Social and Other Determinants of Life Insurance Demand

August | 2022

Diversity, Equity and Inclusion

Social and Other Determinants of Life Insurance Demand

AUTHORS Wilmer Martinez (PhD student)
School of Mathematical and Statistical Sciences
Arizona State University

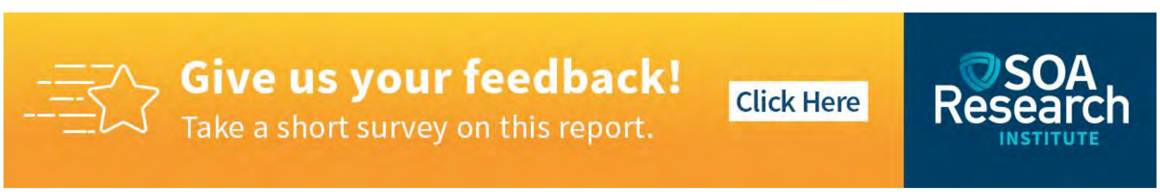
Kyran Cupido, PhD (Co-PI)
Assistant Professor
Mathematics and Statistics Department
St. Francis Xavier University

Petar Jevtic, PhD (PI)
Assistant Professor
School of Mathematical and Statistical Sciences
Arizona State University

Jianxi Su, PhD (Co-PI)
Associate Professor
Department of Statistics
Purdue University

SPONSOR Diversity, Equity and Inclusion
Research Advisory Council

DATA SOURCE LIMRA



CONTENTS

Social and Other Determinants of Life Insurance Demand

Executive Summary

Life insurance demand is a complex phenomenon that can be measured in multiple ways. Understanding it in the context of various potential drivers is of immense interest to insurance companies, insurance markets, regulators, and broader society. Unfortunately, to our knowledge, not much work has been done to address this topic in the current academic literature. That is why in this work, we perform a spatial regression analysis using multi-scale geographically weighted regression (MGWR) approach. As response variables, we consider annual permanent life insurance premiums and annual term life insurance premiums for 2020, which total $12,840,615,055, as proxies for insurance demand in the United States. The covariates considered were broadly classified into two groups: social capital and population composition. Because the COVID-19 pandemic emerged in 2020, results of this study may or may not represent a typical year. Identifying the impact of COVID-19 on results is beyond the scope of this study.

Our findings show that among those covariates found to be statistically significant, not all were relevant on the same spatial scales. Some were globally relevant, meaning they exhibit a relatively equal association with insurance demand across the entire country as a whole, other covariates are regionally relevant, with effects that are realized in certain broad areas of the country, and other covariates operate locally, with effects that specific to small amounts of counties. The spatial scale on which covariates were relevant also depends on whether permanent or term life insurance is considered.

We found that the five most significant covariates associated with permanent insurance sold are household income, percentage of the population that is African American, education, health insurance, and Gini index. All the aforementioned covariates show a positive association with permanent insurance sold. For term insurance sold, the five most significant covariates are household income, education, Gini index, percentage of households with no vehicles, and health insurance. Their relationships with term insurance sold are positive except for the percentage of households with no vehicles.

Table 1 summarizes the mean relationships between the covariates considered and the insurance demand proxies, as well as the associated scales of impacts. The orders of the absolute marginal impacts of the covariates on the response variable are also reported, with lower ranks indicating a stronger average marginal impact across the space.

Table 1

MEAN RELATIONSHIPS BETWEEN COVARIATES AND LIFE INSURANCE DEMAND PROXIES AND SCALE OF IMPACT

	Permanent Insurance Sold			Term Insurance Sold		
Covariate Description	**Global, Regional, or Local**	**Positive or Negative**	**Rank**	**Global, Regional, or Local**	**Positive or Negative**	**Rank**
Percentage of households with yearly income above $75,000	Local	Pos	1	Local	Pos	1
Percentage of the population that is African American*	Regional	Pos	2	Global	Pos	8
Percentage of the population with a bachelor's degree or higher (25 years and older)	Global	Pos	3	Local	Pos	2
Percentage of the population without health insurance	Local	Pos	4	Global	Pos	5
Gini Index (i.e., a statistical measure of wealth inequality)	Regional	Pos	5	Regional	Pos	3
Percentage of the population born in the United States	Global	Pos	6	Global	Neg	10
Percentage of households in poverty	Global	Pos	7	Local	Neg	18
Percentage of the population in the labor force	Global	Pos	8	Global	Pos	14
Percentage of the population that is Asian/Asian American*	Global	Pos	9	Local	Neg	13
Association density (i.e., the number of social institutions present within a county in proportion to its population)	Regional	Pos	10	Local	Pos	15
Percentage of the population that is Hispanic/Latino*	Global	Pos	11	Global	Neg	17
Percentage of the voting-age population that voted in the 2016in the 2016 election	Regional	Pos	12	Global	Pos	6
Percentage of the population that is Indigenous	Local	Pos	13	Global	Pos	18
Percentage of single parent households	Local	Neg	14	Local	Pos	10
Percentage of households with no vehicles	Global	Neg	15	Regional	Neg	4
Response rate for the 2020 census	Local	Neg	16	Regional	Pos	7
Percentage of the population living in the same place since 2009	Global	Pos	17	Global	Neg	9
Unemployment rate	Local	Neg	18	Global	Neg	12
Percentage of housing that is owner occupied	Regional	Pos	19	Regional	Pos	15

* Data use the terms "African American," "Hispanic" and "Asian American," but we use the more inclusive terms "Black/African American," "Hispanic/Latino" and "Asian/Asian American."

The results presented in Table 1 highlight the subtle differences in the regression patterns among the two insurance groups, including:

1. The percentage of the population living in the same place since 2009 is a global covariate. It is positively associated with permanent insurance sold, yet negatively associated with term insurance sold.
2. Association density has a positive relationship with insurance demand on average.

3. Percentage of the voting-age population that voted in the 2016 election has a globally positive relationship with insurance demand, but there are spatial variants of impacts when it comes to permanent insurance sold.
4. Response rate for the 2020 census has spatially varying impacts on insurance demand, The average impact is negative for permanent insurance sold but positive for term insurance sold.
5. Unemployment rate has a negative association with insurance demand overall, yet the scale of impact is global only for term insurance.
6. The percentage of single parent households has local impacts on insurance demand. The average relationship is negative for permanent insurance sold, but it is positive when it comes to term insurance sold.
7. The percentage of households in poverty has a globally positive dependence with permanent insurance sold, but the relationship becomes negative and spatially varying when it comes to term insurance sold.
8. The percentage of households with no vehicles has a negative dependence with the two groups of insurance demands on average. The dependence presents a spatially varying pattern for term insurance sold, but it is spatially consistent for permanent insurance sold.
9. The percentage of the population with a bachelor's degree or higher is positively associated with the two groups of insurance demands, and the association only varies spatially for term insurance sold.
10. The percentage of the population born in the U.S. has a global effect on the two groups of insurance sold. The dependence is positive for permanent insurance sold, but negative for term insurance sold.
11. The percentage of the population that is Black/African American is positively associated with the two groups of insurance sold on average. For permanent insurance sold, the regression patterns contain spatial variants.
12. Both the percentage of the population that is Hispanic/Latino and the percentage of the population that is Asian American have almost the same regression patterns on the two groups of insurance sold. Namely, they are both positively associated insurance demand for permanent insurance sold, but the dependence is negative for term insurance sold. The only difference in regression pattern between these two covariates is that when it comes to term insurance sold, the dependence is global for Hispanic/Latino population, and it is locally varying for the Asian/Asian American population.
13. The percentage of the population that is Indigenous has a positive relationship with the two insurance demand proxies on average. The dependence is locally varying for permanent insurance sold, but spatially consistent for term insurance sold.

Section 1: Introduction

Across geography as vast as the U.S., various determinants may operate at different scales to produce spatially different life insurance demand outcomes. Therefore, measured on state and county levels with different constituents and/or life conditions, it is conceivable that the insurance demand may be improved by sustained actions of insurers, regulators, and agents/brokers based on a comprehensive understanding of these various determinants and their impact across multiple geographical scales.

Socioeconomic determinants include age, education, gender, race, ethnicity, income, etc. In contrast, other determinants such as behavioral and attitudinal characteristics may consist of crime rates or attitudes towards government, climate change, etc. In our paper, considering the most current data, we hypothesize that such determinants across space impact insurance demand, as measured via aggregate premiums, on different spatial scales. For example, income as a determinant of insurance demand might be relevant to the entire U.S., whereas the percentage of representatives of a particular minority community may impact determinants of demand at only the county level. At the same time, education or religious affiliations can have a regional or state-level impact. Thus, insurance demand differences across space may be impacted differently (or not at all) by the determinants from the same group. That is why, in this paper, we aim to identify the spatial scale at which the various determinants of life insurance demand in the U.S. operate. To the best of our knowledge, it is an unexplored topic in the actuarial literature. However, it is important for us to stress that the regression analysis involved in our study only reveals the associations, but not the causalities, between various socio-economic variables and insurance demand.

1.1 LITERATURE REVIEW

Most recent research of (C. T. Trinh et al., 2020) focuses on the OECD countries in the period from 2000 to 2017 and investigates how cultural characteristics impact the demand for property, accident and health insurance. The work of (Letiţia Andronic, 2019) investigates how social and financial macroeconomic variables such as average net salary, the unemployment rate, the enrolment ratio in education and the birth rate, etc. influence the density of the insurance market in Romania in period from 1997 to 2017. The (Sampath Sanjeewa et al., 2019) investigates the determinants of life insurance consumption in emerging insurance markets of South Asian from 1996 to 2017. Also, the (Cavalcante et al., 2018) examines economic growth and financial development as determinants of non-life insurance premium in Brazil. Earlier research of (T. Trinh et al., 2016) investigated the determinants of the demand for non-life insurance in developed and developing countries before and during the global financial crisis and considered 36 developed and 31 developing countries over the period from 2000 to 2011. The (Podoabă, 2015) investigated how economic development was associated with health insurance given sample of 32 European countries observed from 2002 to 2011. The (Kamiya et al., 2014) studied the association between banking crises and non-life insurance consumption using cross-country panel of data from 139 countries from 1988 to 2010. Also, the (Jean Kwon, 2013) investigated significance of regulatory agency structure, key regulatory measures, political stability and cultural dimension in insurance markets of 56 developed and developing countries from 2005 to 2009. The (Outreville, 2013) proposed a review of empirical papers examining the various relationships between insurance and economic development across developed and developing countries. Finally, (Park & Lemaire, 2012)examined impact of culture on the demand for non-life insurance examining 68 countries observed over a ten-year period. Thus, all considered, the present state of the literature suggests that neither current nor comprehensive analysis of determinants of the insurance demand on the level of U.S. states and counties exists.

This paper focuses solely on life insurance demand. From this point forward, for brevity, we will use the term “insurance” to be synonymous with “life insurance,” unless it is explicitly noted otherwise.

Section 2: Data

This study incorporates a variety of datasets relating to the counties of the U.S. The variables included in the models are described below along with justification for their use in the study. The datasets featured geographic location in the form of Federal Information Processing Standards (FIPS) codes. FIPS codes are implemented by the National Institute of Standards and Technology (NIST) as a means of supporting consistent references of geographic areas in the U.S. by defining unique codes for each state, as well as for each county within the state.

2.1 RESPONSES

The data used in this study were obtained from LIMRA which is a worldwide research, consulting, and professional development not-for-profit trade association.[1] The file, which was produced by LIMRA contains county-level data for the year 2020. Because the COVID-19 pandemic emerged in 2020, data from 2020 may not represent a typical year. Our study does not attempt to identify the impact of the pandemic on results. Similar studies of data from before as well as several years subsequent to 2020 would be necessary to identify the impact of the pandemic on results.

In this work, the response variables considered were annual term insurance premiums per county and annual permanent insurance premiums per county. Only individual insurance sold is considered in the LIMRA data. Note that to use premium sold as a proxy of insurance demand, we have to assume that the premium per face amount is identical across the population. Due to the limited data, this is an inevitable assumption we have to make.

2.1.1 NOTE ON COUNT DATA AND THEIR EXCLUSION FROM CONSIDERATION

In addition to the insurance premiums data, the LIMRA data contains counts of the total number of insurance policies sold (as well as term and permanent policies separately) in each county of the U.S. Note that the policy count data is significantly driven by the population size of a given territory. Given that there is a great deal of variability in the population sizes of U.S. counties, the total number of policies sold variable would need to be normalized so that the marginal effects of social and economic factors on insurance demand can be properly captured. As such, to consider this a candidate response variable in the MGWR model, this raw data of total number of policies sold was offset by the population size of each county. When this transformation was complete, preliminary spatial analyses revealed a significant pattern of policies sold across the U.S., with a particularly large cluster of extremely high values being observed in regions around Alabama, Mississippi and Louisiana. Such a strong concentration in this pattern of insurance policies means that it is entirely possible to model this as a function of location alone, without the need for any additional predictor variables. This can be observed in Figure 2.1, which displays the results of a simple regression model where the response variable is total number of policies sold and no predictor variables. Intuitively, this global model does not explain any of the variability in the policies sold. However, it can be observed from the diagnostics from the GWR approach, that a spatial model is able to explain 94.4% of this variability in policies sold (see, the R2 statistic in Table 2.1). This means that to estimate the total normalized number of policies sold in a given county, the best approach would be to simply take an average of neighboring counties. As such, the total number of policies sold was not included as a response variable in any of the MGWR analyses. For comparative purposes, the same normalization approach was taken using the total annualized premiums sold as a response, and the results are presented in Figure 2.2. From this figure, we observe that much less of the variability in premiums can be attributed to geographic location. For more details about the statistics outlined in Tables 2.1 and 2.2, we refer the readers to Fotheringham, et al. (2003).

[1] See: https://www.limra.com/en/.

Table 2.1

ESTIMATION OF INTERCEPT MODEL FOR THE TOTAL NUMBER OF POLICIES SOLD

```
===============================================================================
MGWR Version: 2.2.1
Released on: 03/20/2020
Source code is available at: https://github.com/pysal/mgwr
Development Team: Ziqi Li, Taylor Oshan, Stewart Fotheringham, Wei Kang,
Levi Wolf, Hanchen Yu, Mehak Sachdeva, and Sarah Bardin
Spatial Analysis Research Center (SPARC)
Arizona State University, Tempe, USA
===============================================================================
Model type:                                                            Gaussian
Number of observations:                                                    3073
Number of covariates:                                                         1
Dependent variable:                                          Total.Policies.Sold
Variable standardization:                                                    On
Total runtime:                                                          0:00:03

Global Regression Results
-------------------------------------------------------------------------------
Residual sum of squares:                                               3073.000
Log-likelihood:                                                       -4360.398
AIC:                                                                   8722.796
AICc:                                                                  8724.800
R2:                                                                       0.000
Adj. R2:                                                                  0.000

Variable                                   Est.         SE  t(Est/SE)    p-value
------------------------------------ ---------- ---------- ---------- ----------
Intercept                                 0.000      0.018      0.000      1.000

Geographically Weighted Regression (GWR) Results
-------------------------------------------------------------------------------
Coordinates type:                                                     Spherical
Spatial kernel:                                               Adaptive bisquare
Criterion for optimal bandwidth:                                           AICc
Bandwidth used:                                                          44.000

Diagnostic Information
-------------------------------------------------------------------------------
Residual sum of squares:                                                173.287
Effective number of parameters (trace(S)):                              206.206
Degree of freedom (n - trace(S)):                                      2866.794
Sigma estimate:                                                           0.246
Log-likelihood:                                                          57.745
Degree of Dependency (DoD):                                               0.336
AIC:                                                                    298.922
AICc:                                                                   329.041
BIC:                                                                   1548.458
R2:                                                                       0.944
Adj. R2:                                                                  0.940
Adj. alpha (95%):                                                         0.000
Adj. critical t value (95%):                                              3.674
```

Figure 2.2

ESTIMATION OF INTERCEPT MODEL FOR THE TOTAL NUMBER OF PREMIUMS SOLD

```
===========================================================================
MGWR Version: 2.2.1
Released on: 03/20/2020
Source code is available at: https://github.com/pysal/mgwr
Development Team: Ziqi Li, Taylor Oshan, Stewart Fotheringham, Wei Kang,
Levi Wolf, Hanchen Yu, Mehak Sachdeva, and Sarah Bardin
Spatial Analysis Research Center (SPARC)
Arizona State University, Tempe, USA
===========================================================================
Model type:                                                         Gaussian
Number of observations:                                                 3073
Number of covariates:                                                      1
Dependent variable:                                Total.Annualized.Premium.Sold
Variable standardization:                                                 On
Total runtime:                                                       0:00:06

Global Regression Results
---------------------------------------------------------------------------
Residual sum of squares:                                            3073.000
Log-likelihood:                                                    -4360.398
AIC:                                                                8722.796
AICc:                                                               8724.800
R2:                                                                    0.000
Adj. R2:                                                               0.000

Variable                                   Est.         SE  t(Est/SE)    p-value
------------------------------------ ---------- ---------- ---------- ----------
Intercept                                -0.000      0.018     -0.000      1.000

Geographically Weighted Regression (GWR) Results
---------------------------------------------------------------------------
Coordinates type:                                                  Spherical
Spatial kernel:                                            Adaptive bisquare
Criterion for optimal bandwidth:                                        AICc
Bandwidth used:                                                       48.000

Diagnostic Information
---------------------------------------------------------------------------
Residual sum of squares:                                            2253.946
Effective number of parameters (trace(S)):                           188.766
Degree of freedom (n - trace(S)):                                   2884.234
Sigma estimate:                                                        0.884
Log-likelihood:                                                    -3884.127
Degree of Dependency (DoD):                                            0.347
AIC:                                                                8147.785
AICc:                                                               8172.905
BIC:                                                                9292.153
R2:                                                                    0.267
Adj. R2:                                                               0.219
Adj. alpha (95%):                                                      0.000
Adj. critical t value (95%):                                           3.652
```

2.2 COVARIATES

Seventeen variables were selected to investigate the economic, demographic, and social factors that affect insurance demand within the U.S.. All of the covariates included in this study were standardized to zero mean and unit variance so that we can directly compare their respective effects. Definitions of the covariates are provided in Table 2.1 and maps displaying the values of all the variables are provided in Section 2.2.3. The determinants of insurance demand fall into two general groups: social capital and population composition.

