The Impact of Storm Damage on Homeowners Insurance Premiums Across the United States

Timothy DeBord
Lawrence University

Faculty Advisor: Professor Andrew J. Sage
Faculty Advisor: Professor Jonathan Lhost
Faculty Advisor: Professor Abhishek Chakraborty

March 18, 2025

Abstract

As the United States experiences an increasing number of disasters each year, the effects of climate change are becoming more evident. This paper examines the relationship between storm damage and homeowners insurance premiums. Using damage covered by homeowners insurance as the treatment variable, I applied a spatial error model that uses two-way state and year fixed effects to analyze its impact on premiums. The results indicate a small positive effect of covered storm damage on homeowners insurance premiums.

Introduction

Climate change is a looming threat that all of humanity faces together. In recent years, wildfires, hurricanes, and other deadly natural disasters have become more frequent. In 2024 alone, California was ablaze, while the Atlantic Ocean unleashed an unprecedented number of hurricanes on the United States East Coast. Scientific theory suggests that as global ocean surface temperatures rise, the number of tropical storms and hurricanes will also increase due to the Coriolis effect, which causes storms to rotate and influences their movement based on the Earth's rotation and wind patterns. These events leave destruction in their wake, often resulting in costly damages. As the risks of climate change grow, so should the cost of insurance that covers storm-related damage. Specifically, homeowners insurance premiums reflect an insurer's perceived risk of insuring a client and the surrounding area. The higher the risk, the more premiums should rise. In my analysis I will look at storm data from the National Oceanic and Atmospheric Administration (NOAA) to examine their effect on homeowners insurance premiums between 2008 and 2019 by using a spatial error panel model by maximum likelihood in R.

Intuitively, I expect to find a positive significant relationship between damage covered by homeowners insurance and homeowners premiums. The more damage a state receives, the higher the amount insurance companies will have to pay in claims, and therefore, premiums should be higher to account for the risk of damage. Furthermore, the effect of damage not covered by homeowners insurance is less clear. If the damage is covered by a different type of insurance, I expect to see a negative effect on homeowners premiums due to the economic substitution effect. It is also reasonable to believe that there will be no effect, as the damage should hypothetically not impact homeowners premiums if it is excluded from how insurance companies calculate premiums.

The existing literature supports the idea that homeowners insurance premiums are related to climate risk. In 2024, Benjamin J. Keys and Philip Mulder analyzed the relationship between property insurance and disaster risk and found that premiums have risen sharply since 2020, with this growth concentrated in disaster-prone ZIP codes (Keys et al., 2020). Keys and Mulder split their natural disaster risk variable into disaster risk and climate risk, and I used that idea as a basis for separating damage into coverable and non-coverable damage depending on the storm. Additionally, when accounting for socioeconomic status and demographic variables, they used the percentage of the population that is not white as an explanatory variable. Their reasoning was that lower-income areas see much higher premiums as a share of income, and ZIP codes with larger nonwhite population shares pay higher premiums even after controlling for disaster risk and income (Keys et al., 2020).

Further, in 2012, Randy E. Dumm et al. tested the hypothesis of whether increases in insurance premiums serve as a risk signal and negatively impact house prices. Dumm et al. found a negative and significant relationship between insurance premiums and house prices in each regression (Dumm et al., 2012). Dumm et al.'s use of home prices helped me decide between the importance of average home prices and median household income, and I ultimately chose the former. Additionally, their findings provided intuition behind the relationship between home prices and premium prices. In high-risk areas, they found that places with historically high damage had reductions in home prices but still had to pay higher premiums.

Data

The data in this paper merges information from three different sources: the National Oceanic and Atmospheric Administration (NOAA), Kaiser Family Foundation (KFF), and Zillow. After dropping Alaska and Hawaii, the final spatial panel dataset includes 48 states from 2008 to 2019, totaling 576 observations across 224 variables. It also contains spatial polygons for each state, allowing me to analyze the spatial relationships between homeowners insurance premiums and storm damage. Table 1 presents the first five rows of the dataset, highlighting the key variables used in my analysis: homeowners insurance premiums, total home damage, total natural disaster damage, average home price, percentage of the population that is white, and population density.

Table 1: Selected variables from the first 5 rows in the dataset

State	Year	Home Premiums	Home Damage	Disaster Damage	Avg Home Price	Percent White	PopDensity	geometry
Alabama	2008	845	28643750	584500	133632.5	0.69	86.36	POLYGON ((-85.12733 31.7625...
Arizona	2008	628	522500	12842000	223729.0	0.58	55.82	POLYGON ((-110.7507 37.0030...
Arkansas	2008	788	477033400	220277000	123088.1	0.76	52.08	POLYGON ((-90.95577 34.1187...
California	2008	911	70609200	204602500	377991.5	0.42	219.20	MULTIPOLYGON (((-119.9999 4...
Colorado	2008	842	159140500	6987500	229457.2	0.71	46.27	POLYGON ((-105.155 36.99526...

Looking at Table 1, the geometry variable provides the dataset with its spatial attributes. The NOAA dataset includes every storm recorded in the United States between 1950 and 2024. After filtering to keep only storms that occurred between 2008 and 2019, each row either contained a latitude and longitude coordinate for a storm or had a missing value. To address this, I used a shapefile from the Census Bureau that provided polygon data for each state and merged each storm with the state in which it was recorded. This effectively transformed the spatial data from point data to polygon data, aligning it with the rest of my dataset, which is at the state level. For storms that affected multiple states, each state recorded the event and reflected the corresponding storm data.

After restructuring the dataset, I created the damage variables. Home damage represents damage caused by storms that are typically covered by homeowners insurance, including tornadoes, excessive heat, heavy snow, high winds, hail, winter storms, blizzards, ice storms, strong winds, and lightning. Disaster damage includes damage from events that generally aren't covered by homeowners insurance and often require specialized coverage, such as hurricanes, floods, flash floods, storm surges, wildfires, coastal floods, lakeshore floods, and tsunamis. I separated these damage types because damage typically covered by homeowners insurance should explain some variation in insurance premiums, while damage outside standard coverage may have a less direct impact on pricing. Looking at the left map in Figure 1, we can see that areas like Tornado Alley experience more damage than regions that don't frequently get tornadoes. Historically, Tornado Alley refers to the central United States, specifically the region stretching from Texas through Oklahoma, Kansas, and into the Dakotas. However, the overall distribution of home damage is fairly spread out, with no extreme outliers. When looking at disaster damage, a different pattern emerges. Florida, Louisiana, and especially Texas have sustained significantly more damage than the rest of the country. This is largely due to the devastating hurricanes that hit these regions between 2008 and 2019.

