 2011 and 2013. We begin seeing repercussions of the crash in

2011, a couple of years *after* the stock market crashed and people began to struggle financially.

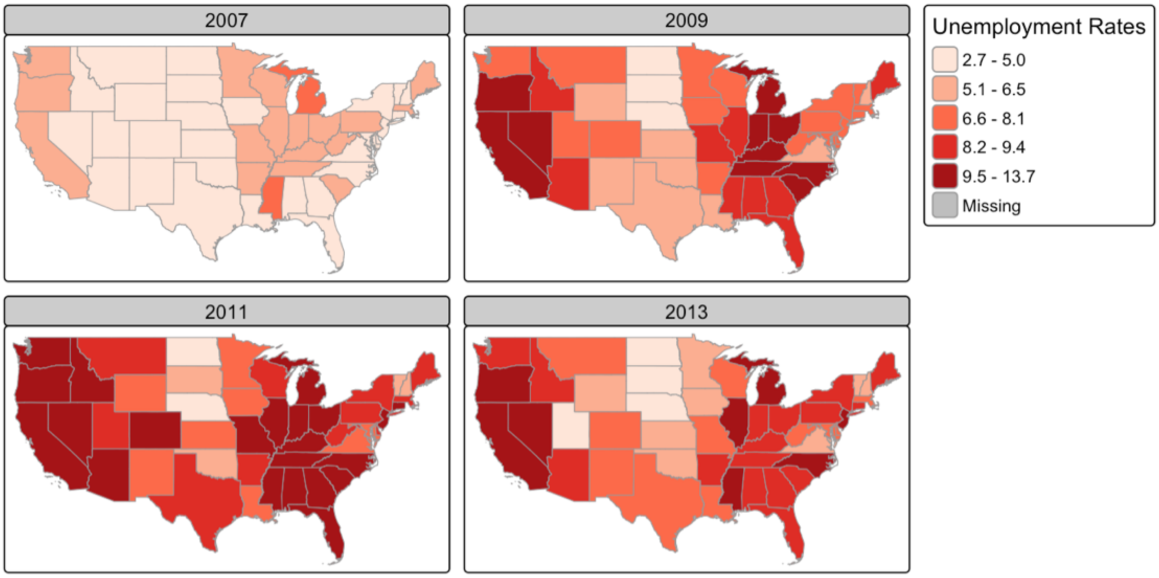

Figure 3: Unemployment Rates Plots

In Figure 3, we can see the implications of the recession reflected in unemployment rates. Rates are very low across the nation in 2007, and they drastically increase (and stay elevated) in 2009. As recession appears across the United States, companies can no longer afford to employ people, leading to layoffs and high unemployment rates. However, unemployment rates seem to spike faster than bank closure rates. This may be due to banks laying off employees to stay open, which provides a *short-term* solution. However, layoffs cannot save banks in an extreme recession like 2008, which leads to the ultimate spike in bank closure rates in 2011 and 2013.

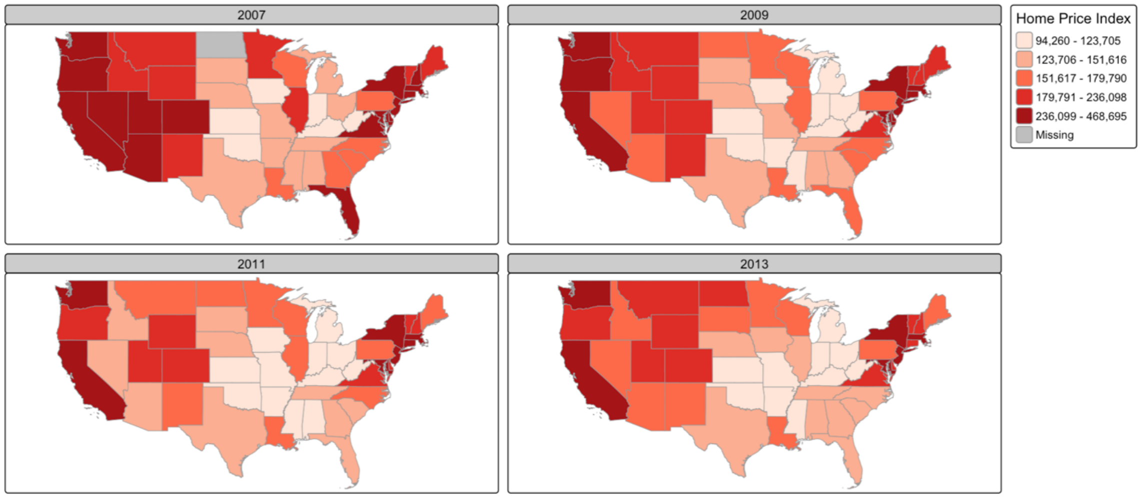

Figure 4: Home Price Index - Calculated by Zillow

The housing price index seems relatively consistent within different regions across the United States over the four years in study. However, we can see prices start to decline slowly as the US enters recession in 2009. Since the 2008 recession was caused by a housing bubble, it is natural to expect home prices to decrease after the bubble "pops". Overall, our data consists of two covariates (unemployment rates and housing price indices), and our target variable is bank closure rate (total bank closures / total branches).