2.2.1 SOCIAL CAPITAL DETERMINANTS

Social capital is defined by the Organization for Economic Cooperation and Development (OECD) as the "networks together with shared norms, values and understandings that facilitate within or among groups" (Centre for Educational Research and Innovation, 2001). Socioeconomic research suggests that social capital has a positive influence on economic growth, in which the promotion of trust within communities allows for effective collective action towards many societal issues (Rupasingha et al., 2006). As such, the introduction of variables associated with social capital are naturally relevant for any investigation of insurance demand. As part of their research to quantify the social connections and networks present within the U.S., (Rupasingha et al., 2006) compiled county-level data to create a database which includes the association density, percentage of the voting-age population that voted in the 2016 election and census completion variables which will be incorporated in this study.[2] The association density variable refers to the number of social institutions present within a county in proportion to its population. These institutions represent organizations where individuals come together for a common purpose, such as local businesses, religious buildings and recreational centers. Further, as motivated by studies such as (Glaeser et al., 2002), which demonstrated that stable neighborhoods imply positive interactions between residents, the percentage of the population living at the same address since the year 2009 and percentage of housing that is owner-occupied were included as variables in the study as measures corresponding to residential stability.[3] For consistency across the covariates, the years associated with these variables of social capital correspond to the timing that the most recently released primary sources of census data, such as the American Community 5-Year Survey, became available.

2.2.2 POPULATION COMPOSITION DETERMINANTS

Population composition predictor variables quantify the effects that social affluence and disadvantage have on the demand for insurance products. The percentage of people with a bachelor's degree, household income above $75,000 and percentage of people in the labor force were included in the study as measures of social affluence. These variables contrast the measures of social disadvantage included in the study, which are the percentage of households with single parents, unemployment rate, percentage of households in poverty, percentage of households with no health insurance and percentage of households with no access to vehicles. Further, the Gini index measure of income inequality has been included to capture the dispersion of wealth among the counties of the U.S. To investigate the predictive effects that the racial and ethnic composition of a county has on insurance demand, the percentage of the population which identifies as Black/African American, Asian/Asian American, Indigenous, and Hispanic/Latino has been selected for study. To avoid collinearity, the percentage of the white

---

[2] The Social Capital Variables for 2014 spreadsheet was obtained on from https://aese.psu.edu/nercrd/community/social-capital-resources/social-capital-variables-for-2014

[3] Data from the American Community Survey 2019 5-Year Estimates were obtained from data.census.gov

population has been excluded. The percentage of the population born in the U.S. was added to evaluate the effects that immigration, or a lack thereof, has on the county level insurance demand.[4]

2.2.3 MAPS OF THE COVARIATES

A visual representation of the covariates in this study are provided in Figures 2.3–20, with the descriptions of the covariates summarized in Table 2.1. The presented quantile maps distribute the entire set of observed values into four equally-sized groups and provide a useful illustration for comparing the differences in the variables across space. The counties which are colored in dark blue correspond to the lowest observed values of the individual covariates, and similarly, counties colored in gold correspond to areas with the highest observed values. From these maps, we observe several clear, underlying trends in the covariates associated with social capital and population composition across space. The presence of clear, visible patterns among the mapped covariates justifies the need for the implementation of modelling techniques which can capture this variability across space.

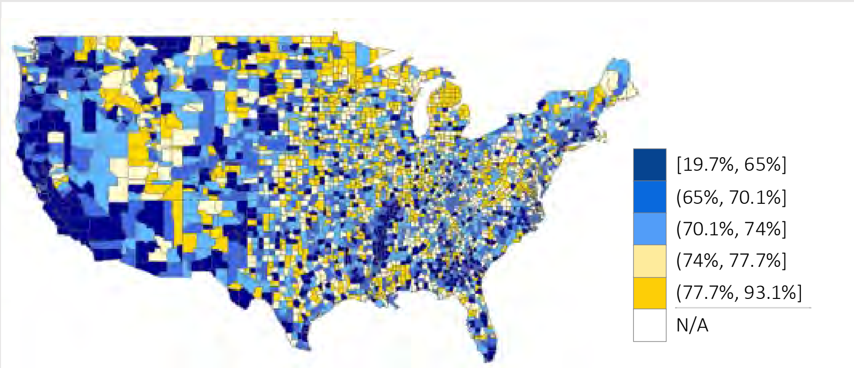
Table 2.1
COVARIATES DESCRIPTION

Group	Figure	Covariate	Description
Social Capital	2.3	OwnOcc	Percentage of housing that is owner occupied*
	2.4	SameHous	Percentage of the population living in the same place since 2009*
	2.5	AssDens	Association density (i.e., the number of social institutions present within a county in proportion to its population)**
	2.6	VoTurn	Percentage of the voting aged population that participated in the 2016 election**
	2.7	CenResp	Response rate for the 2020 census**
Population Composition	2.8	P_Labor	Percentage of the population in the labor force*
	2.9	Unemp	Unemployment Rate*
	2.10	NoHeIns	Percentage of the population without health insurance*
	2.11	Gini	Gini index (i.e., a statistical measure of wealth inequality)*
	2.12	SinPar	Percentage of single parent households*
	2.13	HInc	Percentage of households with yearly income above $75,000*
	2.14	Poverty	Percentage of households in poverty*
	2.15	NoVehi	Percentage of households with no vehicles*
	2.16	BachDe	Percentage of the population with a bachelor's degree or higher (25 years and older)*
	2.17	BornUSA	Percentage of the population born in the U.S.*
	2.18	P_AfriA	Percentage of the population that is African American*
	2.19	P_Hisp	Percentage of the population that is Hispanic*
	2.20	P_Asian	Percentage of the population that is Asian American*
	2.21	P_Indig	Percentage of the population that is Indigenous*

*Data from the American Community Survey 2019, 5-Year Estimates from data.census.gov; terms for races and ethnicities reflect those used in the data source and may differ from SOA Research Institute's preferred terms for inclusivity.
** Data for the Social Capital Variables for 2014 from https://aese.psu.edu/nercrd/community/social-capital-resources/

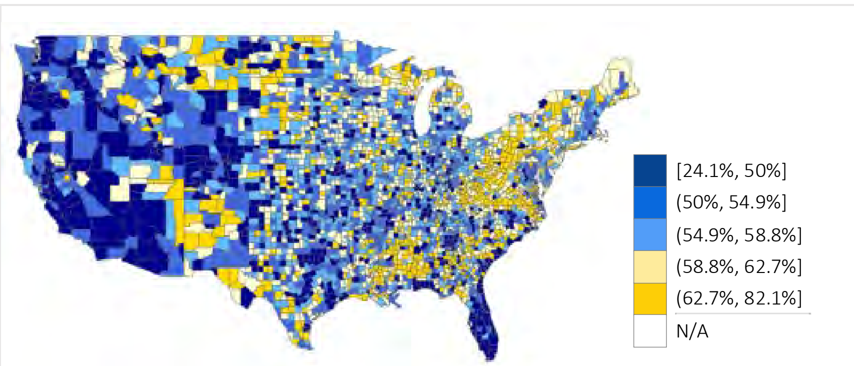
[4] Data from the American Community Survey 2019 5-Year Estimates were obtained from data.census.gov

Figure 2.3
PERCENTAGE OF HOUSING THAT IS OWNER OCCUPIED



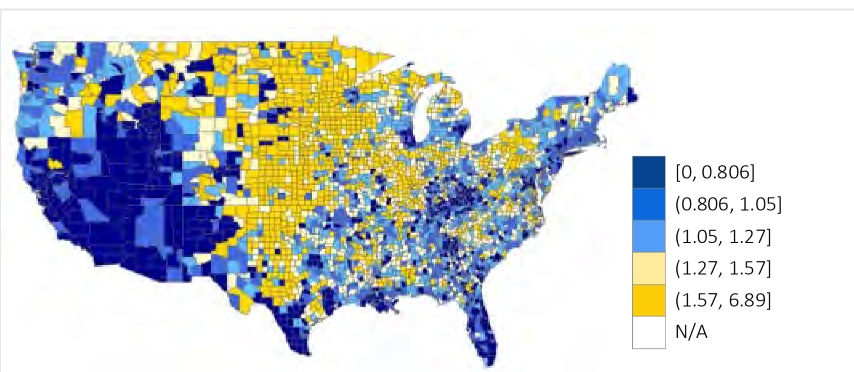
A lower percentage of housing that is owner occupied is observed along the West Coast, while the percentage distribution is rather random across other parts of the U.S.

Figure 2.4
PERCENTAGE OF THE POPULATION LIVING IN THE SAME PLACE SINCE 2009



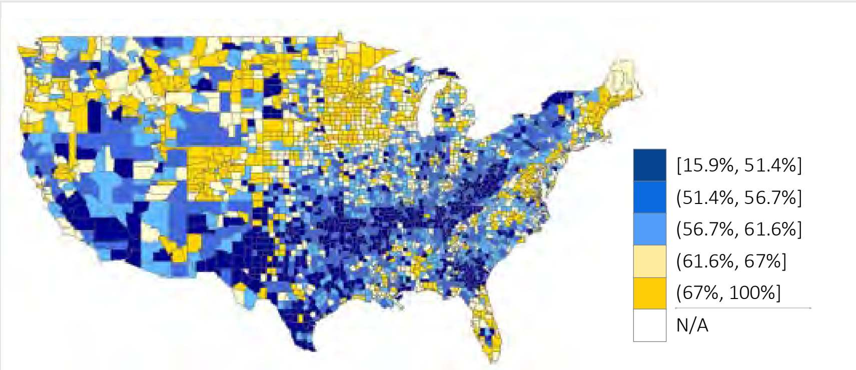
In the western region of the country, a lower percentage of the population have lived in the same place since 2009, while a higher percentage is observed in the Middle Atlantic States.

Figure 2.5
ASSOCIATION DENSITY—NUMBER OF SOCIAL INSTITUTIONS PRESENT WITHIN A COUNTY IN PROPORTION TO ITS POPULATION



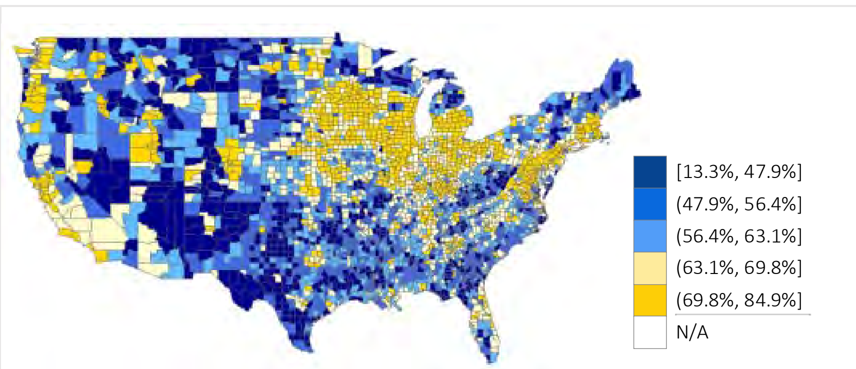
The density of social institutions in proportion to population is lower in the Southwest, while it is higher in the Midwest and Texas.

Figure 2.6
PERCENTAGE OF THE VOTING AGED POPULATION THAT PARTICIPATED IN THE 2016 ELECTION



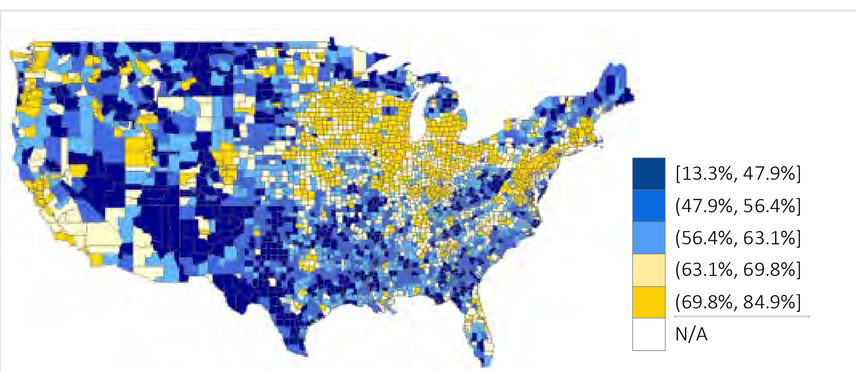
The percentage of the voting-age population that voted in 2016 is higher in Florida, the Pacific Northwest, Rocky Mountains, Midwest, and New England than in other parts of the U.S.

Figure 2.7
RESPONSE RATE FOR THE 2020 CENSUS



The response rate for the 2020 census was higher in the Midwest and Middle Atlantic States than the Southwest.

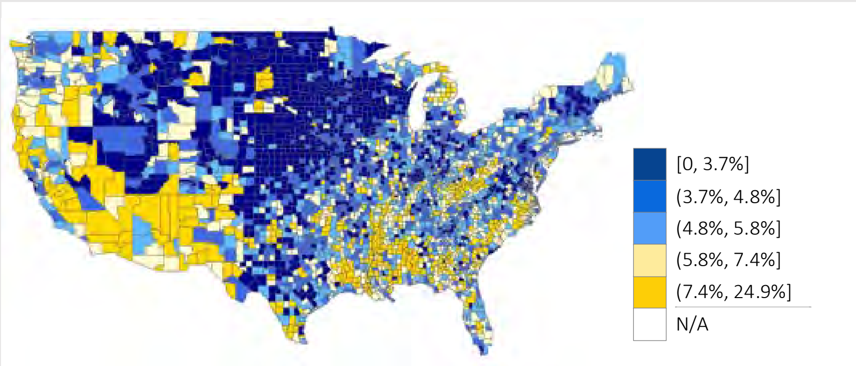
Figure 2.8
PERCENTAGE OF THE POPULATION IN THE LABOR FORCE



The percentage is higher along the western coast, and in the Midwest and Middle Atlantic states than in the Southwest.

Figure 2.9

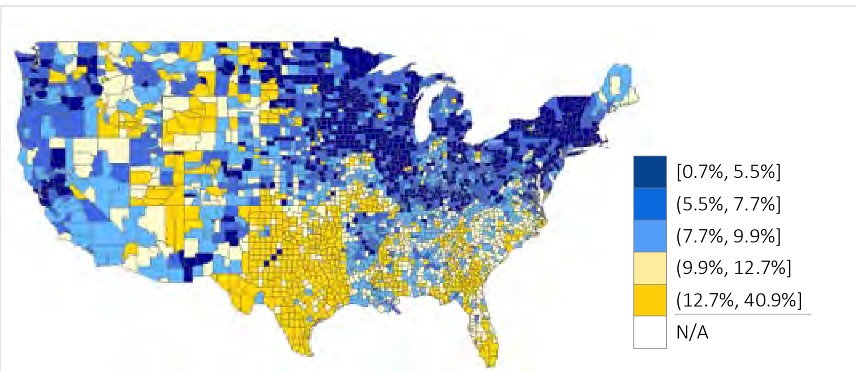
UNEMPLOYMENT RATE



Unemployment rates in the counties of the central U.S. are significantly lower than counties in California and the southwestern U.S.

Figure 2.10

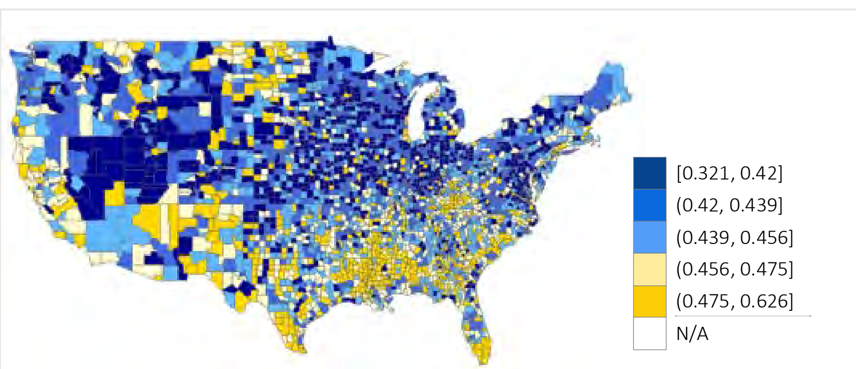
PERCENTAGE OF THE POPULATION WITHOUT HEALTH INSURANCE



A higher proportion of individuals in the midwestern U.S. have health insurance than individuals in Texas and Florida.

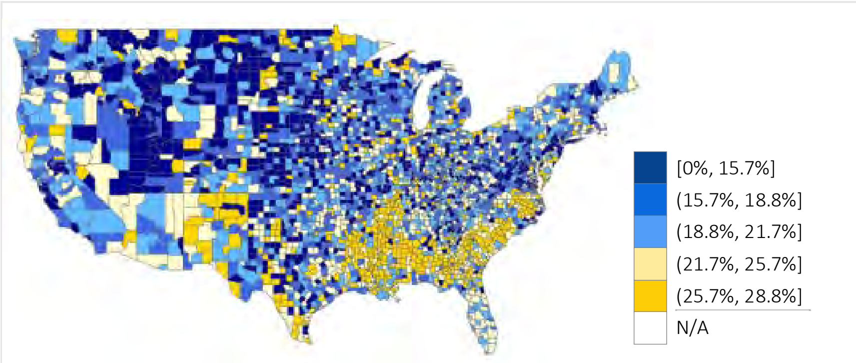
Figure 2.11

GINI INDEX—A STATISTICAL MEASURE OF WEALTH INEQUALITY



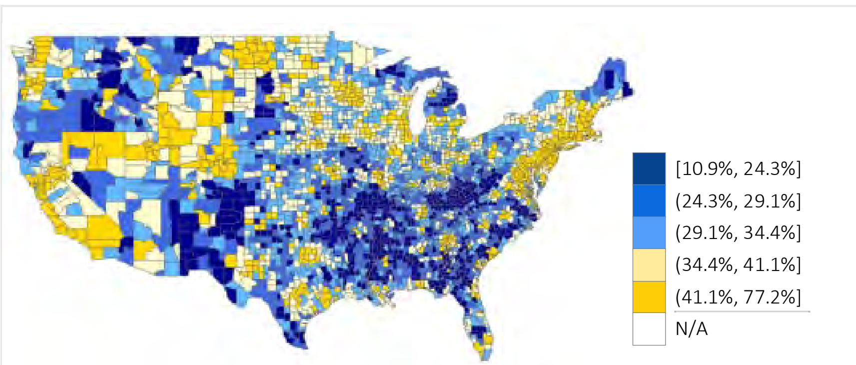
The Gini index is higher in regions of the southern U.S. than in the Midwest.

