# Spatial Dependence and Neighborhood Effects in Mortgage Lending: A Geographically Weighted Regression Approach

Duan Zhuang

LUSK Center for Real Estate
University of Southern California
Los Angeles, CA, 90089
(213)820-8853
dzhuang@usc.edu

July 2006

This research is supported by the LUSK Research Grant. The contents of the paper are the view of the author and do not necessarily reflect the views of the LUSK Center. The author thanks Dr. Stuart Gabriel for valuable advice and comments.

**ABSTRACT**

Current research on mortgage lending disparity is mostly based upon the process-based approach on discrimination, and the outcome-based approach on lending disparity and redlining. In recent years, the outcome-based model has received much attention, particularly on the relationship between intra-metropolitan geography and mortgage lending outcomes. From a policy perspective, the theoretical and empirical evidence on lending disparity is of great importance. However, there exists a mismatch between theoretical models, which focus on racial preferences, and empirical studies, which are essentially reduced form without adequate information. Besides, it may be difficult to unravel the effects of neighborhood race and other attributes. So far most studies conducted with the Home Mortgage Disclosure Act (HMDA) data ignore determinants of geographic variations in lending outcome, or simply attribute them to local variations in risk.

This study intends to investigate spatial dependence and neighborhood effects of mortgage lending disparities in the Southern California Five-county Region. In so doing, it assesses indicators of primary mortgage market activity and their determinants for the region as a whole and for the sub-regions inside it. The study compiles data from the 2002 HMDA and the 2000 U.S. Census to undertake a variety of analyses, including computation, assessment, and mapping of social-economic characteristics, as well as home mortgage origination, denial rates, and secondary market purchase rates by census tracts among sampled areas and population cohorts. Cluster analysis on those social-economic and mortgage parameters show distinctive patterns of spatial clustering among tracts across the region. In observing these blueprints of spatial dependence, the study further undertakes a geographically weighted regression (GWR) to analyze the spatial non-stationarity of the determinants of variability in primary market loan denial rates across locations for the year 2002.

The modeling result reveals that significant spatial non-stationarity exists between mortgage denial rates and the social-economic determinants. Specifically, the study finds that those census tract-level attributes, including income, population, age, racial composition, housing stock, etc., show significant and varying impacts on mortgage denial rate pattern by spatial clusters. In particular, higher values of spatially varying coefficients on racial composition on traditionally underserved areas, such as south-central Los Angeles, and central cities of outer counties, shed lights on the concerns of redlining. The study concludes that mortgage lending pattern is better understood by the geographically-weighted model than traditional Ordinary Least Square (OLS) regression approaches on lending outcome, which ignore the spatial correlation among local determinants.

**Key words**: spatial dependence, geographically weighted regression (GWR), HMDA, redlining

## Introduction

Homeownership is a longstanding subject for both academic and policy concerns because it is widely believed to encourage good communities and citizenship. For decades, the Federal Government has been making tremendous efforts to promote fair housing so as to increase homeownership rate among low-income and minority groups.

The national homeownership rate has been going up in recent years. However, According to HUD's 2002 Annual Performance Plan, while the overall homeownership rate in America reached 67.7 percent in 2000, the rate among households with income less than median family income was only 52.2 percent during the same period. In 2000, only 47.3 percent of African-American households and 46.7 percent of Hispanic households were homeowners, compared with 74.3 percent of non-Hispanic white households.

The analysis in this paper is motivated by trends in homeownership. Huge disparities between white and minority homeownership rates continue to be evident. Most low-income and minority groups are concentrated in underserved areas[1]. Census data show that, geographical concentration of poverty and isolation of low-income households worsened in recent decades. For instance, the average homeownership rate in central cities was only 52.9 percent in 2000. This has drawn considerable policy attentions because their relative low homeownership rates have dragged down the national homeownership rate and hampered the National Homeownership Strategy initiated by the Federal Government.

Our initial computations seek to assess the extent to which those homeownership patterns are mirrored in the mortgage lending data. Table 1 presents summary measures from the HMDA data that provide evidence of primary and secondary mortgage market activity for 2002. It is comprised of conventional conforming purchase (CCP) loan activity by census tracts among sampled areas and population cohorts in the Los Angeles region. The low level of credit found in underserved areas, measured by denial rate of mortgage applications, is an important factor to explain the low level of homeownership rates in those areas.

[1] Underserved areas are metro/non-metro areas (as defined by the Office of Management and Budget) with census tracts having a median income at or below 120 percent of the median income of the metropolitan/state non-metro area and a minority population of 30 percent or greater; or a median income at or below 90/95 percent of median income of the metropolitan/state non-metro area.

**Table 1: 2002 Los Angeles Region Conventional Purchase Conforming Loan Characteristics by Census Tract Income and Race**

| | | # of Tracts | # of Loans | Denial Rate | Origination Rate | Secondary Market Share | Secondary Market Purchase Ratio |
|---|---|---|---|---|---|---|---|
| **Los Angeles** | **Less than 25K** | 354 | 23,072 | 18.2% | 43.6% | 18.5% | 35.1% |
| | **25K to 50K** | 841 | 113,551 | 12.0% | 50.1% | 29.2% | 46.4% |
| Income | **over 50K** | 470 | 64,169 | 8.0% | 54.2% | 36.0% | 52.1% |
| | **Minority < 50%** | 787 | 107,589 | 9.0% | 53.5% | 34.4% | 50.7% |
| | **Minority > 50%** | 878 | 93,203 | 14.2% | 47.3% | 25.1% | 42.8% |
| Race | **Minority < 5%** | 37 | 54 | 9.3% | 33.3% | 41.2% | 77.8% |
| **Orange** | **Less than 25K** | 46 | 1,436 | 16.2% | 43.4% | 23.8% | 42.5% |
| | **25K to 50K** | 234 | 29,113 | 12.1% | 49.2% | 30.2% | 47.8% |
| Income | **over 50K** | 230 | 44,433 | 6.9% | 55.1% | 35.9% | 50.7% |
| | **Minority < 50%** | 416 | 63,827 | 8.0% | 53.9% | 34.8% | 50.2% |
| | **Minority > 50%** | 94 | 11,155 | 15.4% | 45.2% | 25.7% | 44.7% |
| Race | **Minority < 5%** | 39 | 689 | 6.8% | 59.5% | 43.5% | 60.7% |
| **Riverside** | **Less than 25K** | 55 | 2,355 | 16.4% | 46.0% | 25.5% | 44.1% |
| | **25K to 50K** | 93 | 72,028 | 9.7% | 52.1% | 32.7% | 49.8% |
| Income | **over 50K** | 9 | 11,446 | 7.7% | 53.7% | 35.8% | 52.0% |
| | **Minority < 50%** | 126 | 80,821 | 9.2% | 52.6% | 33.4% | 50.2% |
| | **Minority > 50%** | 31 | 5,008 | 15.5% | 45.0% | 25.0% | 44.1% |
| Race | **Minority < 5%** | 34 | 36 | 2.8% | 66.7% | 8.8% | 12.5% |
| **San Bernardino** | **Less than 25K** | 48 | 5,355 | 18.0% | 42.1% | 19.6% | 37.3% |
| | **25K to 50K** | 111 | 44,268 | 11.8% | 49.1% | 29.0% | 45.9% |
| Income | **over 50K** | 24 | 14,616 | 8.2% | 52.5% | 36.6% | 53.3% |
| | **Minority < 50%** | 136 | 51,786 | 10.4% | 50.9% | 32.4% | 49.2% |
| | **Minority > 50%** | 47 | 12,453 | 16.0% | 42.9% | 19.8% | 36.6% |
| Race | **Minority < 5%** | 10 | 15 | 13.3% | 46.7% | 16.7% | 28.6% |
| **Ventura** | **Less than 25K** | 10 | 743 | 9.4% | 54.9% | 35.0% | 52.0% |
| | **25K to 50K** | 64 | 10,921 | 10.3% | 53.9% | 28.5% | 42.5% |
| Income | **over 50K** | 59 | 11,508 | 6.6% | 58.4% | 36.1% | 49.6% |
| | **Minority < 50%** | 104 | 18,273 | 7.8% | 57.2% | 34.6% | 48.5% |
| | **Minority > 50%** | 29 | 4,899 | 11.0% | 52.0% | 24.8% | 38.2% |
| Race | **Minority < 5%** | 6 | 43 | 4.7% | 55.8% | 38.7% | 50.0% |

*Note: Numbers are summarized from the 2002 HMDA data.*

Traditionally, central city areas were regarded underserved in the locational framework because of the concentration of poverty and minority groups in general. However, the definition of underserved areas has changed overtime to employ the concentration of minorities and the level of income within a census tract as the criteria for designation. This is based on the research done by Shear (1995), who modeled the level of credit needs in metropolitan areas and found that not all of areas within central cities suffer from concentration of poverty and high housing costs and that some areas outside of central cities do. Therefore, the categorization should be based on income and concentration of minorities rather than its geographic location. The approach to defining underserved areas was regulated in HUD's final rule (HUD, 1995). Furthermore, studies (McClure, 2001) show that the level of lending in underserved areas is much lower than that in well-served areas. Despite the fact that the definition of underserved areas only includes the income and minority measures, these areas are expected to have lower levels of lending, thus "underserved".

On one hand, ensuring equal opportunity in homeownership in underserved areas is a focus of the Federal Government's strategic goal for housing.  On the other hand, academic research aimed at underserved areas, typically central cities and distressed neighborhood, will fulfill the dual mission of lessening discrimination and providing education to potential lenders and others involved in the home buying process. Consequently, it will ultimately benefit poor families and minorities in extending mortgage credit and achieving homeownership.

## Literature

Current research on mortgage lending disparity is mostly based upon the process-based approach on discrimination, and the outcome-based approach on lending disparity and redlining.  The process model, represented by the seminal Boston Fed Study (Munnell et al. 1996), has received widely varying criticism, largely because of methodological issues such as data errors, sensitivity to outliers, and the exclusion of factors important in the lending process. The study found that after controlling for all objective indicators of applicant characteristics, lenders still rejected minorities 56 percent more often than otherwise identical whites, which provided evidence of disparate treatment discrimination.  However, some argue that unless the stages and corresponding causes of discrimination are clearly distinguished, its overall incidence cannot be properly measured and interpreted.  Most of the empirical work in this line of research only focuses on the early application stage and does not address this issue explicitly.

In recent years, the outcome-based model has received much attention, particularly on the relationship between intra-metropolitan geography and mortgage lending outcomes. For instance, using the HMDA data, Schilll and Wachter (1993) examined whether the racial and ethnic composition of the neighborhood where loan applicants wish to purchase, refinance, or improve their homes is related to lenders' decisions to accept or reject their applications.  The results indicate that neighborhood racial composition does appear related to loan disposition.

