THE SPATIAL PATTERNS AND CAUSES OF U.S. UNEMPLOYMENT RATES: A COUNTY LEVEL ANAYSIS

A Thesis

Presented to the

Faculty of Geography

California State University, Fullerton

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts

in

Geography

By

Michael Ribant

Approved by:

Mark Drayse, Committee Chair Department of Geography	Date
John Carroll, Member Department of Geography	Date
Jonathan Taylor, Member Department of Geography	Date

ABSTRACT

The national unemployment rate tells us little about unemployment differentials at the state level, and even less about what is happening in that regard at the county level. While unemployment rate studies abound, they are most often conducted from a macroeconomic perspective and ignore wide disparities in unemployment rates across smaller geographic regions. This paper attempts to identify the factors that generate such large differences in geographic unemployment rates as well as the spatial patterning of U.S. county unemployment rates. Specifically, the determinants and spatial patterns of county unemployment rates across the contiguous United States are investigated using spatial autocorrelation, spatial regression, multilevel modeling and binary logistic regression techniques. With 3,104 contiguous U.S. counties as the study area, and the end-month of the past two recessions as the study periods (June 2009 and November 2001), this research seeks to answer four main research questions: 1) What are the spatial patterns of county unemployment rates at the end of these two recessions? 2) What are some of the determinants of county unemployment rates at the end of these two recessions? 3) To what extent do state-level factors exert influence upon county-level unemployment rate distributions? 4) Can counties with a high risk of belonging to the upper quartile of counties, as measured by the unemployment rate, be statistically predicted using racial, socioeconomic and industrial mix variables?

Using geostatistical techniques applied to a large-scale database, this study has revealed five major findings: 1) geographic clustering of high-unemployment persists to a high degree at the county level in the U.S.; 2) counties that are manufacturing-dependent or that have a high percentage of non-whites show a positive relationship to the unemployment rate, while self-employment and educational attainment are negatively associated; 3) counties located in the top subprime mortgage states show a positive relationship to the unemployment rate for June 2009; 4) state-level factors account for over half the variance of county unemployment; and 5) high risk factors for belonging in the upper quartile of counties, as measured by the unemployment rate, include manufacturing-dependent counties and a high unemployment rate registered at the end of the previous recession.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to thank the geography department at California State University, Fullerton for moving me forward in my life. In particular, I would like to thank Dr. Carroll for his professionalism in running such a fine program. Also, I would like to thank Dr. Drayse for his guidance; Dr. Taylor for his patience; Dr. Xu for her kindness; and Dr. Miller for his friendship.

This thesis was completed late, very late, through nobody's fault but mine. After leaving CSU Fullerton, I have been fortunate enough to have completed an additional two years of coursework towards the pursuit of a Ph.D. degree in geography. The time required to complete this thesis competed with my studies, and with my personal life. I met my wife, Anastasia, in the summer of 2006 at a coffee cart I owned in front of San Diego's central library. A year later we were married and I was in graduate school. A year after that, our son was born. Anastasia has watched me work, piecemeal, on this thesis for three years. At times, I doubted it would ever be finished, she never did. She has given me nothing but love, support, understanding and joy since I met her. This thesis is dedicated to her.

I wish also to thank the other two women in my life, my daughter, Lauren, and mother, Joanne. I would not have reached this point in my life without their unwavering love and support throughout the years. Lauren, you are my hero. Mom, you are the best.

Finally, I wish to acknowledge the one person who more than any other made my time at CSU Fullerton both memorable and fun, Claire Wilson. Claire and I often shared a train or bike ride to the Fullerton campus, or a beer after our night class. Always principled and gregarious, she became a role model for all during times of extreme personal duress as she fought with grace against the awful cancer that would ultimately steal her life. I am a better person for having known her and shall always treasure our friendship. I miss you, Claire.

CHAPTER 1

INTRODUCTION

Study Overview

The first decade of the 21st century has seen the United States mired in two recessions. As measured by the National Bureau of Economic Research, the first of these two recessions was a brief six-month trough lasting from March through November of 2001, while the more severe second recession ran from December 2007 through June 2009 (NBER, 2011). This paper examines spatial variations in unemployment rates across 3,104 U.S. counties as measured at the end of these two U.S. recessions. These study periods were chosen because the effects of economic shocks and unemployment rate variability are most severe near the end of a recession. Using the statistical techniques of spatial autocorrelation, spatial econometrics, multilevel modeling, and binary logistic regression, this paper seeks to answer four main research questions related to these two recessions: 1) What are the spatial patterns of county unemployment rates at the end of these two recessions? 2) What are some of the determinants of county unemployment rates at the end of these two recessions? 3) To what extent do state-level factors exert influence upon county-level unemployment rate distributions? 4) Can counties with a high risk of belonging to the upper quartile of counties, as measured by unemployment rate, be statistically predicted using racial, socioeconomic and industrial mix variables? The research concludes by discussing implications of the paper's findings.

Regional unemployment differentials are typically considered a benchmark for a region's socio-economic performance. The topic is well explored within the literature. However, as Elhorst (2003) states, most discussions of regional unemployment disparities occur without ever connecting it to geography. The focus of this paper is to add to the growing amount of literature attempting to bridge that gap by analyzing the causes and effects of regional unemployment disparities from a geographic perspective.

Background Information

The organization having the responsibility of officially declaring whether the U.S. has entered a recession is the National Bureau of Economic Research (NBER). It is commonly believed that the NBER defines a recession as two quarters of decline in gross domestic product (GDP), however, this is not a strict definition. The NBER also analyzes declines in broader measures such as real income, employment, industrial production and sales, in addition to GDP, before officially declaring the beginning and end of a recession. Since World War II, there have been eleven contractions in the US business cycle (Table 1). According to NBER data for these post-war business cycles, the mean contraction duration was 10.4 months and the mean expansion duration was 56.9 months.[1] However, as measured by their unemployment rates, some counties and states quickly absorb business cycle shocks while others struggle with unemployment persistence long after a national recession has officially ended.

[1] NBER web site accessed 02/15/11: http://www.nber.org/cycles.html.

In December of 2007, the United States entered its second recession of the 21st century. Going back to the 1930s, only the 1982 recession and the Great Depression have surpassed the severity of this most recent recession as measured by the national unemployment rate. As of June 2009, the national unemployment rate stood at a seasonally adjusted rate of 9.5%, contrasted to the December 1982 peak rate of 10.8%.[2] As of this writing (February 2011), the peak national unemployment rate, post-recession, was 10.1% in October 2009.

According to Claessens, Kose, & Terrones (2008), the average recession lasts about four quarters and the average output decline from peak to trough is around 2%. Their study of recessions, credit crunches and asset busts also finds that a typical credit crunch lasts 2.5 years, a housing bust 4.5 years, and an equity bust 2.5 years, with corresponding declines of 20% in real credit, 30% in housing prices, and a 50% drop in equities. That the United States recently found itself in the midst of all three scenarios is disturbing, but ancillary to the purpose of this paper, which is the analysis of U.S. unemployment rates during the past two recessions.

Figure 1 details the monthly U.S. unemployment rate since January of 1948. An analysis of the past four recessions going back to 1980 shows that the 2007-2009 recession did not quite hit the unemployment rate highs of the 1982 recession, but surpassed the peak highs of the milder recessions of 1990-1991 and 2001. Interestingly, the peak unemployment levels of these two milder recessions, 7.8% and 6.3%,

---

[2] FRB web site accessed 02/15/11: http://research.stlouisfed.org/fred2/series/UNRATE

respectively, were not reached until some 15 to 19 months after the official end of the respective recessions (Table 1). Every other recession going back to 1948 had an unemployment rate peak that more or less coincided with the official end of each business cycle, with the longest lag period being only four months after the recession of 1953-1954. It remains to be seen how long the lag from the 2007-2009 recession will be. So far, the national unemployment rate seems to have peaked only four months after its official end, in October 2009 (10.1%). However, given the severity of this past recession, it seems too early to make this observation official as of the time of this writing (February 2011).

Table 1. Dates and Durations for Postwar Recessions

Peak	Trough	Contraction Duration in Months	Expansion Duration in Months	Unemployment Rate Peak**: Lag Month and Year	Unemployment Rate Lag**: Months after end of recession
November-48	October-49	11	37	Oct-1949	0
July-53	May-54	10	45	Sep-1954	4
August-57	April-58	8	39	Jul-1958	3
April-60	February-61	10	24	May-1961	3
December-69	November-70	11	106	Dec-1970	1
November-73	March-75	16	36	May-75	2
January-80	July-80	6	58	Jul-1980	0
July-81	November-82	22	12	Dec-1982	1
July-90	March-91	8	92	Jun-1992	15
March-01	November-01	8	120	Jun-2003	19
December-07	June-09	19	73	Oct-2009	4
Average		11.7	57		4.4

Source: National Bureau of Economic Research (NBER), 2011. Shaded area denotes study period. ** denotes author calculations.

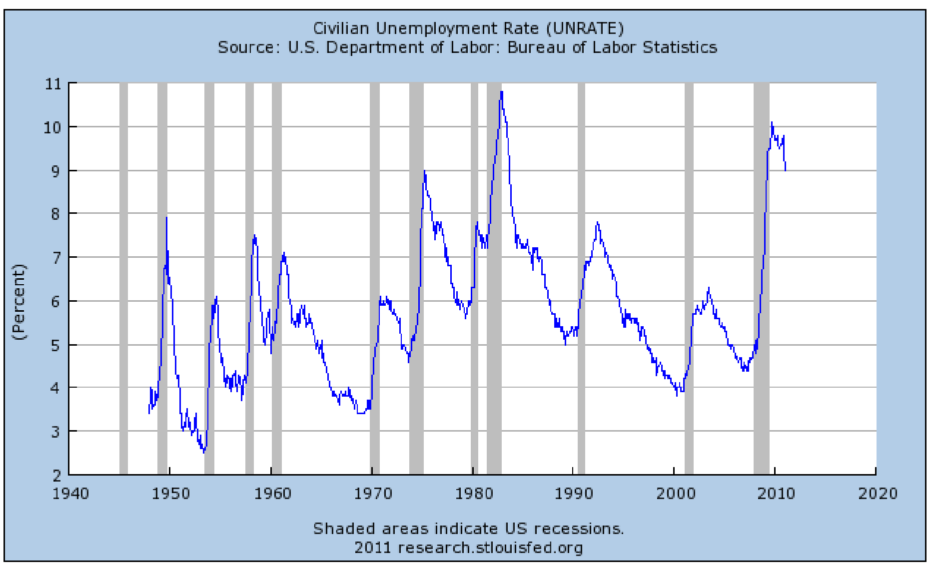

Figure 1. U.S. Unemployment Rate: January 1948 to January 2011. Source: Federal Reserve Bank of St. Louis, 2011.

While this paper is concerned with the past two recessions, it is worthwhile first to briefly examine the two decades that preceded them. The period between 1982 and 1990 saw the largest peacetime economic expansion in U.S. history to that point as real GDP grew at an annual rate of 3.3 percent (Walsh, 1993). This was followed by the 1990-91 recession. As noted by Blanchard (1993), a proximate cause for this recession is hard to define but many have been offered. The deregulation and subsequent collapse of the savings and loan industry, a prolonged economic expansion withering out, the decrease in defense spending brought about by the Cold War, oil price hikes after the Iraqi invasion of Kuwait, and a restrictive monetary policy maintained by the Federal Reserve have all been proposed as contributory reasons for the downturn. Whatever the reasons, the long-term jobless rate for this recession was much lower than the recession of 1980-82. Employment within the manufacturing sector, always a casualty of economic declines, dropped only 3.2 percent during the 1990-91 recession, representing a smaller drop than any of the previous postwar contractions (Gardner, 1994). However, unlike previous recessions, the labor market continued to sag long after this recession ended. It took 15 months (June of 1992) for the unemployment rate to start dropping again, whereas previous postwar recessions saw labor market adjustments almost immediately after the contraction's end. And though job losses came from a much wider spectrum of the work force, 86 percent of those who lost jobs did so permanently, compared to 56 percent in earlier recessions (Gardner, 1994). In summary, the 1990-91 recession was mild, but the labor market recovery was slow.

While the second half of the 20th century generally produced the largest economic expansion in world history, the defining period for such growth was the U.S. decade of the 1990s. The ten years between the end of the 1991 recession and the beginning of the 2001 recession was the longest temporal period of expansion on record, breaking the previous expansion record set after the 1960-61 recession (Kliesen, 2003). America had never experienced such growth in its history, and the decade's economic momentum drove the unemployment rate down to 3.8% by April of 2000, a level the country had not seen since December of 1969. The correction period, if it can be called that, for this expansive economic growth was the short recession of 2001.

The 2001 recession has been dubbed The Mildest Recession by Nordhaus (2002). Indeed, real GDP actually rose 0.2 percent from the first to the fourth quarters of 2001, the peak to trough quarters of the recession as arbited by the NBER (Kliesen, 2003). The unemployment rate in December of 2001 stood at a mere 5.7%, making it the lowest rate recorded at the end of a recession since World War II. The mettle shown by the economy in the aftermath of a stock market bust and the September 11th tragedies was a welcome surprise. However, as with the 1990-91 contraction, the labor market once again took a long time to recover from this recession. In fact, it took 19 months for the unemployment rate to peak after the end of this recession, when it reached 6.3% in June of 2003 (Figure 1). This was four months longer than the previous postwar lag high of 15 months set after the end of the 1990-91 recession. The expansion coming out of this recession was weak, with employment and wage growth less than that of any previous postwar expansion period. Furthermore, this was accompanied by a housing boom (and a subsequent bust) that was partially fueled by a spectacular increase in sub-prime mortgage lending.

Industrial Structure

An important underlying factor accompanying U.S. business cycle history since 1980 is the critical change that occurred within the U.S. industrial structure. Between 1990 and 2007, U.S. manufacturing and farming sectors suffered significant declines in industrial employment shares. These declines were offset by employment increases within the high-tech, finance, insurance and real estate service sectors (Table 2). The manufacturing sector in 1990 employed over 14 percent of the workforce, nearly 20 million Americans. By 2007, over 5 million of those jobs had disappeared, dropping the employment share for manufacturing to a mere 8%. And it has become worse still since 2007. In fact, the 11.65 million manufacturing jobs within the United States as reported by the Bureau of Labor Statistics (BLS) for November of 2009 stands at the lowest level for manufacturing jobs since just before WWII began.[3] However, losses in the manufacturing sector over the past thirty years (the number of manufacturing jobs stood at 19.5 million just before the beginning of the 1980-1982 recession) represent only one side of America's changing industrial makeup. Indeed, the U.S. economy added over 40 million jobs between 1990 and 2000, with spectacular sectoral growth occurring within the high tech and services industries. In fact, by 2014, 4 out of 5 jobs in the United States

[3] Federal Reserve Bank of St. Louis website accessed January 3, 2011: http://research.stlouisfed.org/fred2/series/MANEMP

Table 2. Total Full-Time and Part-Time Employment by Industry

Sector	1990 (SIC)	2001 (NAICS)	2007 (NAICS)	% of Non-farm Employment: 1990	% of Non-farm Employment: 2001	% of Non-farm Employment: 2007
Nonfarm employment	136,227,900	163,958,700	178,102,800	118.46%	116.47%	115.77%
Private employment	114,995,900	140,778,700	153,845,800	100.00%	100.00%	100.00%
Agr. services, forestry, fishing	1,454,000	1,022,500	1,014,400	1.26%	0.73%	0.66%
Mining	1,044,100	811,400	984,900	0.91%	0.58%	0.64%
Transportation and public utilities	6,550,600			5.70%	0.00%	0.00%
Transportation and warehousing		5,474,000	5,887,700	0.00%	3.89%	3.83%
Utilities		618,800	576,500	0.00%	0.44%	0.37%
Construction	7,261,800	9,846,700	11,641,100	6.31%	6.99%	7.57%
Manufacturing	19,694,200	16,994,600	14,512,000	17.13%	12.07%	9.43%
Wholesale trade	6,720,500	6,273,400	6,657,800	5.84%	4.46%	4.33%
Retail trade	22,885,500	18,528,800	19,282,000	19.90%	13.16%	12.53%
Information		4,053,800	3,537,000	0.00%	2.88%	2.30%
Finance, insurance, and real estate	10,714,600			9.32%	0.00%	0.00%
Finance and insurance		7,839,600	8,429,700	0.00%	5.57%	5.48%
Real estate, rental and leasing		5,551,400	8,142,400	0.00%	3.94%	5.29%
Services	38,670,600			33.63%	0.00%	0.00%
Professional, scientific, technical		10,575,800	11,866,300	0.00%	7.51%	7.71%
Management of companies		1,779,300	1,965,200	0.00%	1.26%	1.28%
Administrative/Waste services		9,621,000	11,180,300	0.00%	6.83%	7.27%
Educational services		3,058,300	3,833,000	0.00%	2.17%	2.49%
Health care/Social assistance		15,611,400	18,204,900	0.00%	11.09%	11.83%
Arts, entertainment, and rec.		3,243,100	3,736,900	0.00%	2.30%	2.43%
Accommodation/Food services		10,825,200	12,253,000	0.00%	7.69%	7.96%
Other services except pub admn	9,049,600	10,140,700		7.87%	7.20%	0.00%
Government/Govt. enterprises	21,232,000	23,180,000	24,257,000	18.46%	16.47%	15.77%
Federal, civilian	3,233,000	2,728,000	2,782,000	2.81%	1.94%	1.81%
Military	2,718,000	2,099,000	2,041,000	2.36%	1.49%	1.33%
State and local	15,281,000	18,353,000	19,434,000	13.29%	13.04%	12.63%
State government	4,404,000	5,031,000	5,205,000	3.83%	3.57%	3.38%
Local government	10,877,000	13,322,000	14,229,000	9.46%	9.46%	9.25%

Note: Industry totals are combined from Bureau of Economic Analysis tables CA25 (SIC) and CA25N (NAICS). Source: BEA, http://www.bea.gov/regional/spi/default.cfm?selTable=SA25N&selSeries=NAICS.

are projected to be within the service-providing sectors; with educational, health care, and professional and management sectors expected to perform as the fastest growing sectors (Berman, 2005). These declines and increases within the American industrial mix highlight some of the recent structural changes occurring within the U.S. economy and provide a background for some of the industrial mix variables used later in this paper to help explain regional differences in U.S. unemployment rates.