Figure 2.12
PERCENTAGE OF SINGLE PARENT HOUSEHOLDS



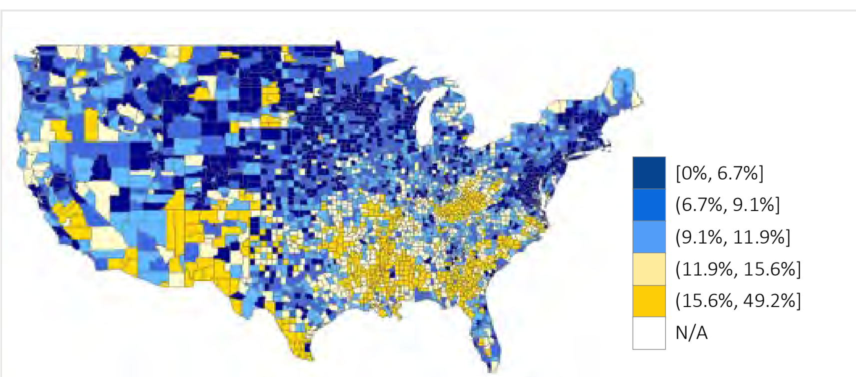
The percentage of single parent households is highest in the counties of the southern U.S.

Figure 2.13
PERCENTAGE OF HOUSEHOLDS WITH YEARLY INCOME ABOVE $75,000



Many counties in the Southeastern and Southwest have lower percentages of households with yearly income above $75,000 than in other parts of the country.

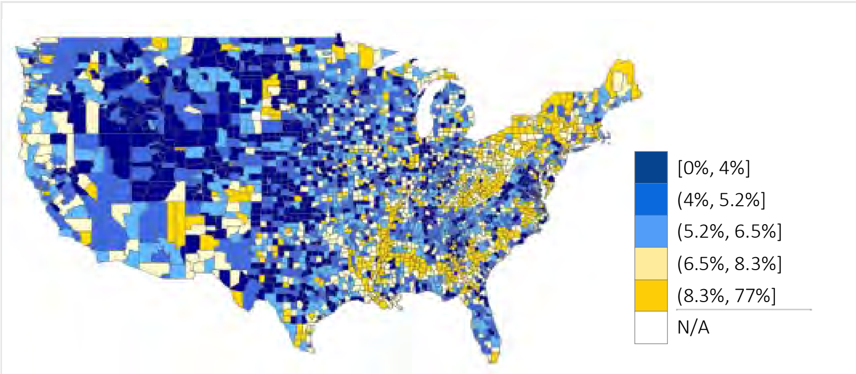
Figure 2.14
PERCENTAGE OF HOUSEHOLDS IN POVERTY



Counties in the southern U.S. show a higher percentage of households in poverty, while counties in the northeastern U.S. show relatively low percentages of poverty.

Figure 2.15

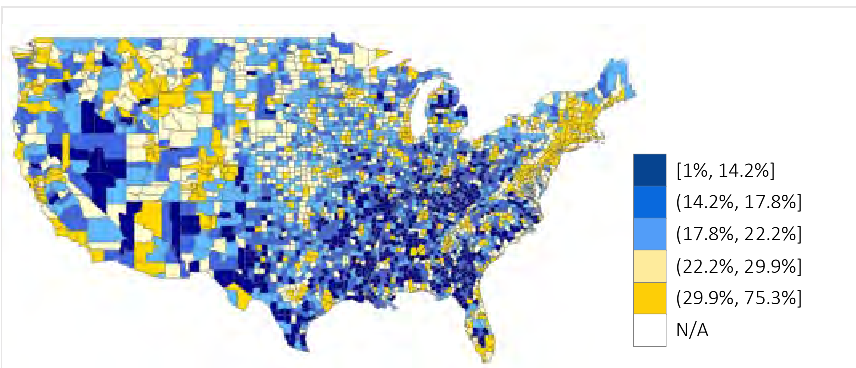
PERCENTAGE OF HOUSEHOLDS WITH NO VEHICLES



The percentage of households with no vehicles is higher in the eastern U.S. than the western U.S.

Figure 2.16

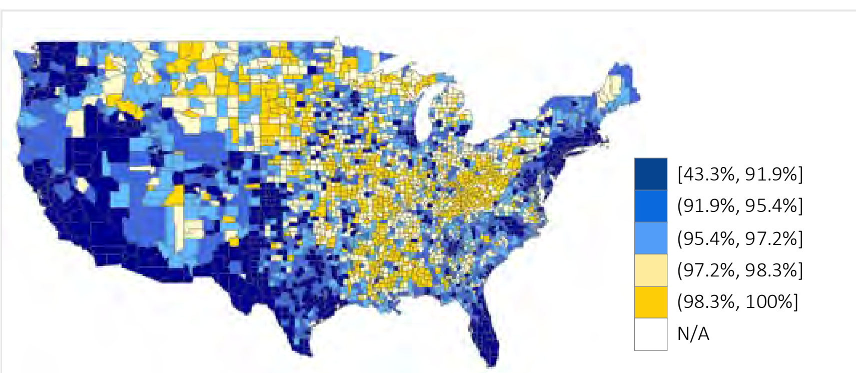
PERCENTAGE OF THE POPULATION WITH A BACHELOR'S DEGREE OR HIGHER (25 YEARS AND OLDER)



The percentage of adults having a bachelor's degree is higher in the Northeastern and some counties in the West where large cities are located.

Figure 2.17

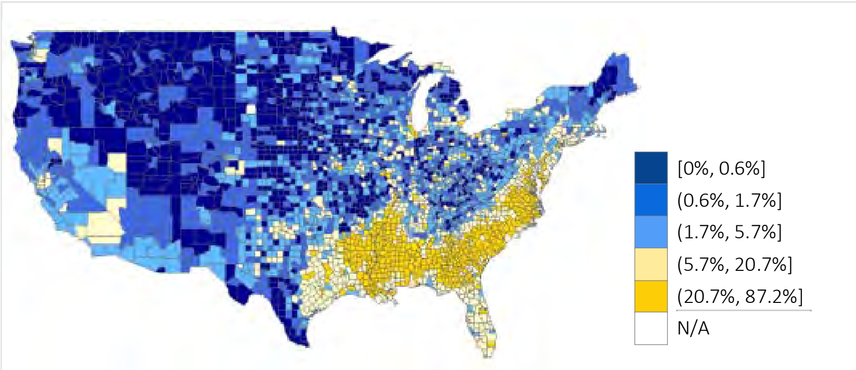
PERCENTAGE OF THE POPULATION BORN IN THE U.S.



There is a higher percentage of the population born in the U.S. in the Midwest than in areas along the Atlantic coast and in regions of the western U.S.

Figure 2.18

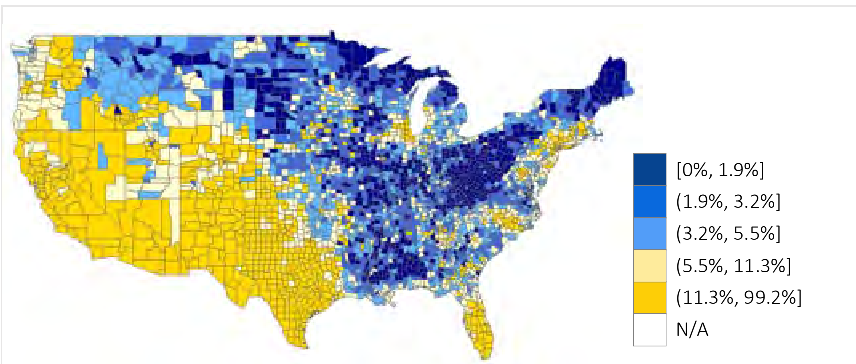
PERCENTAGE OF THE POPULATION THAT IS BLACK/AFRICAN AMERICAN



There is a significantly higher percentage of the population that is Black/African American in the southeastern U.S. than in other parts of the country.

Figure 2.19

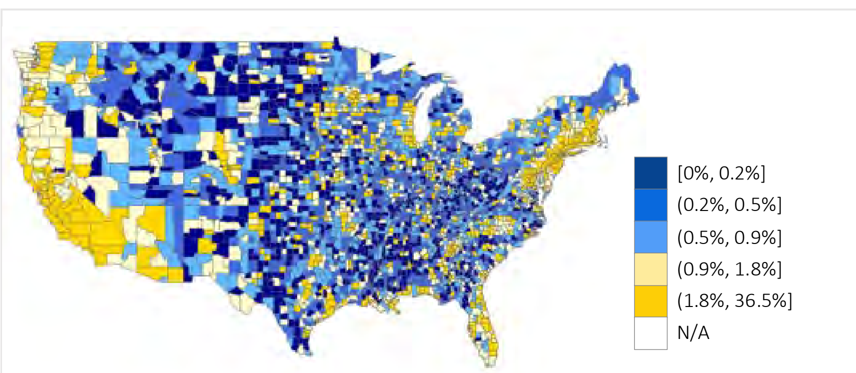
PERCENTAGE OF THE POPULATION THAT IS HISPANIC/LATINO



The percentage of the population that is Hispanic/Latino is higher in Florida, Texas, the Southwest and West than in other areas of the U.S.

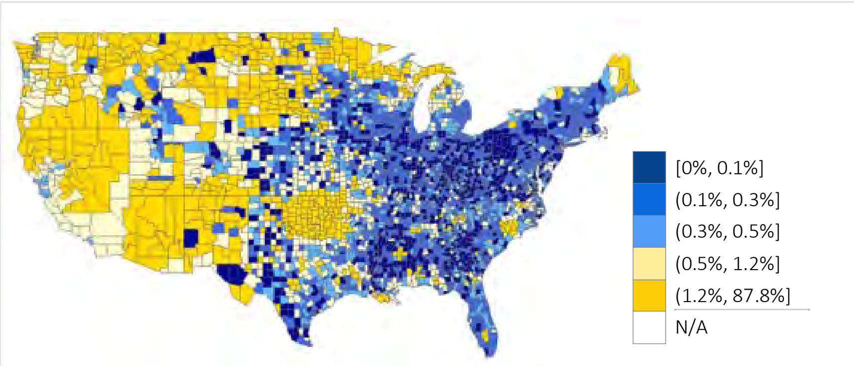
Figure 2.20

PERCENTAGE OF THE POPULATION THAT IS ASIAN/ASIAN AMERICAN



Populations on the western and eastern coasts have greater concentrations of Asian/Asian Americans than other regions of the country.

Figure 2.21
PERCENTAGE OF THE POPULATION THAT IS INDIGENOUS



The western half of the U.S. has a significantly higher percentage of the population that is Native American or Indigenous than the eastern half.

Section 3: Methodology

At the outset, we denote the insurance demand data under investigation by $\{\mathbf{x}_i, y_i, (u_i, v_i)\}_{i=1,\ldots,n}$, where $\mathbf{x}_i = (x_{i1}, \ldots, x_{id})^T$ is the d -dimension covariate containing the potential determinants of insurance demand, and y_i is the demand response variable which can be the number of policies sold, the premium amount of policies sold or the face amount of policies sold per capita, and (u_i, v_i) is the geographic coordinate from which the i -observation is originated.

3.1 STATE OF THE ART REGRESSION ANALYSIS

In a tradition regression analysis, each $(\mathbf{x}_i, y_i)$ is treated as an independently and identically distributed sample generated from a linear regression model:

(2.1) $$y_i = \beta_0 + \sum_{k=1}^{d} \beta_k x_{ik} + \epsilon_i,$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)$ are the regression coefficients, and ϵ_i denotes a normally distributed zero mean error term. Using ordinary least squares (OLS) method, an estimate of $\boldsymbol{\beta}$ can be obtained as

(2.2) $$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y},$$

where $\mathbf{X} = (1, \mathbf{x}_1, \ldots, \mathbf{x}_n)$, and $\mathbf{y} = (y_1, \ldots, y_n)^T$. The regression coefficient estimate provides a global picture about the relationship between the covariates and response including the sign and magnitude in relation to a priori set of hypotheses.

When following the above route to studying the associations between covariates and response, one must be cautious with linear regression because the geographic information in the data and thus the spatial variation in the local observations are completely disregarded. As such, the regression coefficient estimate considered in Equation (2.2) is a global statistic, representing the average relationship over space. This average relationship may not be a representative data pattern in any location under consideration. Instead, it may occur that the associations between covariates and response in two locations are contrasting of each other. In this situation, the differences in local associations may be cancelled out due to the averaging involved in the global statistic calculation. The aforementioned issue becomes a more serious concern as the spatial variance in the local observations increases.

Giving the U.S.'s spatially diverse social and demographic landscape, it is more informative—and even necessary—for us to account for the spatial variation in the data when modeling the determinants of insurance demand. For instance, the statistically non-significant determinant of the insured population in one territory may become significant in another. Fitting a regression model to entire U.S. insurance demand data may be too global in its scale and overlook the subtle differences in the impacts of social factors among different counties or regions, leading to implausible or less useful statistical conclusions. In the next subsection, we describe a more general notion of multiscale geographically weighted regression (MGWR) to study spatially varying relationships.

3.2 MULTISCALE GEOGRAPHICALLY WEIGHTED REGRESSION

MGWR extends the linear regression by allowing the regression coefficients to vary in relation to space. Thereby, we can use the location-specific parameter estimate to examine the local associations between covariates and response in terms of sign and magnitude. Formally, an MGWR model is defined as

(2.3) $y_i = \beta_0(u_i, v_i) + \sum_{k=1}^{d} \beta_k(u_i, v_i) x_{ik} + \epsilon_i$,

where $\beta_k(u_i, v_i)$ now becomes a regression parameter that depends on the geographic coordinate (u_i, v_i), $i = 1, \ldots, n$ and $k = 1, \ldots, d$. MGWR allow the spatial variations in relationships to be recognized by introducing a (usually, continuous) surface of parameter values for each covariate. If $\beta_k(u_i, v_i)$ is set to be a constant, then MGWR (2.3) reduces back to the traditional linear regression (2.1). The intercept β_0 represents the residual spatial variation which remains after controlling for the model covariates.

Fitting MGWR (2.3) to spatial data is considerably more complicated than the fitting of linear regression. Motivated by the fact that the OLS method for estimating linear regression assigns equal weights to all the spatially varying observations, which yields a global regression parameter estimation, a geographically weighted variant of the original OLS method can be applied to obtain a local estimate of MGWR:

(2.4) $\hat{\boldsymbol{\beta}}(u_i, v_i) = (\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{y}$,

where $\mathbf{W}(u_i, v_i) \in \mathbb{R}^{n \times n}$ is a weight matrix with the j-th diagonal elements being the geographical weight of the j-th observation to the i-th observation, and off-diagonal elements being all zero. It is straightforward to see that when the geographical weights are all equal to one, then at any spatial point (u_i, v_i), the regression coefficients are estimated in the same way using all the observations. In this case, the regression estimator in Equation (2.4) collapses to the linear regression estimator in Equation (2.2) which does not account for the spatial variation in relationships. Otherwise, the weighting procedure in Equation (2.4) is similar to fitting a local regression to a subset of data surrounding the spatial coordinate (u_i, v_i) . Thereby, fitting MGWR to spatial data can be also view as fitting an ensemble of local linear regression models at any number of locations.

The subsequent question pertains to how to select the appropriate geographical weights. Denote the geographical weight of the j-th observation to the i-th observation by ω_{ij}, and the distance between the i-th and j-th observations for $i, j = 1, \ldots, n$. A simple way to implement the weighting procedure is to choose

$\omega_{ij} = 1$ if $d_{ij} < d$;

$\omega_{ij} = 0$ otherwise,

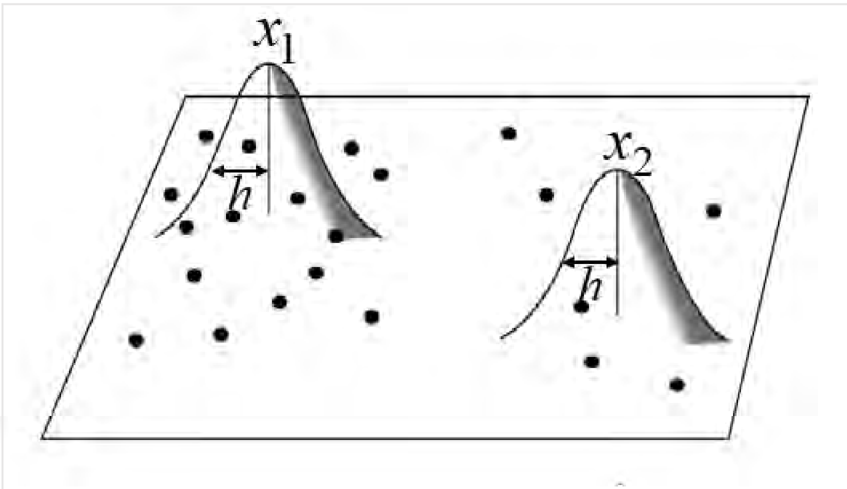
where d is a pre-specified distance threshold. The above setup implies that observations that are further away than d from a regression point are excluded in the fitting of the associated local repression. One drawback of this simple approach is that the fitted parameter surfaces may not be continuous because as the regression point changes, the observations being included in the local regression fitting may also change substantially. Alternatively, we can implement the weighting via bisquare kernel smoothing:

(2.5) $\omega_{ij} = (1 - (d_{ij} / h)^2)^2$ if $d_{ij} < h$;

$\omega_{ij} = 0$ otherwise,

where h is a bandwidth parameter to be tuned to minimize prediction errors. A smaller bandwidth corresponds to a more concentrated density which assigns more weights to those data points closed to the regression point, and vice versa. Figure 2.1.1 illustrates the weighting procedure in an illuminating manner.