Furthermore, some researchers have focused on the uneven pattern of mortgage lending. Compare to the demand-side approach by expanding home purchasing power of low-income households, the place-based supply-side approach tends to direct mortgage credit into urban areas that have been underserved by conventional mortgage lenders.  Vidal (1995) states that the availability of credit is essential to the disadvantaged urban neighborhoods.  She also demonstrated the effectiveness of various place-based programs that are designed to bring credit into underserved areas.

From a policy perspective, the theoretical and empirical evidence on mortgage lending disparity is of great importance.  However, there exists a mismatch between theoretical models, which mostly focus on racial preferences, and empirical studies, which are essentially reduced form without adequate information.  Besides, it may be difficult to unravel the effects of neighborhood race and other attributes.   So far most studies conducted with the HMDA data ignore determinants of geographic variations in lending outcome, or simply attribute them to local variations in risk.

“Spatial Inequality" is defined as differences in income across places. Economic theory suggests that differences in incomes across place should be adjusted automatically; as incomes drop, land values become less expensive, leading to increased investment and rising incomes. However, recent research has indicated a growing gap between rich and poor places in metropolitan America. Orfield (1997) found that in the mid-1990s the fiscal inequality of suburbs increased in 25 large metro areas. Rusk (1993) hypothesized that once per capita income in a central city falls below 70 percent of that of its suburbs, it cannot regain its economic strength. In their recent paper, Swanstrom and Flack (2004) test these theories by examining trends in "spatial inequality" in the 50 largest metros between 1980 and 2000. They find segregation and concentrated poverty help to explain differences in the path of spatial inequality. Based on this analysis, there is a need for market-correcting policies in places where there is tendency towards greater inequality and mismatch between supply and demand.

In the context of mortgage practice, this problem is represented by the uneven loan flow to different neighborhoods, which exacerbates the fundamentals of underserved areas[2]. Therefore, there is a need to draw more mortgage credit to neighborhoods suffering from disparate treatment.

All these call for a reexamination of redlining. Discrimination disadvantages an agent independent of her location; redlining disadvantages agents in a location independent of their individual characteristics. Redlining occurs when a given market transaction costs more or is less likely to be approved in a geographic area with a high minority population (or in an inner-city location) than in a low minority (or suburban) area, even when differences in these areas’ economic characteristics are considered (Ross and Yinger, 2002). The basic theory behind this line of research is as follows. Area economic variables might legitimately affect housing value, and hence mortgage flows; but if mortgage decisions are based solely on economic fundamentals, then area social variables, including neighborhood racial composition, should be insignificant. Redlining arises when area race affects loan flows, even when controlling for economic fundamentals.

The rest of the paper is organized as follows. An investigation and discussion of OLS model and GWR model on lending outcome is presented in the next section (Section 3). It also discusses the data and research approach in detail. Section 4 describes the analytical results and discussions. Section 5 concludes the paper with summary of the analysis and policy implications.

---

[2] Why will loans in minority neighborhoods be riskier? Two lines of theoretical explanations have been proposed. The first argument, refined by Lang and Nakamura (1993), explains redlining as due to neighborhood externalities and information costs. They argue that the return on lending in a neighborhood depends on the total volume of lending there. Therefore, lenders prefer to follow other lenders and focus their lending on locations with less uncertainty. The second argument, initially made by Guttentag and Wachter (1981), argues that since it is costly to gather full information on individual borrowers, if borrowers' race and economic fundamentals are correlated, lenders can reasonably use neighborhood racial composition as a proxy for unknown information of borrowers when making lending decisions.

## Research Approach and Methodology

The research on lending outcome and redlining provides a basic framework for assessing the impacts of locational economic and social characteristics on mortgage lending flows. However, the relationship between the geographic distribution of areas by social-economic factors and variations in lending outcome has not been well explored the current literature. All these call for a careful examination of spatial dependence and neighborhood effects on lending disparity.

Empirical studies on lending outcome and redlining have been a common practice of the housing economics literature. Many applications have been witnessed to use area social-economic factors to analyze determinants of mortgage flow in a reduced form structure. However, although the theoretical basis of such models was logical, specifications for the model and methodology often received criticism. The traditional mortgage flow model is usually in a linear form calibrated by OLS regression in which the regression coefficients represent the implicit locational attributes. Misspecification resulting from missing important area determinants, collinearities among the determinants, and spatial dependence often cast doubt on the traditional model.

Moreover, the traditional OLS model is a typical global model. It assumes that there exists a stationary relationship between locational attributes within a metropolitan area. Such presumption has been long challenged and debated among housing economists (Yu, 2004; Watkins, 2001). As clusters of neighborhood are relatively uniform sub-groups of the metropolitan housing market, relationships may vary in different clusters but remain similar within each cluster.

This paper intends to promote an alternative methodology, the geographically weighted regression (GWR, Fotheringham et al., 2002), in investigating and modeling mortgage lending dynamics. Specifically, the GWR approach does not assume a priori particular patterns of the market non-stationarity. Instead it employs various statistical methods to test whether such non-stationarity exists. It hence might be an ideal tool in both exploring the non-stationarity of the housing market dynamics and identifying the existence of urban housing clusters.

### <u>Data</u>

The sophisticated statistical analysis of mortgage discrimination study requires clear information about the decision of whether to accept or reject a mortgage loan application. The primary information used in these studies is a dataset compiled as a consequence of the 1975 Home Mortgage Disclosure Act (HMDA), which has been widely applied in this field. HMDA controls the annual reporting of information, by mortgage lending institutions with at least $10 million in assets, on the number and dollar amount of both home mortgage and home improvement loans, by Census tract. Since the passage of the Financial Institutions Reform, Recovery, and Enforcement Act of 1989, HMDA data have also included the race, gender, and income of mortgage loan applicants.

HMDA data are routinely used to compare a lender's denial rates for minority and white loan applicants, as a measure of their loan performance with regard to minorities. HMDA evidence can be used to examine the geographic pattern of residential loan applications and denials by neighborhood. Since different geographic areas have different social-economic compositions, matching applications and denials with those patterns of the underlying location may indicate whether area social factors are the significant determinants of loan flows at the local level.

HMDA data contains some information regarding local social-economic indicators by census tract. However, the tract level data with 2002 HMDA is based on 1990 census geography and significant changes could have happened during a period of over a decade. Therefore, it is necessary to combine 2002 HMDA with the most recent 2000 census data with the 1990 census tract boundary.

As a rapidly growing region with rich ethnical diversity, Los Angeles five-county area provides an interesting case study of mortgage lending outcome. The census tract level data used in this study are from the 1990-2000 Comparability File from the California Department of Finance. The file was created to simplify tract-level comparisons of California census data over time and view the historical census data in thematic maps. Census tract boundaries have changed since 1970, making the process of comparing tracts difficult. In order to make comparisons over time, standardized tracts were established. In this case, 1990 census tracts were used as the baseline and the data from 2000 census were converted to their 1990 census tract equivalents.

The census dataset contains information on income, demographics and housing. First, it has information on tract median family income. Second, demographic information includes race/ethnicity composition, age distribution, marital status count, education attainment, occupation division, residence of previous census year. These are important factors for the analysis of tract loan flow. For example, data on the measures of neighborhood racial composition enable assessment of race-related effects associated with loan disposition. Third, the dataset has property-related measures of local housing market, including numbers of housing units by type, median housing value and median housing rent. The following table (Table 2) exhibits the profile of the key tract-level social-economic variables.

**Table 2: Descriptive Statistics for the 2000 Census Tract-level Variables**

| Variable | Mean | Std Dev | Minimum | Maximum |
|---|---|---|---|---|
| Population (000s) | 6.334 | 4.139 | - | 48.842 |
| Median Age | 33.577 | 6.868 | - | 79.000 |
| Percentage of Black | 0.077 | 0.140 | - | 0.911 |
| Percentage of Hispanic | 0.382 | 0.276 | - | 0.984 |
| Percentage Age 5-17 | 0.196 | 0.059 | - | 0.525 |
| Percentage Age 64+ | 0.107 | 0.066 | - | 1.000 |
| Percentage Married | 0.389 | 0.084 | - | 0.863 |
| Percentage College Degree | 0.170 | 0.140 | - | 0.748 |
| Percentage Below Highschool | 0.162 | 0.107 | - | 1.000 |
| Percentage Blue Collar | 0.155 | 0.054 | - | 0.375 |
| Percentage Same Household* | 0.480 | 0.105 | - | 1.000 |
| Percentage Same County* | 0.309 | 0.070 | - | 0.591 |
| Percentage Single Family Units | 0.137 | 0.087 | - | 0.840 |
| Median Family Income (000s) | 50.082 | 23.843 | - | 200.001 |
| Median House Value (000s) | 244.543 | 154.324 | - | 1,000.000 |
| Median Rent (000s) | 0.780 | 0.295 | - | 2.001 |

Number of Tracts: 2479

*Note: Residence of five years ago.*

## Empirical Framework

Based on the assumption that Census tracts of similar characteristics tend to be close to each other, cluster analysis can demonstrates the efficiency of a methodology for improving mortgage flow estimates for Census tracts. For a certain geographic district, such as a county, iterative cluster analysis will be performed on its Census tracts, based on the tract's values for a selection of variables from the census data.

This test results in a typology of different clusters—neighborhood "types" based on demographic and economic variables on racial composition, median income, and housing (Figure 1) Tests of the cluster analysis on the Los Angeles Region Urbanized tracts result in five distinct clusters, and reveal strong spatial correlation of neighboring tracts. The largest cluster includes South-central Los Angeles County, central city of Orange County, San Bernardino County, Riverside County, Ventura County, and central part of San Fernando Valley, where most traditionally underserved tracts are located. The smallest cluster includes coastal and mountain tracts of west Los Angeles County (Malibu and Santa Monica Mountains) and costal area of Orange County. Median clusters contain inner suburban tracts. The segment spatial pattern raises doubts on an independent and uniform pattern, which is the presumption of the OLS regression on tract level variables. The results reinstate the necessity to account for spatial dependence in analyzing the data.

Figure 1 about here.

In observing the non-stationarity among explanatory variables, this study utilizes a geographically GWR technique to unravel the issue of spatial dependence of the data, as discussed in the previous section. GWR is a newly developed statistical methodology specifically in dealing with spatial non-stationarity within traditional regressions. In the past few years, it has received intensive attention among scholars in geography (Fotheringham et al. 1997, 1999, 2002) and urban planning (Fotheringham et al. 1998; Leung et al. 2000a, 2000b; Paez et al. 2002a, 2002b). Most recently, it has been introduced into the urban housing studies (Yu, 2004) incorporated with the hedonic pricing model.