Application of Methods

Before fitting the SAR/CAR models mentioned in the Data Exploration section, we must determine if there is any spatial autocorrelation in our response variable (closure rate). Moran's I is a standard test for spatial autocorrelation that has a null hypothesis of $\rho = 0$ (indicating there is no spatial autocorrelation), and an alternative hypothesis of $\rho \neq 0$ (indicating there is evidence of spatial autocorrelation). The only year that has a significant Moran's I p-value (0.01) is 2013. Therefore, the only year for which we have statistically significant evidence for spatial autocorrelation is 2013. 2007, 2009 and 2011 likely will not drastically benefit from a model with a spatial dependence term and may be adequately described with a regression equation. Within the context of this problem, we somewhat expect for bank closures to take some amount of time to “show up” after the stock market crashes. Therefore, we see steady, nationwide opening of banks during the housing bubble (2007). In 2009 and 2011, banks begin employee layoffs and cost cuts to help stay in business, and many of the newly opened banks are *still* open. Finally in 2013, banks begin to fail more frequently. Therefore, observing spatial autocorrelation (and an increase) in bank closures in 2013 makes contextual sense.

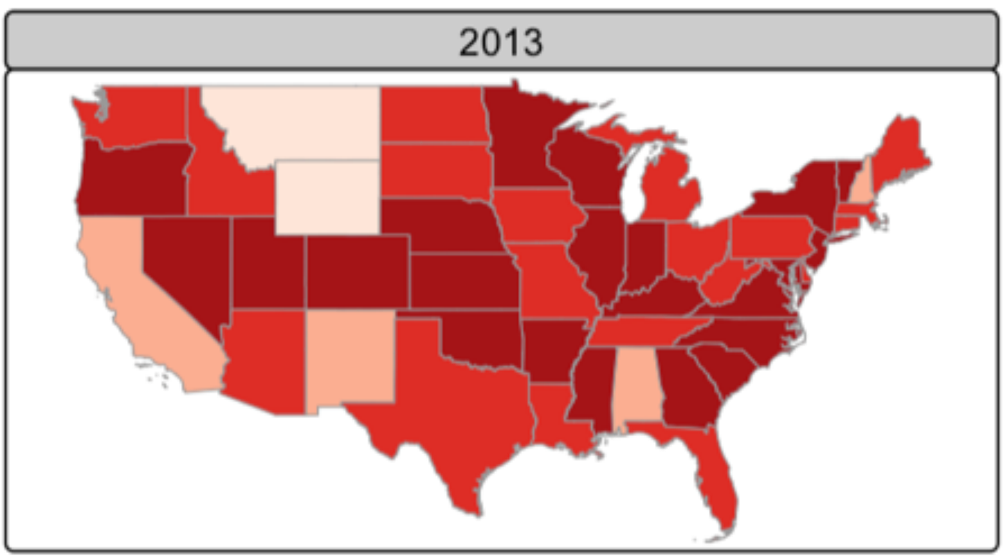

Figure 5: Bank Closure Rate in 2013

Now that we have narrowed our attention to 2013, we can fit the SAR and CAR models with the covariates previously determined (unemployment rates and housing price index). Both SAR and CAR parameters share the same spatial proximity matrices and observation weights. Both models define a “neighboring” state to be one that shares *any* border (even a corner or a point). This is called the queen's contiguity spatial proximity matrix, and it is helpful to give a more comprehensive definition of neighbor when studying the interaction between states[8]. We will also weight observations to give a priority towards observations that exist in a state with many of total banks. This gives us more stable bank closure rates by prioritizing observations that are likely to have a small variance when compared to observations that have sparse data.

Results

While the 2013 SAR and CAR models present similar results on the surface, we can run Moran's I again on the residuals to understand which model performed better.