Figure 2.1.1
ILLUSTRATION OF GWR WITH CONSTANT BANDWIDTH



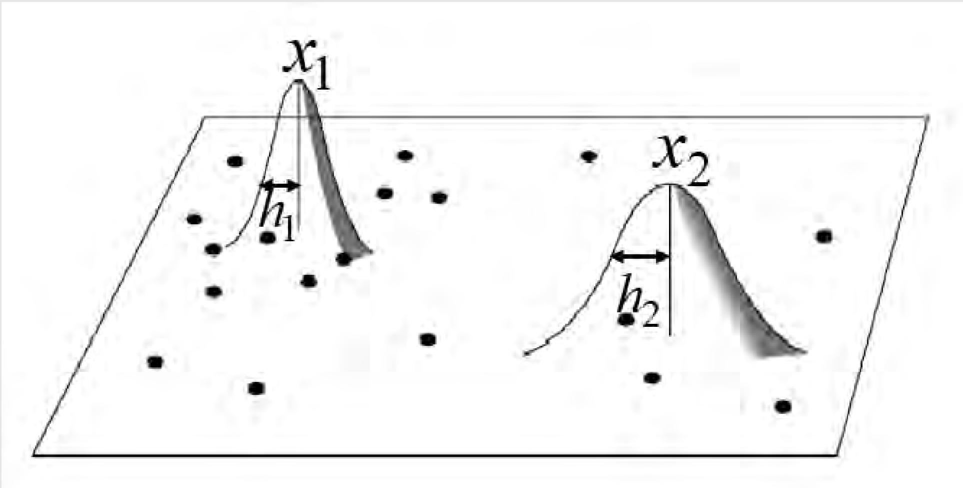
Kernel-based weighting (2.5) uses the same kernel bandwidth for all the regression points. It may become an issue when the number of data points surrounding a study area is substantially less than another area (e.g., the right-hand area around x_1 versus the left-hand area around x_2 in Figure 2.1.1). If there is a significant data imbalance appearing over space, then some local regression models are only calibrated based on very few observations, causing unacceptably large standard errors for the regression estimators. To address this issue, we can modify the kernel-based weights (2.5) such that their bandwidths are adaptive to the denseness of data points around a study area. The aforementioned adaptive spatial kernel method is illustrated in Figure 2.1.2, in which we can see a smaller (resp. large) bandwidth is assigned to the regression point x_1 on the left-hand (resp. x_2 on the right-hand) side of the figure where data are plentiful (resp. scarce). Mathematically, the adaptive version of the bisquare kernel in Equation (2.5) is given by

(2.6) $\omega_{ij} = (1-(d_{ij}/h_i)^2)^2$ if $d_{ij} < h_i$;

$\omega_{ij} = 0$ otherwise,

where h_i is the distance from the i-th observation to its m-th nearest neighbor. However, the use of h_i in normalizes the magnitudes of the distances such that the order of the weights depends on the rank of distance. Namely, the closest data to a given regression point is assigned the highest weight, and the weights decrease according to the increasing rank of the distance. However, depending on the denseness of data points surrounding the i-th observation, ω_{ij} is not necessarily greater than ω_{ij^*} even though $d_{ij} < d_{ij^*}$. This makes the weighting method in Equation (2.6) dependent of the local denseness of data points.

Figure 2.1.2
ILLUSTRATION OF GWR WITH ADAPTIVE BANDWIDTH



So far, we have discussed how to estimate the spatial-varying regression coefficients involved in MGWR. One may be also interested in constructing confidence intervals for the regression coefficients which can be further used for hypothesis testing purpose. In an MGWR model, the standard errors for the coefficient estimators can be computed by inverting the local information matrix. Then the statistical inferences can be made following the same approach used in the classical linear regression context. We refer to the readers to Fotheringham, et al. (2003) for more detailed discussions.

Section 4: Impact of Agents

This section and section 4.2 evaluate the role that the variable number of agents per county may play in the context of insurance demand. Since this data is available at the level of states,[5] we interpolate at the level of counties by using both the population size and the income at the county level. We investigate how well the number of agents (disaggregated either by population size or income) is explained by other covariates under consideration and how well the premiums considered are explained by the number of agents (disaggregated either by population size or income).

4.1 AGENTS (POPULATION) VS. COVARIATES

This section evaluates the variable number of agents interpolated by the population size versus the set of covariates in Table 2.1. From Table 4.1.1 we identify that the following covariates are significant at 5% of significance level:

- Percentage of the population that is Asian/Asian American, percentage of the population that is Hispanic/Latino,
- Percentage of the population living in the same place since 2009,
- Percentage of households in poverty,
- Percentage of the population born in the U.S.,
- Percentage of households with no vehicles,
- Response rate for the 2010 census,
- Gini index, and
- Association density.

Locally, we found that the intercept and eleven of the covariates have a significant effect (we include maps only for these cases). The scale of these maps is different for each covariate and displays negative parameter estimates in blue, positive estimates in yellow, and nonsignificant estimates in grey.

In this case we find that adjusted R2 of global model (0.380, Table 4.1.1) and our spatial model (0.752, Table 4.1.2) are reasonably high, indicating that number of agents in this case is reasonably well-explained by our covariates. This suggest that number of agents should not be included as a covariate in the final spatial regression model.

[5] See: www.bls.gov/oes/current/oes413021.htm

Table 4.1.1

ESTIMATION MODEL

```
===========================================================================
MGWR Version: 2.2.1
Released on: 03/20/2020
Source code is available at: https://github.com/pysal/mgwr
Development Team: Ziqi Li, Taylor Oshan, Stewart Fotheringham, Wei Kang,
Levi Wolf, Hanchen Yu, Mehak Sachdeva, and Sarah Bardin
Spatial Analysis Research Center (SPARC)
Arizona State University, Tempe, USA
===========================================================================
Model type:                                                        Gaussian
Number of observations:                                                3073
Number of covariates:                                                    20
Dependent variable:                                          Int.Nagents.pop
Variable standardization:                                                On
Total runtime:                                                      1:34:56

Global Regression Results
---------------------------------------------------------------------------
Residual sum of squares:                                           1894.097
Log-likelihood:                                                   -3616.867
AIC:                                                               7273.734
AICc:                                                              7276.037
R2:                                                                   0.384
Adj. R2:                                                              0.380

Variable                                   Est.         SE  t(Est/SE)    p-value
------------------------------------ ---------- ---------- ---------- ----------
Intercept                                -0.000      0.014     -0.000      1.000
P_AfriA                                   0.036      0.022      1.618      0.106
P_Indig                                   0.034      0.018      1.901      0.057
P_Asian                                   0.216      0.023      9.368      0.000
P_Hisp                                   -0.063      0.025     -2.529      0.011
BachDe                                    0.008      0.028      0.288      0.773
OwnerOcc                                  0.018      0.029      0.627      0.531
SameHous                                 -0.054      0.024     -2.264      0.024
P_Labor                                   0.032      0.025      1.266      0.206
Unemp                                     0.010      0.020      0.509      0.610
HInc                                     -0.031      0.032     -0.946      0.344
NoHeIns                                  -0.030      0.021     -1.421      0.155
Poverty                                  -0.065      0.029     -2.200      0.028
SinPar                                    0.006      0.021      0.299      0.765
BornUSA                                  -0.357      0.030    -11.946      0.000
NoVehi                                    0.101      0.020      5.162      0.000
CenResp                                   0.097      0.021      4.575      0.000
Gini                                      0.097      0.019      5.119      0.000
AssDens                                  -0.074      0.017     -4.416      0.000
VoTurn                                    0.038      0.020      1.888      0.059
```

Table 4.1.2
ESTIMATION MODEL (CONTINUED)

```
Multiscale Geographically Weighted Regression (MGWR) Results
---------------------------------------------------------------------------------
Coordinates type:                                                       Projected
Spatial kernel:                                                 Adaptive bisquare
Criterion for optimal bandwidth:                                             AICc
Score of change (SOC) type:                                           Smoothing f
Termination criterion for MGWR:                                           1.0e-05
Number of iterations used:                                                     65

MGWR bandwidths
---------------------------------------------------------------------------------
Variable                  Bandwidth      ENP_j   Adj t-val(95%)             DoD_j
Intercept                   396.000     10.278            2.818             0.710
P_AfriA                     211.000     20.167            3.028             0.626
P_Indig                    3071.000      1.139            2.016             0.984
P_Asian                      44.000    141.020            3.576             0.384
P_Hisp                     2188.000      1.418            2.106             0.956
BachDe                     2688.000      1.691            2.177             0.935
OwnerOcc                   1018.000      5.592            2.616             0.786
SameHous                   3054.000      1.219            2.044             0.975
P_Labor                    3071.000      1.195            2.036             0.978
Unemp                      3048.000      1.226            2.047             0.975
HInc                       3071.000      1.179            2.030             0.979
NoHeIns                    3071.000      1.144            2.018             0.983
Poverty                    3071.000      1.126            2.011             0.985
SinPar                     3071.000      1.211            2.042             0.976
BornUSA                      44.000    145.615            3.584             0.380
NoVehi                       44.000    165.211            3.617             0.364
CenResp                    3071.000      1.263            2.059             0.971
Gini                       3071.000      1.256            2.057             0.972
AssDens                    2360.000      2.066            2.255             0.910
VoTurn                     3071.000      1.203            2.039             0.977

Diagnostic Information
---------------------------------------------------------------------------------
Residual sum of squares:                                                  636.639
Effective number of parameters (trace(S)):                                506.220
Degree of freedom (n - trace(S)):                                        2566.780
Sigma estimate:                                                             0.498
Log-likelihood:                                                         -1941.630
Degree of Dependency (DoD):                                                 0.598
AIC:                                                                     4897.700
AICc:                                                                    5098.715
BIC:                                                                     7956.447
R2:                                                                         0.793
Adj. R2:                                                                    0.752
```

Table 4.1.3
ESTIMATION MODEL (CONTINUED)

Summary Statistics For MGWR Parameter Estimates

Variable	Mean	STD	Min	Median	Max
Intercept	0.027	0.099	−0.209	0.024	0.238
P_AfriA	0.312	0.302	−0.481	0.327	1.032
P_Indig	0.035	0.000	0.034	0.035	0.037
P_Asian	0.139	0.362	−1.206	0.065	4.038
P_Hisp	−0.133	0.075	−0.226	−0.146	−0.039
BachDe	0.038	0.006	0.025	0.038	0.052
OwnerOcc	0.019	0.035	−0.071	0.018	0.092
SameHous	−0.024	0.003	−0.031	−0.023	−0.019
P_Labor	0.031	0.001	0.030	0.030	0.033
Unemp	0.003	0.004	−0.006	0.005	0.008
HInc	0.010	0.003	0.007	0.010	0.015
NoHeIns	−0.012	0.003	−0.016	−0.013	−0.006
Poverty	−0.012	0.001	−0.015	−0.012	−0.010
SinPar	0.014	0.001	0.012	0.015	0.016
BornUSA	−0.283	0.371	−2.781	−0.181	0.228
NoVehi	0.142	0.366	−0.338	0.057	4.996
CenResp	0.093	0.003	0.085	0.094	0.096
Gini	0.025	0.002	0.019	0.025	0.029
AssDens	−0.057	0.021	−0.097	−0.051	−0.033
VoTurn	0.036	0.001	0.035	0.037	0.038

Acknowledgement:
We acknowledge the support of the National Science Foundation under Award 1758786 from the Geography and Spatial Sciences Program to A. S. Fotheringham which enabled this software to be written and made freely available.

4.2 AGENTS (INCOME) VS. COVARIATES

This section evaluates the variable number of agents interpolated by the income versus the set of covariates in Table 2.1. From Table 4.2.1, we identify that only the intercept, percentage of the population that is Indigenous, and response rate for the 2020 census are not significant at a 5% of significance level. Comparing with the results in the previous section (number of agents interpolate by the population size), we found that for the global effect, only two covariates are not significant in the current case. In contrast, ten of them are not significant in the previous section.

Locally, we found that the intercept and most of the covariates have a significant effect (we include maps only for these cases). The scale of these maps is different for each covariate and displays negative parameter estimates in blue, positive estimates in yellow, and nonsignificant estimates in grey.

Similarly, to the previous case, we find that adjusted R2 of global model (0.470) and our spatial model (0.944) is reasonably high, indicating that number of agents in this case is reasonably well-explained by our covariates. This suggest that number of agents should not be included as a covariate in the final spatial regression model.

Table 4.2.1

ESTIMATION MODEL

MGWR Version: 2.2.1
Released on: 03/20/2020
Source code is available at: https://github.com/pysal/mgwr
Development Team: Ziqi Li, Taylor Oshan, Stewart Fotheringham, Wei Kang,
Levi Wolf, Hanchen Yu, Mehak Sachdeva, and Sarah Bardin
Spatial Analysis Research Center (SPARC)
Arizona State University, Tempe, USA

Model type:	Gaussian
Number of observations:	3073
Number of covariates:	20
Dependent variable:	Int.Nagents
Variable standardization:	On
Total runtime:	3:41:24

Global Regression Results

Residual sum of squares:	1620.001
Log-likelihood:	-3376.688
AIC:	6793.377
AICc:	6795.680
R2:	0.473
Adj. R2:	0.470

Variable	Est.	SE	t(Est/SE)	p-value
Intercept	-0.000	0.013	-0.000	1.000
P_AfriA	-0.173	0.021	-8.414	0.000
P_Indig	-0.011	0.017	-0.666	0.505
P_Asian	0.130	0.021	6.072	0.000
P_Hisp	-0.072	0.023	-3.164	0.002
BachDe	-0.093	0.026	-3.523	0.000
OwnerOcc	-0.130	0.027	-4.905	0.000
SameHous	0.058	0.022	2.633	0.008
P_Labor	-0.371	0.023	-16.008	0.000
Unemp	0.140	0.019	7.472	0.000
HInc	0.427	0.030	14.327	0.000
NoHeIns	-0.067	0.020	-3.448	0.001
Poverty	-0.233	0.027	-8.568	0.000
SinPar	0.147	0.020	7.536	0.000
BornUSA	-0.385	0.028	-13.948	0.000
NoVehi	0.048	0.018	2.631	0.009
CenResp	-0.012	0.020	-0.630	0.529
Gini	0.078	0.018	4.455	0.000
AssDens	-0.067	0.015	-4.317	0.000
VoTurn	0.126	0.019	6.728	0.000

Table 4.2.1

ESTIMATION MODEL (CONTINUED)

```
Multiscale Geographically Weighted Regression (MGWR) Results
---------------------------------------------------------------------------
Coordinates type:                                                 Projected
Spatial kernel:                                           Adaptive bisquare
Criterion for optimal bandwidth:                                       AICc
Score of change (SOC) type:                                     Smoothing f
Termination criterion for MGWR:                                     1.0e-05
Number of iterations used:                                              157

MGWR bandwidths
---------------------------------------------------------------------------
Variable                  Bandwidth      ENP_j   Adj t-val(95%)       DoD_j
Intercept                    44.000    104.693            3.497       0.421
P_AfriA                     116.000     27.324            3.119       0.588
P_Indig                      44.000    104.783            3.497       0.421
P_Asian                    3071.000      1.412            2.105       0.957
P_Hisp                     1642.000      2.146            2.270       0.905
BachDe                      220.000     17.750            2.989       0.642
OwnerOcc                   3071.000      1.090            1.997       0.989
SameHous                     44.000    144.879            3.583       0.380
P_Labor                      44.000    125.735            3.546       0.398
Unemp                      3071.000      1.123            2.010       0.986
HInc                        917.000      3.344            2.435       0.850
NoHeIns                     128.000     34.641            3.189       0.559
Poverty                     238.000     16.370            2.965       0.652
SinPar                     3071.000      1.158            2.023       0.982
BornUSA                    3071.000      1.116            2.007       0.986
NoVehi                     3071.000      1.173            2.028       0.980
CenResp                    3071.000      1.131            2.013       0.985
Gini                         61.000    102.670            3.492       0.423
AssDens                      61.000     94.732            3.470       0.433
VoTurn                       44.000    130.984            3.556       0.393

Diagnostic Information
---------------------------------------------------------------------------
Residual sum of squares:                                            121.461
Effective number of parameters (trace(S)):                          918.253
Degree of freedom (n - trace(S)):                                  2154.747
Sigma estimate:                                                       0.237
Log-likelihood:                                                     603.756
Degree of Dependency (DoD):                                           0.523
AIC:                                                                630.993
AICc:                                                              1416.915
BIC:                                                               6174.464
R2:                                                                   0.960
Adj. R2:                                                              0.944
```

Table 4.2.1
ESTIMATION MODEL (CONTINUED)

Summary Statistics For MGWR Parameter Estimates

Variable	Mean	STD	Min	Median	Max
Intercept	-0.010	0.712	-0.610	-0.159	4.188
P_AfriA	0.103	0.233	-0.192	0.022	1.026
P_Indig	-0.076	0.466	-2.962	-0.013	1.815
P_Asian	0.022	0.002	0.017	0.022	0.026
P_Hisp	-0.042	0.008	-0.076	-0.041	-0.027
BachDe	0.023	0.088	-0.054	0.002	0.488
OwnerOcc	0.004	0.001	0.003	0.004	0.006
SameHous	0.035	0.195	-0.659	0.001	1.316
P_Labor	0.020	0.163	-0.490	-0.008	1.475
Unemp	0.013	0.001	0.011	0.013	0.014
HInc	0.121	0.033	0.079	0.111	0.204
NoHeIns	-0.054	0.143	-0.650	-0.016	0.105
Poverty	-0.040	0.067	-0.295	-0.016	0.029
SinPar	0.010	0.001	0.006	0.010	0.010
BornUSA	-0.031	0.001	-0.032	-0.031	-0.029
NoVehi	0.022	0.001	0.020	0.022	0.023
CenResp	-0.016	0.002	-0.018	-0.016	-0.013
Gini	0.050	0.113	-0.102	0.020	0.600
AssDens	-0.000	0.104	-0.769	0.012	0.305
VoTurn	0.045	0.172	-0.859	0.048	0.754

Acknowledgement:
We acknowledge the support of the National Science Foundation under Award 1758786 from the Geography and Spatial Sciences Program to A. S. Fotheringham which enabled this software to be written and made freely available.

4.3 PREMIUMS VS. AGENTS (POPULATION)

This section evaluates the premiums versus the number of agents interpolated by the population size per county. Globally, the number of agents has a significant effect over all the premiums, and we observe a very large value of the coefficient of determination statistic, likely due to the usage of population in the creation of this number of agents variable. Given such strong correlation between the number of agents and considered premiums, this variable will be omitted from the final regression model.