In particular, GWR allows regression coefficients to vary across space in terms of Tobler's (1970) first law of geography[3]. Within the framework of GWR, the model of denial rate can be expressed as:

$$Di\ (H) = \beta_{i0} + \beta_{in} N + \varepsilon i$$

Where Di is the denial rate of a census tract, $\beta_{i0}$ is the intercept term, and $\beta_{in}$ is spatially varying coefficients of neighborhood attribute.

Adjustment of the GWR model follows a local weighted least squares approach. Different from OLS, in GWR, a weighting scheme is imposed to specific locations (census tracts in our case) to assign weights. Hence the coefficients on location i are calibrated. This weighting scheme is based on each individual location's spatial proximity to a location i, i.e. near locations have more influence on the calibration than locations farther away.

To obtain the weights, a spatial kernel function must be imposed. Fixed and adaptive kernels are the two typical spatial kernels. In the fixed kernel, an optimum spatial kernel (bandwidth) will be obtained and applied over the study area, which involves less computationally intensity (Yu, 2004). However, according to recent empirical investigation (Paez et al., 2002a, 2002b; Fotheringham et al., 2002), fixed kernel approach can produce large local estimation variance in areas where data are sparse, and may mask subtle local variations in areas where data are dense. In our tests, an adaptive kernel function is used to seek a certain number of nearest neighbors to ensure a constant size of local samples. This kernel might present more reasonable means in representing the degree of spatial non-stationarity in the study area. In this study, the sizes of census tracts vary significantly across space. To account for the impact of neighboring tracts, contiguity plays more important roles than distance. Therefore, the adaptive kernel function is more appropriate and is employed in the GWR model.

## Results and Analysis

In the first attempt, an OLS regression on denial rate is carried out and results are reported in Table 3. The model performs well, and most independent variables show significance on the 95% level. The results indicate that it is plausible to model denial rate with tract social-economic variables, as manifested in previous literature.

---

[3] The first law of geography states that "All things are related, but nearby things are more related than distant things".

In addition, some of the hypothesized relationship between denial rate and social economic variables are supported by the data. In particular, the percentage of black population and percentage of residents with lower than high school degree are positively related to denial rate. Tract population, percentage of residents over 64 year old, percentage of college graduates, percentage of workers, percentage of residents staying in the same household or same county, and percentage of single family units all show negative relationship with denial rate. Surprising results are shown on median income and median house value, which show positive coefficients on denial rate.

**Table 3: Global OLS Results on 2002 Tract Denial Rate**

| Variable | Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|
| Intercept* | 0.467 | 0.111 | 4.200 | <.0001 |
| Population (000s)** | (0.001) | 0.000 | (3.800) | 0.000 |
| Median Age | (0.000) | 0.001 | (0.570) | 0.571 |
| Percentage of Black** | 0.122 | 0.013 | 9.280 | <.0001 |
| Percentage of Hispanic | 0.011 | 0.015 | 0.730 | 0.468 |
| Percentage Age 5-17 | (0.150) | 0.144 | (1.040) | 0.298 |
| Percentage Age 64+ ** | (0.252) | 0.129 | (1.960) | 0.050 |
| Percentage Married | 0.009 | 0.032 | 0.270 | 0.786 |
| Percentage College Degree** | (0.067) | 0.029 | (2.280) | 0.023 |
| Percentage Below Highschool** | 0.287 | 0.036 | 7.990 | <.0001 |
| Percentage Blue Collar** | (0.126) | 0.048 | (2.640) | 0.008 |
| Percentage Same Household** | (0.134) | 0.025 | (5.450) | <.0001 |
| Percentage Same County** | (0.127) | 0.026 | (4.930) | <.0001 |
| Percentage Single Family Units** | (0.048) | 0.010 | (4.900) | <.0001 |
| Median Family Income (000s)* | 0.00030 | 0.00016 | 1.950 | 0.052 |
| Median House Value (000s)** | 0.00005 | 0.00002 | 2.660 | 0.008 |
| Median Rent (000s) | 0.012 | 0.008 | 1.420 | 0.156 |
| Number of Tracts: 2479 | | Adj R-Sq | 0.339 | |

| Source | DF | Sum of Squares | F Value | Pr > F |
|---|---|---|---|---|
| Model | 17 | 5.298 | 75.3 | <.000 |
| Error | 2443 | 10.111 | | |
| Corrected Total | 2460 | 15.410 | | |

*Note: Variables with * are significant on the 90% level; with ** are on the 95% level.*

Comparison of GWR and OLS

The results from the global model (OLS) reveal important relationship between mortgage denial rate and tract social-economic variables. However, the relationship is built upon the theory of a stationary housing market, which is unlikely to obtain. The non-stationarity of the data has been manifested in the cluster analysis. The similar test of the data is used in the following GWR model. A tri-cube weighting method with 40 nearest neighbors is used to create the surfaces.

The ANOVA test for GWR against the global OLS model is presented in Table 4. Since GWR involves a regression for each of the observations in the sample (2479 census tracts), it is not appropriate to report the results in a table form. Instead, the varying coefficients on selected variables are produced using Arcview, which provide a visual interface to understand the determinants of denial rate.

**Table 4: ANOVA test of GWR against OLS Model**

| Source | SS | DF | MS |
|---|---|---|---|
| OLS Residuals | 5.30 | 17 | 90.07 |
| GWR Improvements | 2.31 | 256 | 592.49 |
| GWR Residuals | 2.98 | 2206 | 6,582.41 |

*Note: SS = sum of squares; DF = degree of freedom; MS = residual mean square.*

The ANOVA test (Table 4) indicates that the GWR model has significant improvement over the global OLS model. This indicates that even taking into account the added complexity of the GWR model; it still performs better than the OLS model. Results from Table 4 justify the hypothesis that significant non-stationary relationships between denial rate and neighborhood attributes exist in the Los Angeles Region.

## Results on Individual Covariates

The spatial pattern of such non-stationarity merits further attention. Figure 2 through Figure 10 show the significant surface of selected individual attribute's coefficients[4]. From the maps, the following observations emerge. First, the established relationship between tract denial rate and tract attributes is not necessarily significant everywhere in the region. Unlike the global OLS model, within a local modeling environment, some of the neighborhood attributes do not project significant influence on denial rate in specific areas. For instance, only 700 to 800 tracts out of 2479 have significance on percentage of black, percentage of Hispanic, percentage of below high school degree, percentage of college degree, percentage blue-collar workers and median income; 800 to 950 out of 2479 are significant on percentage resident of same county/household, and median house value Table 5).

**Table 5: Explanation of GWR Results on Selected Coefficients**

| Variable | Number of Tracts with Significance | Percentage | Note |
|---|---|---|---|
| Percentage of Black | 747 | 30.1% | Figure 2: Majority show positive relationship. |
| Percentage of Hispanic | 731 | 29.5% | Figure 3: No clear pattern. |
| Percentage Below Highschool | 788 | 31.8% | Figure 4: Majority show positive relationship. |
| Percentage College Degree | 791 | 31.9% | Figure 5: Majority show negative relationship. |
| Percentage Blue Collar | 707 | 28.5% | Figure 6: Scattered pattern. |
| Percentage Same County | 958 | 38.6% | Figure 7: Majority show negative relationship. |
| Percentage Same Household | 946 | 38.2% | Figure 8: Majority show negative relationship. |
| Median Family Income (000s) | 768 | 31.0% | Figure 9: Majority show negative relationship. |
| Median House Value (000s) | 823 | 33.2% | Figure 10: Majority show negative relationship. |

Number of Tracts: 2479

[4] In all the maps, tracts with colors show significance on 90% level. Tracts without color show no significance.

Second, the varying magnitude of the coefficients reveals interesting patterns. Most tracts of significance show both negative and positive relationship between covariates and denial rate, where one relationship often dominates the other. Overview of the relationships on selected variables is also exhibited in table 5.

The pattern is very distinct on percentage of black population (Figure 2), where tracts near downtown LA, east LA, Long Beach, downtown Anaheim/Santa Ana, and San Fernando Valley show high positive coefficients. In particular, most tracts with the highest positive coefficients are located in central city areas. Therefore, same magnitude of change in percentage of black will bring larger extent of change in denial rate for those central city tracts. This result shed lights on the concerns about redlining on these traditionally underserved areas, where race is an important factor in determining mortgage denial rate.

Figure 2 about here.

OLS results give positively significant coefficient on denial rate for tract median family income (Figure 3) and median house value (Figure 4), which is very counter-intuitive. The GWRs provide different pictures, which appear to be more in detail and convincing. Nearly one-third of tracts for both variables are significant, and most of them are negatively correlated with denial rate. Most of the tracts in the region are with negative coefficient on both median family income and median house value. Some scattered tract clusters in south-central Los Angeles, San Fernando Valley, south Orange County and southwest San Bernardino show positive results on median family income, and very few tracts in southeast Los Angeles and central Orange County also have positive coefficients.

Figure 3 about here.

Figure 4 about here.

Third, while OLS results show no significance on some variables, the estimated coefficients of GWR exhibit a different pattern in some cases. For instance, percentage of Hispanic population (Figure 5) receives no significance in the OLS model; in GWR, nearly one-third of the tracts are significant on 90% level. The overall pattern is more scattered than the results for percentage of black. The visual results in Figure 3 demonstrated how the tracts with high positive coefficient stand out among their neighbors (the tracts with darker color). Surely GWR provides a better understanding of the spatial and local pattern than OLS.

Figure 5 about here.

Some other observations also appear to be interesting. Percentage of blue-collar workers (Figure 6) shows significance on nearly one-third of the tracts, where the pattern is dispersed. OLS gives negative significant coefficient to the same variable. Two variables on residency, the percentage of residents in the same county in 1995 (Figure 7), and the percentage of residents in the same household in 1995 (Figure 8) have similar stories. While OLS give negatively significant coefficient on both of them, the GWR results show more scattered pattern on tracts with significance. In Figure 7, while most tracts show negative relationship, some tract clusters in west, east, and south coast of Los Angeles, coastal Orange County, northern Riverside, and southern Ventura do show positive coefficient. In Figure 8, also a few tracts in west, east and south coast of Los Angeles have positive impact on denial rate.

Figure 7 about here.

Figure 8 about here.

The two variables on education level, percentage of college graduates (Figure 9) and percentage of below high school degree (Figure 10) showed divergent coefficients in the OLS model. The GWR results conform to the conclusion and show a more detailed visual outcome in terms of where these differences occur. Some small tract clusters in Los Angeles have positive coefficient for college degree variable, along with the majority showing negative results. Most of the tracts in the region show positive coefficients for below high school variable on denial rate, which reinstates the result on the OLS model.

Figure 9about here.

Figure 10 about here.