General Problem

While macroeconomic factors affect the general trend of the national unemployment rate, the unemployment rate is not evenly distributed across the country. The unemployment rate of some states is twice that of others, while some counties have unemployment rates more than triple the rates of others (Figures 2 and 3). Adding to the disparity problem is the fact that some areas take longer to recover from financial shocks than others. Still more perplexing are the major geographic clusters of high (and low) county unemployment rates present within the country, the existence of which is hard to eliminate once established. The aim of this paper is to discover and examine local unemployment rate differentials and discover the causes for these spatial and geographic clusters.

The changing industrial structure of the United States is certainly a factor to be studied when trying to explain regional unemployment differentials. This paper examines the relationship between regional unemployment rate levels and industrial mix. In addition to industrial mix, other explanatory variables are examined. According to Elhorst (2003), who extensively surveys the unemployment literature, some of the independent variables used to explain unemployment rate differentials beyond industrial mix are educational attainment, per capita income, the percentage of self-employed and race, which are all examined in some measure within this study. The scope and purpose of this paper preclude the presentation of an exhaustive predictive model for unemployment rate variations. The most important variable to be examined in this study is location. However, many other explanatory variables, substantiated within the literature and readily available at the state and county levels of analysis, are included as well.

Importance of the Problem

The human and social consequences of economic downturns are probably best measured by the unemployment rate. High unemployment and persistence bring human suffering, financial and otherwise. Understanding the geographical manifestation of unemployment helps identify problem areas so that financial, managerial and governmental resources can be applied towards regional solutions.

Motivations

The national unemployment rate tells us little about unemployment differentials at the state level, and even less about what is happening in that regard at the county level. This paper seeks to identify the spatial relationships of U.S. unemployment at the county level. Recently, there has been extensive research from around the globe, particularly Europe, examining the spatial characteristics of regional unemployment rate distributions. This literature is covered in depth in Chapter 2. However, spatial studies on U.S. unemployment differentials remain largely absent from the literature. Most of the U.S.-based literature attempts to predict unemployment rate levels by use of regression models, which fail to capture the spatial correlations of the underlying data. If, as this paper hypothesizes, geography plays a role in determining U.S. regional unemployment differentials, then classic Ordinary Least Squares (OLS) regression models will fail to catch these effects. This paper uses the statistical techniques of spatial autocorrelation, spatial regression, multilevel modeling and binary logistic regression to answer the following four main research questions:

1) What are the spatial patterns of county unemployment rates at the end of these two recessions?

2) What are some of the determinants of county unemployment rates at the end of these two recessions?

3) To what extent do state-level factors exert influence upon county-level unemployment rate distributions?

4) Can counties with a high risk of belonging to the upper quartile as measured by unemployment rate be statistically predicted using racial, socioeconomic and industrial mix variables?

These are the four main research questions that this paper seeks to answer. In the next section, six testable hypotheses are proposed, which are later evaluated by use of a four-step methodological inquiry. The statistical techniques and models which comprise this methodology are summarized later in this chapter.

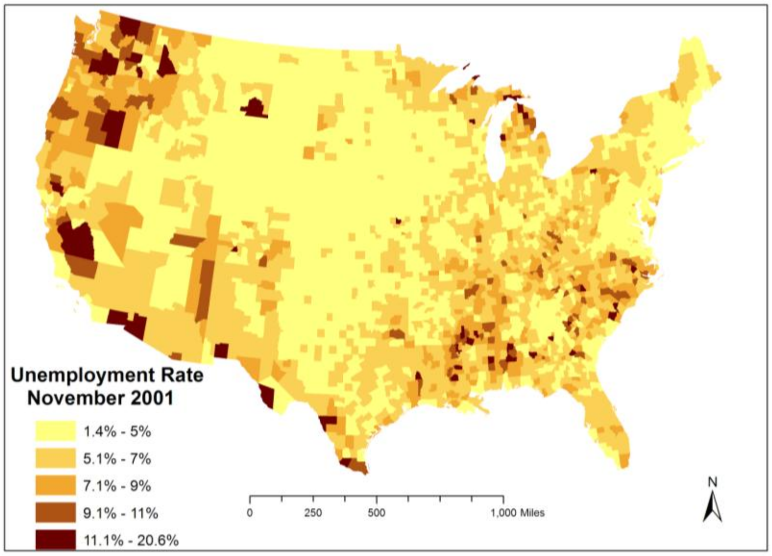

Figure 2. Unemployment Rates by County, November 2001. Data Source: St. Louis Federal Reserve Bank (2011).

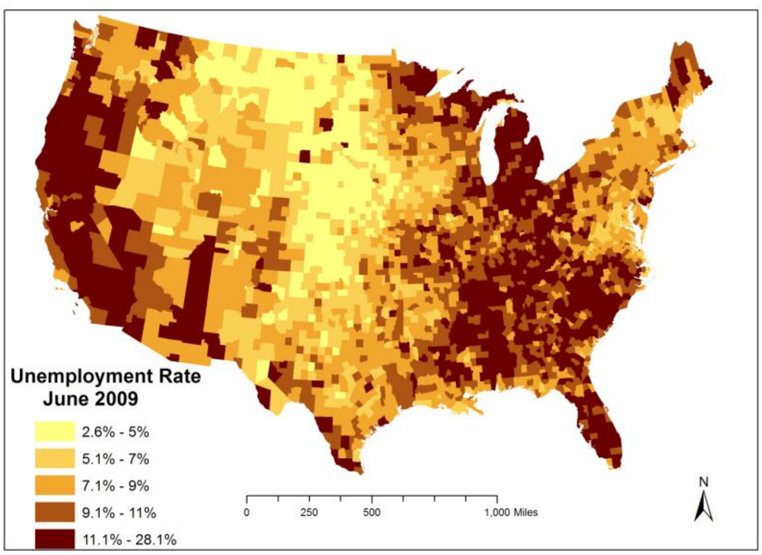

Figure 3. Unemployment rates by county: June 2009. Data Source: St. Louis Federal Reserve Bank (2011).

<u>Hypotheses</u>

This research examines the geographical patterns of county unemployment rates at the end of the past two economic recessions. The hypotheses tested herein are given below:

H1: Specific geographic clusters of high (and low) unemployment rates persist at the county level for both study periods (June 2009 and November 2001).

H2: Place-specific differences in race and ethnicity, educational attainment, and levels of self-employment, are hypothesized to influence unemployment rate variations at the county level.

H3: Geographic location is a significant determinant of unemployment rate variation at the county level.

H4: Industrial mix has a significant effect upon unemployment rate variations. Specifically, whether or not a county is manufacturing-dependent or farming-dependent has a positive effect upon unemployment rate levels.

H5: State-level factors have a strong influence upon county-level unemployment rate variation.

H6: For the June 2009 study period, there is a positive relationship between unemployment rates and whether or not a county is located in one the top states for subprime mortgages.

Definitions

Before proceeding further, some key definitions are provided relating to the theory and methodology which drives this research. This study uses three main theoretical constructs to guide its analyses: *agglomeration economies*, *industry structure* and *path dependence*. Frenken et al. (2007) define *agglomeration economies* as "economies from which a firm can benefit by being located at the same place as one or more other firms." Industry structure simply refers to the number, type and size distribution of firms in an industry. *Path dependence* is what occurs "if the outcome in any period depends on history" (Page, 2006) which is essentially a "dynamic-property allocative processes" (David, 1997).

Methodologically, this paper implements a four-step analysis using the statistical techniques of *spatial autocorrelation, spatial regression, multilevel modeling* and *binary logistic regression*. Getis (2007) provides a thoughtful analysis of the term *spatial autocorrelation*: "(while) correlation statistics were designed to show relationships between variables, autocorrelation statistics are designed to show correlations within variables, and spatial autocorrelation shows the correlation within variables across space." A "*multilevel model* " is a statistical model applied to data collected at more than one level (individual and group) in order to elucidate relationships at more than one level" (Luke, 2004). With regard to this paper, explanatory data will be collected at both the individual county level as well as the group state level, since counties are nested within states. *Spatial regression* (otherwise referred to herein as a spatial lag or spatial error model) is simply a regression model which adds a spatially lagged version of the dependent variable as an explanatory variable within the model to capture the unknown effects of geographic location. The inclusion of this spatially lagged variable often captures the unknown effects of geographic location within the model, thusly reducing, if not eliminating, correlation within the residuals. Finally, *binary logistic regression* is a form of regression that is used when the dependent is a dichotomy and the independents are continuous variables, categorical variables, or both (Suzen & Doyuran, 2004).

Methods

The four research questions proposed for the study are related to Q1) the spatial patterning of county unemployment rates, Q2) the determinants of county unemployment rates, Q3) the influence of state-level groupings upon county unemployment rates, and Q4) the determinants of upper quartile counties as measured by their unemployment rate. To help answer these four questions, a four-step methodology is implemented, using a different statistical technique for each specific question: spatial autocorrelation (Q1), spatial regression (Q2), multi-level modeling (Q3), and binary logistic regression (Q4). After a brief discussion of the variables to be used in the proposed research, a summary of these methods follows.

The dependent variable for the spatial regression and multilevel models is the individual unemployment rate for each of the 3,104 contiguous U.S. counties coinciding with the two study periods (the ending months of past two economic recessions: (November 2001 and June 2009). The dependent variable for the binary logistic regression is a binary variable indicating membership in the upper quartile of counties as measured by the unemployment rate for the June 2009 study periods (0 = no, 1 = yes).

According to Elhorst (2003), who has extensively surveyed the unemployment literature, some of the independent variables commonly used to explain unemployment rate differentials include industry mix, human capital, race and socioeconomic. The descriptive statistics for the variables in the spatial regression, multilevel modeling and binary logistic regression models used are more fully described in Chapter 3. The data for the study was collected from numerous sources including the Bureau of Labor Statistics (BLS), the Bureau of Economic Analysis (BEA), the 2000 decennial census of the U.S. Census Bureau, the American Community Survey of 2005-2007 (ACS), the Economic Research Service of the United States Department of Agriculture (ERS/USDA) and the Federal Reserve Bank of St. Louis.

Spatial Autocorrelation and Clustering Analysis

Step One of the research design is to analyze county unemployment data using spatial autocorrelation techniques, including a Local Indicator of Spatial Association (LISA) analysis, in order to identify spatial patterns and clusters of high (and low) county unemployment. Spatial autocorrelation and Exploratory Spatial Data Analysis (ESDA), as defined in Anselin (1994, 1999), is a collection of techniques to discover patterns of spatial association (clusters or hot spots). Tests for spatial autocorrelation help identify the geographic dimensions of clustering and measure the extent to which areas with high or low unemployment rates are positioned next to another area in space. Following Anselin (2005) this study uses GeoDa software to measure for clustering and identify cluster locations. GeoDa is a multi-featured software program which performs spatial data manipulation, queries, mapping and statistic analysis, among other functions. The specific techniques used for this research include the configuration of a global Moran's I (Moran, 1948) to measure the extent of unemployment clusters, a Moran Scatterplot (Anselin, 1996), and a local univariate LISA (Local Indicator of Spatial Association) analysis (Anselin, 1995) to identify unemployment cluster locations. The operations are performed in this respective order

Spatial Regression

Step Two of the research design utilizes spatial regression techniques to examine the determinants of county unemployment rate distribution. Elhorst (2003) reminds us that a regional unemployment rate model that does not account for serial (time) and spatial autocorrelation when exploring the determinants of unemployment rate levels may be significantly mispecified. Most previous empirical U.S. studies use traditional oridinary least squares (OLS) regression analysis to discern unemployment rate determinants. However, OLS regression analysis assumes that variables are independent and randomly-distributed, an unlikely premise when dealing with regional unemployment rate differentials. In order to investigate the determinants of a spatially dependent variable, OLS techniques must be supplemented by corrective spatial modeling. As described by Anselin (1988), there are two models available to regress for spatial dependence. Known as the spatial lag and spatial error models, they are spatial autoregressive techniques for revealing for the existence of spatial dependence within sample data. The spatial lag model is used to examine the 'existence and strength of spatial interaction' which may be present due to diffusion or spillover effects, while the spatial error model tests for 'biasing influences due to the use of spatial data' (Anselin, 1988). If spatial lag is ignored, then the OLS model will have an omitted variable problem and the estimates will be biased and inconsistent. Spatial error refers to spatial effects in the residuals, or unexplained variation, external to the behavior process, such as data collection errors. If spatial error is ignored, the OLS will remain unbiased, but inefficient, and the standard errors and t-tests will be biased (Anselin, 1988).

Multilevel Modeling

Step Three of the research design implements multilevel modeling techniques to answer the question of how state-level factors influence county-level unemployment rate distribution. A *multilevel model* "is a statistical model applied to data collected at more than one level (individual and group) in order to elucidate relationships at more than one level" (Luke, 2004). With regard to this paper, explanatory data will be collected at both the *individual* county level as well as the *group* state level, since counties are nested within states. Multilevel modeling is a generalization of regression methods using nested data structures (hierarchical) which, in this study, allows for the analysis of individual county variables that are nested within state groupings. By grouping individual county unemployment rates by state, the proposed research can take advantage of what is known about the average county from a particular state, and analyze it along with the actual data collected for each individual county. The advantage of multilevel modeling is that "it can be used for a variety of different purposes, with far fewer and less restrictive assumptions than traditional regression methods" (Micceri, 2007). The basic notion of the multilevel model used for this paper is that unemployment rate variations across counties are a function of their *individual* demographic traits, human capital factors, and industrial mix, as well as the *group* effects of geographic state-level location.

Binary Logistic Regression

Using binary logistic regression, Step Four of the research design attempts to predict whether or not counties will belong in the upper-quartile of U.S. counties based on their unemployment rate (1 = yes, 0 = no). Binary logistic regression is useful for predicting categorical outcomes and has the advantage of not requiring as many assumptions as does OLS. It is a suitable method for this study since it does not assume linearity, homoscedasticity, or normally distributed variables (though it is sensitive to high correlations among the predictor variables, referred to often as multicollinearity) (Pallant, 2007). This is an important relaxation of statistical requirements for this study since the dependent variable indicating whether or not a county is among the top 25% of all counties in terms of the unemployment rate, will naturally produce correlated errors.

Limitations of the Study

The primary focus of this paper is to examine the role of geography on unemployment rate differentials across U.S. counties at the end of the past two recessions. This paper strives for methodological consistency in the presentation of the empirical models contained herein. Due primarily to limitations regarding data availability across all areal units and time periods, minor differences amongst the models do exist, which are noted when they are presented. And while the role of geography is examined by presenting explanatory models of unemployment rate differentials, the explanatory power of these models is deemed subordinate to the primary goal of testing for the role of geography. In other words, the goal of this paper is not to maximize R squared amongst the models when analyzing the determinants of unemployment rate variations but, rather, to examine whether spatial location increases explanatory power. Also, the results of this study should not be interpreted as a proxy for unemployment rate behavior that is specifically characteristic of recession years. I do not test and do not know whether similar studies would find differently or similarly while researching unemployment rate variations during expansionary years. While consistency amongst the models is strived for, some of the independent variables are derived from data with vintage years up to three years lagged from the recession under study. Finally, this paper is broad in scope in that it covers the entire United States. While many explanatory variables are utilized in the models presented, the study's scope and purpose places limits upon the depth of analyses. For instance, industrial mix NAICS data is not available for *every* county, so a binary variable from the ERS/USDA, showing whether a county is dependent upon manufacturing, farming or government employment sectors, is used as a proxy. It is hoped that any results presented herein may assist future research, particularly studies focusing upon case specific studies of a more localized nature.

Organizational Structure

The rest of this paper is structured as follows. Chapter 2 reviews the relevant literature concerning industrial structure, path dependence, agglomeration economies, globalization, and the recent strand of global spatial autocorrelation correlation studies relative to unemployment, before summarizing the explanatory variables covered in the literature. Chapter 3 describes the data and research methodology used for this study.

Chapter 4 presents the results of the four-step methodological inquiries, while Chapter 7 summarizes the results and findings of the study.

CHAPTER 2

LITERATURE REVIEW

Introduction

In order to provide a background for the theoretical explanations of uneven regional development and, hence, unemployment rate distribution, this literature review begins with a discussion of some of the topics germane to economic geography, namely, agglomeration economies, path dependence, and regional industry structure. Later, I cover some of the U.S. regional and global unemployment studies. I end this chapter with a discussion of some of the explanatory variables used to explain regional unemployment disparities.

Path Dependence

As Krugman (1995) has noted, economic development does not evolve evenly over space. The concept of path dependence is utilized within economic geography to help describe the evolutionary process of unequal economic development between regions. Path dependence refers specifically to the inability of a region to escape from its past (Martin & Sunley, 2006). Its conceptual importance pertaining to regional unemployment is that it provides a theoretical framework in which to examine regions with an eye towards their history, in order to explicate causality. As Essletzbichler & Rigby (2007) explain, path dependence "illustrates how established industrial regions can become locked into rigid trajectories because their techno-industrial legacy of the past . . .

has eroded or weakened their ability to adjust to new technology." They further explain that path dependence, unlike mainstream economics with its mathematical models and ahistorical equilibrium assumptions, provides a framework for assessing long-term economic outcomes with the view that there is no contextually independent outcome.

Examples of path dependence abound. An easily recognizable example is how the Detroit area came to be the center of domestic car production within the United States. Henry Ford was born just outside the city of Detroit, so when he launched his auto empire there in 1901, it was by the accident of birth, or path dependence, that the auto industry began in Detroit. This historical decision on the part of Ford echoes throughout Detroit's history even now, particularly now. However, it should be noted that while path dependence helps explain how the auto industry started in Detroit, and why other possible paths were altered because of this auto-centric trajectory, more theory is needed to explain Detroit's development into an auto/industrial complex. As proponents are quick to declare, path dependence is not historical determinism; choices are made along the way (Wilsford, 1994). This is where the theory of agglomeration economies helps to fill the theoretical gaps.

Agglomeration Economies

The starting point for understanding regional differences in unemployment is with a theoretical review of how regions develop economically. The term used within the discipline of geography to describe the clustering processes underlying regional development, including the evolution of industry clusters and aggregate activity within metropolitan areas, is known as agglomeration economies. Agglomeration economies operate at levels of scale ranging from local to regional in context. Within the literature, its origin is usually traced to Alfred Marshall. Marshall (1920) describes the forces leading to the concentration of industry within a geographic region. He cites three external economies resulting from the clustering of firms around specific locations: (1) lower transportation and transaction costs arising from improved access along the supply chain; (2) access to a large and qualified labor pool, and a shared public infrastructure; and (3) knowledge and innovational spillovers resulting from the increased informational exchange between persons within close proximity of each other.