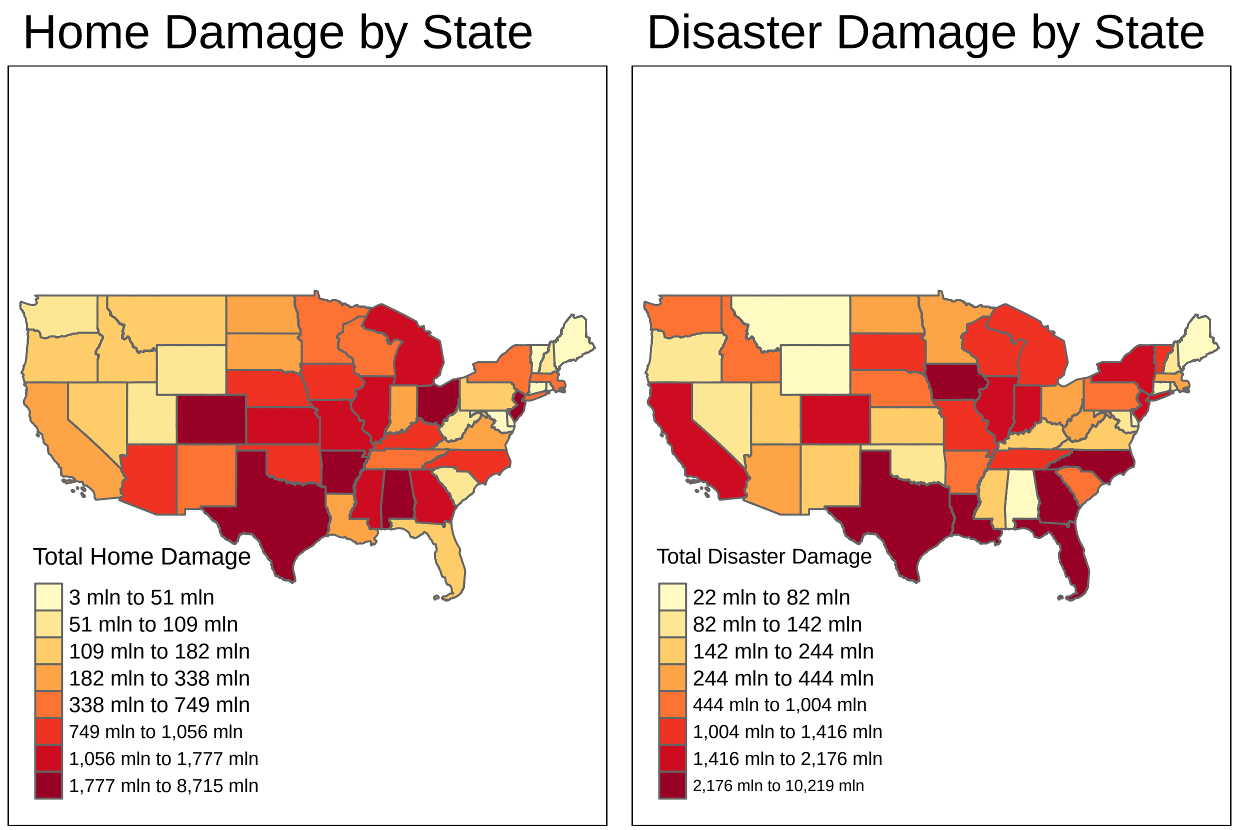


Figure 1: Left map shows home damage, right shows disaster damage (mln = million)

Storms in the United States have historically been classified into nine climate regions: Northeast, Upper Midwest, Ohio Valley, Southeast, Northern Rockies and Plains, South, Southwest, Northwest, and West. Although premiums are not determined at the regional level, states within the same region experience similar risks and hypothetically should have comparable premium prices. I will use regions as a potential fixed effect rather than state fixed effects, which will be further detailed in the methods section. Below, in Figure 2, the left map from NOAA displays the defined climate regions, while the right map shows the average homeowners insurance premiums from 2008 to 2019. Comparing the two maps reveals that the South region, on average, has the highest premiums, suggesting that the idea of regions sharing similar premiums due to regional risk is not unreasonable.

In addition, the other control variables include the percentage of the population that is white, population density, and average home price. The percentage of the population that is white comes from KFF, formerly known as the Kaiser Family Foundation. This variable helps

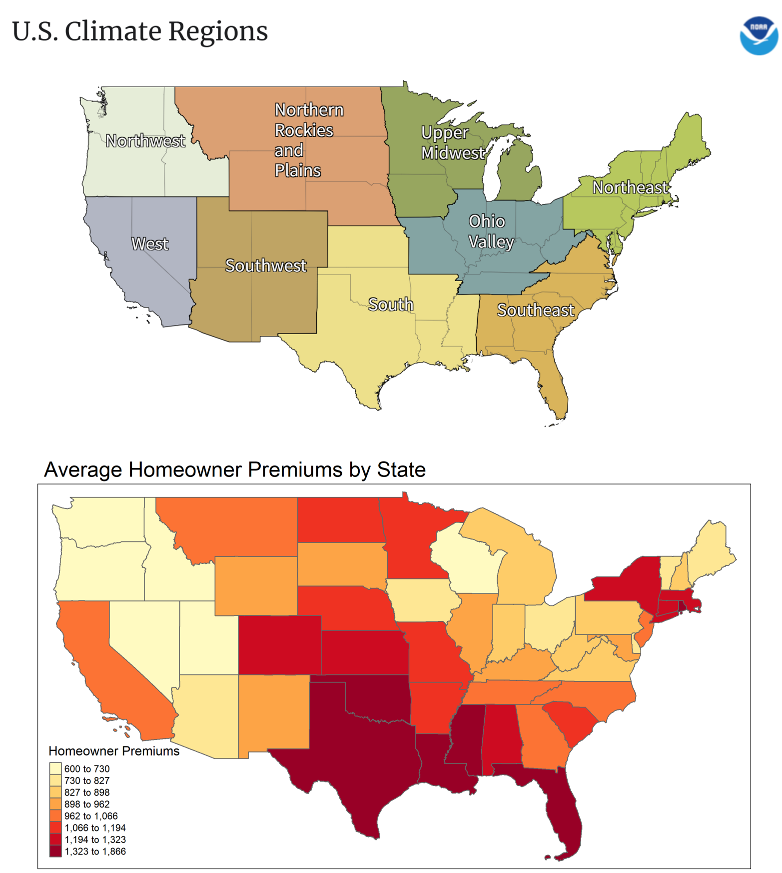


Figure 2: Top map shows regions while bottom map shows homeowners premiums

control for potential racial disparities in insurance pricing, whether due to discrimination or coincidental factors, both of which could influence homeowners insurance premiums. Population density is calculated by dividing the total number of residents in each state by the state's area in square miles, using data from KFF and the Census Bureau. Accounting for each state's population density should impact home insurance because, in densely populated areas, insurance companies have a larger pool of policyholders, which helps distribute risk. Conversely, in less densely populated areas, the risk pool is smaller, meaning premiums may be higher to compensate for the increased financial exposure per policyholder. Finally, the last control variable is average home price, which I got from Zillow. The average home price should directly impact homeowners insurance premiums since the higher the property value, the more there is for an insurance company to cover under a single policy, leading to higher premiums.

Methodology

The main goal of this paper is to examine the spatial relationship between homeowners insurance premiums and storm damage. To do that, we first need to test whether an Ordinary Least Squares (OLS) regression shows evidence of spatial autocorrelation. From there, we have to define the neighbors and weights of the model, which will allow us to incorporate a spatial element into the calculations. Once that's set up, we can build the final model using the spml function from the splm package in R (Millo et al. 2012).

Before deciding whether to build a spatial model, I first need to run an OLS regression, as shown in Table 2. The regression output shows that the treatment variable — damage that would be covered by homeowners insurance — is not statistically significant, meaning that their p-values are greater than 0.1. Likewise, damage that wouldn't be covered by homeowners insurance is also not statistically significant. However, the average home price, the percentage of a state's population that is white, and population density are all statistically significant. That said, these results only matter if there's no evidence of spatial autocorrelation.

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Tue, Mar 18, 2025 - 12:18:15 PM

To test the OLS regression for spatial autocorrelation, I ran a Moran's I test and a Lagrange multiplier test on the OLS output. Spatial autocorrelation measures the extent to which values of a variable at nearby locations are correlated. If there is evidence of spatial autocorrelation in the OLS model, the independence assumption of OLS is violated, meaning we can't rely on those results. The null hypothesis of the Moran's I test states that there is no spatial autocorrelation. Visually, if no spatial autocorrelation is present, the Moran's I statistic would fall within the distribution shown in Figure 3. However, since the I value was significant, it falls outside the distribution—indicating evidence of spatial autocorrelation.