The estimated denial rate from the GWR model reveals a clear pattern of concentrations of high denial rate in the central areas of each county (Figure 11). The most pronounced tract cluster of high denial rate is the traditionally underserved neighborhood in south-central Los Angeles, and others are Orange County central cities, central, San Fernando Valley, east Los Angeles, neighboring tracts of San Bernardino and Riverside Counties, and Los Angeles port area. The smoothed surface on the estimated denial rate (Figure 12) is also presented here, in which the high ridge represents the south-central Los Angels.

Figure 11 about here.

Figure 12 about here.

## Conclusions

The analysis on lending outcome has been a well-accepted framework in understanding mortgage lending patterns in urban housing economics. As most of the studies assume a stationary housing market that is unlikely to obtain, this study takes a geographical approach, which does not presume such stationary across space. Recognizing that there exist spatial non-stationarity among tract-level social-economic characteristics through the visual output of a cluster analysis, this paper develops a GWR lending model utilizing the 2002 HMDA data of the Los Angeles region to investigate such non-stationarity. The Statistical tests reveal that significant spatial non-stationarity exists between mortgage denial rate and most selected spatial and neighborhood attributes. The spatial non-stationarity is further mapped through GIS and reveals interesting patterns. Specifically, on the methodology side, three key conclusions can be drawn based on this study.

First, the established relationship between tract denial rate and tract attributes is not necessarily significant everywhere in the region, and only GWR can reveal this divergence spatially. Second, most spatial units of significance show both negative and positive relationship between covariates and the dependent variable, where one relationship often dominates the other if the same variable is significant in the corresponding OLS model. Third, while OLS results show no significance on some variables, the estimated coefficients of GWR exhibit a different pattern in some cases.

On the policy side, both the GWR model and OLS model show significance on most social-economic variables. While economic variables, such as income and median house value exhibit economic fundamentals of neighborhood locations and are intuitively correlated to lending outcome, the significance on social and ethnical variables shed lights on the concerns about disparate treatment. In particular, redlining arises when area race affects loan flows, even when controlling for economic fundamentals, which is the case in our study. The results from GWR provide a picture on where and to what extent these concerns merit attention.

Overall, the study is unique in terms of both its methodology and policy implications. The results of the study can provide policy makers more powerful insights of mortgage flow to different areas. Policy makers in observing the disparity should hence to promote policies regarding equal opportunity to access housing finance and to help residents living in the underserved areas become homeowners.

## REFERENCES

Fotheringham AS, and Brunsdon C. 1999. "Local forms of spatial analysis". *Geographical Analysis 31: 340-358*

Fotheringham AS, Brunsdon C and Charlton ME. 1997. "Two techniques for exploring non-stationarity in geographical data". *Geographical Systems 4: 59-82*

Fotheringham AS, Brunsdon C and Charlton ME. 1998. "Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis". *Environment and Planning A 30: 1905-1927*

Fotheringham AS, Brunsdon C and Charlton ME. 2002. "Geographically Weighted Regression: the Analysis of Spatially Varying Relationships". West Sussex: John Wiley & Sons Ltd

Leung Y, Mei C-L and Zhang W-X. 2000a. "Statistical tests for spatial nonstationarity based on the geographically weighted regression model". *Environment and Planning A 32: 9-32*

Leung Y, Mei C-L and Zhang W-X. 2000b. "Testing for spatial autocorrelation among the residuals of the geographically weighted regression". *Environment and Planning A 32: 871-890*

McClure, Kirk. 2001. "The Twin Mandates Given to the GSEs: Which Works Best, Helping Low-Income Homebuyers or Helping Underserved Areas?" *City Shape: A Journal of Policy Development and Research*, *Vol. 5, No. 3. 2001*

Megbolugbe, Isaac. 1993. "Understanding Mortgage Lending Patterns." *Journal of Housing Research, Vol. 4, Issue 2*. Fannie Mae Foundation.

Munnell, Alicia H. Lynn E. Brown, James McEneaney, and Geofferey Tootell. 1996. "Mortgage Lending in Boston: Interpreting HMDA Data". *American Economic Review, Vol. 86 (1), 25-53.*

Orfield, Myron. 1997. "Metropolitics". Brookings Institution Press.

Paez A, Uchida T and Miyamoto K. 2002a. "A general framework for estimation and iInference of geographically weighted regression models: 1. location-specific kernel bandwidths and a test for local heterogeneity". *Environment and Planning A 34: 733-754*

Paez A, Uchida T and Miyamoto K. 2002b. "A general framework for estimation and inference of geographically weighted regression models: 2. spatial association and model specification tests". *Environment and Planning A 34: 883-904*

Ross Stephen, and Yinger, John. 2002. "The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair-Lending Enforcement", The MIT Press.

Rusk, David. 1993. "Cities without Suburbs", Woodrow Wilson Center Press

Schill, Michael and Wachter, Susan. 1993. "A Tale of Two Cities: Racial and Ethnic Geographic Disparities in Home Mortgage Lending in Boston and Philadelphia". *Journal of Housing Research 4:245-75.*

Shear, William B. 1995. "Unmet Housing Needs: The Role of Mortgage Markets," *Journal of Housing Economics 4, 3:291-306.*

Swanstrom, Todd and Flack, Robert. 2004. "Spatial Inequality in Metropolitan America: A Vicious Cycle?" *Working Paper*, The Brookings Institute.

Vidal, Avis. 1995. "Reintegrating Disadvantaged Communities into the Fabric of Urban Life: The Role of Community Development". *Housing Policy Debate 6, 1:169-230*

U.S. Department of Housing and Urban development. 2002. 2002 Annual Performance Plan

Watkins, CA. 2001. "The definition and identification of housing submarkets*". Environment and Planning A 33: 2235-2253*

Yu, Dalin. 2004. "Modeling Housing Market Dynamics in the City of Milwaukee: a Geographically Weighted Regression Approach". *Working Paper, University of Wisconsin, Milwaukee.*