Marshall's work has proven quite durable over the years and remains quite useful for explaining industrial spatial structure and the unevenness of economic development across regions. Over the years, definitional extensions and changes have been made to Marshallian theory (Feser, 1998). Hoover (1937) makes the distinction between localization and urbanization economies, where localization refers to the economies benefiting local firms within the same industry, while urbanization economies benefit all firms within a region. Furthermore, not all of the Marshallian tenets are universally accepted as absolute, for instance, Breschi and Lissoni (2001) have lamented the vagueness of the term knowledge spillovers. No matter the conceptual divisions, the overall usefulness of Marshallian theory in explaining the effects of agglomeration economies upon the location of economic activity is tangible. As Rosenthal and Strange (2004) point out, 75% of Americans within the contiguous U.S. live in cities whose borders account for only 2% of the land area. This aggregation of industry, capital and labor is punctuated by the industrial concentrations found within numerous regions, for example; furniture in North Carolina, high-tech in Boston, software in Silicon Valley, wine in Napa Valley, and autos in Detroit, to name but a few. Understanding the existence and location of these clusters is of paramount importance to the analysis of the models presented in this paper.

Historically, these agglomeration forces exert influence on the economic development of regions. The relevance of agglomeration theory to contemporary studies of regional unemployment disparities is highlighted by research conducted by Desmet & Falschamps (2005). In a study of the spatial concentration of employment in U.S. counties between 1970 and 2000, they examine the spatial development of the non-service and service sectors. They find that while economic activity remains concentrated at the aggregate level, the non-service sector, particularly manufacturing, has spread out, while service jobs have become increasingly concentrated, confirming that the U.S. economy is being driven by the service sector rather than manufacturing. Agglomeration theory is useful for understanding this transformation from a manufacturing economy to a service economy. The explanation by Desmet & Falchamps (2005) for this transformation goes as follows. Manufacturing, which began clustering in the late 1800s, began to de-concentrate in response to falling transportation costs and rising land costs. As globalization became reality, manufacturing, which is a land-intensive use, took advantage of the lower costs for land, and labor, in the hinterlands. Services, which were formerly spread out, took advantage of the lower transportation costs and began to agglomerate. In fact, according to Kolko (1999), the high-tech service sector *requires* agglomeration in order to be near a qualified labor pool.

Scott and Storper (2003) explain agglomeration economies within the context of development theory in order to help us better understand how regional economic differentials develop and persist. While admitting the role of macroeconomic factors in the development process, they assert that the development process is also "strongly shaped by processes that occur on the ground, in specific regions…characterized by significant variations in the intensity and character of economic order from one place to another" (Scott & Storper, 2003). Cities, for instance, often benefit from their "capital-intensive infrastructure" which create important economies of scale. Scott and Storper also list three other benefits of agglomeration: 1) the backward and forward linkages of firms, 2) the development of dense local labor pools, and 3) the learning and innovation effects which evolve from such interconnectedness. The authors take issue with neoclassical development theories which hold that inter-regional economic differences tend to dissipate in time through convergence. Divergence, they assert, is an oft-found process leading to uneven spatial economic development.

Industrial Structure

The concept of industrial structure begins with some elementary geographic concepts of spatial development as developed by Cronon (1991). First-nature geography, he states, is concerned with the physical geography of an area, and, in many cases, can be considered a determinant of the location decisions made regarding early economic centers (for example, the initial development of St. Louis or Cincinnati as river trading posts). However, as time evolves, first-nature geography quickly exhausts its explicative power relative to industrial development. Consider the scores of initial settlements along

Lake Michigan, yet only Chicago grew into an economic powerhouse. Second-nature geography carries on where first-nature leaves off by concerning itself with the actors and agents involved with each other in geographic space.

Using Cronon's concepts as a jumping off point, I borrow from Krmenec & Esparza (1999) to provide a brief history of U.S. industry structure in order to explain the causality of divergent economic regions. They argue that the diffusion of industrialization, which had spread to the Northeast and Midwest regions of the United States by the mid 19th century, did not take hold in the southern and western states. They cite Meyer (1983) to explain that the lack of population density in the rural south, and the dominant market mechanisms already in place in the north (which more than met unmet demand for goods in the south), prevented the diffusion of industrialization to the south. The location of American manufacturing in 1929 is illustrated in Figure 4. Returning to Krmenec & Esparza (1999), the mountain ruggedness, expansive aridness, and lack of navigable waterways prevented industrialization from taking hold in the western U.S. as well. As the country grew, the lack of industrialization within the West and South served to amplify the industrial significance of the Northeast and Midwest, and launched their trajectories beyond mere agrarian and mercantile functions. The industrial complexes of the Midwest and Northeast, with their dependence upon manufacturing, continued to flourish for decades until reaching their apex during the Fordist period of the 1940s-1960s. What followed, is best described by Thomas Lassman:

> By the 1970s, the American economy was in the midst of a wrenching transformation that eviscerated once-venerable manufacturing industries on a

scale not seen since the Great Depression. The extent of the wreckage was unprecedented, as Pittsburgh, Buffalo, Detroit, Baltimore, and scores of other communities across the country experienced plant shutdowns and massive employee layoffs. No longer able to compete effectively in an increasingly global economy dominated by more nimble foreign firms, American producers of steel, automobiles, and other capital-intensive goods closed aging factories and shifted their resources to new locales outside the Rust Belt. (Lassman, 2005)

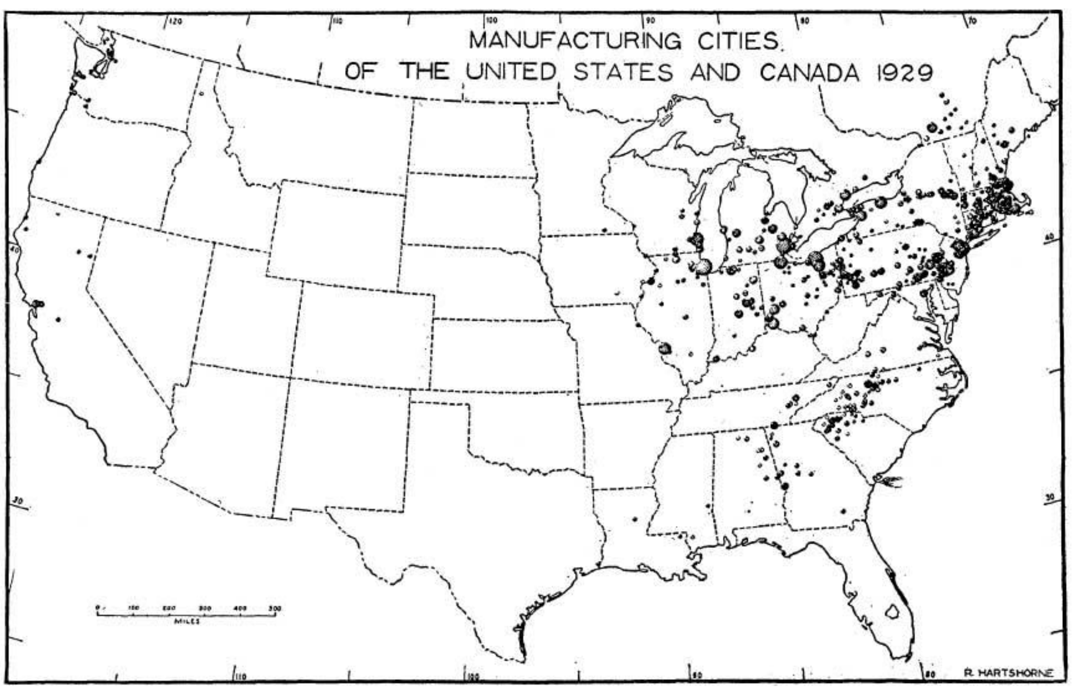

Figure 4. Manufacturing Cities, 1929 (Hartshorne, 1936).

Deindustrialization had taken full hold by the 1980s. Green & Sanchez (2007) attribute the effects of globalization, technological change, and the decline of the unions as the primary determinants of deindustrialization. Many firms in the North either closed, or moved to the South in an effort to remain competitive. Commonly used reasons offered for the decline of the Northeast and Midwest in favor of the Sun Belt include the lack of unions, lower wages, weather, and business climate (Crandall, 1988). However, there is evidence that the spectacular economic growth enjoyed recently by the Sun Belt states may soon end. Caselli and Coleman (2001) document how relative wages in the South, which in 1900 were less than half that of wages in the Midwest and Northeast, had converged to near parity by 1980.

Historically, as noted by Green & Sanchez (2007), manufacturing had provided the middle-class opportunities for upward mobility, and had supported agglomerated economies with its multiplier effects. While cities may have morphed from manufacturing centers to service centers, Green & Sanchez (2007) remind us that the manufacturing sector still pays an important role within the U.S. economy. Further light on this topic is shed by Hanson (2001) when he states that while the death of manufacturing plants may feed the decline of a particular region, the birth of new plants help recreate a new cluster in a different region, proving that second-nature agglomeration processes have replaced a region's first-nature endowments as the primary determinant of firm locational choice.

This literature review has just concluded a discussion about the concepts relevant to economic geographers, such as path dependence, agglomeration economies and industrial structure. These provide the theoretical lenses with which to examine the geographical differences of regional unemployment rates. This paper attempts to explain regional variations in unemployment rates within the framework of the historical processes and agglomerative conditioning effects described herein, without adopting a fully deterministic approach to either. This discussion now turns to the unemployment studies, both regional and global, found within the literature.

U.S. Regional Unemployment Studies

Elhorst (2003) surveys the unemployment literature and lists and reviews the variables being used to explain regional unemployment. A discussion of those variables, and others found in the literature, appears later in this chapter. However, this portion of the literature review begins by specifically exploring the U.S.-based studies examined by Elhorst.

Of the 41 empirical studies surveyed by Elhorst (2003) where unemployment was explained with the help of regional data, 16 of the studies cover U.S. unemployment. Most of the remaining studies covered by Elhorst are from Europe. Of the U.S. studies, nine were conducted at the state level of analysis, four at the metropolitan level, and three were localized studies. The publication years of these 16 studies ranged from 1976 to 1996.

One of the state-level studies reviewed by Elhorst (2003) was Summers (1986), which uses dummy variable groupings corresponding to the nine U.S. Census divisions. He found that much of the difference in regional unemployment rates was the result of industrial performance rather than industrial composition differences. However, there are two points to be made about this study. First, he did indeed find significant differences between states for manufacturing sector employment, only he found more significant differences for industrial performance than industrial composition and thus emphasized the former as a determinant. Second, his thirty-year study period, 1954 to 1985, was analyzed using a time-series approach that may have underestimated the erosive effects of manufacturing job losses as service jobs were created to help replace them. Indeed, manufacturing employment viewed in a particular snapshot of time, as in this study, might well yield more conclusive results relative to spatial unemployment rate variations. As with Summers (1986), this study uses the nine U.S. Census divisions in its Step Four predictive model. It should be noted that Neumann & Topel (1991) studied the determinants of geographical unemployment differentials for the period 1950–1985 and their main finding was that industrial diversification leads to persistently lower unemployment rates. Naturally, theirs was also a time-series-based study.

Vedder & Gallaway (1996) analyze variations in state unemployment rates for the period 1960-1991 and find large differentials that tend to persist over time. They note the fact that West Virginia’s unemployment rate is almost always above the national average, while Nebraska’s is always below it. They explain that these sizable persistent variations amongst states can, depending upon the ascribed theoretical school, be considered evidence of regional variations in the magnitude of path dependence related to historical unemployment rates.

Elhorst (2003) finds many studies that control for path dependence by using the unemployment rate lagged in time as an explanatory variable (Blackaby & Manning, 1992; Blanchard & Katz, 1992, Descressin & Fatas, 1995; Holzer, 1991; Hyclak, 1996; Vedder & Gallaway, 1996). Borrowing from their methodology, this paper uses each county's unemployment rate, measured at the end of the previous recession, as an explanatory variable in the Step Four predictive model. For instance, when examining county-level unemployment rates for June 2009, the unemployment rate for November 2001 is used as a variable in the model to account for persistence.

Most of the previously mentioned U.S. studies use regression analysis to explain regional differentials in unemployment rates. Some of these analyses utilize dummy variable modeling or covariance decomposition to further isolate and explore the effects of geography. However, as proposed by Cliff & Ord (1981) and Anselin (1988), the use of a spatial weight matrix is the preferred method for investigating spatial dependence and heterogeneity between regions. Such spatial modeling allows researchers to better gauge geographical influences from neighboring areal units. The aforementioned U.S. studies represent significant efforts to determine the causes of regional unemployment rate variations. However, in the 21st century, the main body of literature testing for spatial dependence of unemployment rate variations primarily comes from outside the United States.

Global Spatial Autocorrelation Studies

While this paper examines U.S. data only, the following strand of global unemployment literature is included because their methodologies and analyses offer insights into the geographical examination of regional unemployment rates. Technological advances in the development of software able to test for spatial autocorrelation and conduct spatial econometric analyses (as with the Geoda software used within this paper) has led to more extensive studies focusing upon the role of geography in regional unemployment variations.[4]

There are an increasing number of studies using spatial data and spatial regression techniques to investigate regional unemployment rate differentials (Aragon et al., 2003; Cracollici, Cuffaro, & Nijkamp, 2007; Lopez-Bazo, del Barrio, & Artis, 2002; Niebuhr, 2003; Mitchell & Bill, 2004; Patacchini & Zenou, 2008). Lopez-Bazo et al. (2002) analyze the spatial dimension of unemployment distribution in the Spanish provinces for the years 1985 and 1997. Their study uses employment growth, net migration, unit labor costs, industrial mix (percent share of agriculture and manufacturing), educational attainment, age structure (working age population 16-25), and gender as explanatory variables. They found that spatial effects upon unemployment play a role for both years, but that the determinant variables change for the two different time periods. For 1985, the significant determinants of unemployment were unit labor costs, industrial mix and educational attainment. However, these variables did not help explain unemployment differentials in 1997, where age structure and gender proved to be explanatory. A primary reason for this change in determinants within Spanish regions is that within the 12-year interim of the study periods the Spanish economy underwent significant structural change. In 1985, Spain had just ended a decade of crisis and industrial restructuring, unemployment rates stood at record levels, and the country was a year away from joining the European Union, which brought about market deregulation and

[4] For a brief and concise history of the evolution of spatial autocorrelation within the geographic literature, see Getis, 2007.

reforms. By 1997, the Spanish economy, after enjoying significant growth within the late eighties and early nineties, saw unemployment levels rise to the previously high levels of 1985, thus completing another business cycle. Notably important in this study is that the two study periods both represent the end of a business cycle and therefore, as its authors point out, the research is not tainted by disparate regional responses to the different phases of the business cycle. It is worth pointing out here that the two study periods examined within this paper also correspond to the end of a business cycle.

Niebuhr (2003) analyzes the spatial association of regional unemployment for European countries from 1986 to 2000. Using measures of spatial autocorrelation and spatial econometric methods, he finds that regions with high unemployment levels tend to cluster in space, as do regions with low unemployment levels. Niebuhr relates his findings to those of Overman & Puga (2002) who had previously concluded that the unemployment rates of European regions are much closer to the rates of adjacent regions than to the average rate of other regions within the same EU country because of polarization effects driven by employment changes. These polarization effects refer to the phenomenon where regions with high (or low) initial unemployment rates undergo little change while nearby regions with initial mid-level rates of unemployment move towards the extreme values of their neighbors. Overman & Puga (2002) attribute this polarization effect, whereby neighboring regions share similar outcomes, to spatially related changes in labor demand, which is partly due, in the case of high unemployment regions, to initial regional clustering of low-skilled labor pools and uncompetitive industries. Importantly, even after controlling for labor skill and industry mix, this neighboring effect remains, which shows the importance of geographical location as a determinant of unemployment rate differentials. The authors attribute the significance of geographical location to the agglomeration effects of economic integration.

.In a similar study, control variables used by Niebuhr (2002) to study regional unemployment include employment growth, population density, industrial sectoral composition (manufacturing and services) and country dummies. Country dummy variables are used to allow for country-specific labor market regulations, labor supply effects, and generalized conditions affecting the demand/supply matching process within each country's labor market. The results show that across the 359 European regions studied, substantial neighboring effects exist and that changes in regional unemployment levels are associated with the concentration of labor market problems in spatial clusters. Niebuhr explains that this geographical concentration corresponds to the polarization processes found by Overman & Puga (2002). All explanatory variables were significant the 5% level, with manufacturing, services and employment growth having negative signs while population density was positive.

Aragon et al. (2003) use data for 174 French districts and build two spatial regression models to examine whether unemployment rate differentials are the result of equilibrium or disequilibrium factors. The equilibrium view of unemployment holds that substantial geographic differentials in unemployment will exist even after the effect of short-run shocks have faded, a concept commonly referred to as the natural rate of employment. Alternatively, the disequilibrium view holds that geographic unemployment rate differentials will dissipate over time. Similar to this study, their methodology is taken from Anselin (1988) and relies upon standard linear regression models to capture spatial dependence by use of a spatial lag model or a spatial error model. Spatial lag models are appropriate for assessing the strength of spatial interactions, while spatial error models are used for correcting the potential biasing influences of spatial autocorrelation, due to the use of spatial data.[5] The results presented by Aragon et al. (2003) suggest that rising unemployment in one region spreads quickly to its neighbors and are largely supportive of the equilibrium view of unemployment rather than the disequilibrium view.

Mitchell & Bill (2004) analyze metropolitan and nonmetropolitan labor markets in Australia for the period 1991-2001. Motivated by the persistence of regional unemployment rate levels since the early 1990s, they use measures of spatial autocorrelation and spatial econometric techniques to analyze the spatial dependence of regional unemployment rates in Australia. They find that the geographic distribution of Australian unemployment has become more clustered in the latter half of the 1990s, due partly because most regions began the study period with high unemployment and the amelioration provided by the growth processes of the mid-nineties did not take effect until later, which made levels of spatial dependence more pronounced. Even after controlling for population composition and economic growth rates, they conclude that significant spatial interactions occurred between regions in regard to unemployment levels and that these spillovers magnify regional responses to national economic shocks.