The OLS regression has more issues than just failing to account for spatial autocorrelation — it also doesn't account for the correlation between states and years since I'm using panel data. To address the structure of the data, I'll use a Panel Data Estimator (plm), also known

Table 2: Regression Models

	Dependent variable:
	homePrems OLS
log(home_damage)	2.485
	p = 0.451
log(disaster_damage)	−0.235
	p = 0.936
log(AvgHomePrice)	-129.963^{***}
	p = 0.004
log(Percent_white)	-504.555^{***}
	p = 0.000
log(PopDensity)	29.355^{**}
	p = 0.015
Constant	$2{,}260.145^{***}$
	p = 0.00004
Observations	576
R^2	0.126
Adjusted R^2	0.118
Note:	*p<0.1; **p<0.05; ***p<0.01

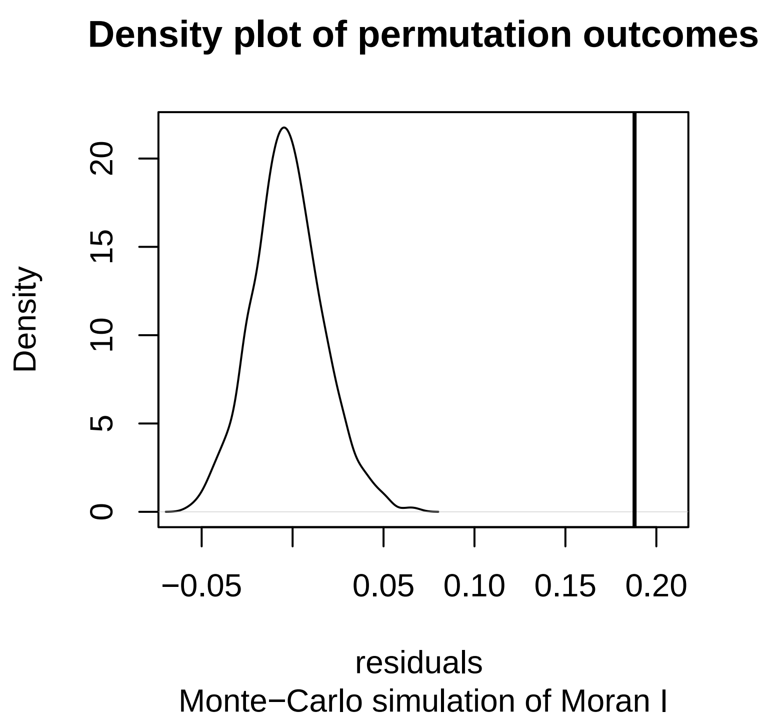


Figure 3: The Moran's I test results

as a fixed effects model. Applying a fixed effects model to my panel data effectively creates a dummy variable for every state and every year, meaning there are dummies for all 48 states and 11 years in my dataset. This overfits the data, and when I run a Moran's I test on the fixed effects model, it produces a statistically insignificant I statistic. To properly test for spatial autocorrelation, I instead use a Lagrange Multiplier test.

A Lagrange Multiplier test checks for spatial autocorrelation in the form of spatial lag and spatial error. Spatial lag occurs when the y-value of one state influences the y-value of a neighboring state—meaning that if spatial lag exists, homeowners insurance premiums in one state affect premiums in neighboring states. Spatial error happens when the error terms of one state are correlated with those of neighboring states. Both cases violate the independence assumption of OLS. Using the Lagrange Multiplier test, I found evidence of both spatial lag and spatial error. (The regression results for the fixed effects model are included in the appendix in section a.1.) Looking at Table 3, both tests are significant based on their p-values. However, despite both being significant, I chose to use a spatial error model since I'm focused on explaining the relationship between homeowners insurance premiums and storm damage rather than how premiums in one state influence those in neighboring states.

Table 3: Spatial Dependence Tests for Fixed Effect Model

Test	Statistic	P_Value
Error Test	4.743595	0.0294076
Lag Test	127.762395	0.0000000

As stated in the data section, I transformed the geometry from points representing latitude and longitude to polygons representing states. This allows my model to fully utilize the state-level data I've collected while maintaining spatial information. Using these polygons, I can create a neighbors matrix that feeds into an equation to generate the spatial weights required for the model. I chose to combine two different methods to construct the neighbors matrix.

The first method is distance-based neighboring, where neighbors are determined by measuring the physical distance between points. In this paper, I used the centroid — the center point of a polygon — as the reference point. If a centroid falls within the radius of another centroid, the two states are considered neighbors. However, if any part of a polygon falls within the radius but its centroid does not, they are not considered neighbors. This distinction is illustrated section a.2 of the appendix. In Figure 4, we see the results of using a 525-kilometer radius to define the neighbors matrix. I chose 525 kilometers because any smaller radius would leave Texas without neighbors, which causes an error in R. I tried to keep the radius as small as possible to properly account for the Northeast, where states are densely packed. If I were to have made the radius too big, Maine, for example, would incorrectly be considered a neighbor of West Virginia. While the East Coast appears to be accounted for appropriately, issues arise further west, where many states end up with only one neighbor. To address this, I combined an additional method to refine the neighbors matrix.

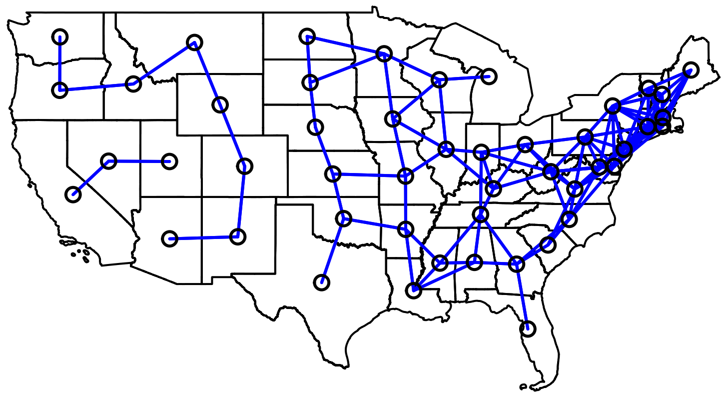


Figure 4: Map of neighbors with only distance-based neighboring

The second method is rook contiguity, where two states are considered neighbors if they share a side — similar to the movement of a rook in chess. If I had also allowed states that share only a vertex to be neighbors, this would be called queen contiguity. (For a visual reference, see section a.3 in the appendix.) In Figure 5, we see the result of combining both methods, which best represents the United States. This approach ensures that the highly clustered Northeast is properly connected while also maintaining appropriate connections across the rest of the country.

The neighbors matrix alone isn't enough for the model to account for the spatial nature of the data, so I also needed to specify a weight matrix to incorporate spatial weights. A spatial weight quantifies the relationship between neighbors based on their proximity or connectivity, defining how much influence one unit has on another in a spatial analysis. The most common approach is to assign equal weights by dividing 1 by the total number of neighbors a polygon has. However, this method doesn't make sense when using distance-based neighbors. For example, New Jersey's effect on New York shouldn't be the same as its effect on West Virginia. To better capture this relationship, I use an inverse distance weight matrix, which penalizes states that are farther from the base neighbor while also accounting for the number of neighbors a state has (the equation is provided in section a.4 of the appendix). This ensures that the model receives a weighted matrix that more accurately reflects the spatial relationships across the United States. Looking at Figure 6, we can see

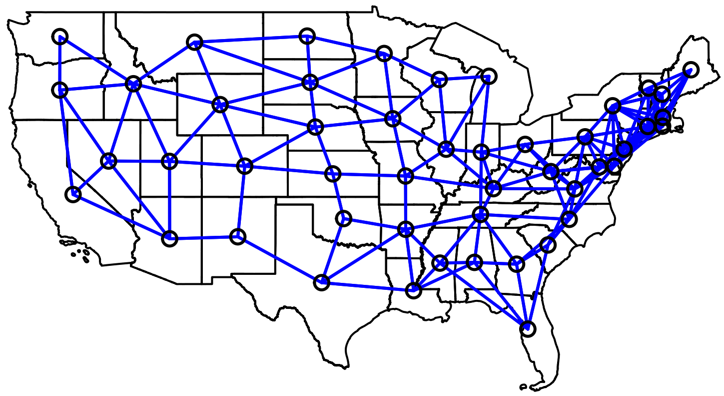


Figure 5: Complete neighbor map using both distance-based neighbors and rook contiguity

Wisconsin and its four neighbors: Michigan, Iowa, Illinois, and Minnesota. Since Michigan's centroid is the closest to Wisconsin's, it has the largest weight.

With the spatial weights defined, the only decision left before building the final model is the functional form of each variable. Except for homeowners insurance premiums, all variables will be logged to address outliers. The graphs showing both the logged and unlogged versions of these variables can be found in section a.5 of the appendix. The remaining variables will be kept in their original forms.