## 2002 Los Angeles Region Clusters on Census Characteristics

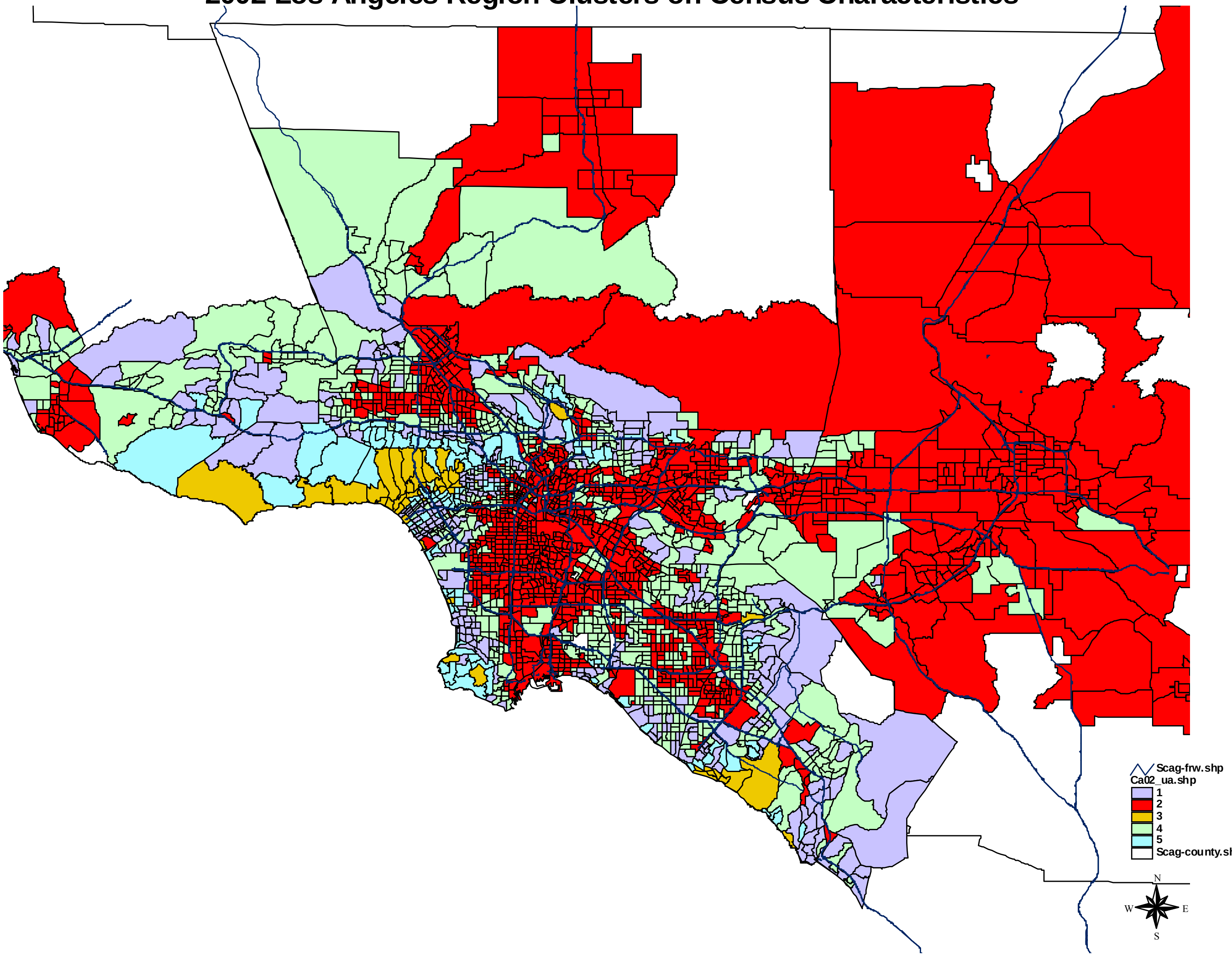



Figure 1

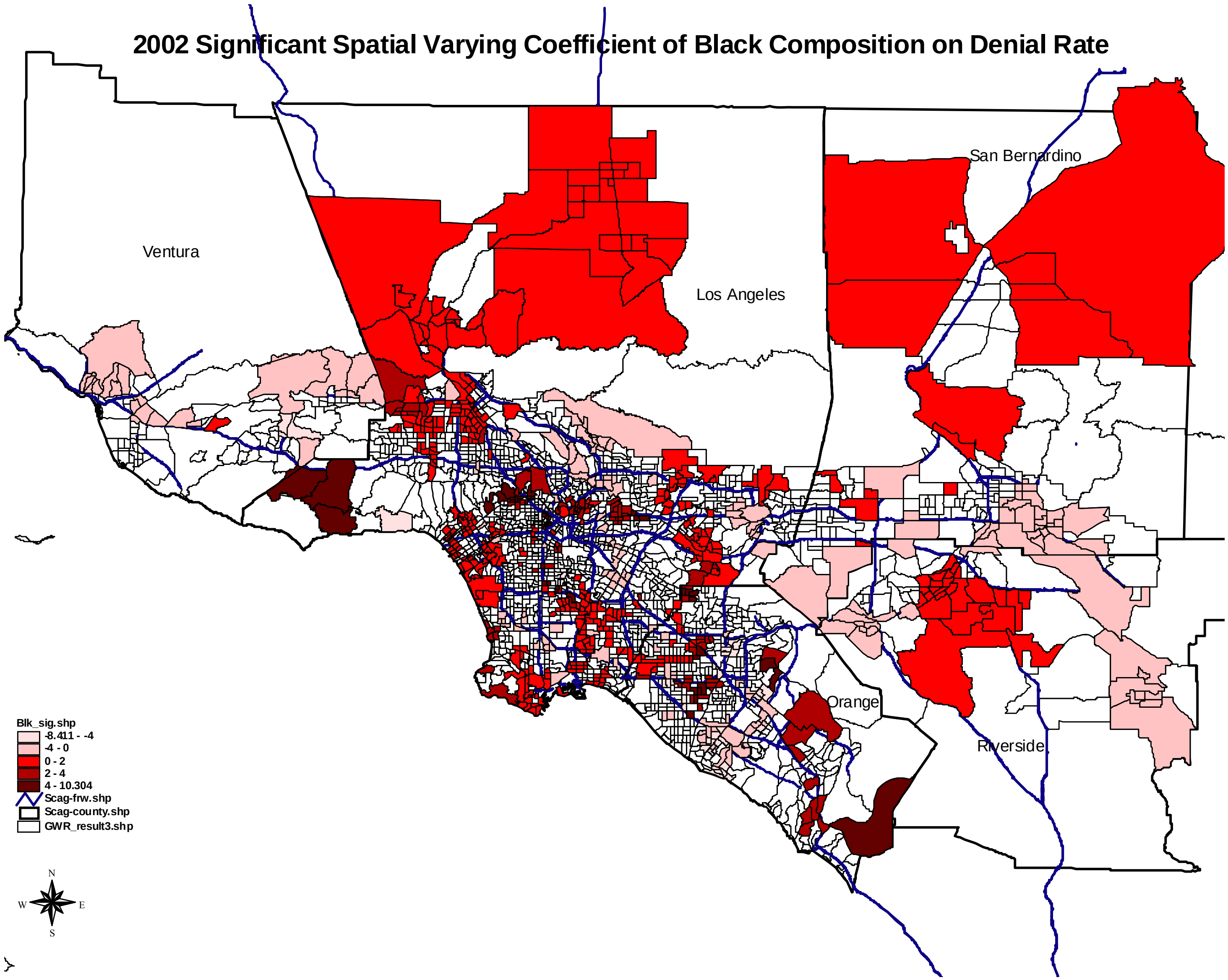

2002 Significant Spatial Varying Coefficient of Black Composition on Denial Rate
Ventura
Los Angeles
San Bernardino
Orange
Riverside
Blk_sig.shp
-8.411 - -4
-4 - 0
0 - 2
2 - 4
4 - 10.304
Scag-frw.shp
Scag-county.shp
GWR_result3.shp
N
W
E
S

Figure 2

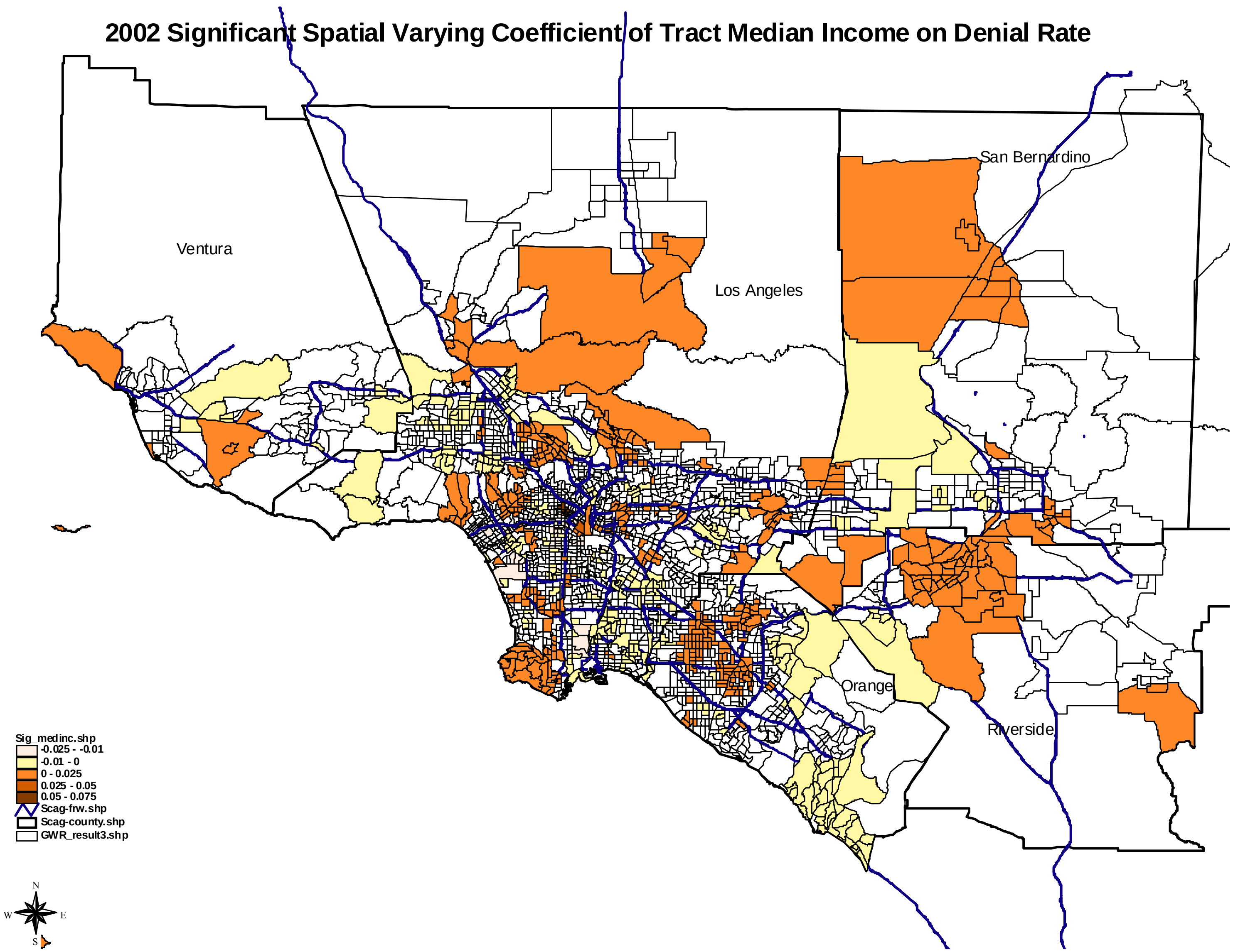

2002 Significant Spatial Varying Coefficient of Tract Median Income on Denial Rate
Ventura
Los Angeles
San Bernardino
Orange
Riverside
Sig_medinc.shp
-0.025 - -0.01
-0.01 - 0
0 - 0.025
0.025 - 0.05
0.05 - 0.075
Scag-frw.shp
Scag-county.shp
GWR_result3.shp
N
W
E
S