---

[5] This paper incorporates both types of models and a more in-depth discussion of these methods appear in Chapter 3.

Beyond allowing for model misspecification, the authors provide no explicative basis for these spatial interactions, suggesting that the causes require further examination via econometric techniques.

Cracollici et al. (2007) use spatial econometric methods based on spatial autocorrelation techniques to explore the geographical distribution of unemployment rates for 103 Italian provinces for the years 1998 and 2003. Similar to Niebuhr (2003), the results show that both high and low unemployment areas tend to cluster together indicating unemployment persistence across space and time. Furthermore, this clustering is related to increasing wage gaps. Among the independent variables used for their initial statistical regression model include were industrial mix (percent shares of manufacturing, agriculture and services) and educational attainment of working age population, which are also used in this study. They found that regional differences in unemployment rates were related to employment variables rather than demographic variables, leading to their conclusion that unemployment differentials were mainly due to labor demand.

Explanatory Variables Associated with Regional Unemployment

According to Nistor (2009), the best approach for modeling regional unemployment rate determinants is to group theoretical model variables with those used within applied research. Some of the applied variables she includes for consideration include population birth rates, age structure, educational attainment, migration, commuting, wages, government social security policy, and population characteristics (e.g., race, ethnicity, age). In deciding which explanatory variables to include in this study, Nistor's work is a good starting point. However, within the literature, the work of Elhorst (2003) best analyzes and summarizes the explanatory variables used to explain regional unemployment differences. Yet, ultimately, the availability of the data across all 3,104 counties was a primary factor in deciding which data to include in the analyses.

One of the main goals of this research is to present a useful regression model to explain unemployment rates. Most of the variables chosen for this study are ones commonly found in both the U.S. studies of regional unemployment, and the European spatial autocorrelation studies. They include, but are not limited to, industrial mix, educational attainment, unemployment lagged in time, average wages (per capita income), and ethnicity and race. However, due to data collection or collinearity issues, not all of the variables presented in this section are used in the empirical regression models used in this study. A description of potential explanatory variables and their place in the literature follows (names and expected signs in parenthesis)

Race and Ethnicity

The issue of race and unemployment is well represented in the literature. Using nearly a hundred years of census data, Fairlie & Sundstrom (1999) found that an unemployment gap between whites and blacks exists and that it can be narrowed by educational attainment or increased by regional economic shifts, such as decreases in the demand for less skilled labor. Hyclak & Stewart (1995) found that the unemployment rate for blacks is more responsive to demand growth than that of whites, inferring that business cycle shocks vary across demographic groups. In his examination of employment opportunities in Cleveland, Ohio, Kaplan (2000) found that residents of low to moderate income black neighborhoods have fewer job opportunities than residents of predominately white neighborhoods. In the U.S. it is commonly known that the unemployment rate for young black males is considerably higher than those of their white counterparts. Ewing et al. (2005), among others, confirm this by finding that the percentage of weeks spent unemployed by blacks, aged 18-34, was more than twice that of similarly aged whites

A study of spatial job search patterns in Los Angeles by Stoll & Raphael (2000) found similar results. Lynch & Hyclak (1984) and Ewing, Levernier & Malik (2005) found that economic downturns adversely affect the unemployment rates of blacks and males more so than for whites and females. They cite differing levels of educational attainment, age structure and discrimination as possible reasons for a certain group to face unemployment difficulties during a recession. Baskaya & Mbiti (2006) found that the employment probability of blacks is significantly decreased if there was a recession in the previous year, and that blacks are 150 percent more likely to endure higher unemployment than whites during a recession. Using wages as a measure of productivity, they explain that these results are not due lower productivity levels for blacks, but rather the result of discrimination and unknown racial differences in labor supply and demand over the business cycle. In addition to blacks, unemployment studies show that Hispanics are also at a disadvantage. Blau and Ferber (1992) and Clogg and Sullivan (1983) both found these two groups to have higher unemployment problems. Asians seem to fare better in this regard. De Jong & Madamba (2001) found that Asian males were less likely to be unemployed compared with non-Hispanic white males, but more likely to be underemployed. For these reasons, a variable is included within this study which gauges the percentage of the non-white population (NONWHT, +).

Path Dependence Variable

Path dependence, referred to as hysteresis in the economic literature, refers to the long-lasting influence of history upon the natural rate of unemployment (Blanchard & Summers, 1987). In the 1990s, alternative unemployment theories were developed enveloping the notion that the equilibrium unemployment rate depends upon the history of the actual unemployment rate. Song & Wu (1997) and Roberts & Morin (1999), among others, have found that path dependence is unimportant in the long term when predicting US unemployment rates. However, Clemente, Lanaspa, & Montanes (2005) acknowledge that the debate over the importance of path dependence is far from settled and highly theoretical. The path dependence variable used within this paper is sometimes referred to as unemployment rate lagged in time (UR_MONYY, +). This variable, used in the empirical models presented later in this paper, is calculated using a county's previous unemployment rate from the end of the most recent recession. For instance, the June 2009 models use the unemployment rate as of November 2001, while the November 2001 models use the unemployment rate as of March 1991. Elhorst (2003) finds many studies that use the unemployment rate lagged in time to explain itself (Holzer, 1991; Blackaby & Manning, 1992; Blanchard & Katz, 1992, Descressin & Fatas, 1995; Vedder & Gallaway, 1996; Hyclak, 1996). Naturally, the relationship is positive.

Educational Attainment

Educational attainment is a classic explanatory variable typically found in unemployment rate studies. Theoretically, the relationship is expected to be negative and empirical studies confirm the significance of this relationship (Burridge & Gordon, 1981; Holzer, 1993; Malizia & Ke, 1993; Partridge & Rickman, 1995; Siegers, 1983; Simon, 1988). Better educated persons are not restricted by geography, while a less educated person will be uncompetitive both at home and in another labor market, and will therefore tend not to migrate for employment opportunities. Better educated persons are also able to take advantage of technology, conduct more efficient job searches, and are less prone to layoffs. Finally, better educated workers are better able to search for new opportunities while still employed. For these reasons, educational attainment is considered to have a significant negative relationship to unemployment. The data used in this study for educational attainment represent the percent of people aged 25 or over who possess either a bachelor’s, master’s, doctoral or professional degree (EDUCYY, -).

Self Employment

Golpe & van Stel (2008) investigated the relation between self-employment and regional unemployment in Spain from 1979-2001. They find that higher levels of self-employment have a negative effect on unemployment and their results were empirically supported both in higher income and lower income regions. Audretsch, Carree, Thurik, & Van Stel (2001) also find that self employment lowers the unemployment rate. They elaborate that the business cycle affects the level of entrepreneurial activity on two levels. The growth opportunities which accompany economic upturns induce people into self-employment, and the unemployed will turn to self-employment in economic downturns simply for reasons of sustainability, creating what is termed both a "prosperity-pull" and a "recession-push" effect. This papers gauges self-employment using a percentage of households with self-employment income (SELEMPYY, -).

Wages/Per Capita Income

According to Elhorst (2003), the effect of higher wages on unemployment is often found to be negative within the empirical studies, a result which differs from theoretical expectations. In his literature survey, he found many studies which used a wage variable to explain unemployment (Blackaby & Manning, 1992; Burridge & Gordon, 1981; Gripaios & Wiseman, 1996; Murphy, 1985; Hyclak & Johnes, 1987; Hofler & Murphy, 1989; Molho, 1995a, 1995b; Partridge & Rickman, 1995). The effect of wages upon unemployment is expected to be positive because as wages increase, labor supply increases and labor demand decreases. However, of these nine studies, only Molho (1995a, 1995b) found a positive relationship between wages and unemployment, and it was insignificant. This paper uses per capita income (PERCAP, +) to explore this variable's effect upon U.S. unemployment rates.

Industrial Mix

Regions with high levels of declining industries, like agriculture and manufacturing, are expected to exhibit higher levels of unemployment. Unfortunately for the researcher, this can be a misleading assumption. As noted in Elhorst (2003), the industrial mix variables used in many empirical studies, even when calculated using shift share analysis or with more complex formulaic expressions of industry mix, have little or no explanatory power (see Dixon & Thirlwall, 1975; Summers, 1986; Taylor & Bradley, 1983). While exceptions do exist within the literature (Hyclak & Johnes, 1987, and Lopez-Bazo et al., 2002, to name a few), the reliability of industrial mix as an explanatory variable for regional unemployment rate differentials is often dependent upon finding the right industry matched with the right region with the right time period. To better understand the complicated mechanisms at work underlying the use of industry mix as a variable, consider the following hypothetical situation. An area undergoes a significant loss of agricultural and manufacturing jobs, while at the same time enjoying a sizeable increase in its services sector. Within this scenario, it is possible; perhaps likely, that the resultant structural shift will produce a correlation linking the higher share of services with higher unemployment levels.

To summarize the findings of Elhorst (2003), industrial mix matters, but not by the traditional sense of employment shares. Lopez-Bazo et al. (2002) found explanatory values in manufacturing and agriculture shares relative to regional unemployment levels. Further justifications for the use of industrial mix variables are twofold. First, despite its spotty explanatory value within the empirical literature, manufacturing share is not even close to being dismissed from the theoretical realm. Second, the growth in U.S. service sector employment requires examination and since this paper analyzes two cross-sectional time periods at the county level, it is not unrealistic to expect industrial mix in general, and the percentage employment share of manufacturing, farming or services in particular, to have an effect in at least one of these models.

Unemployment Lagged in Time and Space.

Elhorst (2003) points out that regional unemployment rates are highly correlated in time and in space, changing only in small amounts and usually within the same direction, and that any regional unemployment rate model built without controlling for serial and spatial effects may be seriously misspecified. As described earlier in this section, this paper uses unemployment lagged in time as a variable (UR_MONYY). Controlling for unemployment lagged in time, or serial autocorrelation, is mostly a statistical matter. The test for serial autocorrelation is called the Durbin-Watson statistic and is easily calculated within any SPSS regression run. Testing for unemployment lagged in space, or spatial autocorrelation, is the methodology which brings geography to the forefront of this study. Techniques such as cluster analysis and spatial autoregressive modeling help determine the effects of geography upon unemployment rate levels in the United States. These methodologies are elaborated upon in great detail later in this study.

Conclusion

This literature review has covered some key theories in economic geography which help provide the overarching framework in which to examine the spatial patterning and determinants of U.S. county unemployment rate variation. Specifically, the concepts of path dependence, agglomeration economies, and industrial structure will be relied upon to explain the results of this study. Additionally, this section has covered both the global and regional unemployment studies which have served to drive this paper's methodology. Finally, this section has described some of the explanatory variables used in the literature to help explain unemployment rate variation, specifically: race and ethnicity, path dependence, educational attainment, per capita income and industrial mix.

CHAPTER 3

DATA AND METHODS

Introduction

This paper examines regional unemployment rate variations occurring at the end of the two most recent business contractions within the United States, namely, June 2009 and November 2001. Using a four-step research design, empirical models are developed at the county level for both periods. This first part of this chapter summarizes the variables used for the empirical models, identifies data sources, describes the areal units, and discusses data aggregation and other relevant data issues. The second part of the chapter describes the four statistical techniques used to conduct the analysis. Specifically, the techniques of spatial autocorrelation, spatial regression, multilevel modeling and binary logistic regression are presented.

Dependent Variables

The dependent variable for the spatial regression and multilevel models is the county unemployment rate for the two study periods, June 2009 and November 2001, respectively. These data are provided by the Bureau of Labor Statistics and were downloaded from the Federal Reserve Economic Data (FRED) website of the Federal Reserve Bank of St. Louis[6]. County unemployment rates are not reported on a seasonally

[6] Website accessed December 20, 2010 at: at http://research.stlouisfed.org/fred2/

adjusted basis. The dependent variable for the binary logistic models is a dichotomous categorical variable (BLRJUN09) and (BLRNOV01) indicating whether a county, as of the June 2009 study period, belonged to the upper quartile of counties as measured by the unemployment rate (0 = no, 1 = yes).

Independent Variables

According to Elhorst (2003), who has extensively surveyed the unemployment literature, some of the independent variables commonly used to explain unemployment rate differentials include industry mix, human capital and race. Industrial mix data is normally available from the North American Industrial Classification System (NAICS). At the state level, this data set is nearly universally complete across the various two-digit classification sectors used to identify employment figures for manufacturing, farming, government, retail and many other sectors. However, at the county level of analysis, industrial employment data are missing for hundreds of U.S. counties, since the NAICS data set is often truncated due to confidentiality concerns. Across the large scale and scope of this study, using NAICS county data is problematic because the lack of spatial congruity interferes with the creation of the spatial weights necessary to conduct the spatial autocorrelation and spatial regression analyses used in this study. Fortunately, a proximate county representation of the basic industrial sectors (manufacturing, farming and government) is available from the Economic Research Service of the United States Department of Agriculture (ERS/USDA). The ERS/USDA county data set for 2004 provides a separate binary variable for each county indicating whether or not a county's economy is dependent upon the manufacturing (USDAMANF), farming (USDAFARM), or government (USDAGOV) employment sectors (0 = no, 1 = yes). Dependence for each of these three sectors is defined by the ERS/USDA as follows: 1) farming-dependent counties have "either 15 percent or more of average annual labor and proprietors' earnings derived from farming during 1998-2000 or 15 percent or more of employed residents worked in farm occupations in 2000"; 2) manufacturing-dependent counties have "25 percent or more of average annual labor and proprietors' earnings derived from manufacturing during 1998-2000"; and 3) government-dependent counties have "15 percent or more of average annual labor and proprietors' earnings derived from Federal and State government during 1998-2000" (ERS/USDA, 2011).

In addition to these industrial mix variables, certain economic and demographic variables are used in the spatial regression models: per capital income (PCPI_08 and PCPI01), the percentage of the population with a bachelor's degree or higher (EDUC00), the percentage of a county's nonwhite population (NONWHT07 and NONWHT00), and the percentage of self-employed living within a county (SELEMP00).

Data Sources

The data for the study was collected from numerous sources including the Bureau of Economic Analysis (BEA), the U.S. Census Bureau's 2000 decennial census, the 2005-2007 American Community Survey, and the Bureau of Labor Statistics (BLS). County unemployment rate data for all periods were provided by the Bureau of Labor Statistics and downloaded from the Federal Reserve Economic Data (FRED) website of the Federal Reserve Bank of St. Louis. Demographic data was collected from the Census

2000 Summary Files 1 and 3, and the 2005-2007 American Community Survey (ACS).[7] Industrial mix data were obtained from the Economic Research Service division of the United States Department of Agriculture (ERS/USDA). Per capita income and state-level industry mix data used for the multilevel model was obtained from the BEA. The education and self-employment variables were collected from the 2000 decennial U.S. Census. The percentage of the nonwhite population was calculated from data provided by the 2000 decennial U.S. Census for the November 2001 study period, and the 2005-2007 American Community Survey (ACS) for June 2007 period. The per capita income variable was collected from the Bureau of Economic Analysis (BEA) for the years 2001 and 2008. A summary of the variable descriptions and data sources appears in Table 3.

[7] Data from the U.S. Census Bureau and the ACS were accessed online on April 10, 2010 at: http://factfinder.census.gov/home/saff/main.html?_lang=en

Table 3. Variable Descriptions and Data Sources for all Models

Model	Variable Label	Description	Source	Mean	St. Dev.	Min.	Max.
		DEPENDENT VARIABLES:					
SR1, MM	UR_JUN09	County unemployment rate as of June 2009 (official end of 2007-09 Recession per the NBER)	Federal Reserve Bank of St. Louis	9.34	3.31	2.60	28.10
SR2	UR_NOV01	County unemployment rate as of November 2001 (official end of 2001 Recession per the NBER)	Federal Reserve Bank of St. Louis	5.29	1.94	1.40	20.60
BLR	JUN09QTR	Binary variable denoting whether county was in the upper quartile of counties as measured by the unemplyment rate, 0=no, 1-yes	Federal Reserve Bank of St. Louis	5.29	1.94	1.40	20.60
		INDEPENDENT VARIABLES:					
	Industrial Mix						
SR1, SR2	USDAFARM	Farm-dependent county indicator. 0=no 1=yes	ERS 2004	0.14	0.35	0.00	1.00
SR1, SR2, MM, BLR	USDAMANF	Manufacturing-dependent county indicator. 0=no 1=yes	ERS 2004	0.29	0.45	0.00	1.00
SR1	MIXGOV07	Percentage of government sectoral employment of total employment in 2007 (REIS code 2000)	BEA	0.16	0.07	0.03	0.89
SR2	MIXGOV01	Percentage of government sectoral employment of total employment in 2001 (REIS code 2000)	BEA	0.17	0.07	0.03	0.89
	Economic/Demographic						
SR1	PCPI_08	Per capita income, 2008	BEA	33016	8468	12558	140275
SR2	PCPI_01	Per capita income, 2001	BEA	24772	5815	10461	82230
SR1, SR2, MM	EDUC00	Percentage of population with a bachelor's, graduate or professional college degree or higher	2000 Census	0.16	0.08	0.05	0.64
	EDUC07d	Binary variable denoting whether county was located in one of the top ten states as measured by the percentage of the college grads, 0=no, 1=yes					
SR1, SR2, MM	SELEMP00	Percentage of population that is self-employed	2000 Census	0.16	0.07	0.03	0.52
SR1	NONWHT07	Percentage of non-White population in 2007	ACS	0.13	0.15	0.00	0.91
SR2	NONWHT00	Percentage of non-White population in 2007	2000 Census	0.15	0.16	0.00	0.95
	Surrounding Variables						
Model 1	LP1K08D	The top ten states in subprime mortgages for August 2008, 0=no, 1-yes …	Federal Reserve Bank of New York	0.11	0.32	0.00	1.00

Levels of Analysis

The primary unit of analysis for this study was at the county level whereby data for 3,108 contiguous counties, excluding Hawaii and Alaska, were collected at the individual county level. Alaska and Hawaii were excluded from the study since 1) their business cycles differ significantly from other states (Owyang, Piger, & Wall, 2005), and 2) their noncontiguous nature is problematic when creating the necessary weight matrices necessary for spatial analysis. Due to recent county compositional changes, as documented by the Change Notices issued by the U.S. Census Bureau (2009), the following counties are excluded from all county data sets: Broomfield County, CO; Yellowstone National Park, MT; Clifton Forge, VA; South Boston, VA.[8] It should be noted that Dade County, Florida, which officially changed its name to Miami-Dade County effective November 13, 1997, is referenced as such in all county data sets. Although there are 3,108 contiguous counties within the 48 contiguous states, four of these counties are "islands", which are problematic when creating the requisite weight matrices for spatial autocorrelation and spatial regression. As such, the following four counties were also excluded from the analysis: Dukes, Massachusetts; Nantucket, Massachusetts: Richmond, New York; and San Juan, Washington. As a result, the total number of counties analyzed for this study stands at 3,104 counties for both study periods (June 2009 and November 2001). Finally, while the spatial autocorrelation, spatial

[8] County compositional changes are documented by the U.S. Census Bureau and are available from their website at: http://www.census.gov/geo/www/ansi/changenotes.html.

regression and binary logistic models presented herein are conducted at the county level, it should be noted that the multilevel models necessarily include state-level data. Specifically, observations collected for the multilevel analyses include data sets for forty-nine group-level observations, consisting of the forty-eight contiguous U.S. states and the District of Columbia.