The final model is a spatial panel model by maximum likelihood (spml) from the splm package in R (Millo et al. 2012). Using a spml allows me to specify the weights, lags, errors, and effects for the model. Given my research question about the relationship between homeowners insurance premiums and storm damage, the spml is specified as a spatial error model, as I am not interested in analyzing the effect of a state's premiums on nearby states' premiums. The inverse distance weighted matrix will be used to measure spatial relationships. After indexing the data by state and year to capture the panel aspect of the data, I need to decide between random effects and fixed effects. Random effects should be used if the variable needs to be controlled for to remove its influence, while fixed effects should be used if the variable is expected to influence the response. I expect state and year to influence homeowners insurance premiums because state characteristics that do not change over time should have some unobserved effect on premiums, and the year should capture shocks that affect all states.

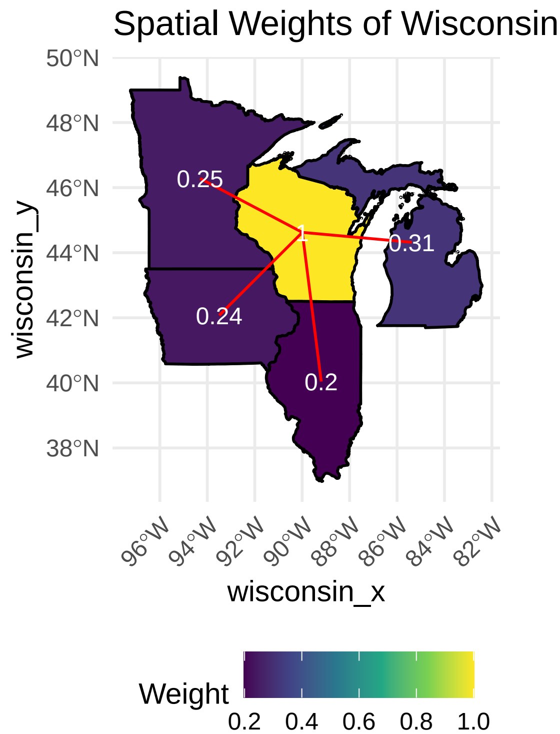


Figure 6: Map of the weights of Wisconsin

I compared four different fixed effects models using home damage covered by homeowners insurance as the treatment variable. Each model included the same explanatory variables to determine the best final model. Fixed effects demean the data, allowing us to measure the effect of the variables relative to the fixed effects mean rather than just their direct impact. Demeaning centers the data around the variable's average. For example, if State A's homeowners insurance premium is at its average in 2008, the model assigns it a value of zero because it aligns with its mean. If in 2009, State A's premium rises \$5.70 above its average, then the value becomes 5.70. I use a combination of state and year fixed effects. The example above illustrates how the model demeans the data when using only state fixed effects. If I were to use only year fixed effects, then all 48 states would be centered around the national average across all states. When I include both state and year fixed effects, the model applies both demeaning processes simultaneously.

The econometric intuition behind using fixed effects is to control for unobservables within and between groups. An unobservable is an effect that isn't directly included as a variable in our data, but we can still account for it using fixed effects at the state and year levels. Starting with Model 1, I include only state fixed effects, which control for unobserved characteristics within a state that remain constant over time. For example, a state's location on the coast or in Tornado Alley increases homeowners insurance premiums due to heightened risk of damage, making it a fixed characteristic. Model 2 applies year fixed effects, which account for unobservable shocks that affect all states equally over time. By combining state and year fixed effects, I control for both state-specific and time-specific unobservables. Region fixed effects work similarly to state fixed effects but apply to larger geographic areas. However, because regions encompass more diverse states, it's harder to assume that unobserved factors remain constant within each region over time. Given this limitation, my final model will use state and year fixed effects. In my analysis, I will compare all fixed effects models to evaluate how they account for unobserved factors and influence the results.

When specifying the spml model in R to create a spatial error model, you can choose between the `b` (Baltagi) or `kkp` (Kapoor) methods (Millo et al., 2012). In my final model, I used Baltagi's error method, but using Kapoor's error method instead would not have changed the results. This is demonstrated in section a.6 of the appendix, where I re-estimate all four models using the `kkp` error method. In the spml model, I also specified a two-way state and year fixed effects model. Additionally, I set lag to equal FALSE since I only want to control for spatial autocorrelation in the error terms. As mentioned earlier, my focus is on how storm damage affects homeowners insurance premiums, rather than how one state's premium prices influence those of nearby states.

The final model is a spatial error model that includes state and year fixed effects. ϵ_{it} is a vector of spatially autocorrelated observations that follow a spatial autoregressive process (Millo et al. 2012). The Kronecker product of the identity matrix I_T (which represents the time dimension) and the spatial weights matrix W_N (which describes the spatial neighbors) ensures proper structuring of the spatial dependence. The identity matrix is an 11×11 matrix with ones along its diagonal, meaning that if the values match (i.e., the years align), there is a one. This identity matrix ensures that the model does not allow for temporal correlation. The weight matrix, W_N, is a 48×48 matrix since there are only 48 states in the data. The matrix consists of either zeros or weight values, with each row summing to one. When we take the Kronecker product $I_T \otimes W_N$, the resulting matrix is a massive 528×528 matrix with zeros everywhere except along the diagonal, where each block is an identical 48×48 copy of W_N. This ensures that the same spatial weights are applied for each state across all years. The spatial autoregressive parameter, ρ, captures the strength of spatial dependence between state i and state j, constrained to be between -1 and 1. In other words, ρ represents the correlation in the error terms. If there is no spatial correlation in the errors, then ρ will be zero. α_i represents the state fixed effect, while α_t represents the year fixed effect. ν_{it} is the independent error term, uncorrelated with the spatial error term, capturing any randomness not influenced by spatial factors.

$$
\begin{aligned}
\text{homePrems}_{it} = {} & \alpha_i + \alpha_t + \beta_1 \log(\text{home damage}_{it}) + \beta_2 \log(\text{disaster damage}_{it}) + \beta_3 \log(\text{AvgHomePrice}_{it}) \\
& + \beta_4 \log(\text{Percent white}_{it}) + \beta_5 \log(\text{PopDensity}_{it}) + \epsilon_{it}, \\
& \text{where } \epsilon_{it} = \rho(I_T \otimes W_N)\epsilon_{it} + \nu_{it}, \quad |\rho| < 1, \\
& \nu_{it} \sim \mathcal{N}(0, \sigma_\nu^2), \quad \epsilon_{it} \sim \mathcal{N}(0, \sigma_\epsilon^2).
\end{aligned}
$$

Results

First, let's examine the results from different fixed effects models using home damage covered by insurance as the treatment variable. The state fixed effects model produces the smallest coefficient (0.947) and is the only statistically insignificant coefficient for home damage, with a statistically significant p-value being less than 0.1. This suggests that controlling only for time-invariant differences across states does not fully capture the effect of home damage. The year fixed effects model, on the other hand, yields the largest coefficient (4.497) and the most statistically significant result. However, a larger coefficient is not a valid way to determine a better fit, as this model only accounts for common shocks affecting all states and ignores unobserved differences between states. It is likely that the larger coefficient is due to an overestimated, biased model—meaning the model finds a larger coefficient than it should due to improper data fitting. Introducing region and year fixed effects partially addresses this issue, reducing the coefficient to 3.438 while maintaining statistical significance, though to a lesser degree.

Controlling for regions helps account for some unobserved heterogeneity between states, but this assumes states within the same region share identical insurance markets or policy implementation, which is incorrect. Therefore, this approach may still produce an overestimated, biased coefficient. This concern is confirmed when applying state and year fixed effects, which yield a statistically significant coefficient of 1.302. Given the theoretical reasoning behind fixed effects and my research question, controlling for both state and year fixed effects ensures that coefficients reflect within-state variation over time while accounting for nationwide shocks that effect all states equally. For this reason, I will use this model to analyze the rest of the variables.

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Tue, Mar 18, 2025 - 12:18:20 PM

Upon examining the final model in Table 4, we see the expected result: damage that should be covered by home insurance has a positive and statistically significant effect on the price of homeowners premiums. Specifically, when holding damage not covered by home insurance, average home price, percent white, and population density constant, for every 1 percent increase in damage that would be covered by home insurance, we expect to see an average 1.3 cent increase in homeowners insurance premiums. Although this is not a large monetary value, it demonstrates that storms do have an effect on premium prices.