Table 4.3.3

ESTIMATION MODEL

```
===============================================================================
MGWR Version: 2.2.1
Released on: 03/20/2020
Source code is available at: https://github.com/pysal/mgwr
Development Team: Ziqi Li, Taylor Oshan, Stewart Fotheringham, Wei Kang,
Levi Wolf, Hanchen Yu, Mehak Sachdeva, and Sarah Bardin
Spatial Analysis Research Center (SPARC)
Arizona State University, Tempe, USA
===============================================================================
Model type:                                                            Gaussian
Number of observations:                                                    3074
Number of covariates:                                                         2
Dependent variable:                                Total.Annualized.Premium.Sold
Variable standardization:                                                    On
Total runtime:                                                          0:01:19

Global Regression Results
-------------------------------------------------------------------------------
Residual sum of squares:                                                241.449
Log-likelihood:                                                        -451.571
AIC:                                                                    907.143
AICc:                                                                   909.151
R2:                                                                       0.921
Adj. R2:                                                                  0.921

Variable                                   Est.         SE  t(Est/SE)    p-value
------------------------------------ ---------- ---------- ---------- ----------
Intercept                                 0.000      0.005      0.000      1.000
Int.Nagents.pop                           0.960      0.005    189.840      0.000

Multiscale Geographically Weighted Regression (MGWR) Results
-------------------------------------------------------------------------------
Coordinates type:                                                     Projected
Spatial kernel:                                               Adaptive bisquare
Criterion for optimal bandwidth:                                           AICc
Score of change (SOC) type:                                         Smoothing f
Termination criterion for MGWR:                                         1.0e-05
Number of iterations used:                                                    9
```

Table 4.3.3
ESTIMATION MODEL (CONTINUED)

```
MGWR bandwidths
-------------------------------------------------------------------------------
Variable                  Bandwidth      ENP_j   Adj t-val(95%)            DoD_j
Intercept                   144.000     49.194            3.289            0.515
Int.Nagents.pop              44.000    163.929            3.615            0.365

Diagnostic Information
-------------------------------------------------------------------------------
Residual sum of squares:                                                 128.717
Effective number of parameters (trace(S)):                               213.122
Degree of freedom (n - trace(S)):                                       2860.878
Sigma estimate:                                                            0.212
Log-likelihood:                                                          515.270
Degree of Dependency (DoD):                                                0.419
AIC:                                                                    -602.295
AICc:                                                                   -570.071
BIC:                                                                     689.019
R2:                                                                        0.958
Adj. R2:                                                                   0.955

Summary Statistics For MGWR Parameter Estimates
-------------------------------------------------------------------------------
Variable                        Mean        STD        Min     Median        Max
--------------------      ---------- ---------- ---------- ---------- ----------
Intercept                     -0.014      0.082     -0.145     -0.026      0.234
Int.Nagents.pop                0.891      0.281      0.426      0.851      1.781
===============================================================================
Acknowledgement:
We acknowledge the support of the National Science Foundation under Award 1758786
from the Geography and Spatial Sciences Program to A. S. Fotheringham which
enabled this software to be written and made freely available.
===============================================================================
```

Table 4.3.4
ESTIMATION MODEL

```
===========================================================================
MGWR Version: 2.2.1
Released on: 03/20/2020
Source code is available at: https://github.com/pysal/mgwr
Development Team: Ziqi Li, Taylor Oshan, Stewart Fotheringham, Wei Kang,
Levi Wolf, Hanchen Yu, Mehak Sachdeva, and Sarah Bardin
Spatial Analysis Research Center (SPARC)
Arizona State University, Tempe, USA
===========================================================================
Model type:                                                        Gaussian
Number of observations:                                                3074
Number of covariates:                                                     2
Dependent variable:                                    Term.Ins.Premium.Sold
Variable standardization:                                                On
Total runtime:                                                      0:01:18

Global Regression Results
---------------------------------------------------------------------------
Residual sum of squares:                                            215.214
Log-likelihood:                                                    -274.779
AIC:                                                                553.558
AICc:                                                               555.566
R2:                                                                   0.930
Adj. R2:                                                              0.930

Variable                                   Est.         SE  t(Est/SE)    p-value
------------------------------------ ---------- ---------- ---------- ----------
Intercept                                 0.000      0.005      0.000      1.000
Int.Nagents.pop                           0.964      0.005    202.007      0.000

Multiscale Geographically Weighted Regression (MGWR) Results
---------------------------------------------------------------------------
Coordinates type:                                                 Projected
Spatial kernel:                                           Adaptive bisquare
Criterion for optimal bandwidth:                                       AICc
Score of change (SOC) type:                                     Smoothing f
Termination criterion for MGWR:                                     1.0e-05
Number of iterations used:                                               10
```

Table 4.3.4

ESTIMATION MODEL (CONTINUED)

```
MGWR bandwidths
---------------------------------------------------------------------------
Variable                  Bandwidth      ENP_j   Adj t-val(95%)          DoD_j
Intercept                    65.000    115.821            3.524          0.408
Int.Nagents.pop              44.000    151.178            3.594          0.375

Diagnostic Information
---------------------------------------------------------------------------
Residual sum of squares:                                                79.243
Effective number of parameters (trace(S)):                             266.999
Degree of freedom (n - trace(S)):                                     2807.001
Sigma estimate:                                                          0.168
Log-likelihood:                                                       1260.852
Degree of Dependency (DoD):                                              0.391
AIC:                                                                 -1985.705
AICc:                                                                -1934.303
BIC:                                                                  -369.474
R2:                                                                      0.974
Adj. R2:                                                                 0.972

Summary Statistics For MGWR Parameter Estimates
---------------------------------------------------------------------------
Variable                        Mean        STD        Min     Median        Max
--------------------      ---------- ---------- ---------- ---------- ----------
Intercept                     -0.013      0.098     -0.165     -0.033      0.519
Int.Nagents.pop                0.866      0.278      0.428      0.817      1.673
===========================================================================
Acknowledgement:
We acknowledge the support of the National Science Foundation under Award 1758786
from the Geography and Spatial Sciences Program to A. S. Fotheringham which
enabled this software to be written and made freely available.
===========================================================================
```

4.4 PREMIUMS VS. AGENTS (INCOME)

This section evaluates the premiums sold versus the number of agents interpolated by income. The global models presented in Tables 4.4.2 to 4.4.5 reveal that when using income to disentangle the number of agents in each county, the resulting model is only able to explain a negligible amount of the variation in premiums sold.

Table 4.4.2
ESTIMATION MODEL

```
===========================================================================
MGWR Version: 2.2.1
Released on: 03/20/2020
Source code is available at: https://github.com/pysal/mgwr
Development Team: Ziqi Li, Taylor Oshan, Stewart Fotheringham, Wei Kang,
Levi Wolf, Hanchen Yu, Mehak Sachdeva, and Sarah Bardin
Spatial Analysis Research Center (SPARC)
Arizona State University, Tempe, USA
===========================================================================
Model type:                                                        Gaussian
Number of observations:                                                3074
Number of covariates:                                                     2
Dependent variable:                               Permanent.Ins.Premium.Sold
Variable standardization:                                                On
Total runtime:                                                      0:02:04

Global Regression Results
---------------------------------------------------------------------------
Residual sum of squares:                                           3066.664
Log-likelihood:                                                   -4358.145
AIC:                                                               8720.289
AICc:                                                              8722.297
R2:                                                                   0.002
Adj. R2:                                                              0.002

Variable                                  Est.         SE  t(Est/SE)    p-value
------------------------------------ ---------- ---------- ---------- ----------
Intercept                               -0.000      0.018     -0.000      1.000
Int.Nagents                             -0.049      0.018     -2.711      0.007

Multiscale Geographically Weighted Regression (MGWR) Results
---------------------------------------------------------------------------
Coordinates type:                                                 Projected
Spatial kernel:                                           Adaptive bisquare
Criterion for optimal bandwidth:                                       AICc
Score of change (SOC) type:                                     Smoothing f
Termination criterion for MGWR:                                     1.0e-05
Number of iterations used:                                               16
```

Table 4.4.2
ESTIMATION MODEL (CONTINUED)

```
MGWR bandwidths
-------------------------------------------------------------------------------
Variable                        Bandwidth      ENP_j   Adj t-val(95%)            DoD_j
Intercept                          48.000    188.826            3.652            0.347
Int.Nagents                      1548.000      4.206            2.517            0.821

Diagnostic Information
-------------------------------------------------------------------------------
Residual sum of squares:                                               2367.075
Effective number of parameters (trace(S)):                              193.033
Degree of freedom (n - trace(S)):                                      2880.967
Sigma estimate:                                                           0.906
Log-likelihood:                                                       -3960.161
Degree of Dependency (DoD):                                               0.431
AIC:                                                                   8308.387
AICc:                                                                  8334.676
BIC:                                                                   9478.548
R2:                                                                       0.230
Adj. R2:                                                                  0.178

Summary Statistics For MGWR Parameter Estimates
-------------------------------------------------------------------------------
Variable                        Mean        STD        Min     Median        Max
--------------------      ---------- ---------- ---------- ---------- ----------
Intercept                     -0.042      0.361     -0.804     -0.105      2.787
Int.Nagents                   -0.307      0.273     -0.706     -0.352      0.004
===============================================================================
Acknowledgement:
We acknowledge the support of the National Science Foundation under Award 1758786
from the Geography and Spatial Sciences Program to A. S. Fotheringham which
enabled this software to be written and made freely available.
===============================================================================
```

Table 4.4.3

ESTIMATION MODEL

```
===========================================================================
MGWR Version: 2.2.1
Released on: 03/20/2020
Source code is available at: https://github.com/pysal/mgwr
Development Team: Ziqi Li, Taylor Oshan, Stewart Fotheringham, Wei Kang,
Levi Wolf, Hanchen Yu, Mehak Sachdeva, and Sarah Bardin
Spatial Analysis Research Center (SPARC)
Arizona State University, Tempe, USA
===========================================================================
Model type:                                                        Gaussian
Number of observations:                                                3074
Number of covariates:                                                     2
Dependent variable:                          Total.Annualized.Premium.Sold
Variable standardization:                                                On
Total runtime:                                                      0:03:19

Global Regression Results
---------------------------------------------------------------------------
Residual sum of squares:                                           3064.614
Log-likelihood:                                                   -4357.117
AIC:                                                               8718.234
AICc:                                                              8720.242
R2:                                                                   0.003
Adj. R2:                                                              0.003

Variable                                  Est.         SE  t(Est/SE)    p-value
------------------------------------ ---------- ---------- ---------- ----------
Intercept                               -0.000      0.018     -0.000      1.000
Int.Nagents                             -0.055      0.018     -3.067      0.002

Multiscale Geographically Weighted Regression (MGWR) Results
---------------------------------------------------------------------------
Coordinates type:                                                 Projected
Spatial kernel:                                           Adaptive bisquare
Criterion for optimal bandwidth:                                       AICc
Score of change (SOC) type:                                     Smoothing f
Termination criterion for MGWR:                                     1.0e-05
Number of iterations used:                                               20
```

Table 4.4.3

ESTIMATION MODEL (CONTINUED)

```
MGWR bandwidths
---------------------------------------------------------------------------------
Variable                        Bandwidth      ENP_j   Adj t-val(95%)       DoD_j
Intercept                          48.000    188.789            3.652       0.347
Int.Nagents                      1405.000      4.650            2.552       0.809

Diagnostic Information
---------------------------------------------------------------------------------
Residual sum of squares:                                                 2216.939
Effective number of parameters (trace(S)):                                193.439
Degree of freedom (n - trace(S)):                                        2880.561
Sigma estimate:                                                             0.877
Log-likelihood:                                                         -3859.445
Degree of Dependency (DoD):                                                 0.431
AIC:                                                                     8107.769
AICc:                                                                    8134.172
BIC:                                                                     9280.380
R2:                                                                         0.279
Adj. R2:                                                                    0.230

Summary Statistics For MGWR Parameter Estimates
---------------------------------------------------------------------------------
Variable                        Mean        STD        Min     Median         Max
--------------------      ---------- ---------- ---------- ---------- ----------
Intercept                     -0.049      0.408     -0.896     -0.117      2.593
Int.Nagents                   -0.364      0.319     -0.840     -0.389      0.002
=================================================================================
Acknowledgement:
We acknowledge the support of the National Science Foundation under Award 1758786
from the Geography and Spatial Sciences Program to A. S. Fotheringham which
enabled this software to be written and made freely available.
=================================================================================
```

Table 4.4.5
ESTIMATION MODEL

```
===========================================================================
MGWR Version: 2.2.1
Released on: 03/20/2020
Source code is available at: https://github.com/pysal/mgwr
Development Team: Ziqi Li, Taylor Oshan, Stewart Fotheringham, Wei Kang,
Levi Wolf, Hanchen Yu, Mehak Sachdeva, and Sarah Bardin
Spatial Analysis Research Center (SPARC)
Arizona State University, Tempe, USA
===========================================================================
Model type:                                                        Gaussian
Number of observations:                                                3074
Number of covariates:                                                     2
Dependent variable:                                     Term.Ins.Premium.Sold
Variable standardization:                                                On
Total runtime:                                                      0:02:39

Global Regression Results
---------------------------------------------------------------------------
Residual sum of squares:                                           3059.174
Log-likelihood:                                                   -4354.386
AIC:                                                               8712.772
AICc:                                                              8714.780
R2:                                                                   0.005
Adj. R2:                                                              0.004

Variable                                   Est.         SE  t(Est/SE)    p-value
------------------------------------ ---------- ---------- ---------- ----------
Intercept                                -0.000      0.018     -0.000      1.000
Int.Nagents                              -0.069      0.018     -3.859      0.000

Multiscale Geographically Weighted Regression (MGWR) Results
---------------------------------------------------------------------------
Coordinates type:                                                 Projected
Spatial kernel:                                           Adaptive bisquare
Criterion for optimal bandwidth:                                       AICc
Score of change (SOC) type:                                     Smoothing f
Termination criterion for MGWR:                                     1.0e-05
Number of iterations used:                                               33
```

Table 4.4.5

ESTIMATION MODEL (CONTINUED)

```
MGWR bandwidths
---------------------------------------------------------------------------
Variable                  Bandwidth     ENP_j   Adj t-val(95%)          DoD_j
Intercept                    44.000   176.041            3.634          0.356
Int.Nagents                  44.000   150.735            3.593          0.375

Diagnostic Information
---------------------------------------------------------------------------
Residual sum of squares:                                             1156.011
Effective number of parameters (trace(S)):                            326.776
Degree of freedom (n - trace(S)):                                    2747.224
Sigma estimate:                                                         0.649
Log-likelihood:                                                     -2858.624
Degree of Dependency (DoD):                                             0.365
AIC:                                                                 6372.800
AICc:                                                                6451.311
BIC:                                                                 8349.530
R2:                                                                     0.624
Adj. R2:                                                                0.579

Summary Statistics For MGWR Parameter Estimates
---------------------------------------------------------------------------
Variable                        Mean        STD        Min     Median        Max
--------------------      ---------- ---------- ---------- ---------- ----------
Intercept                     -0.239      0.525     -1.256     -0.320      1.778
Int.Nagents                   -1.254      1.388     -7.033     -0.923      2.209
===========================================================================
Acknowledgement:
We acknowledge the support of the National Science Foundation under Award 1758786
from the Geography and Spatial Sciences Program to A. S. Fotheringham which
enabled this software to be written and made freely available.
===========================================================================
```

Section 5: Premiums vs. Covariates

In this section, we consider all available premiums as response variables in our spatial regression models which include all introduced covariates excluding the number of agents. At the outset, let us address the potential collinearity issue among the covariates embedded in our regression analysis. Table 5.1 outlines the variance inflation factor (VIF) of the regression variables under investigation. Typically, a VIF value that is great than 10 suggests a significant multicollinearity that needs attentions (see, e.g., Kutner et al., 2005; Mendenhall et al., 2003). As shown, the VIF's for all the covariates considered in our regression analysis are below 10, ranging from 1.38 to 5.15. The VIF statistics suggest that there are no significant multicollinearities involved, and thus the estimates of the marginal effects of predictors on the response variable are creditable.

Table 5.1

SUMMARY TABLE OF THE VIF'S OF THE REGRESSION VARIABLES

Covariate	Variance Inflation Factor (VIF)
OwnOcc	4.07
SameHous	2.80
AssDens	1.38
VoTurn	2.04
CenResp	2.20
P_Labor	3.11
Unemp	2.04
NoHeIns	2.22
Gini	1.80
SinPar	2.21
HInc	5.15
Poverty	4.28
NoVehi	1.91
BachDe	4.00
BornUSA	4.42
P_AfriA	2.45
P_Hisp	3.04
P_Asian	2.64
P_Indig	1.62

5.1 PERMANENT INSURANCE PREMIUMS SOLD

We start by considering the permanent insurance premiums sold as a response variable. Among the predictor variables, the Gini index, percentage of population with at least bachelor's degree, and the percentage of Black/African American population, were found to be positively associated with permanent insurance premiums sold across the contiguous United States (Figure 5.1.2, Figure 5.1.3 and Figure 5.1.8). The Gini index, which is a regional variable, displayed varying amounts of spatial association as indicated by a slightly weaker coloring among the midwestern states. All other global variables were positively associated with permanent insurance premiums sold.

Significant variables that were positively associated with permanent insurance premiums sold in specific areas within the United States include:

- The percentage of population born in the U.S.: along the West Coast (Figure 5.1.4)
- The unemployment rate: most of New Mexico and western Texas (Figure 5.1.7)

- Association density: eastern New Mexico and western Texas, stretching northeast to southeastern Colorado and most of Kansas (Figure 5.1.9)

The percentage of households with yearly income above $75,000 had a rather noticeable spatially varying sign of association with the amount permanent insurance sold (Figure 5.1.5). In some places a greater percentage of higher-income households was associated with greater permanent insurance premiums sold, and in other places it was associated with less permanent insurance sold.

The no health insurance covariate was negatively associated with permanent insurance premiums sold; however, it was only statistically significant in an area around southeast New Mexico and east Texas (Figure 5.1.6).