Data Limitations and Problems

Some of the variables used for modeling the 2001 recession come from the decennial 2000 U.S. Census and, as such, are not available for the models used to analyze the 2007-2009 recession. Where possible, ACS sample data is used for the 2005-2007 recession models to replicate 2000 Census data. For instance, the percentage of a county's nonwhite population is provided for both the decennial census and the ACS estimate of 2005-2007. However, at the county level, data for the percentage of a county's self-employed population (SELEMP00) and the percentage of a county's population with a bachelor's degree or higher (EDUC00) are only available from the 2000 census and are thus used to model both the June 2009 and November 2001 study periods. Also, the same industrial mix sector-dependent variables gathered from the 2004 vintage of the ERS/USDA data is used for modeling both the June 2009 an November 2001 study periods with the obvious caveat that the county-specific variables provided as of 2004 split the two study years of 2001 and 2007. Finally, there is incongruity between the various vintages of independent variable data and the exact month of the two study periods (June 2009 and November 2001). For instance, per capita income is collected for 2001 and 2008, industry mix is collected only for 2004, education and self-employment data for both study periods come from the 2000 census, while the percentage of the nonwhite population come from both the 2000 census and the 2005-2007 ACS. These facts are ameliorated somewhat by the fact that the vintage years chosen for all explanatory variables correspond as closely as possible to the June 2009 and November 2001 study periods.

The remainder of this chapter outlines the four-step research design used in this study. The four research questions proposed for the study are related to Q1) the spatial patterning of county unemployment rates, Q2) the determinants of county unemployment rates, Q3) the influence of state-level groupings upon county unemployment rates, and Q4) the determinants of upper quartile counties as measured by their unemployment rate. To help answer these four questions, a four-step methodology is implemented, using a different statistical technique for each specific question: spatial autocorrelation (Q1), spatial regression (Q2), multi-level modeling (Q3), and binary logistic regression (Q4).

Step One: Spatial Autocorrelation and Clustering Analysis

Exploratory Spatial Data Analysis (ESDA) as defined in Anselin (1994, 1999), is a collection of techniques to discover patterns of spatial dependence, (i.e., clusters or hot spots). ESDA is a data-driven analysis, not theory-driven, so while it is useful for identify spatial patterns, it does not possess the capability to explain why they occur. Nonetheless, it is useful for identifying patterns which can be later examined through a theoretical lens. Tests for spatial autocorrelation help identify the geographic dimensions of clustering and measure the extent to which areas with high or low unemployment rates are positioned next to another area in space. Following Anselin (2005) this paper uses

GeoDa software to measure for clustering and identify cluster locations. GeoDa is a multi-featured software program which performs spatial data manipulation, queries, mapping and statistic analysis, among other functions. The specific features to be used for this research include the configuration of a global Moran's I (Moran, 1948) to measure the extent of unemployment clusters, a Moran Scatterplot (Anselin, 1996), and a local univariate LISA (Local Indicator of Spatial Association) analysis (Anselin, 1995) to identify high-unemployment (and low) cluster locations. The operations are performed in this respective order.

Following Anselin (2003), the first step in the analysis is to construct spatial weights. GeoDa offers three choices of contiguous weighting; rook, bishop and queens. Each of these different contiguity methods evaluates the spatial characteristics of neighboring areal units in much the same way as their chess position equivalents on a checkerboard. Rook contiguity checks vertical and horizontal neighbors in straight-line fashion, bishop checks only for diagonal relationships, and queen contiguity analyzes both straight and diagonally. This paper utilizes the queen contiguity weighting.

After weighting, Moran's I is calculated to determine geography's effect upon the shape of county unemployment distribution. The Moran's I statistic, a global measure of spatial autocorrelation (clustering), is defined as

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$$

where N is the number of spatial units indexed by i and j; X is the variable of interest; $\bar{X}$ is the mean of X; and w_{ij} is a matrix of spatial weights. Positive spatial autocorrelation is shown with positive values, negative correlation with negative values, and values approaching zero indicate the absence of autocorrelation (O'Sullivan & Unwin, 2003).

The Moran Scatterplot, as suggested in Anselin (1996), is in x-y graph form and shows the regression slope and the quadrant relationships between a variable (the unemployment rate for a given county) and the mean value of that same value's neighbors (spatial lag). Clusters in the upper right quadrant are high in both the variable value and its spatial lag; clusters in the lower left are low in both values. Both the upper right (high-high relationship) and the lower left (low-low relationship) represent positive spatial autocorrelation. The other two quadrants contain outliers and have negative spatial autocorrelation (Anselin, 2003).

Depending on its value, Moran's I may indicate the existence of spatial clustering. However, as it is a global measure, it does not indicate where the clusters exist. To do so within GeoDa, a univariate LISA (Local Indicator of Spatial Association) analysis must be performed. The LISA cluster analysis outputs a new integer variable between 1 and 4 for each county. Positively correlated high-high values are assigned 1, and low-low values are assigned 4. The outliers are assigned either the values of 2 or 3. GeoDa also creates a map showing the values according to this newly created field. The LISA table and map results are then exported to a shapefile and opened in Arcmap, where the table is opened and a new field created, with the global values set to zero. Afterwards, a Definition Query in ArcMap is performed to select only the high-high (1) and low-low (4) values. The Field Calculator is also used to generate an integer value of 1 for these positively associated counties, and a final map output is created showing only the counties with a positive spatial autocorrelation. This process is repeated for both study periods, June 2009 and November 2001. The result is a more detailed understanding of the geographical patterning of unemployment variations than is possible from other known measures.

Step Two: Spatial Regression Models

The origins of the *term* spatial autocorrelation can be traced to a regional science paper published by Cliff and Ord in 1969, though the conceptual underpinnings for the concept had already been established previously (Getis, 2007). One of the many contributions of Cliff and Ord (1973; 1981) to the deeper understanding of spatial autocorrelation was the recognition that traditional OLS models were inadequate if spatial autocorrelation was present. However, as noted by Getis (2007), it was the contribution of Luc Anselin's book, *Spatial Econometrics: Methods and Models*, that "called attention to the responsibility of spatial scientists—geographers, urban planners, but especially economists—to take into account spatial effects when faced with regional-type georeferenced models and data."

Step Two of the research design begins with specifications for the proposed OLS regression models so that baseline models can be created for comparative purposes. The OLS model is a good starting point for this paper's analyses. Anselin (2005) recommends using such traditional models before testing for spatial autocorrelation and determining the need for spatial models. The explanatory variables included in the models are those suggested within the literature as described in Chapters 2 and 3. In their most basic form, the OLS models to be estimated are as follows:

$$Y = X\beta + \varepsilon \qquad (1)$$

where Y is the dependent variable, unemployment rate, X is the matrix of independent variables, β is the vector of unknown parameters to be estimated, and ε is the disturbance term following the assumption of no autocorrelation. The matrix of independent variables is different depending upon the specific model. The model for June 2009 is specified as follows:

$$\text{URJUN09} = b0 + b1\ \text{SELEMP00} + b_2\ \text{EDUC00} + b3\ \text{NONWHT07} + b4\ \text{USDAfarm} + b5\ \text{USDAmanf} + b6\ \text{MIXGOV07} + b7\ \text{PCPI\_08} + b8\ \text{LP1K08D} + \text{e} \qquad (2)$$

While the model for November 2001 is specified as follows:

$$\text{URNOV01} = b0 + b1\ \text{SELEMP00} + b_2\ \text{EDUC00} + b3\ \text{NONWHT00} + b4\ \text{USDAfarm} + b5\ \text{USDAmanf} + b6\ \text{MIXGOV00} + b7\ \text{PCPI\_01} + \text{e} \qquad (3)$$

Two OLS regression models have been presented so far. However, regression analysis assumes that variables are independent and distributed randomly, an unlikely premise when dealing with regional unemployment rate differentials. In order to investigate the role of geographic location in unemployment rate differentials, the OLS assumption of no spatial dependence requires corrective spatial modeling, by way of the spatial lag and spatial error models. Following the workflow advice of Anselin (2005), a cluster analysis is performed first to test for the presences of clusters (global spatial autocorrelation) and their locations (local spatial autocorrelation). This step is performed in Step One of the research design.

There is an oft-quoted law within geographical research commonly referred to as Tobler's First Law, which catches the purpose and essence of spatial analyses: "Everything is related to everything else, but near things are more related than distant things" (Tobler, 1970, p. 236). Elhorst (2003) reminds us that a regional unemployment rate model that does not account for serial (time) and spatial autocorrelation may be significantly misspecified. Fortunately, capturing the effects of time and space are not as difficult as it might first appear. The test for serial autocorrelation, thanks to advances in software development (SPSS is used for this paper), is now simply a matter of ensuring that the Durbin-Watson statistic, which is commonly computed along with the regression results, falls within an acceptable statistical range. Similarly, the test for spatial autocorrelation is nearly as simple, again because of spatial software advances. It is interesting to note that the Moran's I statistic is the most commonly used specification test for spatial autocorrelation and is, not coincidentally, remarkably similar in form to the Durbin-Watson statistic for serial autocorrelation (Anselin, 1988).

OLS models are inadequate for capturing the spatial dependence of sample data across neighboring regions because of their assumption of independence across observations. However, as described by Anselin (1988), there are two models available to regress for spatial dependence, and they both incorporate the weight matrix used for the spatial cluster analysis described above. Known as the spatial lag and spatial error models, they are spatial autoregressive techniques for revealing for the existence of spatial dependence within sample data. The spatial lag model is used to examine the 'existence and strength of spatial interaction' which may be present due to diffusion or spillover effects, while the spatial error model tests for 'biasing influences due to the use of spatial data' (Anselin, 1988, p.11). If spatial lag is ignored, then the OLS model will have an omitted variable problem and the estimates will be biased and inconsistent. Spatial error refers to spatial effects in the residuals, or unexplained variation, external to the behavior process, such as data collection errors. If spatial error is ignored, the OLS will remain unbiased, but inefficient, and the standard errors and t-tests will be biased (Anselin, 1988).

The spatial lag model is formally specified as follows:

$$\gamma = \rho W\gamma + X\beta + \varepsilon \tag{4}$$

where **y** is the dependent variable vector, ρ is a spatial autoregressive coefficient, Wγ is the spatially weighted dependent variable, X is the matrix of independent variables, β is a vector of unknown parameters to be estimated, and ε is a vector of error terms.

The spatial error model is formally specified as:

$$y = X\beta + u, \quad \text{and} \quad u = \lambda Wu + \varepsilon \tag{5}$$

where λ is the spatial autocorrelation parameter, and *W* is the spatial-weight matrix. *X* is the matrix of independent variables, *β* is a vector of unknown parameters to be estimated, *ε* is a vector of spatially autocorrelated error terms, and *u* a vector of i.i.d errors.

According to Anselin and Florax (1995), the decision of whether to specify the spatial lag or spatial error model depends on a series of Lagrange Multiplier (LM) test statistics. If the LM-lag test is significant and the LM-error is not, then the spatial lag model is specified. Conversely, if the LM-error test is significant while the LM-lag test is not, then the spatial error model is specified. If both tests are significant, then Anselin et al. (1996) suggest examining the robust LM-lag and LM-error measures.

Step Three: Multilevel Modeling

Multilevel modeling is used to answer the third research question: How do state-level groupings of counties influence county unemployment rates? The advantage of multilevel modeling is that "it can be used for a variety of different purposes, with far fewer and less restrictive assumptions than traditional regression methods" (Micceri, 2007). While spatial autocorrelation and spatial regression techniques are more commonly known, there are, as yet, no unemployment studies which have made use of multilevel modeling. While this study borrows from the previously mentioned U.S.-based research of regional unemployment rates, and from the new strand of spatial unemployment studies recently occurring in Europe, it is one of the first unemployment studies to make use of multilevel modeling.

A *multilevel model* "is a statistical model applied to data collected at more than one level (individual and group) in order to elucidate relationships at more than one level" (Luke, 2004). Multilevel models are referred to by various names: multilevel linear models, mixed-effects models, random-effect models, hierarchical models, generalized linear mixed models, nested models, variance component models and random-coefficient regression models (Heck & Thomas, 2000). The use of multilevel modeling as a research methodology became common in the 1990s as a result of increased computing power and methodological advances in statistical techniques (Ivanović & Baldigara, 2006). However, as noted by Luke (2004), a common problem found within the literature is that even with the formulation of more advanced multilevel modeling approaches, researchers continued to rely upon the use of more simplistic statistical techniques. He attributes the continued use of such single-level techniques to "the long reach of the positivist tradition" and the continued emphasis upon "the use of modeling techniques that statistically remove or control for the effects of covariates" (Luke, 2004). As Luke points out, the positivist approach initially took hold because it was most effective for modeling within those sciences dealing with 'closed' systems. He uses the example of planetary movement, where context is not that important since such movement can be explained by the use of only a few variables, namely, mass and velocity. However, unlike other sciences, the social sciences deal with more "open" systems, involving many influencing factors which cannot be contextually controlled

Multilevel models are useful when analyzing data in which observations are contextually nested within groups. Frequently within the social sciences, including economics, such nesting occurs in a hierarchical geographic structure. Thus, multilevel modeling is appropriate for analyzing the contextual influence of processes occurring at a group level upon those in operation at the individual level. A commonly used example of this is taken from education, whereby the variance of students' test scores is found to be related both to the student (individual) and the student's school (group). This "multilevel" effect cannot effectively be measured by traditional statistical techniques, such as the oft-used Ordinary Least Squares (OLS) regression, owing to violations of standard required assumptions. Multilevel modeling adds a dimension of understanding that is not available with OLS when working with nested data.

In this paper, the *group* level is U.S. states while the *individual* level consists of individual county unemployment rates. A set of explanatory variables is used to regress the dependent variable at both the individual county level and the state group level. According to Hox (1995), the model can be written as follows:

$$Y_{ij} = \beta_{0j} + \beta_{ij} X_{ij} + e_{ij} \quad (6)$$

where:

− *Yij*: response variable (county unemployment rate)

− *Xij*: explanatory variable on first level(county unemployment rate)

− β_{0j} is the intercept (model constant)

− β_{ij} is the regression coefficient (regression slope)

− e_{ij} is the residual error term with $e_{ij} N(0,\sigma^2)$

− *j* is for the group state level

− *i* is for the individual county unemployment rate level

With regard to this paper, explanatory data was collected at both the individual county level as well as the group state level, since counties are nested within states. By grouping individual county unemployment rates by state, this research takes advantage of what is known about the average county from a particular state, and analyzes it along with the data collected for each individual county. In other words, unemployment rate variations across counties are modeled as a function their *individual* demographic traits, human capital factors, and industrial mix, as well as the *group* effects of geographic state-level locational factors. In particular, this paper analyzes uses state-level data regarding the amount of subprime mortgages within each state to examine its effects upon county unemployment rates for the period June 2009. This is important since subprime mortgage data is not available at the county level across the entire United States.

The county data used here have a hierarchical relationship with state level factors. Counties are considered the first level, and state variables that counties belong to are located in the upper group level (level-2). The aim of the HLM modeling presented here is to examine how the individual county-level variables of race, metropolitan status, self-employment and educational attainment affect county unemployment rate levels after controlling for the state level effects of two industrial mix variables, farming and manufacturing. Four HLM models are specified, beginning with the estimation of an "empty model". The HLM analysis begins with a preliminary estimate of the empty model (the model with no independent variables) to split the total variation of county unemployment rate levels into *within and between* variance, then a block entry approach is used whereby first and second level covariates are gradually added.

Step Four: Binary Logistic Regression

Binary logistic regression is useful for predicting categorical outcomes. It also has the advantage of not requiring as many assumptions as does OLS. It is suitable for this study since it does not assume linearity, homoscedasticity, or normally distributed variables, however, it is sensitive to high correlations among the predictor variables, referred to often as multicollinearity (Pallant, 2005). Furthermore, it does require a normal distribution of the error terms of the dependent variable (Menard, 2010). This is an important relaxation of statistical requirements for this study since the dependent variable indicating whether or not a county belongs in the upper quartile of high-unemployment counties will, most likely, produce correlated errors.

Binary logistic regression is an appropriate method when the dependent variable is dichotomous in nature (Menard, 2010). Once the results of the LISA analysis were completed, each county was assigned a binary dependent variable describing whether or not it fell within the upper quartile of high-unemployment counties (1 = yes, 0 otherwise). The binary logistic analysis was conducted for the June 2009 study period.

The binary logistic model used herein tests the null hypothesis that the set of explanatory variables have no capacity to predict whether or not a county belongs within the upper quartile of U.S. counties as measured by the unemployment rate. The explanatory variables for the binary logistic model include many of the same variables used in Step Two spatial regression models as described earlier in this section: USDAMANF, PCPI_08, and SELEMP00. In addition, a categorical variable indicating which census division a county is located in (CENRGN), and another indicating whether or not a county is located in one of the highest-educated states, are also used in the binary logistic regression models. Also, a path dependent variable is used for the model to test the relationship between current county unemployment rate levels and past levels (UR_NOV01). The dependent variable for the model is a binary variable indicating whether or not a county is in the upper quartile of counties as measured by the unemployment rate (BLRJUN09).