Further, the only other statistically significant coefficient is percent white, which has a negative effect on home premiums. Specifically, for every one percent increase in the white population, homeowners premiums are expected to decrease by $4.67 on average. This finding suggests that racial bias could be a factor, with non-white populations potentially facing higher premiums. However, while discrimination is illegal, it is not impossible. More likely, this result reflects income and wealth disparities or lower credit scores. Coastal states tend to have more diverse populations but also face higher risks of damage and increased demand for coastal living, which could explain why areas with lower percentages of white residents have higher premiums on average.

Population density, average home price, and damage not covered by home insurance are all not statistically significant. Despite being significant under year and region fixed effects, population density having a p-value of 1 suggests virtually no variation within a specific state over time. This is not surprising, as population growth rates across states are generally similar. At first glance, the lack of a statistically significant result for average home price

Table 4: Regression Models

	Dependent variable:			
	State FE	Year FE	Region FE Year FE	State FE Year FE
rho	0.906***	0.859***	0.654***	0.702***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
log(home_damage)	0.947	4.497**	3.438*	1.302*
	p = 0.146	p = 0.016	p = 0.071	p = 0.053
log(disaster_damage)	−0.566	2.469	2.027	−0.706
	p = 0.319	p = 0.119	p = 0.211	p = 0.233
log(AvgHomePrice)	99.984**	34.890	80.103*	29.385
	p = 0.041	p = 0.433	p = 0.068	p = 0.562
log(Percent_white)	−806.182***	−200.056***	−191.767***	−469.499**
	p = 0.0002	p = 0.0003	p = 0.002	p = 0.033
log(PopDensity)	−138.935	116.205***	118.100***	0.217
	p = 0.442	p = 0.000	p = 0.000	p = 1.000
RegionNorthern Rockies and Plains			341.547***	
			p = 0.00000	
RegionNorthwest			−97.546	
			p = 0.170	
RegionOhio Valley			206.137***	
			p = 0.00004	
RegionSouth			462.366***	
			p = 0.000	
RegionSoutheast			238.464***	
			p = 0.00000	
RegionSouthwest			1.133	
			p = 0.986	
RegionUpper Midwest			165.099***	
			p = 0.006	
RegionWest			−21.349	
			p = 0.788	
TRUE				
Note:				*p<0.1; **p<0.05; ***p<0.01

may seem unexpected. However, after demeaning the data, the rate of change in average home price within a given state over time is likely uniform, reducing the likelihood of statistical significance. Lastly, the insignificance of damage not covered by homeowners insurance is expected, as I explicitly separated damage variables into those that should influence premiums and those that should not. The model finds no evidence of a relationship between uncovered damage and homeowners insurance premiums.

To visualize the relationships between the explanatory variables and homeowners insurance premiums, I will use partial dependency graphs, as shown in Figure 7. A partial dependency graph holds all other explanatory variables at their means, except for the variable of interest in that specific graph. Since all the variables are logged, I unlogged the x-axis while keeping the line itself logged. This is why the x-axes in the graphs appear to grow exponentially.

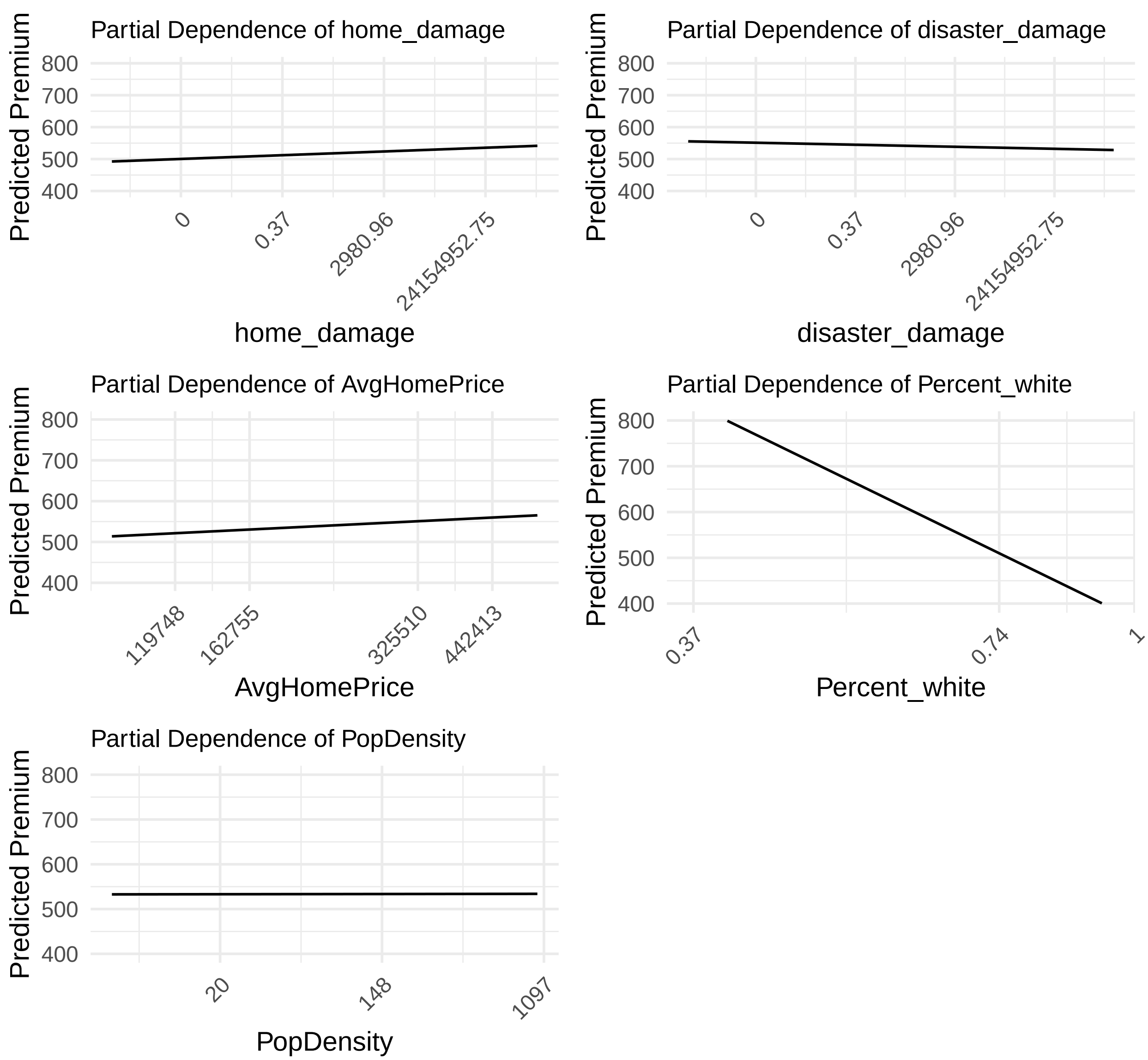


Figure 7: Partial Dependency Plots

Percent white has by far the largest effect, showing an approximate total impact of $400 on premiums between the minimum and maximum values of percent white, holding other explanatory variables constant. In contrast, population density has little influence, with a potential effect of only about a dollar. Disaster damage, which is negatively correlated with premiums, could reflect the need for additional insurance plans, such as flood or hurricane insurance, leading to reduced premiums. However, since disaster damage is statistically insignificant, there is no clear relationship, suggesting it does not affect premiums. The positive effect of average home price on premiums makes intuitive sense, but it is also statistically insignificant, indicating no evidence of a relationship between home prices and homeowners insurance premiums. As mentioned earlier, this is likely due to the uniform increase in home prices across the U.S. and the lack of variability remaining after accounting for state and year fixed effects. Damage covered by home insurance has a statistically significant positive effect on premiums, with the potential to increase the price by $50. However, while this effect is small compared to percent white, damage that would be covered by homeowners insurance can affect the price of premiums, though it mostly appears to be significant only in extreme cases.