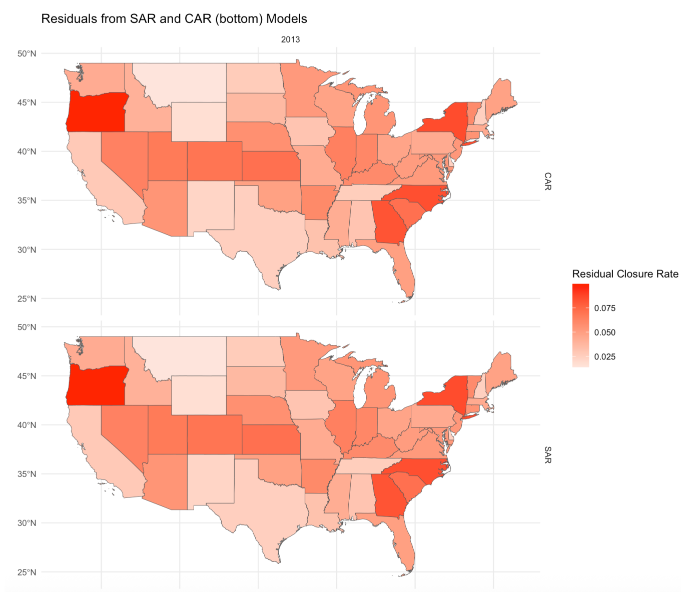

Figure 6: Residuals from SAR (top) and CAR (bottom) models

[8] Spatial Proximity Matrices

When looking at a choropleth plot of the residuals, both models look nearly identical. However, when we turn to a statistical test on the residuals (like Moran's I), we can see more clear results.

Model	**Lambda**	**Significant Coefficient**	**Residual Variance**	**Moran's I of residuals**
SAR	0.17165	HPI	0.26324	Moran's I: 0.131 p-value : 0.06
CAR	0.09591	HPI	0.26707	Moran's I: 0.166 p-value: 0.028

Figure 7: Statistical Results from SAR and CAR Model Fits

After assessing the Moran's I test on the residuals, we choose the SAR model for 2013 as the "best" model. The p-value for the SAR model is larger than the p-value for the CAR model ($0.06 > 0.028$), indicating there is *less* evidence for spatial autocorrelation in the residuals of the SAR model (since we fail to reject the null hypothesis of no spatial autocorrelation, or $\rho = 0$). This means that the SAR model removes more spatial autocorrelation than the CAR model and therefore is the more descriptive model. Now that we have decided the SAR model is our final model, we will unpack the equation:

$$\mathbf{closure\ rate} = \mathbf{0.013} + \mathbf{0.002}(\mathbf{unemployment\ rate_i}) - (\mathbf{9.27}\boldsymbol{e} - \mathbf{8})\mathbf{HPI_i} + \mathbf{0.172}\sum_{\boldsymbol{j}} \boldsymbol{w_{ij}}\mathbf{closure\ rate} + \boldsymbol{\epsilon_i}$$

To interpret our final SAR model, we will step through the equation term-by-term. Starting with the intercept, 0.013, we learn that when all predictors are zero, the closure rate is

0.013. The coefficient of unemployment rate, 0.002, tells us a one percentage point increase in unemployment rate results in a 0.002 increase in closure rate. The Housing Price Index coefficient, $9.27e-8$, tells us a one percentage point increase in the average home price will cause a decrease in closure rate by $9.27e-8$. Contextually, this makes sense since lowering housing prices can indicate an economy is struggling, and hence banks could be struggling. The remaining term, $0.172 \sum_j w_{ij} \text{closure rate}$, is the spatial lag term, which differentiates the SAR model from a standard regression model. The closure rate is observed values in surrounding areas, and w_{ij} is an iterative term from the spatial proximity matrix that was previously mentioned. Therefore, the spatial lag term is a weighted average of neighboring bank closure rates. It is important to note that the only statistically significant coefficient, as referenced in *Figure 7*, is the housing price index, which is also the coefficient with the smallest magnitude. However, since this SAR model removes spatial autocorrelation in 2013, this is a reasonable final model.

Applications

Understanding the spatial autocorrelation of bank closures in the 2008 housing crisis has several applications across many industries. Perhaps most prominently, a SAR model (as discussed in this paper) may be used to understand risk of a real estate investment and determine the economic stability of investing in a specific region. Second, spatial data analysis of the 2008 crisis may be used in financial and credit risk modeling. That is, the 2008 financial crisis failed to perform adequate underwriting on loans, but spatial analysis can help incorporate risk in credit applicants based on regional “spillover” risk. Finally, spatial models may help inform public policy to avoid negative implications of such drastic business cycles in the future. There are many further applications that this analysis may be used for, but the instances mentioned previously are the most prominent.

Conclusions and Future Work

This project may be extended in several ways to reach additional conclusions about the 2008 financial crisis. First, this data does not account for mergers, banks splitting, or other strategic shifts in banks. Rather, if a bank goes out of business, it simply will disappear from the next FDIC report. A more descriptive project may track specifically what happened to banks: if they went out of business, changed names, split, etc. Beyond accounting for intricacies in this dataset, an extension of this project may look like comparing the 2008 financial crisis to other financial crises throughout history. This may give us indicators to what happens before drastic recessions and how to mitigate risk.

Appendix

Zillow Home Value Index (ZHVI) **-**

A measure of the typical home value and market changes across a given region and housing type. It reflects the typical value for homes in the 35^{th} to 65^{th} percentile range. Available as smoothed, seasonally adjusted measure and raw measure. Used as measurement from June each year in question (2007, 2009, 2011, 2013).