Conclusion

This chapter has presented the data and the four-step research design used for this study. The first part of this chapter summarized the variables used for the empirical models, identified data sources, described the areal units, and discussed data aggregation and limitation issues. The second part of the chapter described the four statistical techniques used to conduct the analysis. The four-step research design uses the techniques of spatial autocorrelation, spatial regression, multilevel modeling and binary logistic regression, respectively, to answer, the study's four main research questions relating to Q1) the spatial patterning of county unemployment rates, Q2) the determinants of county unemployment rates, Q3) the influence of state-level groupings upon county unemployment rates, and Q4) the determinants of upper quartile counties as measured by their unemployment rate. Before turning to the results chapter, the descriptive statistics for the variables used in the models are presented in Table 4.

Table 4. Descriptive Statistics for Model Variables

	N	Mean	Std Dev	Min	Max	Type	0Count	1Count
UR_JUN09	3104	9.34	3.31	2.6	28.1	C		
SELEMP00	3104	0.16	0.07	0.03	0.52	C		
EDUC00	3104	0.16	0.08	0.05	0.64	C		
NONWHT07	3104	0.13	0.15	0	0.91	C		
USDAfarm	3104	0.14	0.35	0	1	B	2,604	500
USDAmanf	3104	0.29	0.45	0	1	B	2,204	900
MIXGOV07	3104	0.16	0.07	0.03	0.89	C		
PCPI_08	3104	33,016	8,468	12,558	140,275	C		
LP1K08d	3104	0.11	0.32	0	1	B		
UR_NOV01	3104	5.29	1.94	1.4	20.6	C		
NONWHT00	3104	0.15	0.16	0	0.95	C		
MIXGOV01	3104	0.17	0.07	0.03	0.89	C		
PCPI_01	3104	24,772	5,815	10,461	82,230	C		

Note. B = Binary; C = Continuous; D = Dummy.

CHAPTER 4

RESULTS OF THE FOUR-STEP RESEARCH DESIGN

Introduction

This chapter presents the results of the four-step research design. The results of the spatial autocorrelation and LISA cluster analysis performed in Step One are presented first. Next, the Step Two results of the spatial regression models are discussed, followed by the results of the multilevel models used in Step Three. Finally, the Step Four results of the binary logistic regression models are presented.

Step One Results: Spatial Autocorrelation and Cluster Analysis

For the two study periods of June 2009 and November 2001, Moran's I statistics were generated to test for spatial autocorrelation and prepare cluster maps. The Moran's I for county unemployment rates are .6817 (June 2009), and .5600 (November 2001). These values indicate strong clustering of counties, as measured by the unemployment rate, for both time periods (Figures 5 and 6). However, the Moran's I statistic is a global statistic and therefore not able to identify cluster locations. The task of identifying cluster locations is performed with local indicators of spatial association, known as LISA (Anselin, 1995). A LISA analysis was run for both county models and the LISA maps appear in Figures 7 and 8. When analyzing clusters, Anselin (2005) reminds us that the spatial clusters shown on the LISA maps only refer to the core of the cluster. The county

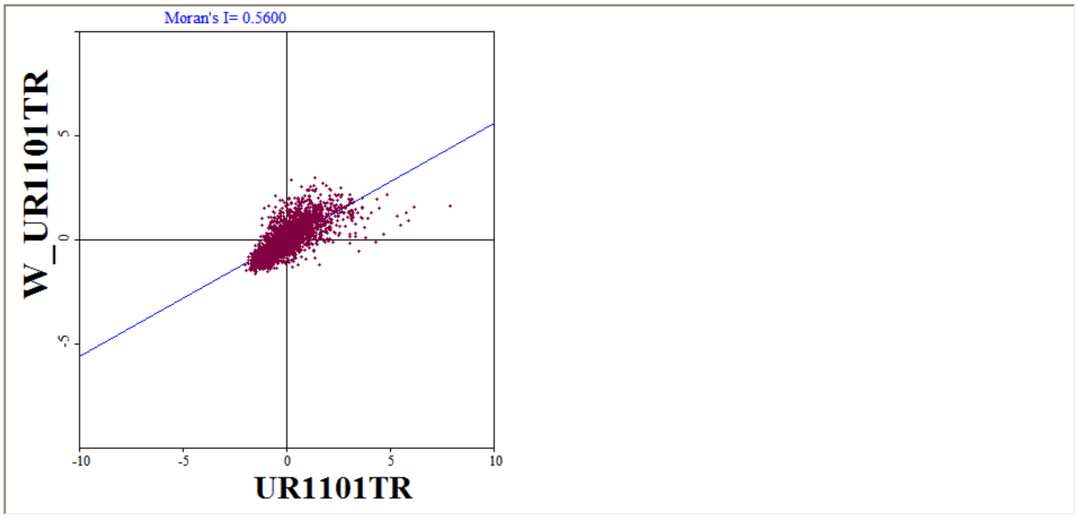

Figure 5. Moran’s I for November 2001. Source: Author’s calculation using Geoda.

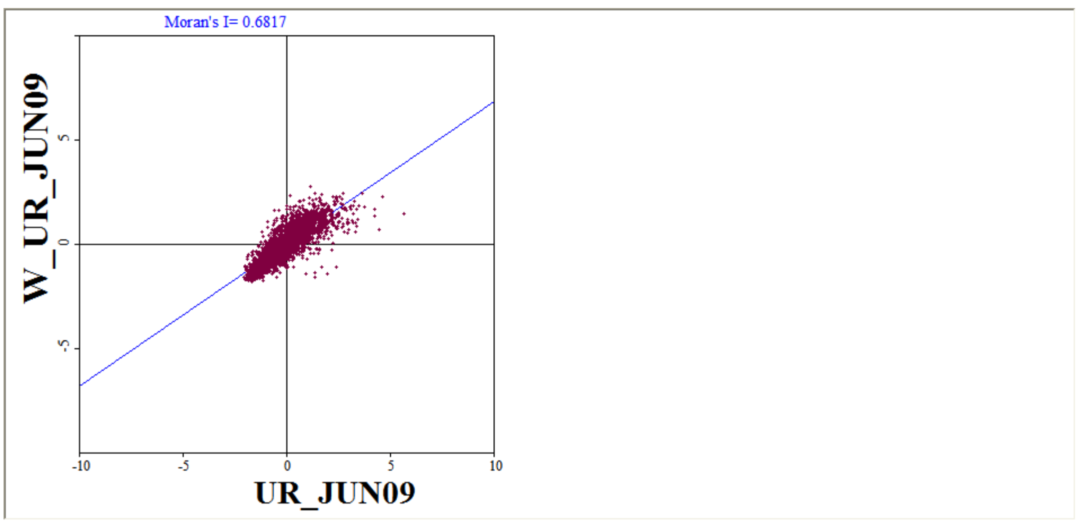

Figure 6. Moran’s I for June 2009. Source: Author’s calculation using Geoda.

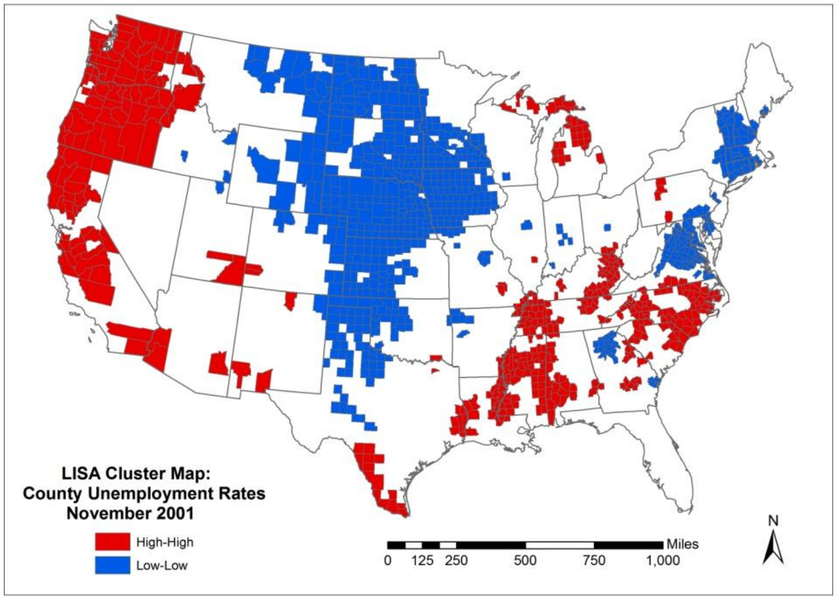

Figure 7. LISA Cluster Map for U.S. Counties, November 2001 (High N = 449).

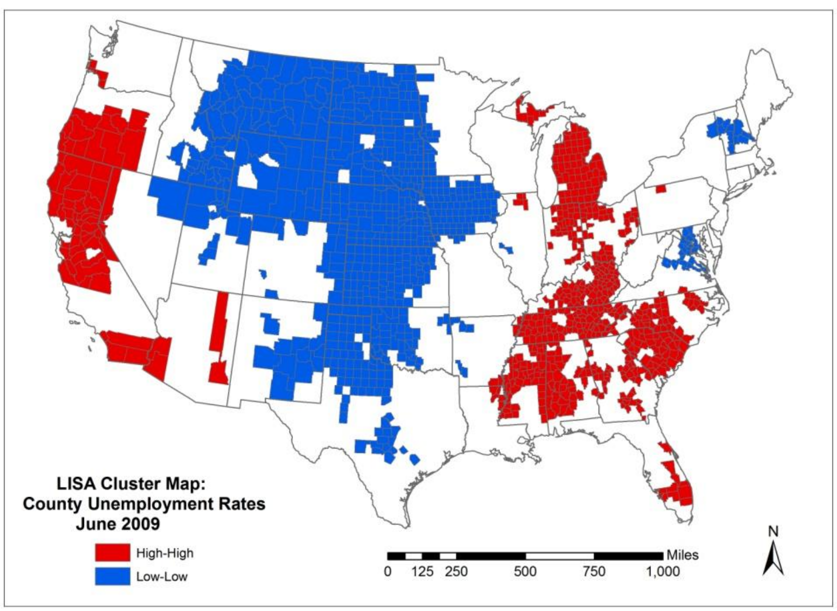

Figure 8. LISA Cluster Map for U.S. Counties, June 2009 (High N = 594).

cluster maps reveal the depth of the spatial relationships for November 2001 and June 2009.

Persistent Cluster Identification

As Vedder & Gallaway (1996) found at the state level, and has been noted within many regional unemployment studies in Europe (e.g., Aragon et al., 2003; Cracolici et al., 2007; Lopez-Bazo et al., 2002), unemployment differentials can persist within the same locations over long periods of time. Anselin (2003) suggests formulating a hypothesis about the presence of stable clusters over time. Following that advice, it was hypothesized early in this paper that specific geographic clusters of high (and low) unemployment rates would persist at the county level for both study periods (November 2001 and June 2009). The LISA cluster analysis presented in this section confirms this hypothesis. However, as pointed out by Delicado & Broner (2008), while LISA maps are useful for detecting clusters of a one-dimensional variable (unemployment), it is not able to guarantee that the regions appearing in the LISA cluster map are spatially connected. Indeed, the LISA analysis conducted for this study identifies 278 counties which register as high-high unemployment for both study periods, and 483 counties which register as low-low.

Figure 9 shows the location of these persistent high-high and low-low unemployment rate clusters. A visual inspection shows clustering of low unemployment rates in and around the Great Plains regions, as well as two small pockets of low unemployment clusters on the East Coast, near Washington, D.C. and New England. The number of counties identified as high-high in unemployment rate is 278 and the descriptive statistics for both the high-high (red) cluster and the low-low Great Plains cluster appear in Table 5.

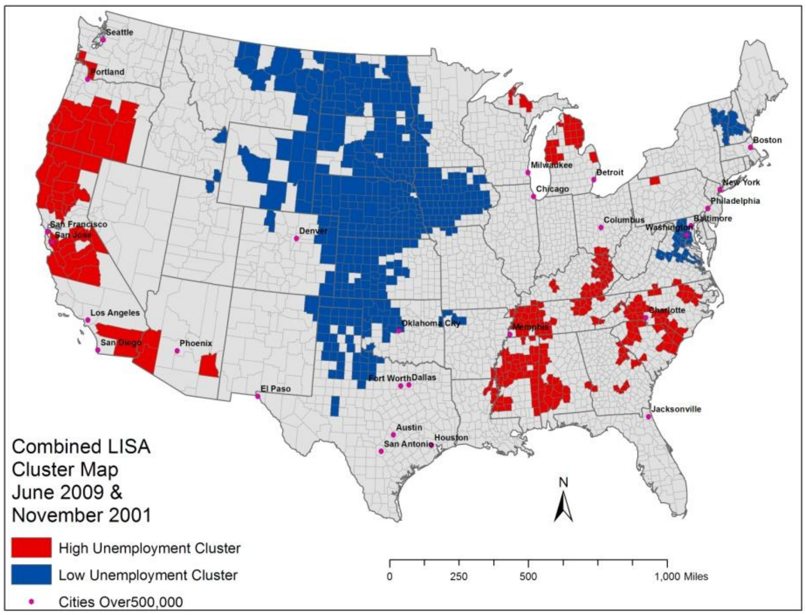

Figure 9. Combined LISA Cluster Map: June 2009 and November 2001 (High N = 278).

This paper is not as concerned with low-low unemployment counties as it is with high-high unemployment counties since the latter indicates positive spatial autocorrelation and are of greater concern to the researcher. This notwithstanding, some brief comments about the large Great Plains cluster is warranted. First, the vast majority of the counties in the Great Plains cluster have a population less than 20,000 people, versus 96,352 people for all U.S. counties. Second, the percentage of self-employed in counties in the U.S. is 29%, only 9% of the counties in the Great Plains cluster are manufacturing-dependent. Fourth, the racial mix of the Great Plains cluster is 95%

white, versus 87% for the rest of the country. Furthermore, the Great Plains cluster is only one percent African-American, versus nine percent for the average American county. Finally, and most notably, the average unemployment rate in the Great Plains cluster is only 5.15% for June 2009, versus the national county average of 9.34.

Table 5. Descriptive Statistics for the Major Clusters

VARIABLE	ALL	PLAINS	RED
	(N=3104)	(N=447)	(N=278)
UR_JUN09	9.34	5.16	14.09
UR_NOV01	5.29	3.27	8.14
AVGWGE07	31,708	28,634	29,985
SQMI	968	1,088	1,006
EDUC00	0.16	0.17	0.12
MIXFRM01	0.09	0.19	0.07
MIXFRM07	0.09	0.15	0.06
AVGPOP07	96,352	18,787	70,878
NONWHT07	0.15	0.05	0.23
RACBLK07	0.09	0.01	0.20
RACHISP07	0.07	0.07	0.06
RACWHT07	0.87	0.95	0.77
SELEMP00	0.16	0.25	0.12
USDAMANF	0.29	0.09	0.46
USDAFARM	0.14	0.54	0.08
MIXMFG91	0.15	0.07	0.22
MIXMFG01	0.11	0.06	0.15
MIXMFG07	0.09	0.06	0.12
% US_Land	100%	16.2%	9.3%
% US_Pop 07	100%	2.8%	6.6%
POPDEN(sq.mi.)	100	17	70
LQFARM09			1.72
LQMFG09			1.34

this cluster is about 25%, versus 16% for all counties. The percentage of self-employed is often associated with rural farming areas and this is supported by the fact that 54% of the counties in the Great Plains cluster are farming-dependent, versus only 14% for the average American county. Third, while the average number of manufacturing-dependent

As mentioned above, there are 278 counties identified as being persistently in the combined high-high unemployment rate cluster for both June 2009 and November 2001 (red cluster). The average unemployment rate for these 278 counties in June 2009 is 14.09%, versus 9.34% for the national county average. As of the 2000 census, only 12% of those residing in the red cluster had a college degree, versus 16% for the U.S. Perhaps the most important difference is the fact that while only 9% of American counties were manufacturing-dependent, 46% of the red cluster counties were such. Also, the percentage of the African-American population in the red cluster (20%) was more than twice that of the national county average (9%).

Persistent Clusters. Turning to the red cluster of high-high unemployment, while it is clear that these counties in the aggregate display characteristics vastly different from the national averages, it would be a mistake to treat them as one separate cluster with common identifiable traits. Admittedly, the identification of clusters is an inexact science. However, some clear differentiations can be made among the red cluster as to geographic location, industrial mix, and race. As such, the following five clusters are identified: the West, Michigan, Appalachia, the Carolinas, and the Deep South. The locations of these clusters are shown in Figure 10, and their descriptive statistics are shown in Table 6. A more in-depth discussion of these clusters, including socioeconomic characteristics drawn from this study's panel data, follows.

Pacific West. The Pacific West cluster is comprised primarily of agricultural counties in Washington, Oregon, California and Arizona. Naturally, sub-clusters within this area could be further identified, but since the component counties are primarily agricultural, and identification of micro clusters is beyond the scope of this paper, the Pacific West is treated as a single cluster herein. Included in the Pacific West cluster is a ten-county area in Central California known as the San Joaquin Valley which is the most productive agricultural region in the country. While not all agricultural counties in this cluster are as economically productive as this region, the economic problems plaguing the San Joaquin Valley are emblematic of those affecting other counties in the Pacific West. Foremost among these problems is an increasing rural population competing for a shrinking number of agricultural jobs. Technological advances in the agricultural field have helped shrink the demand for agricultural labor. The location quotient for agricultural sector employment in the Pacific West cluster is 1.30, which indicates that the economy of this cluster is more specialized in the agricultural sector than the rest of the county. However, that is only part of the story. Perhaps a more relevant characteristic of the Pacific West cluster with regard to its higher than average unemployment rates lies with its growing population. The Great Plains region is also an agricultural region, yet unemployment there is far below the national average. One of the primary differences between the two clusters is that the Great Plains region has lost population over the past decade, while the Pacific West cluster has experienced population growth. Furthermore, the average county size in the low-unemployment Great Plains cluster is 1,088 square miles, while the average county size in the Pacific West cluster is 3,020 square miles. Yet, the population density of the Great Plains cluster is a mere 17 persons per square mile, but the Pacific West cluster has a population density of 75 persons per square mile. Clearly, the effect of more people seeking a limited number of agricultural jobs has pressured unemployment rates in the Pacific West cluster. It should also be noted that due to the nature of its agricultural production, employment within the Pacific West clusters tends to be highly seasonal. County unemployment figures from the Bureau of Labor Statistics are unadjusted for seasonal changes, which often leads to higher reporting of unemployment figures in Western agricultural counties.