Figure 3

## 2002 Significant Spatial Varying Coefficient of Tract Median House Value on Denial Rate

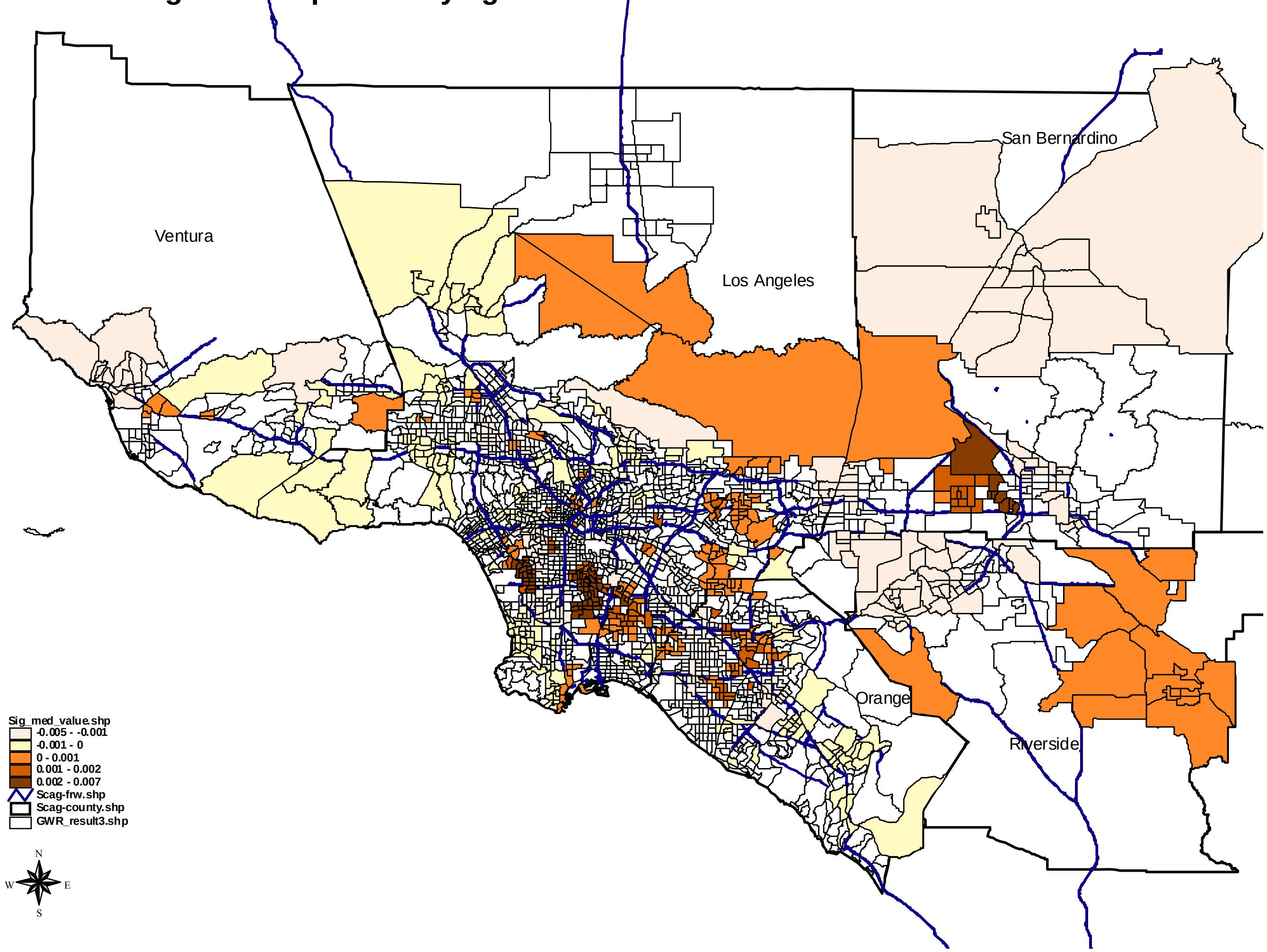



Figure 4

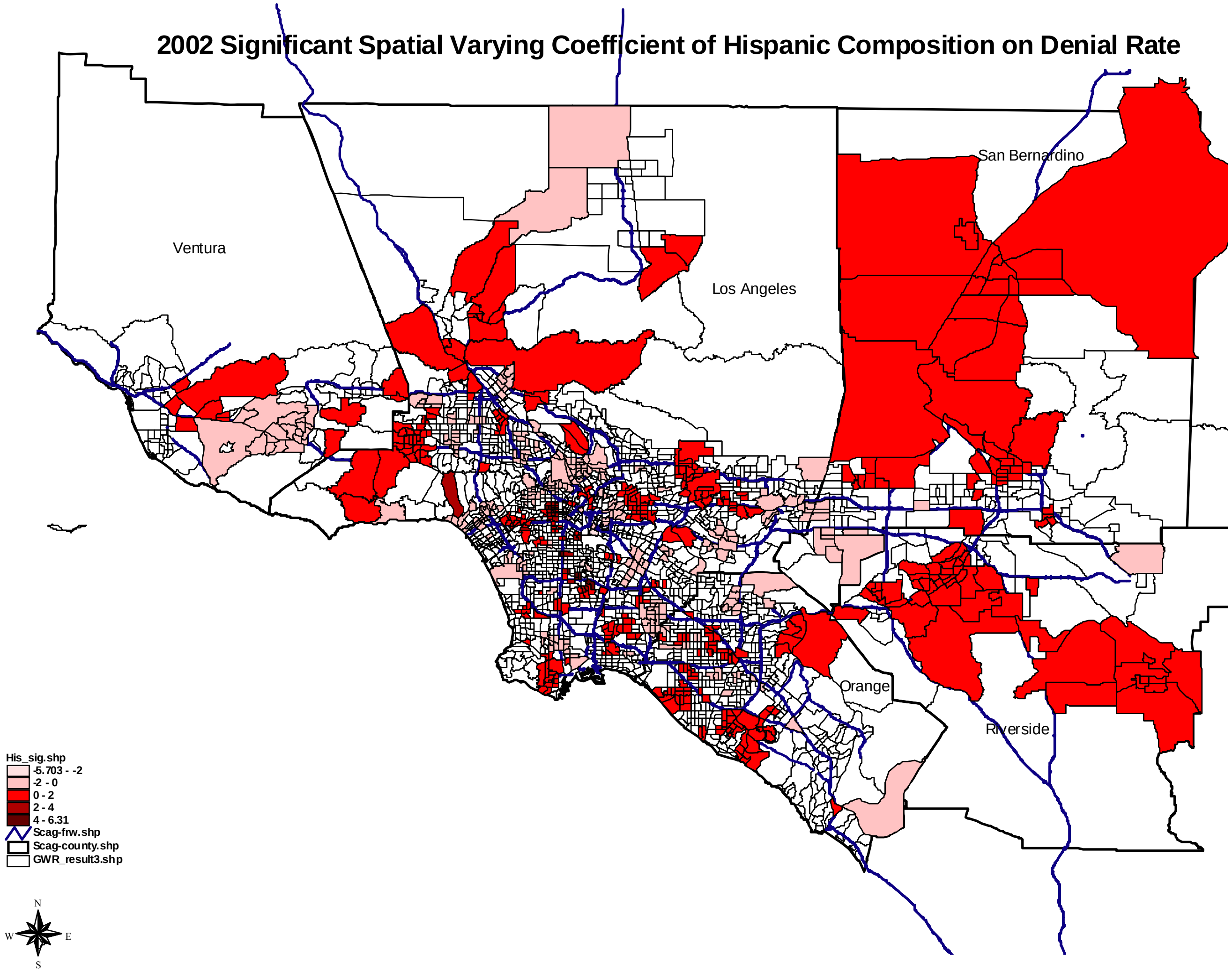

2002 Significant Spatial Varying Coefficient of Hispanic Composition on Denial Rate
Ventura
Los Angeles
San Bernardino
Orange
Riverside
His_sig.shp
-5.703 - -2
-2 - 0
0 - 2
2 - 4
4 - 6.31
Scag-frw.shp
Scag-county.shp
GWR_result3.shp
N
W
E
S

Figure 5

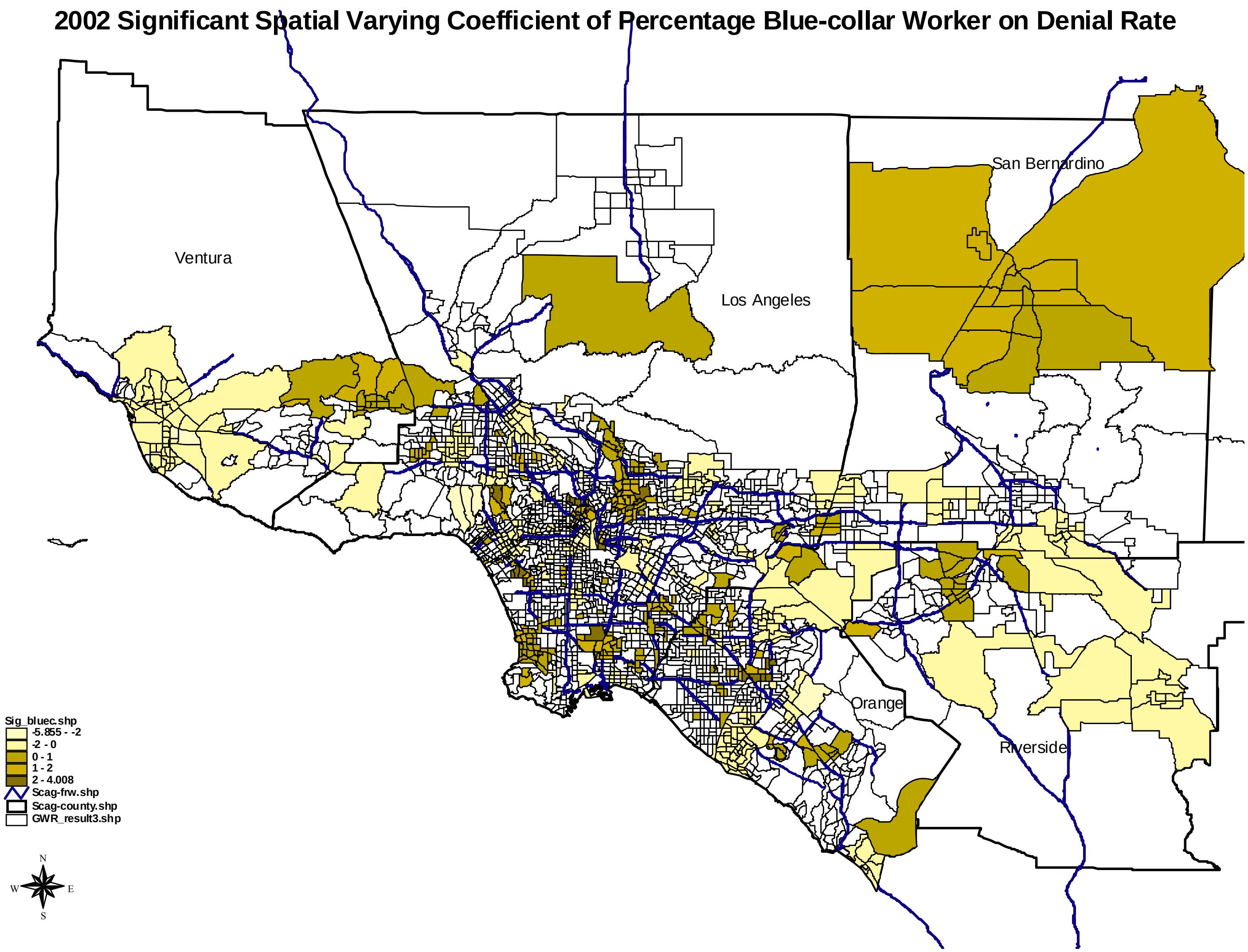
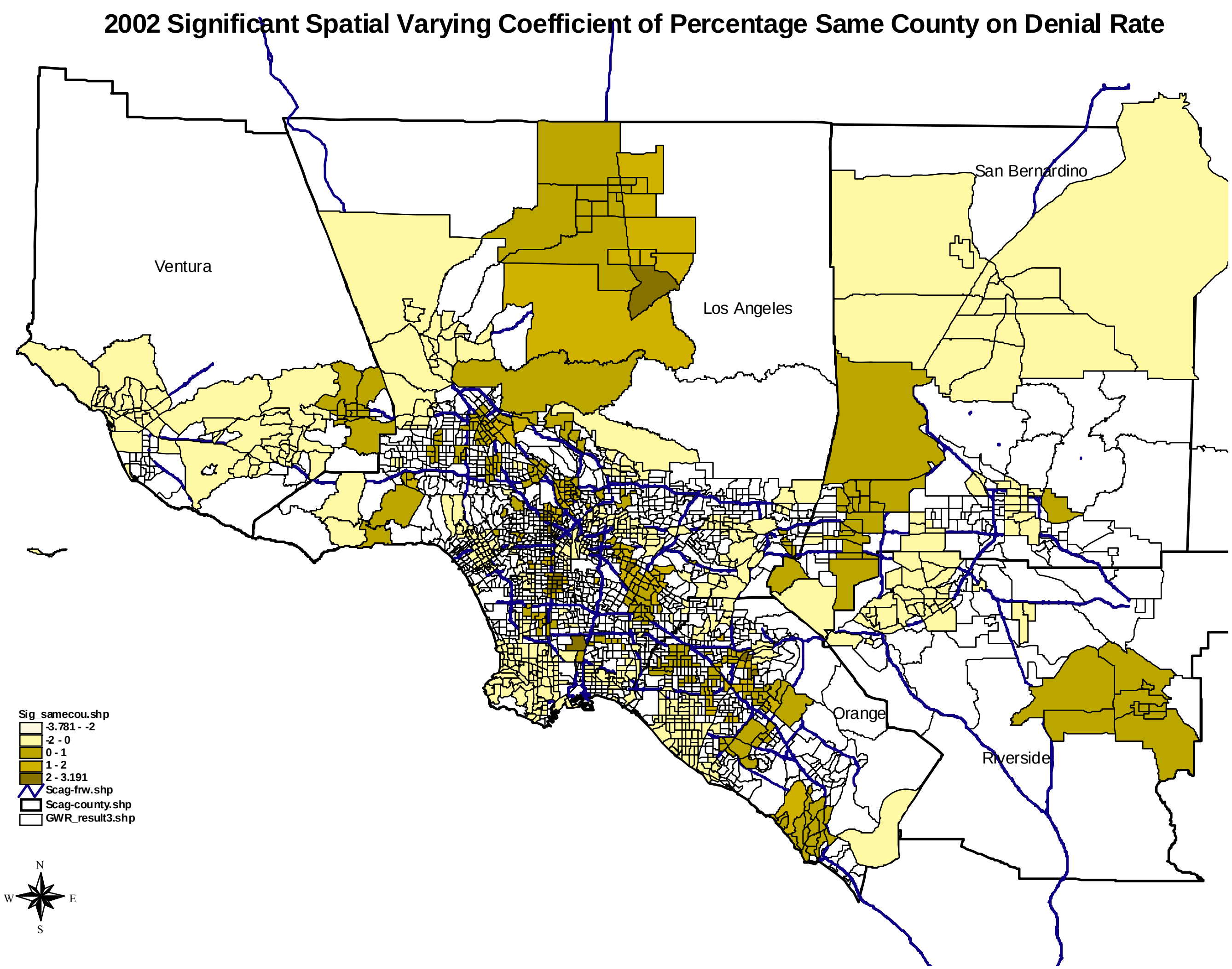