Conclusion

Using a spatial error panel model by maximum likelihood while controlling for state and year fixed effects best fits the data and my research question of whether there is a positive statistically significant relationship between homeowners insurance premiums and damage caused by storms between 2008 and 2019. The results provide evidence of a positive relationship. With the growing literature on the increasing risks from climate change, policymakers should focus on ways to reduce the costs homeowners face and find solutions to mitigate the effects of climate change. Areas like California are becoming too risky to insure, leading insurance companies to pull their services from the state. Steps need to be taken to ensure that people in high-risk areas can access affordable homeowners insurance.

The results I found align with my intuitive expectations, but these results are not causal. Lack of access to county-level data limits the accuracy of the estimates, and not having data on reinsurance rates also affects the analysis. Reinsurance, often referred to as insurance for insurance companies, significantly impacts premium prices (Investopedia). Omitting this variable weakens the causal claims this paper can make, but I expect to find similar results with its inclusion.

The model was also unable to account for temporal correlation due to the nature of the identity matrix, where the weights remained the same over time. However, to fully capture the spatial and temporal relationship between homeowners insurance premiums and storm damage, there should be weights that change over time. Future research should explore methods that incorporate both temporal and spatial parameters, as doing so would likely produce a larger effect from damage covered by homeowners insurance.

References

Baltagi, Badi H., Seuck Heun Song, and Won Koh. Testing Panel Data Regression Models with Spatial Error Correlation. Department of Economics, Texas A&M University, 2001, https://econpapers.repec.org/cpd/2002/24_Baltagi.pdf.

Baltagi, Badi H., Peter Egger, and Michael Pfaffermayr. A Generalized Spatial Panel Data Model with Random Effects. 2007, https://economics.ucr.edu/wp-content/uploads/2019/11/Baltagi.pdf.

Bivand, Roger, et al. splm: Spatial Panel Data Models. CRAN, https://cran.rproject.org/web/packages/splm/splm.pdf.

Burchfield, Emily. Spatial Regression Lab. https://www.emilyburchfield.org/courses/gsa/spatial_regression_lab.

Investopedia. Reinsurance. https://www.investopedia.com/terms/r/reinsurance.asp.

Keys and Mulder. Property Insurance and Disaster Risk: New Evidence from Mortgage Escrow Data. NBER Working Paper No. 32579, National Bureau of Economic Research, 2024, https://www.nber.org/system/files/working_papers/w32579/w32579.pdf.

Federal Emergency Management Agency (FEMA). Expected Annual Loss. https://hazards.fema.gov/nri/expected-annual-loss.

Millo, Giovanni, and Gianfranco Piras. splm: Spatial Panel Data Models in R. Journal of Statistical Software, vol. 47, no. 1, 2012, pp. 1–38. https://www.jstatsoft.org/article/view/v047i01.

Moscone, Fosca, and Giuseppe Tosetti. "GMM Estimation of Spatial Panels with Fixed Effects and Unknown Heteroskedasticity." Journal of Econometrics, vol. 41, no. 5, 2011, pp. 487-497. ScienceDirect, https://www.sciencedirect.com/science/article/abs/pii/S0166046211000445.

Princeton University Library. R Guide: Panel Data Analysis. https://libguides.princeton.edu/R-Panel. Stack Overflow. R plm Time Fixed Effect Model. Stack Exchange, https://stackoverflow.com/questions/28359491/r-plm-time-fixed-effect-model.

Yang, Zhenlin. "Quasi-Maximum Likelihood Estimation for Spatial Panel Data Regressions." Research Collection School of Economics, Singapore Management University, 2013, https://ink.library.smu.edu.sg/cgi/viewcontent.cgi?article=2574&context=soe_research

Data Sources

National Centers for Environmental Information. Storm Events Database. National Oceanic and Atmospheric Administration, https://www.ncdc.noaa.gov/stormevents/.

Kaiser Family Foundation. Total Number of Residents. https://www.kff.org/other/state-indicator/total-residents/?activeTab=graph&currentTimeframe=0&startTimeframe=14&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D.

Kaiser Family Foundation. Distribution by Race/Ethnicity. https://www.kff.org/other/state-indicator/distribution-by-raceethnicity/currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D.

Zillow. Zillow Research Data. https://www.zillow.com/research/data/.

U.S. Census Bureau. State Area Measurements and Internal Point Coordinates. U.S. Department of Commerce, https://www.census.gov/geographies/reference-files/2010/geo/state-area.html.

Appendix

a.1

Table 5 shows the regression results of the base fixed effects model, where the effects of both damage variables are similar to their effects in our final model. However, the major differences are observed in the control variables. Percent white now has a positive but insignificant effect, while both average home price and population density are statistically significant. The significant variables make sense as well. Home prices in high-risk areas tend to decrease, while premiums increase (Dumm et al., 2012). Additionally, more densely populated areas are expected to have higher risk, which justifies the significance of population density.

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Tue, Mar 18, 2025 - 12:18:21 PM

Table 5: Regression Models

	Dependent variable:
	homePrems Fixed Effects Model
log(home_damage)	1.757^{*}
	p = 0.060
log(disaster_damage)	-1.168
	p = 0.162
log(AvgHomePrice)	-126.858^{**}
	p = 0.039
log(Percent_white)	354.812
	p = 0.199
log(PopDensity)	449.306^{**}
	p = 0.037
Observations	576
R^2	0.019
Adjusted R^2	-0.102
Note:	*p<0.1; **p<0.05; ***p<0.01

a.2

Figure 8 illustrates how the radius of a polygon determines whether nearby polygons are considered neighbors. As the figure shows, the centroid of a polygon must fall within the radius to be counted as a neighbor.

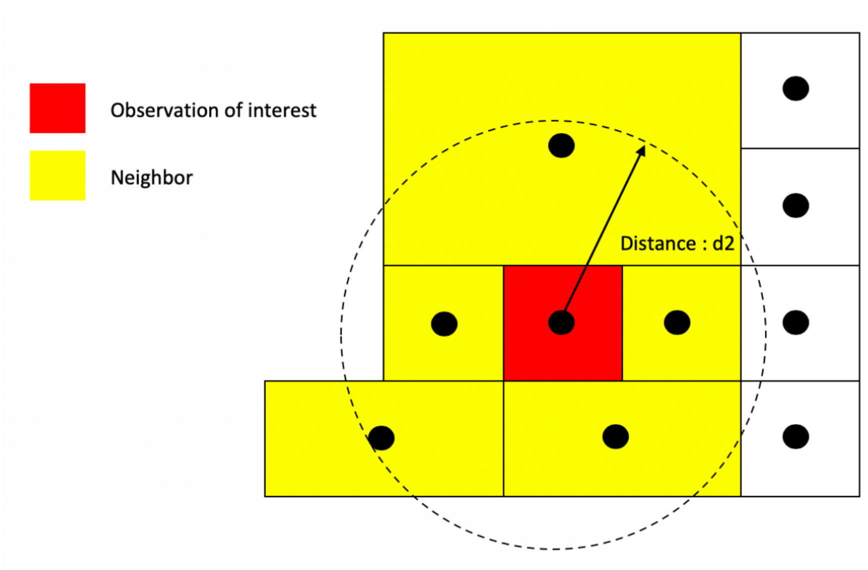


Figure 2: Distance based

Figure 8: Distance based neighbors

a.3

Figure 9 shows the differences between rook and queen contiguity. This paper used manual rook contiguity in combination with distance-based neighbors.