Michigan. The counties in the Michigan cluster are located in Northern Michigan. This cluster is rural in nature; none of the component counties have a population over 60,000 and the average county population is only around 27,000. Normally, when one thinks of Michigan, one thinks of manufacturing. However, the Michigan cluster presented herein has a location quotient for manufacturing sector employment of only 1.11, not much more specialized in manufacturing than the rest of the nation. The telling statistic for this cluster is its extraordinarily high location quotient for agricultural sector employment, measuring 2.54. Thus, problems associated with agricultural sector employment exist at higher magnitude in this cluster than the rest of the country's regions, save for Appalachia.

Appalachia. The 44 counties in the Appalachia cluster are amongst the most economically distressed counties in the country and are historically amongst the most impoverished in the United States. Of all the identified high-unemployment clusters, Appalachia has the lowest percentage of college-educated residents (9%), the highest farming-sector location quotient (3.60), and the lowest average wage ($28,569). In addition to its farming sector dependence, the Appalachia cluster also has a high location quotient for manufacturing sector employment (LQ = 1.40). None of the counties in this cluster contain large populations: the average county population for this region is only 28,569. The population of the Appalachian cluster is over 97% white, which is the highest level for that demographic amongst all of the regional unemployment clusters.

Deep South. The 94 counties comprising this cluster represent a collective population that is over a third African-American (36%). Nearly a quarter of the jobs within this cluster are in manufacturing, and they are low-paying jobs. The location quotient for manufacturing sector employment is 1.79. Like the Appalachia cluster, unemployment in the South cluster is affected by its dual reliance upon manufacturing and agricultural sector jobs. The location quotient for agricultural sector employment is 3.56, amongst the highest levels in the country. Only 11% of persons residing in the Deep South cluster have college degrees.

Carolinas. There are 66 counties in the Carolinas cluster and its geographic center is near the city of Charlotte, North Carolina. As such, the population density of this cluster is the highest of the five identified high-unemployment clusters (116 persons/ square mile). This cluster is highly dependent upon the manufacturing sector, as indicated by its manufacturing location quotient of 1.84, the highest of all identified clusters. Only 12% of the population in the Carolinas cluster have college degrees. Historically, the furniture and textile manufacturing industries have been the leading employers in this region, but globalization effects have moved many of those jobs overseas over the past few decades (Drayse, 2008).

Summary of Identified Clusters. The data show significant differences in cluster composition and some interesting results. For instance, four of the five clusters are below the national county average for college-educated (16%): Michigan (13%), Appalachia (9%), the Carolinas (12% and the Deep South (11%). The West cluster approximates the national average with 17% of its population having a college degree. The clusters of Appalachia (3%) and Michigan (4%) are far below the national county average of non-white population (15%), while the Carolinas (33%) and the Deep South (37%) far exceed the national average. The West (12%) approximates the national non-white average. The racial and ethnic numbers become even more interesting when they are broken down further into the percentage of African-Americans (Figure 11) and Hispanics (Figure 12). The South cluster (36%) and the Carolinas cluster (30%) contain three times the number of African-Americans than the national average (9%), while the Michigan (1%) and Appalachia (2%) clusters are far below this average. With regard to Hispanics, The West cluster has more than double (22%) the national percentage (9%) living in it, while none of the other clusters are more than 4% Hispanic. But perhaps most interesting are the differences among the clusters in their mean number of manufacturing-dependent by the ERS/USDA as any county that has “25 percent or more of average annual labor counties and proprietors' earnings derived from manufacturing during 1998-2000" (ERS/USDA, 2011). The national county average for this figure is 29%, but four of the five clusters exceed this figure: Michigan (35%); Appalachia (32%); the Carolinas (70%) and the Deep South (60%). Only the West falls below the national average in manufacturing-dependent counties (12%). In fact, the Carolina and South clusters have lost more than half of their manufacturing jobs since 1991, while the Appalachia cluster has lost more than a third.

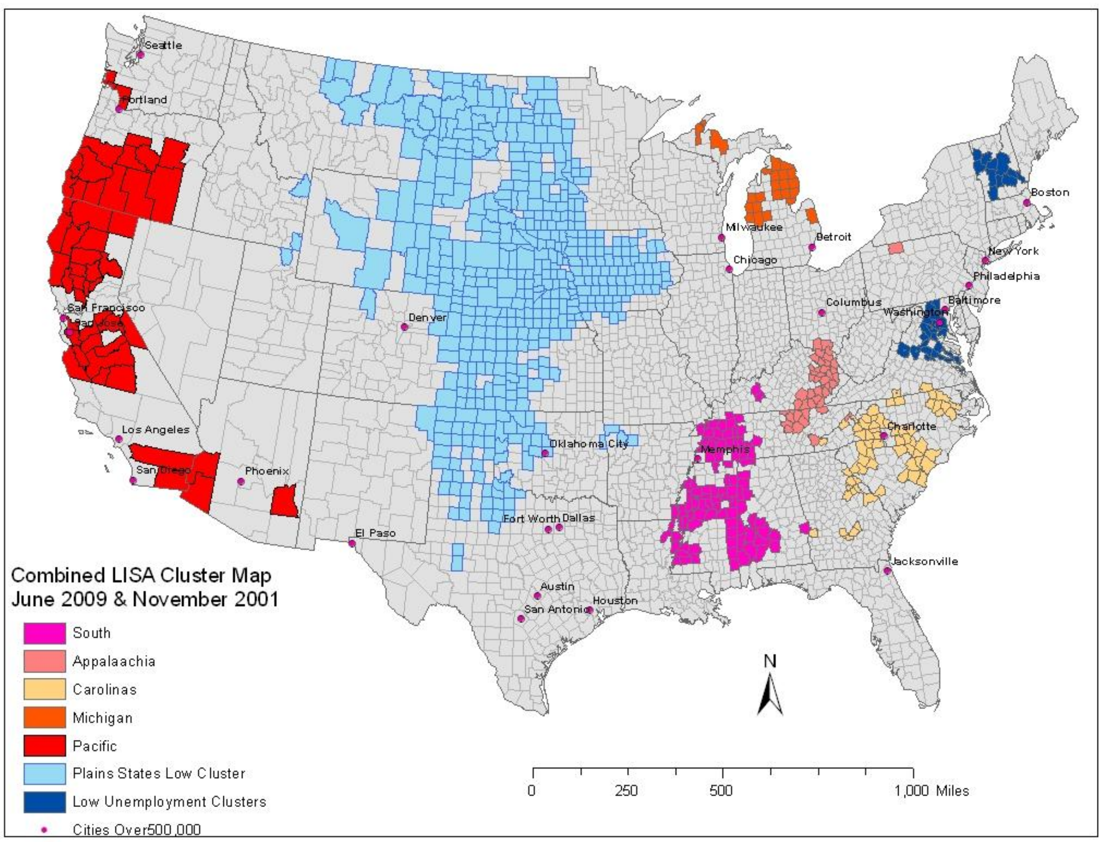

Figure 10. Persistent Cluster Identification for June 2009 and November 2001.

Table 6. Descriptive Statistics for the Identified High-High Clusters

VARIABLE	ALL	WEST	MICHIGAN	APPALACHIA	CAROLINAS	SOUTH
	(N=3104)	(N=51)	(N=23)	(N=44)	(N=66)	(N=94)
UR_JUN09	9.34	14.08	14.78	13.59	13.95	14.26
UR_NOV01	5.29	8.78	7.93	7.54	8.11	8.13
AVGWGE07	31,708	34,350	29,340	28,569	29,455	28,810
SQMI	968	3,020	678	405	527	610
EDUC00	0.16	0.17	0.13	0.09	0.12	0.11
MIXFRM01	0.09	0.07	0.04	0.09	0.05	0.09
MIXFRM07	0.09	0.05	0.04	0.07	0.04	0.07
AVGPOP07	96,352	226,534	26,757	25,995	61,301	24,955
NONWHT07	0.15	0.12	0.04	0.03	0.33	0.37
RACBLK07	0.09	0.02	0.01	0.01	0.30	0.36
RACHISP07	0.07	0.22	0.02	0.01	0.04	0.02
RACWHT07	0.87	0.88	0.96	0.97	0.67	0.63
SELEMP00	0.16	0.15	0.13	0.11	0.10	0.10
USDAMANF	0.29	0.12	0.35	0.32	0.7	0.6
USDAFARM	0.14	0.12	0.00	0.00	0.04	0.05
MIXMFG91	0.15	0.11	0.13	0.16	0.31	0.26
MIXMFG01	0.11	0.08	0.10	0.11	0.20	0.18
MIXMFG07	0.09	0.07	0.09	0.09	0.15	0.14
% US_Land	100%	5.1%	0.5%	0.6%	1.2%	1.9%
% US_Pop 07	100%	3.9%	0.2%	0.4%	1.4%	0.8%
PopDensity	100	75	39	64	116	41
LQFARM09		1.30	2.54	3.60	1.35	3.56
LQMFG09		1.10	1.11	1.40	1.84	1.79

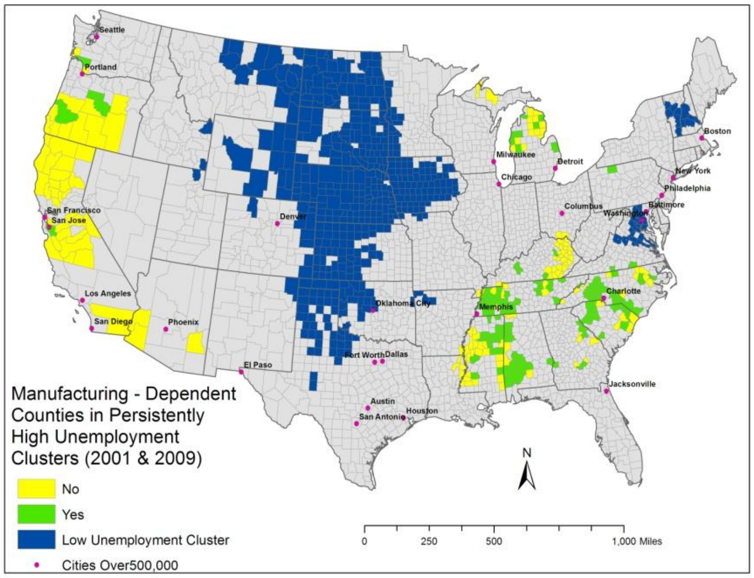

Figure 11. Manufacturing-Dependent Counties in Persistent Clusters.

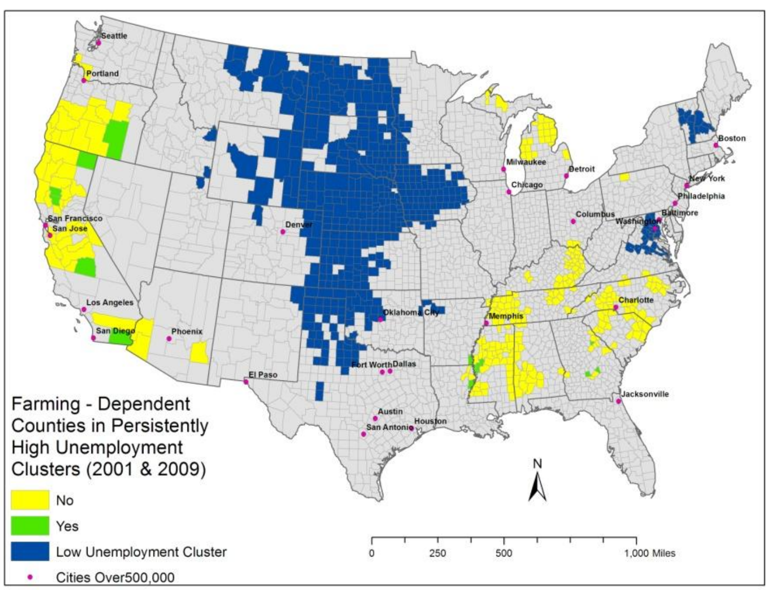

Figure 12. Farming-Dependent Counties in Persistent Clusters.

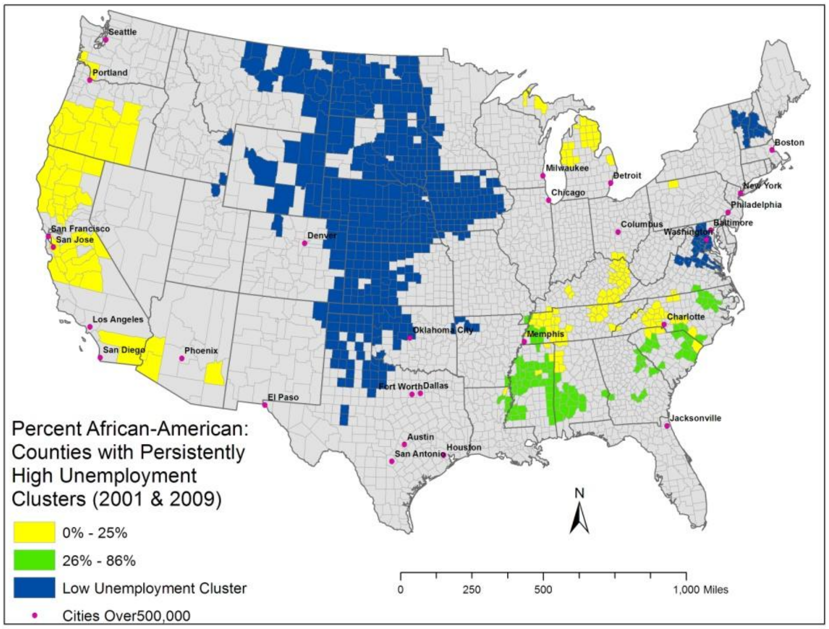

Figure 13. Percent African-American in Persistent Clusters.

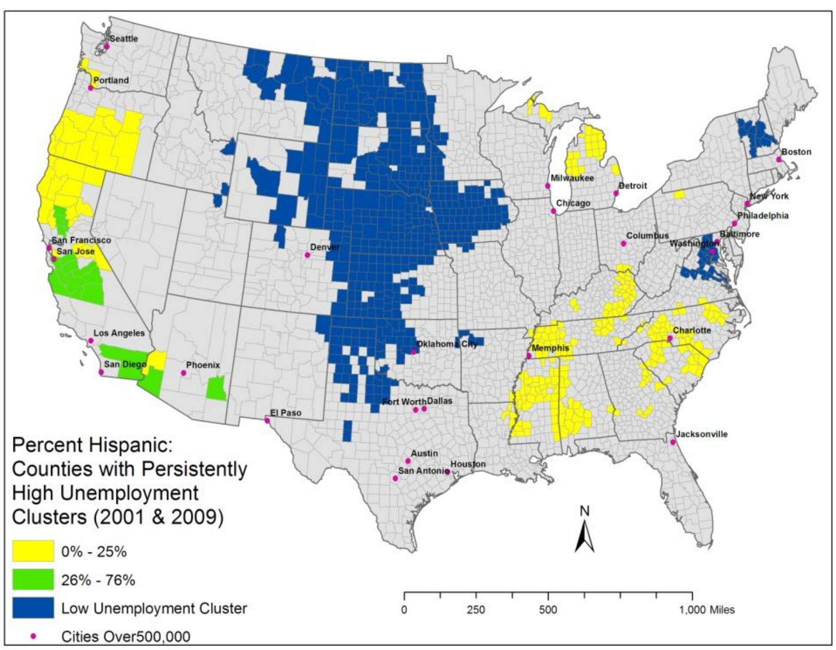

Figure 14. Percent Hispanic in Persistent Clusters.

Step Two Results: Spatial Regression Models

The purpose of the Step Two research design is to examine the determinants of county unemployment rate variation. Three types of models are presented for both study periods: 1) OLS, 2) spatial lag, and 3) spatial error. As discussed in Chapter 3, the June 2009 and November 2001 models are specified as follows:

June 2009: URJUN09 = *b0* + *b1* SELEMP00 + b_2 EDUC00 + *b3* NONWHT07 + *b4* USDAFARM + *b5* USDAMANF + *b6* MIXGOV07 + *b7* PCPI_08 + *b8* LP1K08D + e (7)

November 2001: URNOV01 = *b0* + *b1* SELEMP00 + b_2 EDUC00 + *b3* NONWHT00 + *b4* USDAFARM + *b5* USDAMANF + *b6* MIXGOV00 + *b7* PCPI_01 + e (8)

The June 2009 models were run with eight independent variables while the November 2001 models were run with seven independent variables. The binary variable indicating whether or not a county was located within one of the top states with subprime mortgages per 1,000 residents (LP1K08D: 1 = yes, 0 otherwise) was not used for the November 2001 analysis since it lacks any theoretical purpose for that time period. The explanatory variables common to both time periods include binary variables for each county indicating whether or not a county's economy is dependent upon manufacturing (USDAMANF: 1 = yes, 0 = no, Figure 15), dependent upon farming (USDAFARM: 1 = yes, 0 = no, Figure 16), the percentage of a county's employment

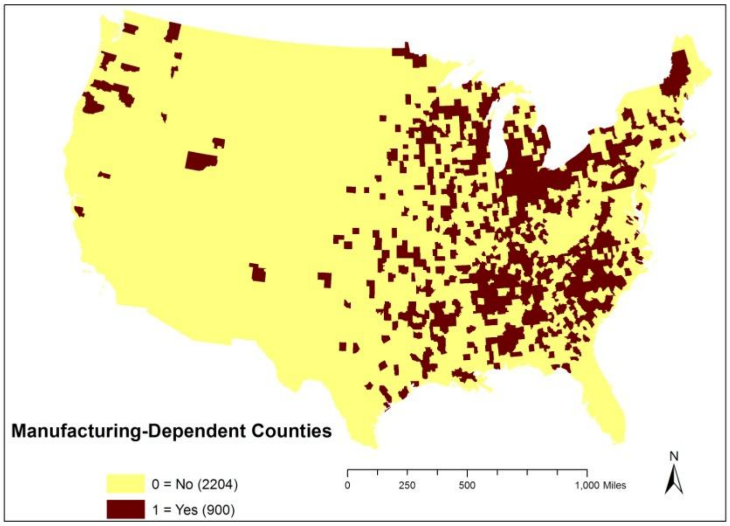

Figure 15. Manufacturing-Dependent Counties.