When considering the map of the United States as a whole, we notice that the mean of response parameters was consistently the highest across space for the percentage of Black/African American population (Figure 5.1.8). This suggests that compared to all of the variables studied, a change in the proportion of a population that is Black/African American will likely have a greater impact on the amount of permanent insurance sold to that population than would a change of the same magnitude in any of the other variables. Also, the value of this parameter has low spatial variability as measured by their standard deviation (see Table 5.1.1).

Table 5.1.1
ESTIMATION MODEL

```
================================================================================
MGWR Version: 2.2.1
Released on: 03/20/2020
Source code is available at: https://github.com/pysal/mgwr
Development Team: Ziqi Li, Taylor Oshan, Stewart Fotheringham, Wei Kang,
Levi Wolf, Hanchen Yu, Mehak Sachdeva, and Sarah Bardin
Spatial Analysis Research Center (SPARC)
Arizona State University, Tempe, USA
================================================================================
Model type:                                                             Gaussian
Number of observations:                                                     3073
Number of covariates:                                                         20
Dependent variable:                                   Permanent.Ins.Premium.Sold
Variable standardization:                                                     On
Total runtime:                                                           0:43:40

Global Regression Results
--------------------------------------------------------------------------------
Residual sum of squares:                                                2591.183
Log-likelihood:                                                        -4098.364
AIC:                                                                    8236.728
AICc:                                                                   8239.031
R2:                                                                        0.157
Adj. R2:                                                                   0.152

Variable                                   Est.         SE  t(Est/SE)    p-value
------------------------------------ ---------- ---------- ---------- ----------
Intercept                                 0.000      0.017      0.000      1.000
P_AfriA                                   0.198      0.026      7.631      0.000
P_Indig                                  -0.068      0.021     -3.231      0.001
P_Asian                                  -0.007      0.027     -0.250      0.803
P_Hisp                                   -0.022      0.029     -0.767      0.443
BachDe                                    0.059      0.033      1.789      0.074
OwnerOcc                                  0.089      0.034      2.660      0.008
SameHous                                 -0.034      0.028     -1.238      0.216
P_Labor                                   0.106      0.029      3.603      0.000
Unemp                                     0.003      0.024      0.133      0.895
HInc                                      0.315      0.038      8.361      0.000
NoHeIns                                   0.114      0.025      4.592      0.000
Poverty                                   0.097      0.034      2.815      0.005
SinPar                                   -0.038      0.025     -1.540      0.124
BornUSA                                   0.019      0.035      0.550      0.582
NoVehi                                   -0.054      0.023     -2.328      0.020
CenResp                                   0.017      0.025      0.699      0.485
Gini                                      0.146      0.022      6.575      0.000
AssDens                                   0.041      0.020      2.098      0.036
VoTurn                                   -0.045      0.024     -1.902      0.057
```

Table 5.1.1
ESTIMATION MODEL (CONTINUED)

```
Multiscale Geographically Weighted Regression (MGWR) Results
---------------------------------------------------------------------------------
Coordinates type:                                                       Projected
Spatial kernel:                                                 Adaptive bisquare
Criterion for optimal bandwidth:                                             AICc
Score of change (SOC) type:                                           Smoothing f
Termination criterion for MGWR:                                           1.0e-05
Number of iterations used:                                                     42

MGWR bandwidths
---------------------------------------------------------------------------------
Variable                  Bandwidth      ENP_j   Adj t-val(95%)             DoD_j
Intercept                   429.000     12.920            2.891             0.681
P_AfriA                    2530.000      1.268            2.061             0.970
P_Indig                    2984.000      1.175            2.029             0.980
P_Asian                    3071.000      1.620            2.160             0.940
P_Hisp                     3071.000      1.120            2.009             0.986
BachDe                     3071.000      1.223            2.045             0.975
OwnerOcc                   2897.000      1.583            2.151             0.943
SameHous                   3071.000      1.292            2.068             0.968
P_Labor                    3071.000      1.190            2.034             0.978
Unemp                       742.000      8.148            2.742             0.739
HInc                         82.000     88.458            3.451             0.442
NoHeIns                      77.000     95.609            3.472             0.432
Poverty                    3071.000      1.112            2.006             0.987
SinPar                     3067.000      1.267            2.060             0.971
BornUSA                    3071.000      1.301            2.071             0.967
NoVehi                     3071.000      1.310            2.074             0.966
CenResp                    3071.000      1.265            2.060             0.971
Gini                       2423.000      2.332            2.301             0.895
AssDens                    1538.000      3.772            2.478             0.835
VoTurn                     2328.000      2.272            2.291             0.898

Diagnostic Information
---------------------------------------------------------------------------------
Residual sum of squares:                                                 1908.185
Effective number of parameters (trace(S)):                                230.235
Degree of freedom (n - trace(S)):                                        2842.765
Sigma estimate:                                                             0.819
Log-likelihood:                                                         -3628.253
Degree of Dependency (DoD):                                                 0.696
AIC:                                                                     7718.977
AICc:                                                                    7756.784
BIC:                                                                     9113.422
R2:                                                                         0.379
Adj. R2:                                                                    0.329
```

Table 5.1.1

ESTIMATION MODEL (CONTINUED)

Summary Statistics For MGWR Parameter Estimates

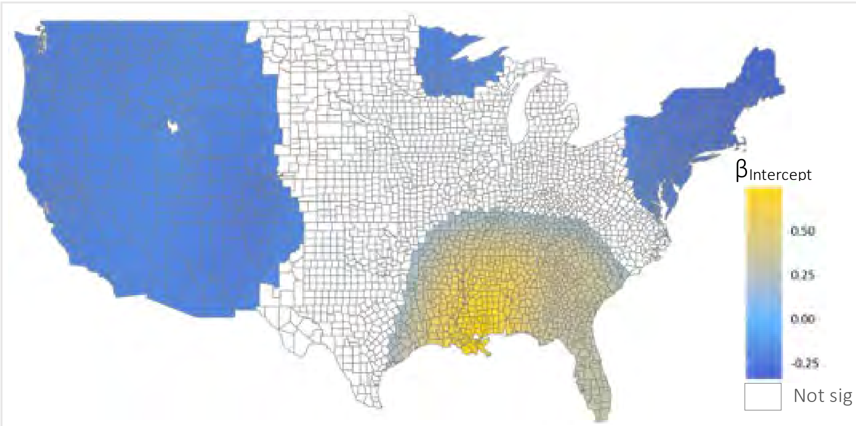
Variable	Mean	STD	Min	Median	Max
Intercept	0.035	0.261	-0.338	-0.059	0.743
P_AfriA	0.121	0.020	0.091	0.120	0.162
P_Indig	0.010	0.003	0.002	0.011	0.017
P_Asian	0.039	0.003	0.035	0.038	0.044
P_Hisp	0.025	0.004	0.015	0.026	0.033
BachDe	0.099	0.002	0.096	0.100	0.102
OwnerOcc	0.000	0.009	-0.021	0.006	0.008
SameHous	0.002	0.002	-0.001	0.001	0.007
P_Labor	0.040	0.001	0.039	0.040	0.043
Unemp	-0.001	0.060	-0.055	-0.020	0.222
HInc	0.411	0.292	-0.090	0.362	3.395
NoHeIns	0.082	0.236	-1.035	0.051	2.263
Poverty	0.061	0.001	0.057	0.061	0.063
SinPar	-0.008	0.003	-0.014	-0.009	-0.002
BornUSA	0.064	0.008	0.053	0.061	0.082
NoVehi	-0.006	0.003	-0.010	-0.005	-0.000
CenResp	-0.004	0.002	-0.005	-0.005	0.003
Gini	0.071	0.015	0.039	0.073	0.102
AssDens	0.032	0.023	-0.037	0.031	0.076
VoTurn	0.024	0.020	-0.010	0.022	0.064

Acknowledgement:
We acknowledge the support of the National Science Foundation under Award 1758786 from the Geography and Spatial Sciences Program to A. S. Fotheringham which enabled this software to be written and made freely available.

SIGNIFICANT MGWR LOCAL PARAMETER ESTIMATES FOR PERMANENT INSURANCE PREMIUM SOLD

Figure 5.1.1

PERM INS PREM SOLD VS. INTERCEPT

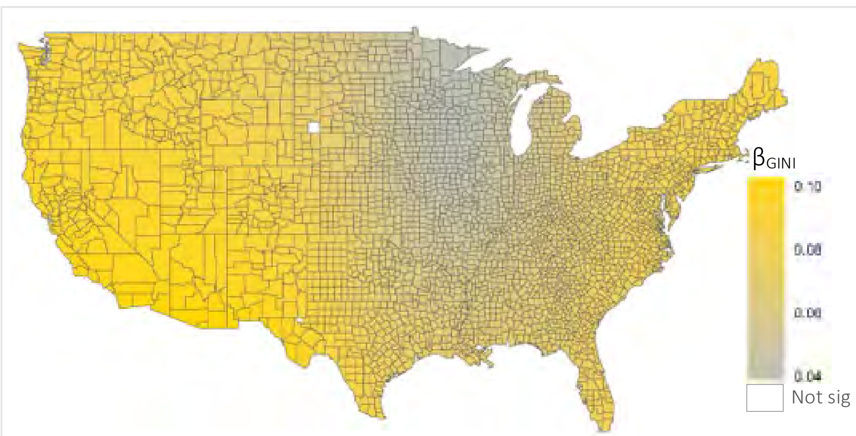


Holding the covariates of the MGWR model constant, there were intrinsically more permanent insurance premiums insurance sold in the counties of the South than in the West and Northeast.

Beta coefficient characteristics: Mean = 0.106, Std = 0.343

Figure 5.1.2

PERM INS PREM SOLD VS. GINI INDEX

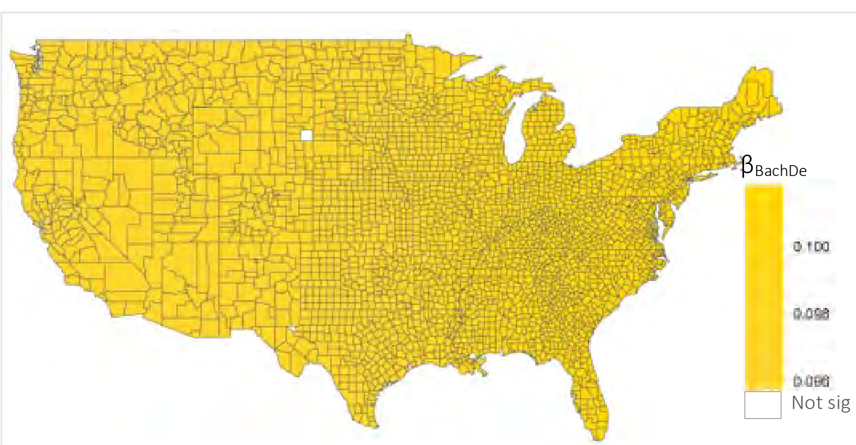


The impact of Gini index on the amount of permanent insurance sold was rather homogenous across most of the country and only slightly weaker in the Midwest.

Beta coefficient characteristics: Mean = 0.12, Std = 0.022

Figure 5.1.3

PERM INS PREM SOLD VS. % WITH BACHELOR'S DEGREE OR HIGHER



The percent of population (age 25 and greater) that have a bachelor's degree or higher was globally statistically significant and positively associated. Throughout the country, higher proportions of the population with a bachelor's degree or higher were associated with greater amounts of permanent insurance premiums sold.

Beta coefficient characteristics: Mean = 0.099, Std = 0.002

SIGNIFICANT MGWR LOCAL PARAMETER ESTIMATES FOR PERMANENT INSURANCE

Figure 5.1.4

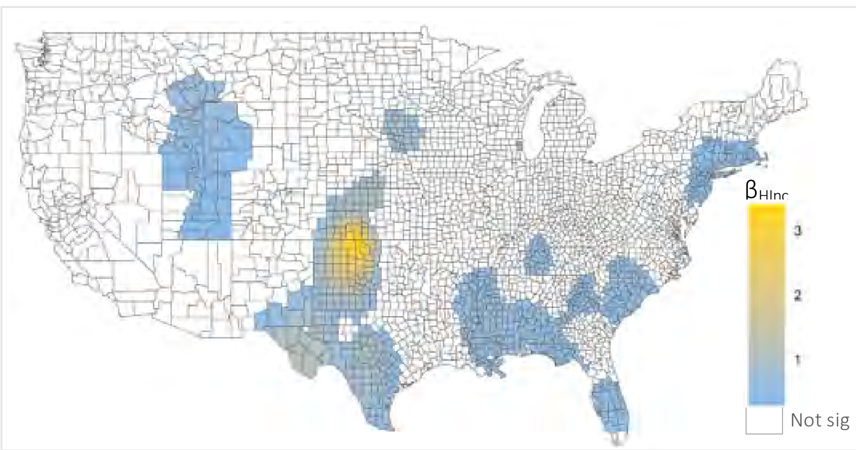
PERM INS PREM SOLD VS. % BORN IN THE USA

Beta coefficient characteristics: Mean = 0.064, Std = 0.008

The percentage of population born in the U.S. was only statistically significant for permanent insurance demand along the West Coast, where it was positively associated with permanent insurance premiums sold.

Figure 5.1.5

PERM INS PREM SOLD VS. % OF HOUSEHOLDS WITH INCOME ABOVE $75,000

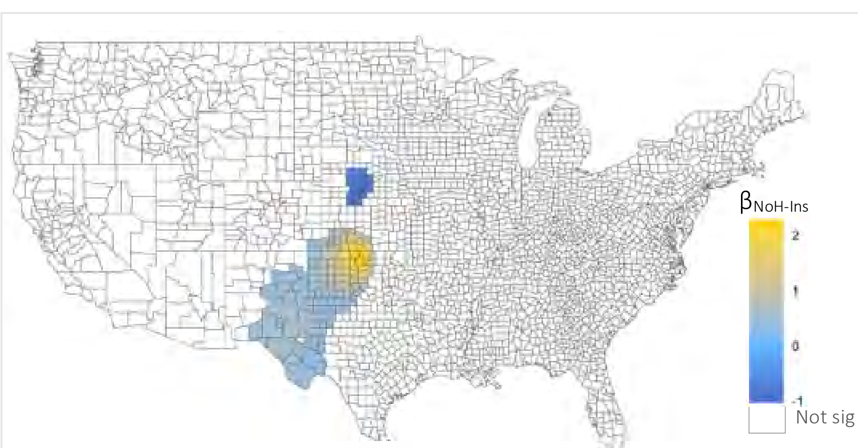


Beta coefficient characteristics: Mean = 0.411, Std = 0.292

The percentage of households with yearly income above $75,000 was not significant across most of the country. In areas where it was significant, the association was typically negative. However, the association was positive in a fairly small area of southwestern Kansas, western Oklahoma and northwestern Texas.

Figure 5.1.6

PERM INS PREM SOLD VS. % OF POPULATION WITH NO HEALTH INSURANCE



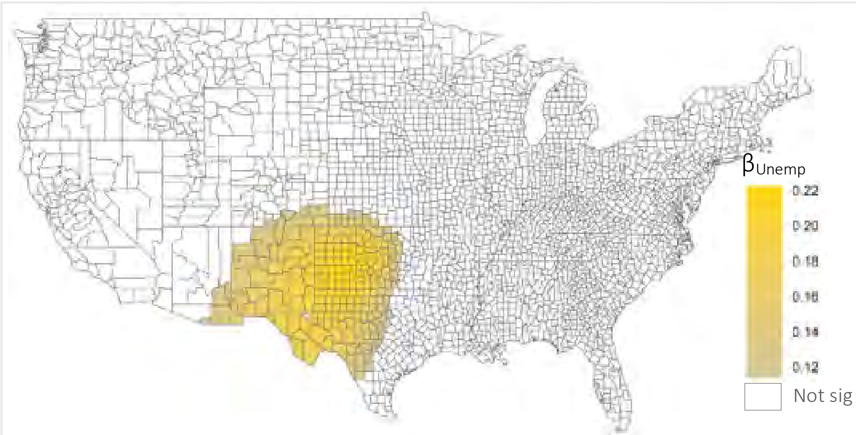
Beta coefficient characteristics: Mean = 0.082, Std = 0.236

For most of the country, the proportion of households without health insurance was not statistically significant to permanent insurance premiums sold. However, it was negatively associated with permanent insurance premiums in southeast New Mexico and western Texas, while it is positively associated in western Oklahoma and southwestern Kansas.

SIGNIFICANT MGWR LOCAL PARAMETER ESTIMATES FOR PERMANENT INSURANCE

Figure 5.1.7

PERM INS PREM SOLD VS. UNEMPLOYMENT RATE

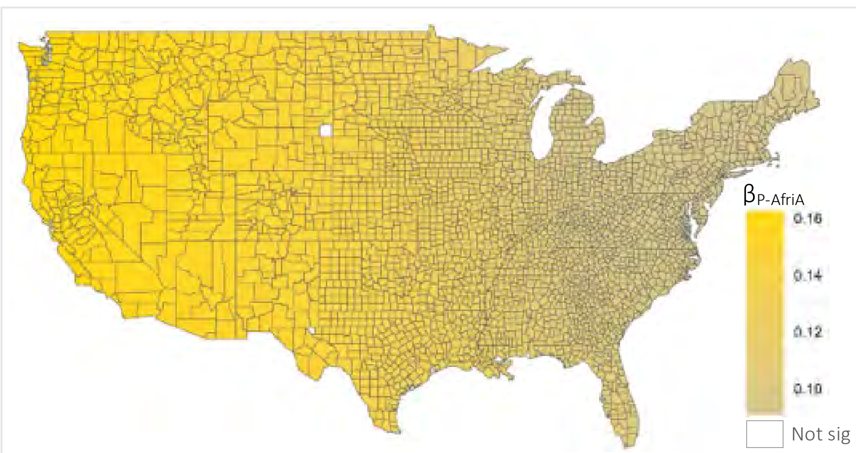


The unemployment rate was statistically significant only in New Mexico and western Texas, where permanent insurance was more likely to be sold in areas with greater unemployment rates (positive association).

Beta coefficient characteristics: Mean = −0.001, Std = 0.060

Figure 5.1.8

PERM INS PREM SOLD VS. % OF BLACK/AFRICAN AMERICAN POPULATION



Across the U.S., greater proportions of Black/African Americans were associated with greater sales of permanent life insurance premiums.