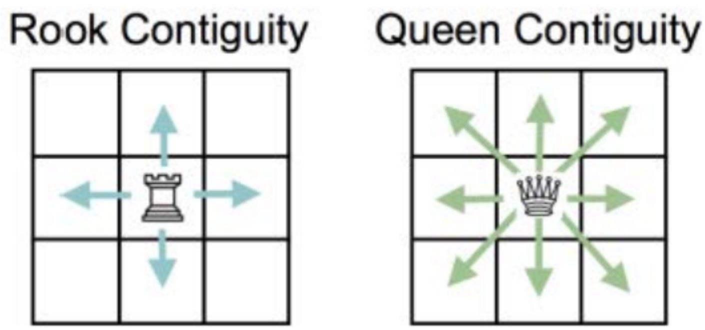


Figure 9: Rook vs. Queen Contiguity

a.4

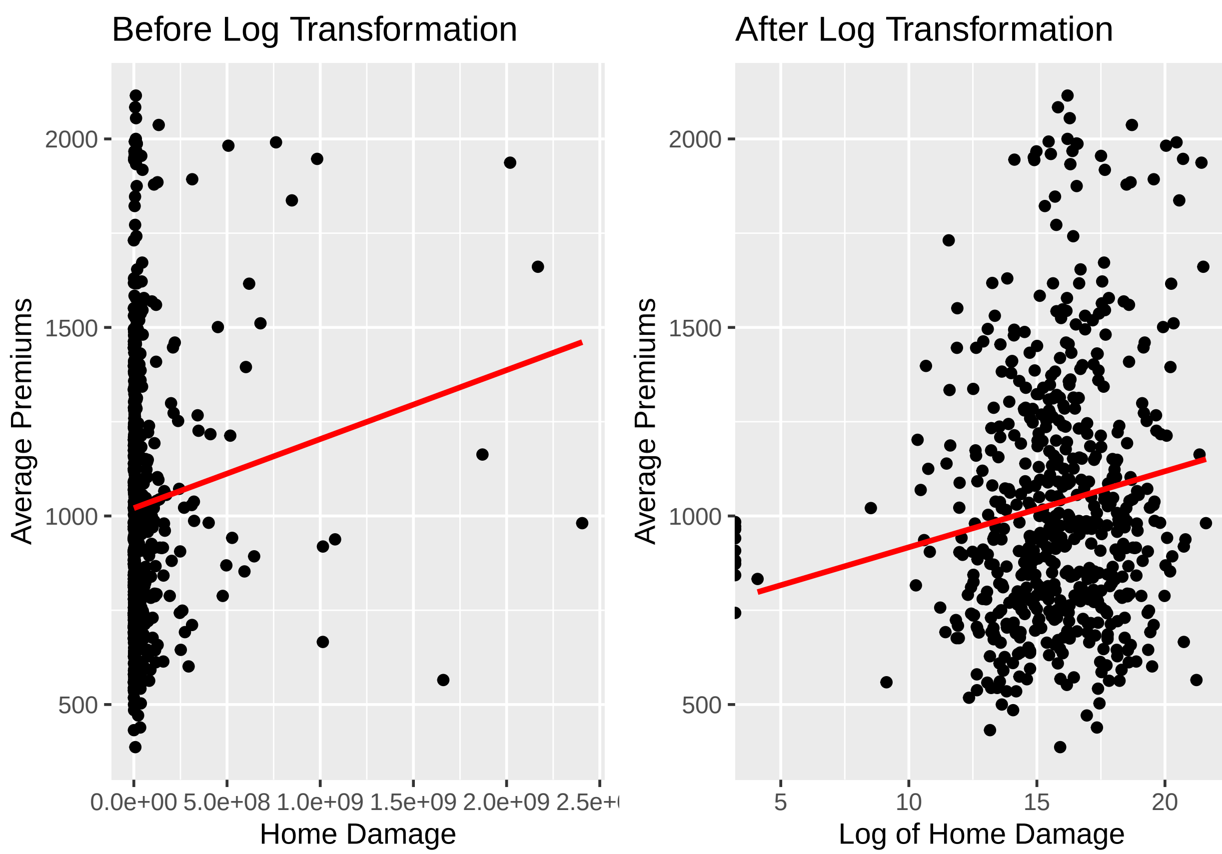
The inverse distance weighting equation, where the numerator is the inverse distance between state i's centroid and neighbor j's centroid, and the denominator is the sum of all the inverse distances between state i and all of its k neighbors. Dividing by the denominator ensures that the sum of all the weights for state i equals 1.

$$w_{ij} = \frac{\frac{1}{d_{ij}}}{\sum\limits_{k \in N_i} \frac{1}{d_{ik}}}$$

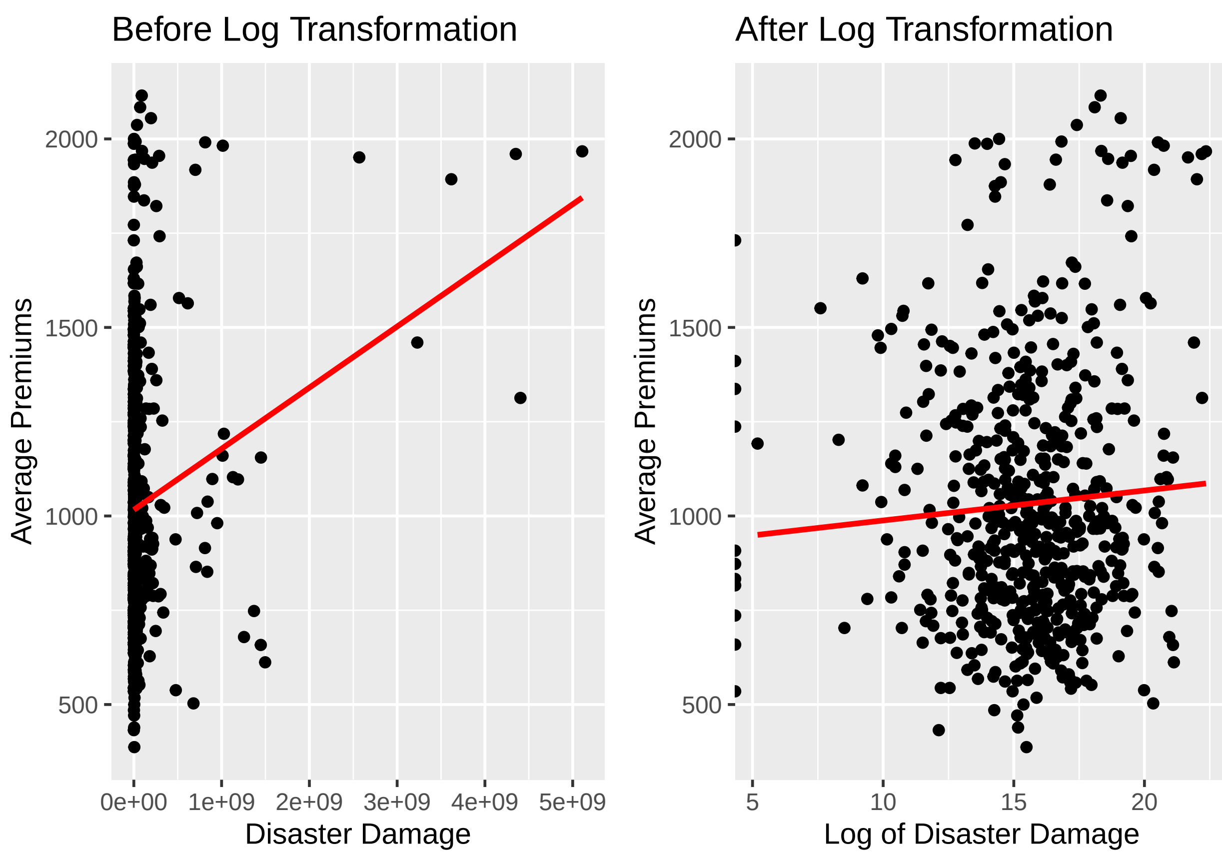
a.5

Data visualizations of both the logged and unlogged versions of each variable.