2002 Significant Spatial Varying Coefficient of Percentage Blue-collar Worker on Denial Rate
Ventura
Los Angeles
San Bernardino
Orange
Riverside
Sig_bluec.shp
-5.855 - -2
-2 - 0
0 - 1
1 - 2
2 - 4.008
Scag-frw.shp
Scag-county.shp
GWR_result3.shp
N
W
E
S

Figure 6



2002 Significant Spatial Varying Coefficient of Percentage Same County on Denial Rate
Ventura
Los Angeles
San Bernardino
Orange
Riverside
Sig_samecou.shp
-3.781 - -2
-2 - 0
0 - 1
1 - 2
2 - 3.191
Scag-frw.shp
Scag-county.shp
GWR_result3.shp
N
W
E
S


Figure 7

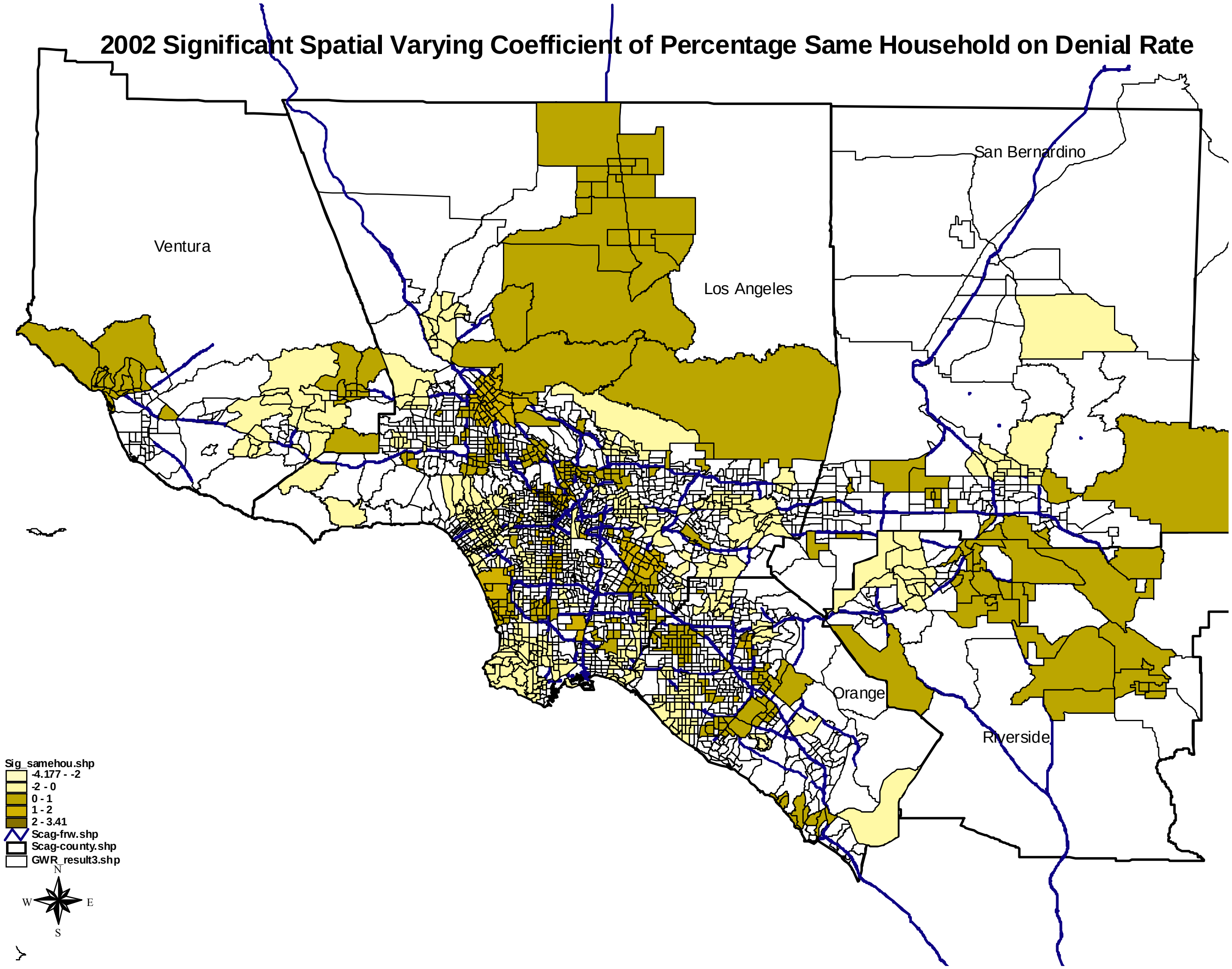

2002 Significant Spatial Varying Coefficient of Percentage Same Household on Denial Rate
Ventura
Los Angeles
San Bernardino
Orange
Riverside
Sig_samehou.shp
-4.177 - -2
-2 - 0
0 - 1
1 - 2
2 - 3.41
Scag-frw.shp
Scag-county.shp
GWR_result3.shp
N
W
E
S