Beta coefficient characteristics: Mean = 0.121, Std = 0.020

Figure 5.1.9

PERM INS PREM SOLD VS. ASSOCIATION DENSITY

Association density showed a locally significant, positive association with permanent insurance premiums sold in an area stretching from southeastern New Mexico and western Texas, northeast to southeastern Colorado and most of Kansas.

Beta coefficient characteristics: Mean = 0.032, Std = 0.023

5.2 TERM INSURANCE PREMIUMS SOLD

Now we, consider term insurance premiums sold as the response variable. For term insurance, we see more statistically significant covariates than we saw for permanent insurance, specifically:

- The percentage of population with a bachelor's or more advanced degree is positively associated with the response variable and of global significance.
- The percentage of population with no vehicle was negatively associated with only regional significance.
- The unemployment rate was negatively associated and of global significance.
- Household income shows varying signs of association and almost global statistical significance. The household income covariate is one of the most significant covariates in this part of the analysis, which is intuitive because the face value of insurance that is recommended for people to purchase is often a multiple of their income.

In contrast to permanent insurance, for term insurance premiums sold:

- Living at the same place was statistically significant , at least regionally, with negative association.
- Association density was positively associated and locally statistically significant.
- The rate of census completion was positively associated and of regional statistical significance.

As was the case with permanent insurance, the Gini index, percentage of the voting-age population that voted in the 2016 election and percentage of single parent households were all globally significant covariates positively associated with term insurance premiums sold. However, while the percentage of Black/African American population remains positively associated with term insurance premiums sold, it was only regionally significant for term insurance. Finally, the percentage of Asian/Asian American population shows only local statistical significance for term insurance sold and negative association.

When it comes to regression parameters, we notice that the mean of response parameters across space is by far highest for household income and percentage of the population with a bachelor's degree or higher. This suggests that change of these covariates, on average, will have the highest impact on the amount of term insurance premiums sold.

Table 5.2.1

ESTIMATION MODEL

```
===========================================================================
MGWR Version: 2.2.1
Released on: 03/20/2020
Source code is available at: https://github.com/pysal/mgwr
Development Team: Ziqi Li, Taylor Oshan, Stewart Fotheringham, Wei Kang,
Levi Wolf, Hanchen Yu, Mehak Sachdeva, and Sarah Bardin
Spatial Analysis Research Center (SPARC)
Arizona State University, Tempe, USA
===========================================================================
Model type:                                                        Gaussian
Number of observations:                                                3073
Number of covariates:                                                    20
Dependent variable:                                   Term.Ins.Premium.Sold
Variable standardization:                                                On
Total runtime:                                                      1:02:36

Global Regression Results
---------------------------------------------------------------------------
Residual sum of squares:                                           1706.993
Log-likelihood:                                                   -3457.058
AIC:                                                               6954.115
AICc:                                                              6956.418
R2:                                                                   0.445
Adj. R2:                                                              0.441

Variable                                   Est.         SE  t(Est/SE)    p-value
------------------------------------ ---------- ---------- ---------- ----------
Intercept                                 0.000      0.013      0.000      1.000
P_AfriA                                   0.396      0.021     18.774      0.000
P_Indig                                  -0.007      0.017     -0.393      0.694
P_Asian                                   0.040      0.022      1.827      0.068
P_Hisp                                   -0.087      0.024     -3.713      0.000
BachDe                                    0.042      0.027      1.553      0.120
OwnerOcc                                  0.172      0.027      6.336      0.000
SameHous                                 -0.143      0.023     -6.344      0.000
P_Labor                                   0.118      0.024      4.954      0.000
Unemp                                    -0.051      0.019     -2.667      0.008
HInc                                      0.417      0.031     13.619      0.000
NoHeIns                                   0.156      0.020      7.757      0.000
Poverty                                   0.106      0.028      3.809      0.000
SinPar                                   -0.033      0.020     -1.622      0.105
BornUSA                                  -0.011      0.028     -0.395      0.693
NoVehi                                   -0.174      0.019     -9.315      0.000
CenResp                                   0.057      0.020      2.860      0.004
Gini                                      0.253      0.018     13.981      0.000
AssDens                                   0.032      0.016      1.995      0.046
VoTurn                                    0.007      0.019      0.363      0.716
```

Table 5.2.1

ESTIMATION MODEL (CONTINUED)

```
Multiscale Geographically Weighted Regression (MGWR) Results
---------------------------------------------------------------------------------
Coordinates type:                                                        Projected
Spatial kernel:                                                  Adaptive bisquare
Criterion for optimal bandwidth:                                              AICc
Score of change (SOC) type:                                            Smoothing f
Termination criterion for MGWR:                                            1.0e-05
Number of iterations used:                                                      50

MGWR bandwidths
---------------------------------------------------------------------------------
Variable                  Bandwidth      ENP_j   Adj t-val(95%)              DoD_j
Intercept                    44.000    194.413            3.659              0.344
P_AfriA                    3071.000      1.026            1.972              0.997
P_Indig                    3071.000      1.164            2.025              0.981
P_Asian                    3067.000      1.628            2.162              0.939
P_Hisp                     3071.000      1.106            2.003              0.987
BachDe                     3025.000      1.263            2.059              0.971
OwnerOcc                   2209.000      2.682            2.354              0.877
SameHous                   3071.000      1.224            2.046              0.975
P_Labor                    3071.000      1.164            2.025              0.981
Unemp                      3071.000      1.205            2.039              0.977
HInc                        144.000     46.287            3.272              0.522
NoHeIns                    3071.000      1.162            2.024              0.981
Poverty                    3069.000      1.122            2.010              0.986
SinPar                     3033.000      1.292            2.068              0.968
BornUSA                    3071.000      1.279            2.064              0.969
NoVehi                     1581.000      3.891            2.489              0.831
CenResp                    1572.000      3.469            2.448              0.845
Gini                       2626.000      1.942            2.231              0.917
AssDens                     348.000     17.513            2.985              0.643
VoTurn                     3071.000      1.205            2.040              0.977

Diagnostic Information
---------------------------------------------------------------------------------
Residual sum of squares:                                                   628.710
Effective number of parameters (trace(S)):                                 286.037
Degree of freedom (n - trace(S)):                                         2786.963
Sigma estimate:                                                              0.475
Log-likelihood:                                                          -1922.372
Degree of Dependency (DoD):                                                  0.669
AIC:                                                                      4418.817
AICc:                                                                     4478.190
BIC:                                                                      6149.765
R2:                                                                          0.795
Adj. R2:                                                                     0.774
```

Table 5.2.1

ESTIMATION MODEL (CONTINUED)

Summary Statistics For MGWR Parameter Estimates

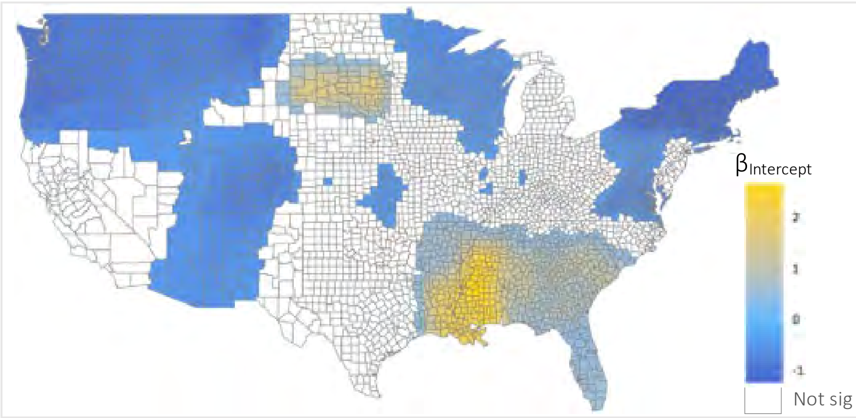
Variable	Mean	STD	Min	Median	Max
Intercept	0.023	0.736	-1.236	-0.176	2.642
P_AfriA	0.046	0.001	0.045	0.046	0.047
P_Indig	0.011	0.001	0.009	0.011	0.013
P_Asian	-0.033	0.004	-0.045	-0.032	-0.024
P_Hisp	-0.022	0.001	-0.024	-0.022	-0.019
BachDe	0.163	0.005	0.148	0.166	0.167
OwnerOcc	0.026	0.012	0.009	0.024	0.048
SameHous	-0.045	0.002	-0.047	-0.046	-0.038
P_Labor	0.027	0.001	0.023	0.027	0.030
Unemp	-0.035	0.001	-0.038	-0.035	-0.031
HInc	0.477	0.155	0.101	0.459	0.888
NoHeIns	0.064	0.000	0.063	0.064	0.064
Poverty	-0.011	0.003	-0.017	-0.010	-0.006
SinPar	0.039	0.004	0.034	0.038	0.048
BornUSA	-0.039	0.001	-0.043	-0.039	-0.036
NoVehi	-0.072	0.035	-0.150	-0.067	-0.024
CenResp	0.050	0.025	0.007	0.047	0.107
Gini	0.083	0.006	0.073	0.082	0.095
AssDens	0.026	0.043	-0.053	0.022	0.129
VoTurn	0.055	0.001	0.051	0.055	0.056

Acknowledgement:
We acknowledge the support of the National Science Foundation under Award 1758786 from the Geography and Spatial Sciences Program to A. S. Fotheringham which enabled this software to be written and made freely available.

SIGNIFICANT MGWR LOCAL PARAMETER ESTIMATES FOR TERM INSURANCE PREMIUMS SOLD

Figure 5.2.1

TERM INS PREM SOLD VS. INTERCEPT

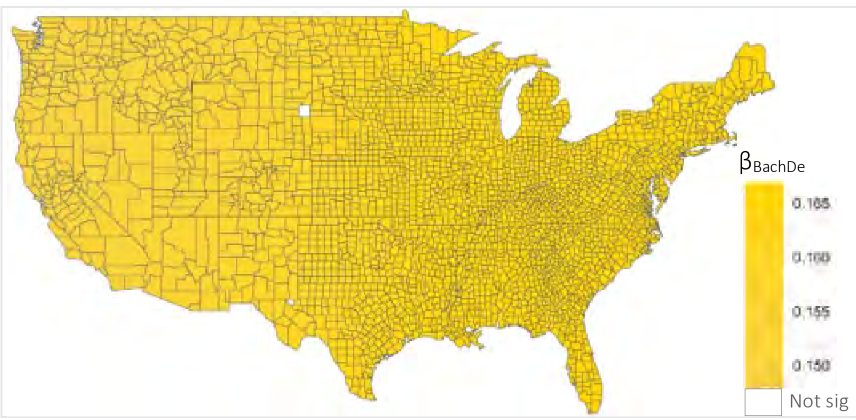


Beta coefficient characteristics: Mean = 0.152, Std = 1.013

Holding the covariates of the MGWR model constant, there were intrinsically more term insurance premiums sold in the counties of the South and areas of the Dakotas than in the Northwest and Northeast.

Figure 5.2.2

TERM INS PREM SOLD VS. % WITH BACHELOR'S DEGREE OR HIGHER

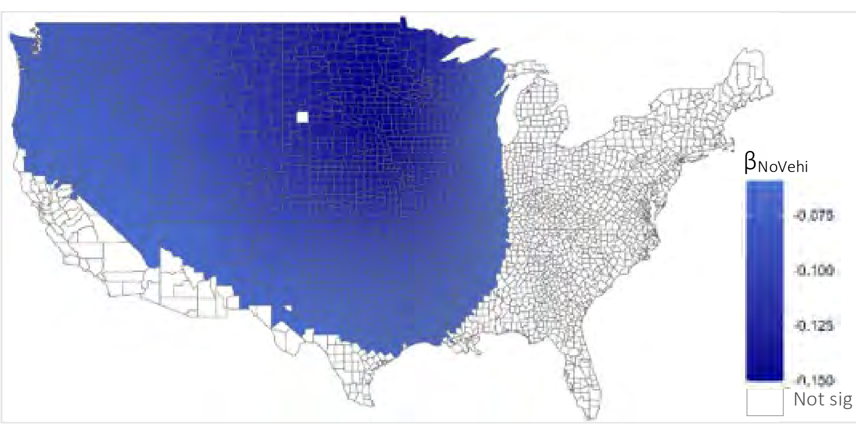


Beta coefficient characteristics: Mean = 0.163, Std = 0.005

The percentage of individuals having a bachelor's degree or higher had a globally positive association with the amount of term insurance sold.

Figure 5.2.3

TERM INS PREM SOLD VS. % WITH NO VEHICLE



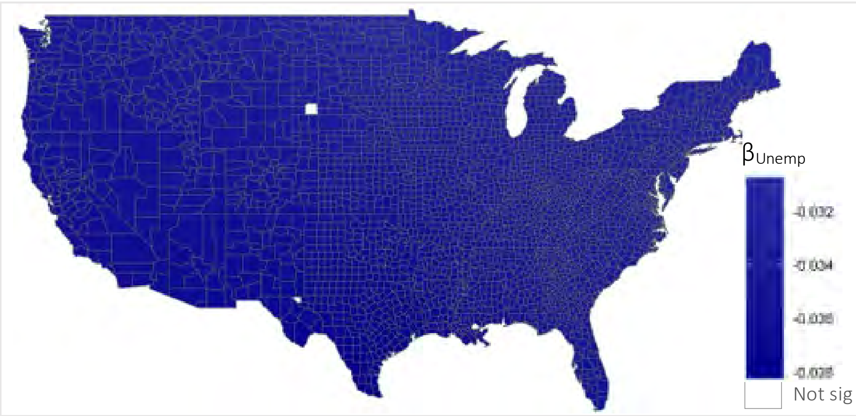
Beta coefficient characteristics: Mean = −0.072, Std = 0.035

The percentage of households with no vehicle was a locally significant covariate. Its relationship with term insurance premiums sold was negative across most of the counties in the western U.S. except California.

SIGNIFICANT MGWR LOCAL PARAMETER ESTIMATES FOR TERM INSURANCE PREMIUMS SOLD

Figure 5.2.4

TERM INS PREM SOLD VS. UNEMPLOYMENT RATE

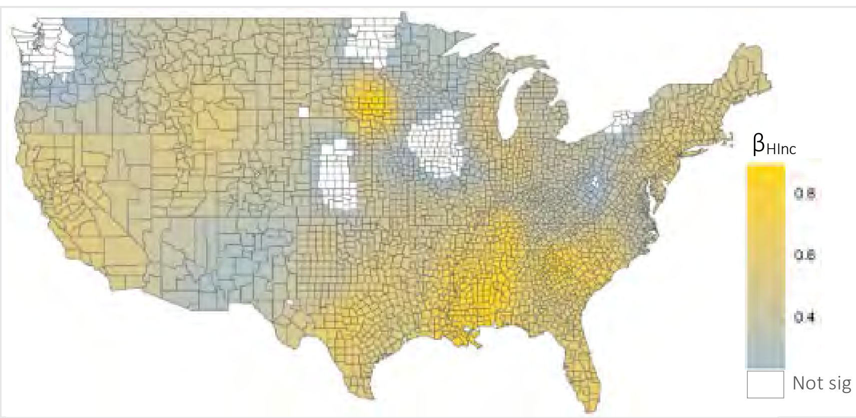


Beta coefficient characteristics: Mean = −0.035, Std = 0.001

Unemployment rate was a global covariate of term insurance demand, and it had a homogenously negative effect on term insurance sold across the country.

Figure 5.2.5

TERM INS PREM SOLD VS. % OF HOUSEHOLDS WITH INCOME ABOVE $75,000

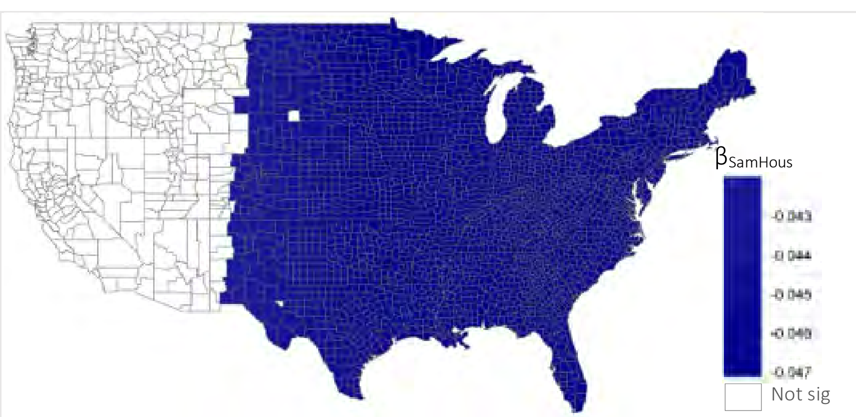


Beta coefficient characteristics: Mean = 0.477, Std = 0.155

The percentage of households having income above $75,000 had an overall positive association with term insurance sold, while the extent of impact presented some spatial variation across the country.

Figure 5.2.6

TERM INS PREM SOLD VS. % LIVING IN THE SAME PLACE



Beta coefficient characteristics: Mean = −0.045, Std = 0.002

The percentage of households living in the same place had a significantly negative effect on term insurance sold in the central and eastern parts of the U.S., but the association was not significant in the western U.S.

SIGNIFICANT MGWR LOCAL PARAMETER ESTIMATES FOR TERM INSURANCE PREMIUMS SOLD

Figure 5.2.7

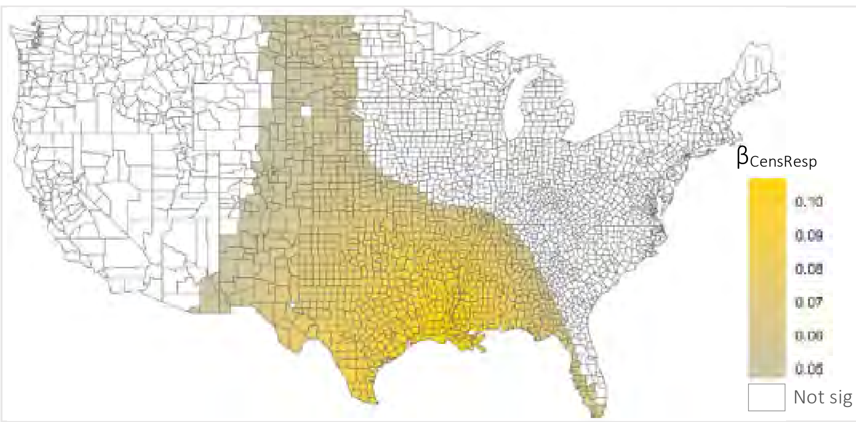
TERM INS PREM SOLD VS. ASSOCIATION DENSITY

Association density was locally significant to term insurance sold in only some counties of Montana and North Dakota.