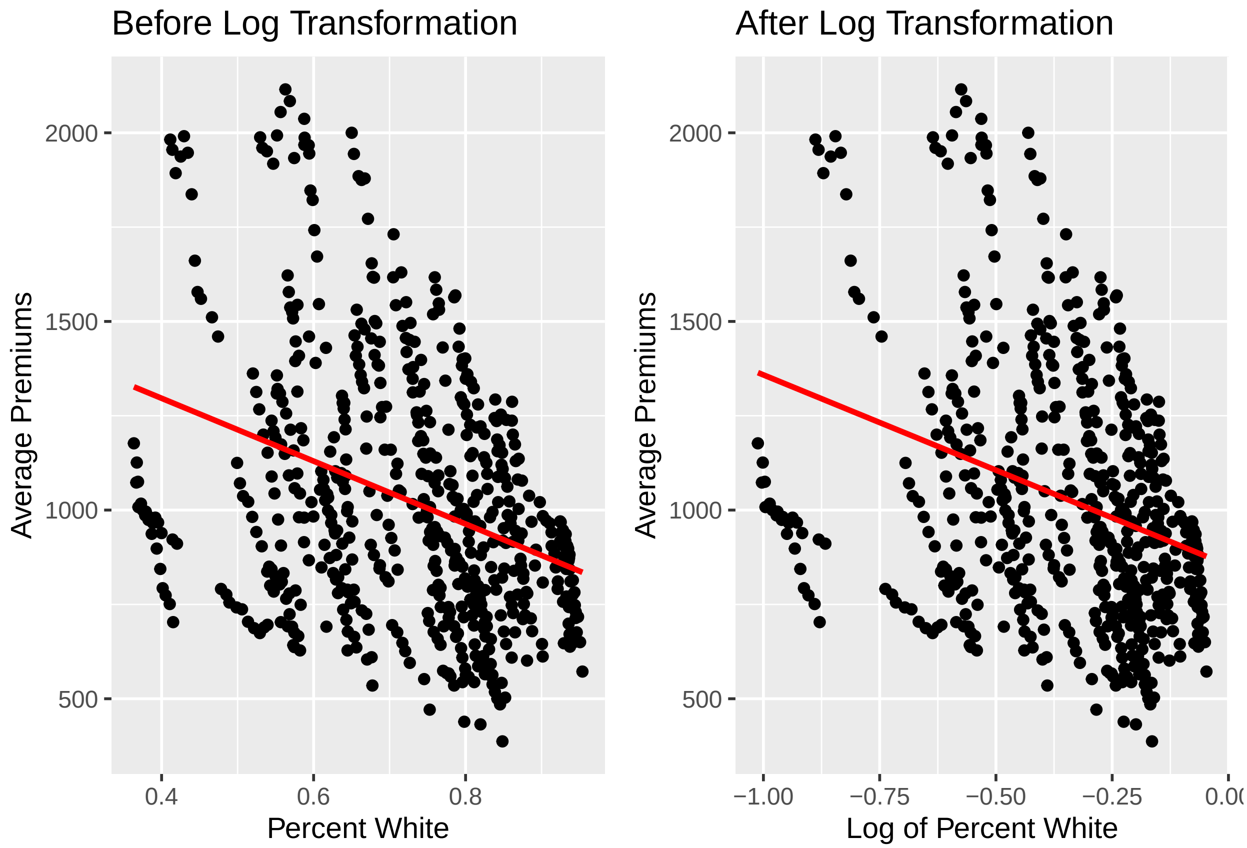
Home Damage


Before Log Transformation
Average Premiums
2000
1500
1000
500
0.0e+00 5.0e+08 1.0e+09 1.5e+09 2.0e+09 2.5e+
Home Damage
After Log Transformation
Average Premiums
2000
1500
1000
500
5
10
15
20
Log of Home Damage

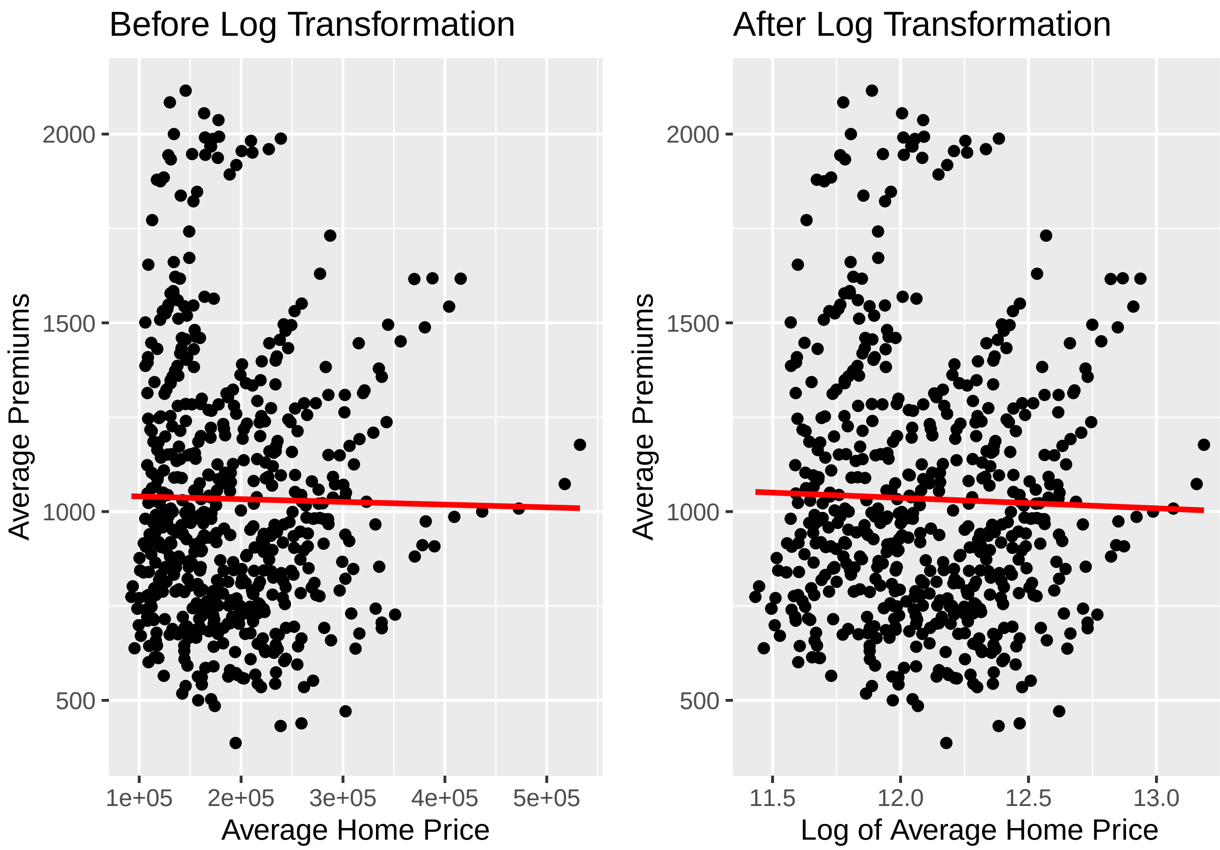
Disaster Damage



Percent White



Average Home Price



Population Density

a.6

The same models were used in the paper; however, instead of using the Baltagi (b) error method, I used the Kapoor (kkp) error method. This approach produces almost identical results to when I used Baltagi's error method.

% Table created by stargazer v.5.2.3 by Marek Hlavac, Social Policy Institute. E-mail: marek.hlavac at gmail.com % Date and time: Tue, Mar 18, 2025 - 12:18:22 PM

Table 6: Regression Models

	Dependent variable:			
	State FE	Year FE	Region FE Year FE	State FE Year FE
rho	0.906***	0.859***	0.654***	0.702***
	p = 0.000	p = 0.000	p = 0.000	p = 0.000
log(home_damage)	0.947	4.497**	3.438*	1.302*
	p = 0.146	p = 0.016	p = 0.071	p = 0.053
log(disaster_damage)	−0.566	2.469	2.027	−0.706
	p = 0.319	p = 0.119	p = 0.211	p = 0.233
log(AvgHomePrice)	99.984**	34.890	80.103*	29.385
	p = 0.041	p = 0.433	p = 0.068	p = 0.562
log(Percent_white)	−806.182***	−200.056***	−191.767***	−469.499**
	p = 0.0002	p = 0.0003	p = 0.002	p = 0.033
log(PopDensity)	−138.935	116.205***	118.100***	0.217
	p = 0.442	p = 0.000	p = 0.000	p = 1.000
RegionNorthern Rockies and Plains			341.547***	
			p = 0.00000	
RegionNorthwest			−97.546	
			p = 0.170	
RegionOhio Valley			206.137***	
			p = 0.00004	
RegionSouth			462.366***	
			p = 0.000	
RegionSoutheast			238.464***	
			p = 0.00000	
RegionSouthwest			1.133	
			p = 0.986	
RegionUpper Midwest			165.099***	
			p = 0.006	
RegionWest			−21.349	
			p = 0.788	
TRUE				
Note:				*p<0.1; **p<0.05; ***p<0.01