Figure 8

## 2002 Significant Spatial Varying Coefficient of Percentage College Degree on Denial Rate

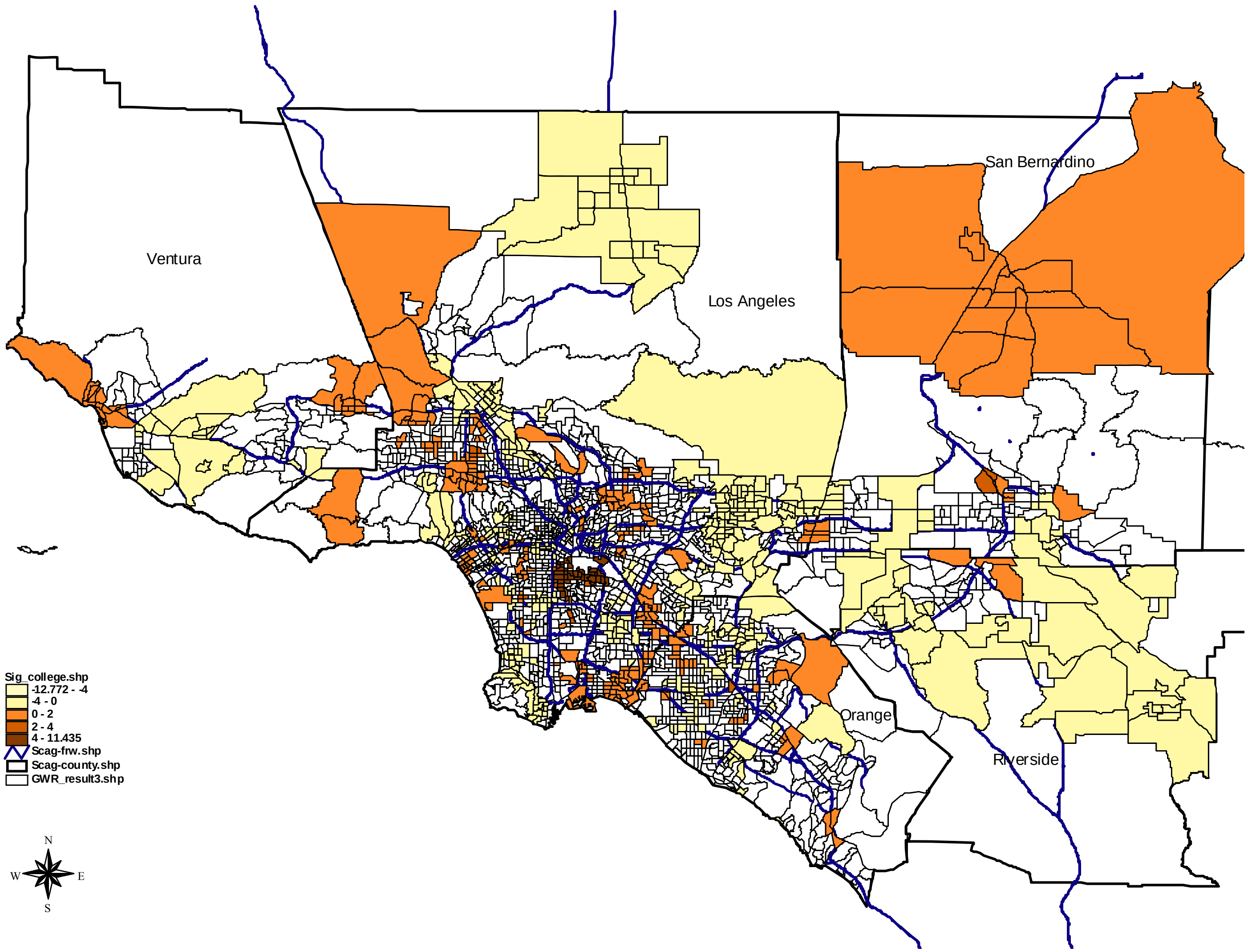



Figure 9

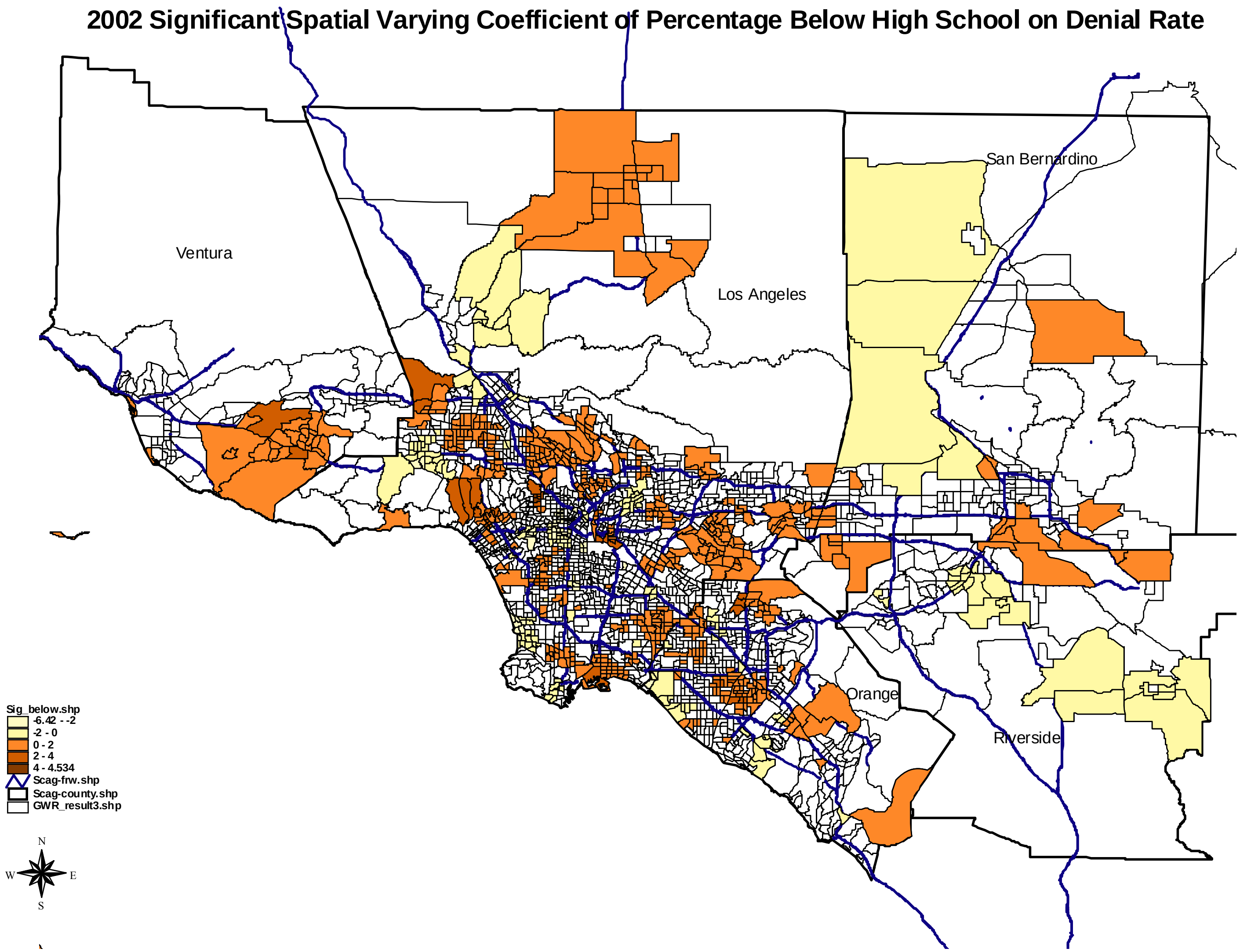

2002 Significant Spatial Varying Coefficient of Percentage Below High School on Denial Rate
San Bernardino
Ventura
Los Angeles
Orange
Riverside
Sig_below.shp
-6.42 - -2
-2 - 0
0 - 2
2 - 4
4 - 4.534
Scag-frw.shp
Scag-county.shp
GWR_result3.shp
N
W
E
S

Figure 10

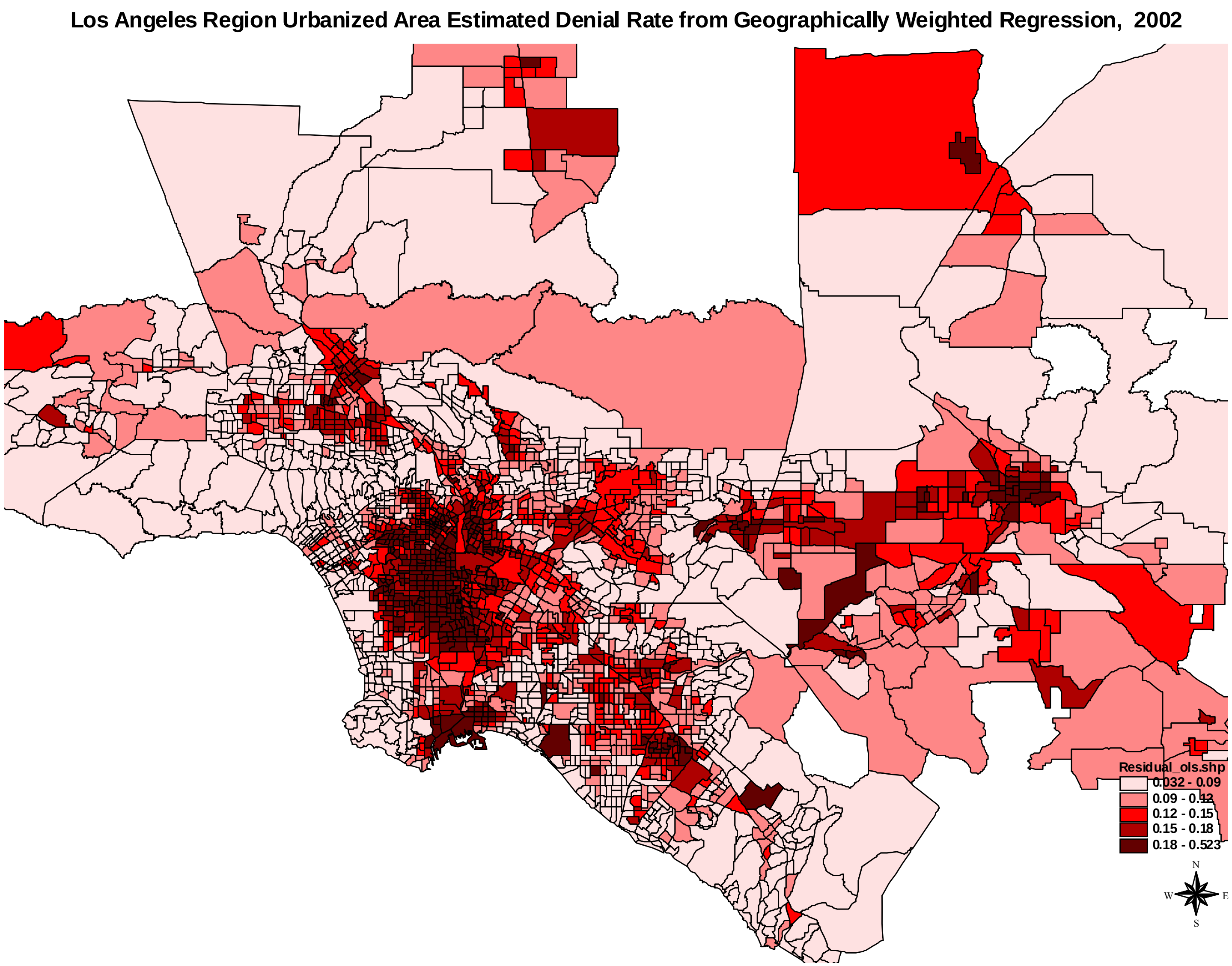

Los Angeles Region Urbanized Area Estimated Denial Rate from Geographically Weighted Regression, 2002
Residual_ols.shp
0.032 - 0.09
0.09 - 0.12
0.12 - 0.15
0.15 - 0.18
0.18 - 0.523
N
W
E
S

Figure 11

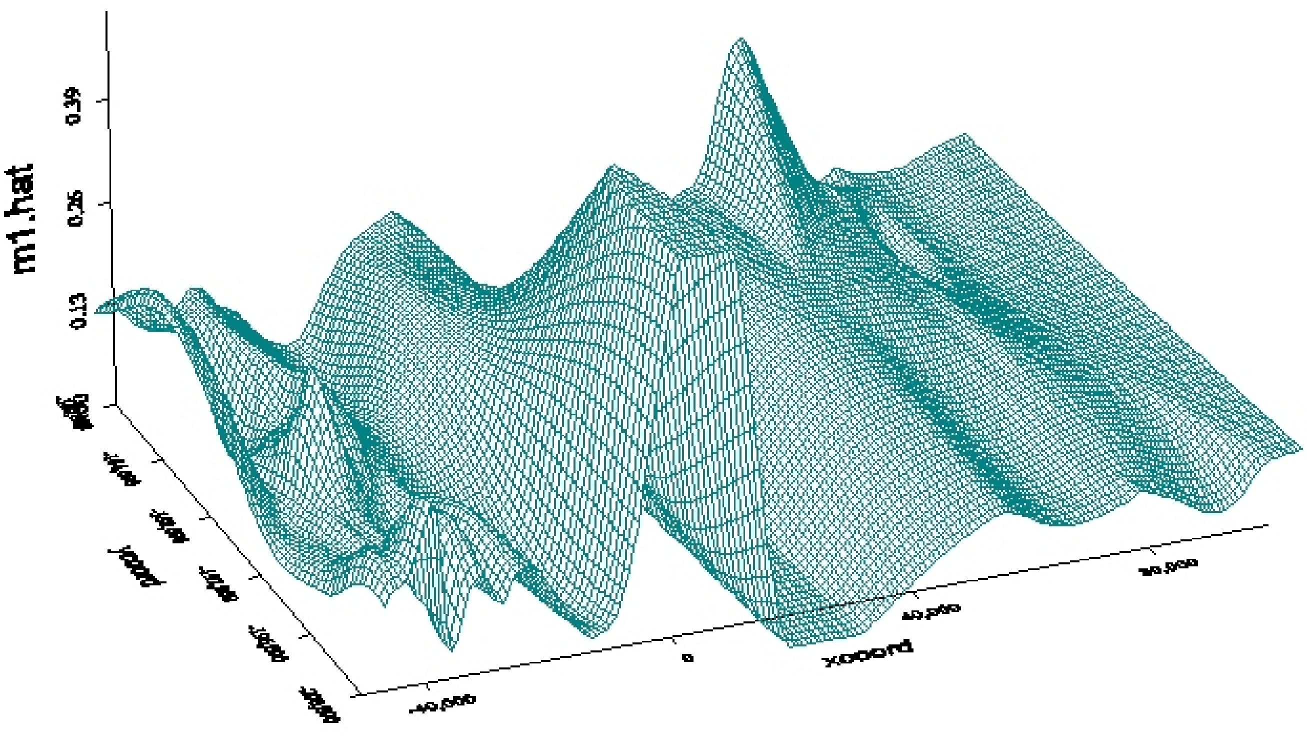



Figure 12