Beta coefficient characteristics: Mean = 0.026, Std = 0.043

Figure 5.2.8

TERM INS PREM SOLD VS. CENSUS RESPONSE RATE

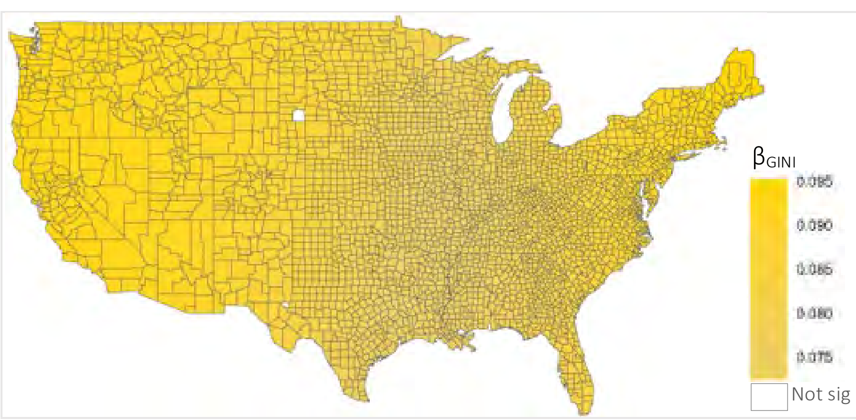


The census response rate was a locally significant covariate of term insurance sold along the southern coast and the middle part of U.S. The variable was insignificant across other parts of the country.

Beta coefficient characteristics: Mean = –0.089, Std = 0.024

Figure 5.2.9

TERM INS PREM SOLD VS. GINI INDEX



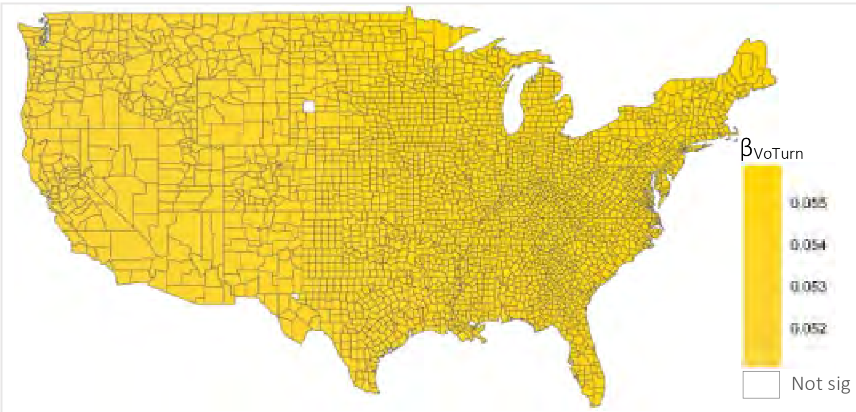
The Gini index had a globally significant, positive impact on term insurance sold.

Beta coefficient characteristics: Mean = 0.083, Std = 0.006

SIGNIFICANT MGWR LOCAL PARAMETER ESTIMATES FOR TERM INSURANCE PREMIUMS SOLD

Figure 5.2.10

TERM INS PREM SOLD VS. VOTER TURNOUT

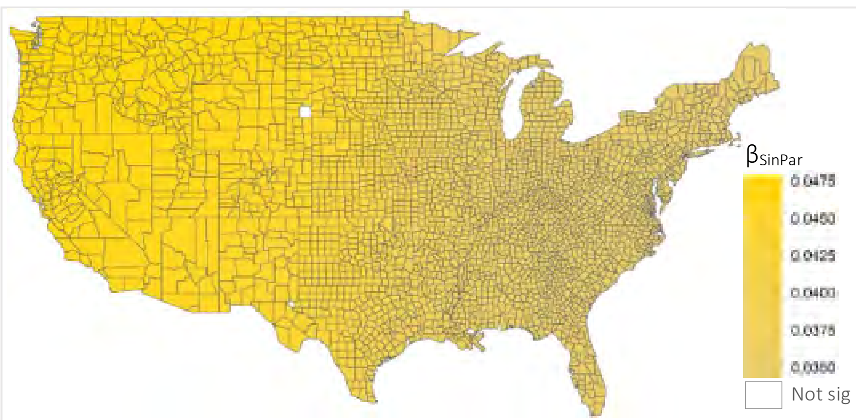


The percentage of the voting-age population that voted in the 2016 election was a globally significant covariate of term insurance sold, and the relationship was positive.

Beta coefficient characteristics: Mean = 0.055, Std = 0.001

Figure 5.2.11

TERM INS PREM SOLD VS. % OF SINGLE PARENT HOUSEHOLDS

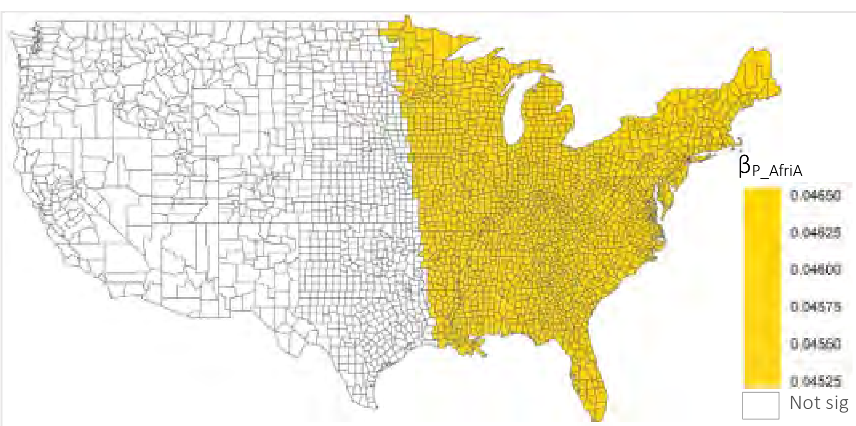


The impact of the percentage of single parent households was rather similar across the U.S. The association was slightly weaker in the eastern U.S. than the western part of the country.

Beta coefficient characteristics: Mean = 0.039, Std = 0.004

Figure 5.2.12

TERM INS PREM SOLD VS. % BLACK/AFRICAN AMERICAN



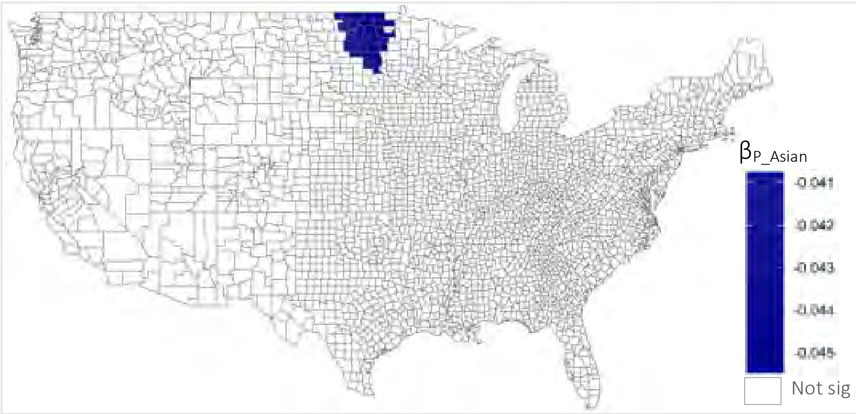
The percentage of the population that is black/African American had a positive relationship with term insurance sold in the eastern U.S., while the relationship is not significant in the middle and western parts of the country.

Beta coefficient characteristics: Mean = 0.046, Std = 0.001

SIGNIFICANT MGWR LOCAL PARAMETER ESTIMATES FOR TERM INSURANCE PREMIUMS SOLD

Figure 5.2.13

TERM INS PREM SOLD VS. % ASIAN/ASIAN AMERICAN



Beta coefficient characteristics: Mean = −0.033, Std = 0.004

The percentage of the population that is Asian/Asian American was negatively associated with term insurance sold in some counties in North Dakota and Minnesota, but no significant association was detected in other parts of the country.

Section 6: Conclusion

In this work, in a spatial regression context of the spatial MGWR model, we investigated the drivers of insurance demand across counties of the contiguous U.S. Our proxies for insurance demand were annual permanent insurance premiums sold and annual term insurance premiums sold for 2020. Because the COVID-19 pandemic emerged in 2020, results of this study may or may not represent a typical year. Identifying the impact of COVID-19 on results is beyond the scope of this study.

Overall, our findings show that various spatial determinants associated with social capital and population composition were statistically significant on different spatial scales or not significant at all. Further, we observe that permanent insurance premiums and term insurance premiums exhibited largely different drivers and spatial patterns.

When it comes to permanent insurance, we find that the most impactful parameter across the United States as a whole was the percentage of Black/African American population, which displays a positively association and a low standard deviation across space. The last point suggests that there was homogeneity in response over space, and an increase in the percentage was associated with higher demand for permanent insurance.

When it comes to term insurance, we find that the most impactful parameters were the percentage of households with at least $75,000 yearly income and the percentage of population with a bachelor's degree or higher. The percentage of Black/African American population was only regionally statistically significant. Specifically, the percentage of the population that was black/African American had a positive relationship with term insurance sold in the eastern U.S., while the relationship was not significant in the middle and western parts of the country.

It would be interesting to conduct a further investigation via surveys or other tools to understand why demand for term insurance was positively associated with specific determinants only regionally. The same survey questionnaire, which may include hypothesized causal reasons for the demand for insurance products, or lack thereof, could be given to representative samples of the population of interest, both in regions where there was an observed positive association and where this positive association was not present. Any statistical differences in these responses might reveal further causal mechanisms, which could be confirmed or rejected with the appropriate statistical testing.

Table 6.1 summarizes the mean relationships between the covariates considered and the insurance demand proxies, as well as the associated scales of impacts. The absolute marginal impacts of the covariates on the response variable are ranked, where a lower rank indicates a stronger average marginal impact across the space.

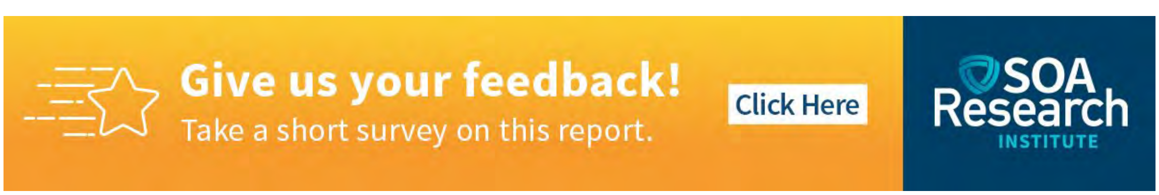
There could be a plethora of potential applications of our findings. Our results can support decision-making in insurance companies, by assessing the impacts that changing social and economic factors have on insurance demand. For example, marketing strategies could be tailored from these results, with further research conducted to inform why specific determinants only act on specific scales with potentially varying intensities. For public policymakers, the estimated marginal effects of covariates can help identify geographical locations for potential future focus groups where more research is needed to create effective strategies for addressing potential inequalities in life insurance.

Further research could include a spatiotemporal analysis of drivers of insurance demand. This would necessitate a dataset containing a demand for premiums across counties and for multiple years (decades). The findings of such research would the stability of determinants over time. This would help further understanding of the impact of determinants under consideration in contexts of the insurance market, diversity, and social science.

Table 6.1

SUMMARY OF THE IMPACT SCALES AND MEAN RELATIONSHIPS BETWEEN THE COVARIATES AND INSURANCE DEMAND

Covariate	Description	Permanent Insurance Sold (Spatial scale, direction, rank)	Term Insurance Sold (Spatial scale, direction, rank)
OwnOcc	Percentage of housing that is owner occupied	Regional, positive, 19	Regional, positive, 15
SameHous	Percentage of the population living in the same place since 2009	Global, positive, 17	Global, negative, 9
AssDens	Association density (i.e., the number of social institutions present within a county in proportion to its population)	Regional, positive, 10	Local, positive, 15
VoTurn	Percentage of the voting-age population that voted in the 2016 election	Regional, positive, 12	Global, positive, 6
CenResp	Response rate for the 2020 census	Local, negative, 16	Regional, positive, 7
P_Labor	Percentage of the population in the labor force	Global, positive, 8	Global, positive, 14
Unemp	Unemployment Rate	Local, negative, 18	Global, negative, 12
NoHeIns	Percentage of the population without health insurance	Local, positive, 4	Global, positive, 5
Gini	Gini index (i.e., a statistical measure of wealth inequality)	Regional, positive, 5	Regional, positive, 3
SinPar	Percentage of single parent households	Local, negative, 14	Local, positive, 10
HInc	Percentage of households with yearly income above $75,000	Local, positive, 1	Local, positive, 1
Poverty	Percentage of households in poverty	Global, positive, 7	Local, negative, 18
NoVehi	Percentage of households with no vehicles	Global, negative, 15	Regional, negative, 4
BachDe	Percentage of the population with a bachelor's degree or higher (25 years and older)	Global, positive, 3	Local, positive, 2
BornUSA	Percentage of the population born in the U.S.	Global, positive, 6	Global, negative, 10
P_AfriA	Percentage of the population that is Black/African American	Regional, positive, 2	Global, positive, 8
P_Hisp	Percentage of the population that is Hispanic/Latino	Global, positive, 11	Global, negative, 17
P_Asian	Percentage of the population that is Asian/Asian American	Global, positive, 9	Local, negative, 13
P_Indig	Percentage of the population that is Indigenous	Local, positive, 13	Global, positive, 18



Section 7: Acknowledgments

The researchers' deepest gratitude goes to those without whose efforts this project could not have come to fruition: the Project Oversight Group and others for their diligent work overseeing questionnaire development, analyzing and discussing respondent answers, and reviewing and editing this report for accuracy and relevance.

Project Oversight Group members:

Patricia Fay, FSA, MAAA

Karen T. Jiang, FSA, CERA, MAAA

Simin Ma, FSA, CERA, MAAA

Belinda Nguyen, FCAS

Matthew S. Wolf, FSA, CERA, MAAA

At the Society of Actuaries Research Institute:

Lisa Schilling, FSA, EA, FCA, MAAA, Senior Research Actuary

Special thanks to LIMRA for providing the data that made this study possible.

Bibliography

Cavalcante, R. T., Sobreiro, V. A., & Kimura, H. (2018). Determinants of the non-life insurance market in Brazil. *Review of Development Finance*, *8*(2), 89–95. https://doi.org/10.1016/J.RDF.2018.11.004

Centre for Educational Research and Innovation. (2001). *The Well-being of Nations THE ROLE OF HUMAN AND SOCIAL CAPITAL EDUCATION AND SKILLS The Well-being of Nations EDUCATION AND SKILLS*. http://www.oecd.org/site/worldforum/33703702.pdf

Fotheringham, A. S., Brunsdon, C., & Charlton, M. (2003). *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. John Wiley & Sons.

Glaeser, E. L., Laibson, D., & Sacerdote, B. (2002). An Economic Approach to Social Capital*. *The Economic Journal*, *112*(483), F437–F458. https://doi.org/10.1111/1468-0297.00078

Jean Kwon, W. (2013). The significance of regulatory orientation, political stability and culture on consumption and price adequacy in insurance markets. *Journal of Risk Finance*, *14*(4), 320–343. https://doi.org/10.1108/JRF-03-2013-0018/FULL/XML

Kamiya, S., Li, J. K. K., & Zanjani, G. H. (2014). The Effect of Banking Crises on Non-Life Insurance Consumption. *SSRN Electronic Journal*. https://doi.org/10.2139/SSRN.2512808

Kutner, M. H., Nachtsheim, C. J., Neter, J., & Li, W. (2005). *Applied Linear Statistical Models*. McGraw-Hill, page 409.

Letiţia Andronic, M. (2019). A STUDY ON THE INSURANCE DENSITY DETERMINANTS IN ROMANIA. *Bulletin of the Transilvania University of Braşov Series V: Economic Sciences*, *12*(61). https://doi.org/10.31926/but.es.2019.12.61.2.9

Mendenhall, W., Sincich, T., & Boudreau, N. S. (2003). *A Second Course in Statistics: Regression Analysis* (Vol. 6). Prentice Hall, page 374.

Outreville, J. F. (2013). The Relationship Between Insurance and Economic Development: 85 Empirical Papers for a Review of the Literature. *Risk Management and Insurance Review*, *16*(1), 71–122. https://doi.org/10.1111/J.1540-6296.2012.01219.X

Park, S. C., & Lemaire, J. (2012). The Impact of Culture on the Demand for Non-Life Insurance. *ASTIN Bulletin: The Journal of the IAA*, *42*(2), 501–527. https://doi.org/10.2143/AST.42.2.2182806

Podoabă, L. (2015). An Empirical Research Regarding Density and Penetration Insurance on Accident & Health Premiums in Future Implementation of Solvency II. *Procedia Economics and Finance*, *32*, 1517–1524. https://doi.org/10.1016/S2212-5671(15)01539-7

Rupasingha, A., Goetz, S. J., & Freshwater, D. (2006). The production of social capital in US counties. *The Journal of Socio-Economics*, *35*(1), 83–101. https://doi.org/10.1016/J.SOCEC.2005.11.001

Sampath Sanjeewa, W., Ouyang, H., & Hashmi, S. H. (2019). DETERMINANTS OF LIFE INSURANCE CONSUMPTION IN EMERGING INSURANCE MARKETS OF SOUTH-ASIA. *International Journal of Information, Business and Management*, *11*(4).

Trinh, C. T., Nguyen, X., Sgro, P., & Pham, C. S. (2020). Culture, financial crisis and the demand for property, accident and health insurance in the OECD countries. *Economic Modelling*, *93*, 480–498. https://doi.org/10.1016/J.ECONMOD.2020.09.002

Trinh, T., Nguyen, X., & Sgro, P. (2016). Determinants of non-life insurance expenditure in developed and developing countries: an empirical investigation. *Http://Dx.Doi.Org/10.1080/00036846.2016.1181834*, *48*(58), 5639–5653. https://doi.org/10.1080/00036846.2016.1181834

About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its strategic research programs: aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of topical research available, including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org