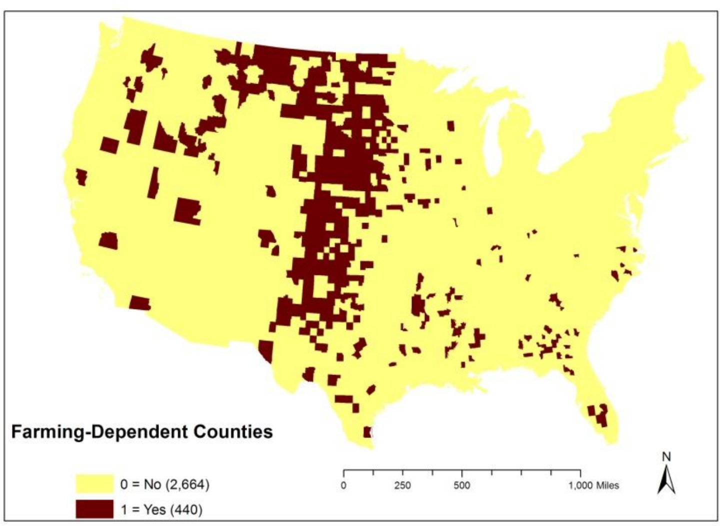

Figure 16. Farming-Dependent Counties.

share in the government sector (MIXGOV07 and MIXGOV01), per capita income (PCPI_08 and PCPI_01; Figures 17 and 18), the percentage of the population with a bachelor's degree or higher (EDUC00; Figure 19), the percentage of the self-employed living within a county (SELEMP00; Figure 20), and the percentage of a county's nonwhite population (NONWHT07 and NONWHT00; Figure 21).

Two OLS regression models are specified, primarily for diagnostic examination and baseline comparative purposes, and their results appear in the first columns of Tables 7 and 8. A diagnostic examination of the OLS results shows no problems with multicollinearity. The Condition Index is only 18.55 for June 2009, and 23.27 for November 2001. Both values are below the suspect value of 30. However, there are diagnostic problems with the OLS regressions. The Jarque Bera test is significant, indicating that the residuals do not follow the normal distribution, though this is quite common with large numbers of observations. Also, the Breusch-Pagan and Koenker-Basset statistics are both significant, indicating heteroskedasticity problems associated with non-constant variance in errors (Anselin, 2005). The heteroskedasticity problem may be partially attributable to the spatial dependence within the data.

According to the specification strategy laid out by Anselin (2005), the next step after the estimation of the OLS models is to test the null hypothesis of no spatial dependence using a global measure of spatial autocorrelation. For this study, the Moran's I statistic is used to test for the presence of spatial autocorrelation. The expected null value of Moran's I equals [1 / N – 1]. Thus, since N = 3104 for both models, the

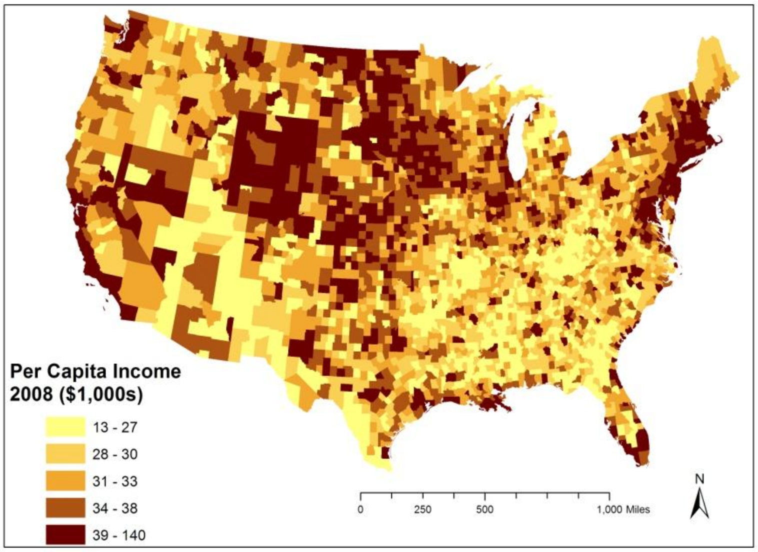

Figure 17. Per Capita Income by County, 2008

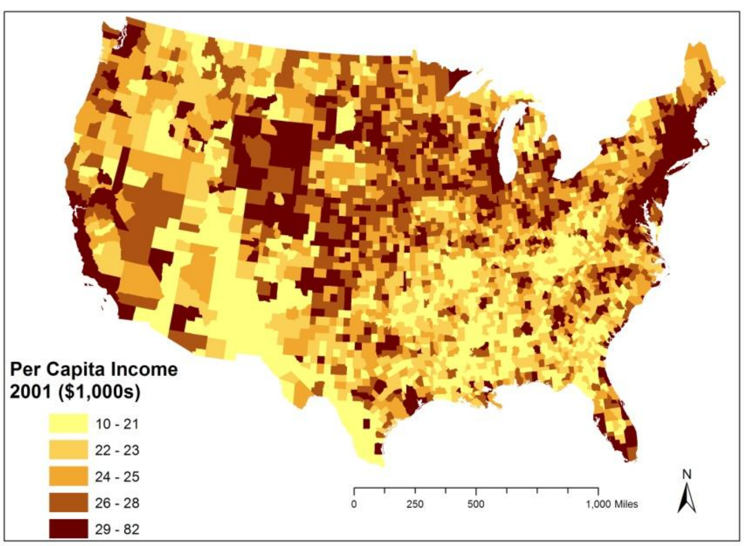

Figure 18. Per Capita Income by County, 2001

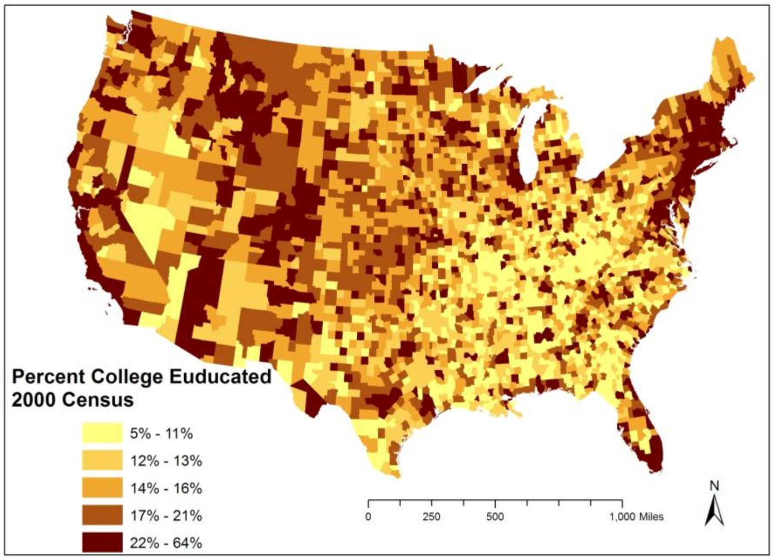

Figure 19. Percent of County Population with a College Degree, 2000 (EDUC00)

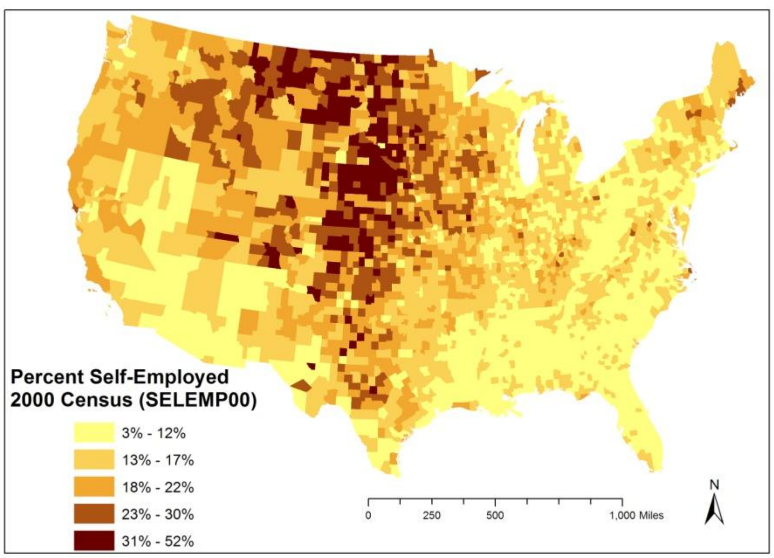

Figure 20. Percent of County Population that is Self-Employed (SELEMP00)

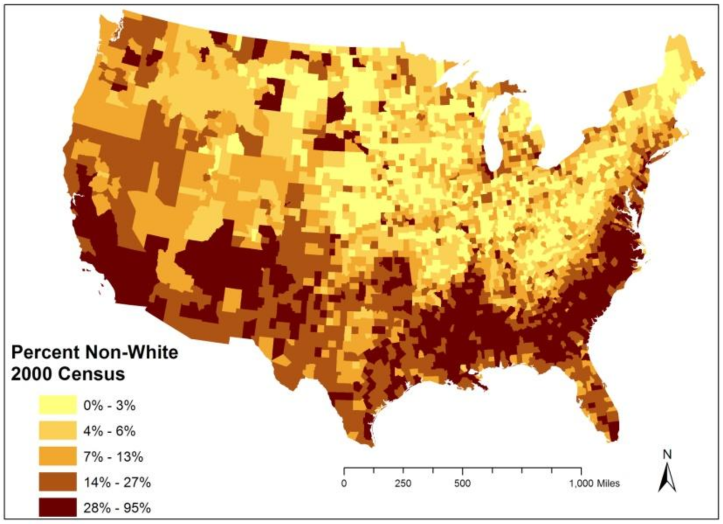

Figure 21: Percentage of Non-White Population, 2000.

Table 7: OLS and Spatial Regression Results for County Models (DV = UR_JUN09)

VARIABLE	OLS		SPATIAL LAG		SPATIAL ERROR	
	Coeff.	Probability	Coeff.	Probability	Coeff.	Probability
Constant	14.8517	0	6.2682	0	12.8199	0
SELEMP00	-16.292	0	-6.5051	0	-9.2064	0
EDUC00	-8.0848	0	-6.5471	0	-9.7956	0
NONWHT07	3.0057	0	2.103	0	4.7966	0
USDAFARM	-0.9511	0	-0.4983	0	-0.1811	0.1275
USDAMANF	1.2816	0	0.7705	0	0.6142	0
MIXGOV07	-3.456	0	-0.6832	0.1794	-1.3372	0.0079
PCPI_08	-6E-05	0	-3E-05	0	-3E-05	0
LP1K08D	2.8485	0	1.1291	0	1.8788	0
W_UR_JUN09			-0.6139	0		
LAMBDA					0.7636	0
R2	0.54		0.72		0.75	
Log likelihood	-6905		-6256		-6218	
Condition Index	18.55					

Table 8. OLS and Spatial Regression Results for County Model (DV = UR_NOV01)

VARIABLE	OLS		SPATIAL LAG		SPATIAL ERROR	
	Coeff.	Prob.	Coeff.	Prob.	Coeff.	Prob.
Constant	8.3415	0	3.1311	0	6.7371	0
SELEMP00	-7.6508	0	-2.8378	0	-4.1375	0
EDUC00	-4.6908	0	-4.4775	0	-6.3968	0
NONWHT00	2.456	0	1.7495	0	3.5641	0
USDAFARM	-0.4051	0.0002	-0.1662	0.0468	-0.037	0.665
USDAMANF	1625	0.0232	0.1614	0.0038	0.2049	0.0003
MIXGOV01	-0.1972	0.6789	-0.5949	0.1097	0.0401	0.9121
PCPI_01	-6E-05	0	-2E-05	0.0045	-1E-05	0.0518
W_UR_NOV01			0.6405	0		
LAMBDA					0.7309	0
R2	0.32		0.59		0.62	
Log likelihood	-5859		-5231		-5153	
Condition Index	23.27					

Table 9. Diagnostics for Spatial Dependence

UR_JUN09	MI/DF	Value	Prob.
Moran's I (error)	0.4223	39.481	0.000
Lagrange Multiplier: lag (LML)	1	1523.046	0.000
Robust LM: lag (LMLR)	1	190.226	0.000
Lagrange Multiplier: error (LME)	1	1536.366	0.000
Robust LM: error (LMLE)	1	203.546	0.000
UR_NOV01	MI/DF	Value	Prob.
Moran's I (error)	0.468	43.691	0.000
Lagrange Multiplier: lag (LML)	1	1641.690	0.000
Robust LM: lag (LMLR)	1	34.659	0.000
Lagrange Multiplier: error (LME)	1	1887.115	0.000
Robust LM: error (LMLE)	1	280.084	0.000

expected value of Moran's I for both models is 0.00032. However, as expected, the Moran's I value for June 2009 (.4233) and November 2001 (.4680) far exceeds the expected Moran's I value, indicating the strong presence of spatial autocorrelation in the models (Table 9). As a result, the OLS regressions are not appropriate models for explaining unemployment rates variations as specified herein. Therefore, in order to further investigate county unemployment rate differentials, the OLS assumption of no spatial dependence requires corrective spatial modeling, by way of the spatial lag and spatial error models.

The results of the spatial lag and spatial error models are presented in the second and third columns of Tables 7 and 8. The explanatory value of the OLS models, as measured by R^2, is a moderately acceptable 0.54 for June 2009, and a moderately low 0.32 for November 2001. For the spatial lag and spatial error models of June 2009, the R^2 value increases from 0.54 to 0.72 and 0.75, respectively. The R^2 for the spatial lag and spatial error models for November 2001 increase from 0.32 to 0.59 and 0.62, respectively. However, it should be noted that the R^2 results from spatial regressions are pseudo R^2 measurements and are not directly comparable to the R^2 from OLS regressions. A better model fit statistic is the Log-likelihood (LL) value, where the best goodness of fit is considered to be the model with the highest LL (Anselin, 2005, p. 207). For both time periods studied, the LL measure is lowest for the OLS models, increases drastically in the spatial lag models, and is higher still in the spatial error models. Therefore, the LL criteria indicate that the spatial error model is the best specification for both June 2009 and November 2001.[9] However, there are other methods used within the literature to decide model choice between spatial lag and spatial error. The prescribed method for choosing between the models is dependent upon whether the Lagrange Multiplier tests (LML and LME), and their robust counterparts (LMLR and LMLE), are significant (Table 9). In cases where they are all significant, as they are here, Anselin, Bera, Florax & Yoon (1996) suggest using the higher of the robust values of LMLR or LMLE to indicate which spatial model to specify. The results of this analysis indicate that the LMLE value is higher than the LMLR value for both time periods and, as such, also points to the spatial error model as the preferred specification. However, Anselin (2005) also points out that there are no absolute rules for choosing between models when all are significant. Within the literature, the results of both are often specified, which is the procedure followed in this paper. Hence, the results of the OLS, spatial lag and spatial error model runs are presented together for comparative purposes.

All of the eight explanatory variables are highly significant for the June 2009 OLS model, while six of the seven explanatory variables are highly significant in the November 2001 model. The percentage employment share in the government sector is not significant in the November 2001 OLS model. The results of the spatial regressions indicate strong evidence of spatial dependence operating at the county level for both study periods. The inclusion of a spatially lagged variable for June 2009 markedly

[9] The Log- likelihood values are not comparable across different dependent variables, only across models for the same dependent variable (i.e., the county unemployment rate models for either June 2009 or November 2001.

increases the overall explanatory value of the spatial lag model. Also, adding the lambda coefficient for spatially correlated errors increased the explanatory value of the spatial error model. The R^2 value of .54 in the OLS model version increased to .72 in the spatial lag model, and .75 in the spatial error model. In other words, geographic location accounts for nearly 18% of county unemployment rate differentials in the spatial lag model for June 2009; and 21% in the spatial error models. Similar results were found for the November 2001 models. The R^2 for the November 2001 OLS model was only .32, however, it increased to .59 in the spatial lag model, and .62 for the spatial error model.

The results show that per capita income (PCPI_08 and PCPI_01), the percentage of self-employed (SELEMP00), and the percentage of county residents with a college degree or higher (EDUC00), had a *negative* and significant relationship to county unemployment rates across all periods and models. Alternatively, the percentage of non-white population (NONWHT07 and NONWHT00) had a *positive* and significant relationship to county unemployment rates across all periods and models. As for the binary variables, it was found that location within a manufacturing-dependent county (USDAMANF) was related to an increase in the unemployment rate and that relationship was significant across all periods and models. Also, location within a farming-dependent county was related to a decrease in the unemployment rate, though that relationship was not significant in the spatial error models for either June 2009 or November 2001. The industry mix variable showing the percentage of workers employed in the government sector (MIXGOV07 and MIXGOV01) had a negative relationship to county unemployment rates. However, this relationship was not significant in any of the

November 2001 models, nor for the June 2009 spatial lag model. Finally, the results show that county location within one of the top states for subprime mortgages (LP1K08D – Figure 22) was significantly related to an increase in the unemployment rate for all of the June 2009 models (no such variable was used for November 2001).

Though the spatial lag and spatial error models improved model fit relative to the OLS models, it should be noted that some problems remain. First, the Breusch-Pagan statistic remains significant after running the spatial regression models, indicating continuing problems with heteroskedasticity. Second, while spatial regression removed most of the spatial effects from the residuals, a residual check of the spatial error models still show the statistically significant presence of spatial autocorrelation (albeit, a very slight presence). The Moran's I and residual maps for November 2001 and June 2009 (Figures 23 through 30) confirm this. Third, the June 2009 models include a binary variable for the top states in subprime mortgages which was shown to be highly significant for June 2009. However, this variable was collected at the state level and disaggregated down to the county level, leaving the models open to the effects of ecological fallacy.

In summary, the results of this section show a high degree of spatial dependence in county unemployment rates. When spatial lag and spatial error models are specified to correct for this, certain significant relationships appear across the models. Explanatory variables showing a positive relationship with unemployment rate variation include the percentage of non-whites, manufacturing-dependent counties, and counties located in high subprime mortgage states. Negative relationships were found for self-employment and education. This section now turns to the results of the multilevel modeling techniques conducted in Step Three of the research design, which are free of many of the statistical assumptions required by regression techniques and, therefore, able to correct for some of the shortcomings described here.

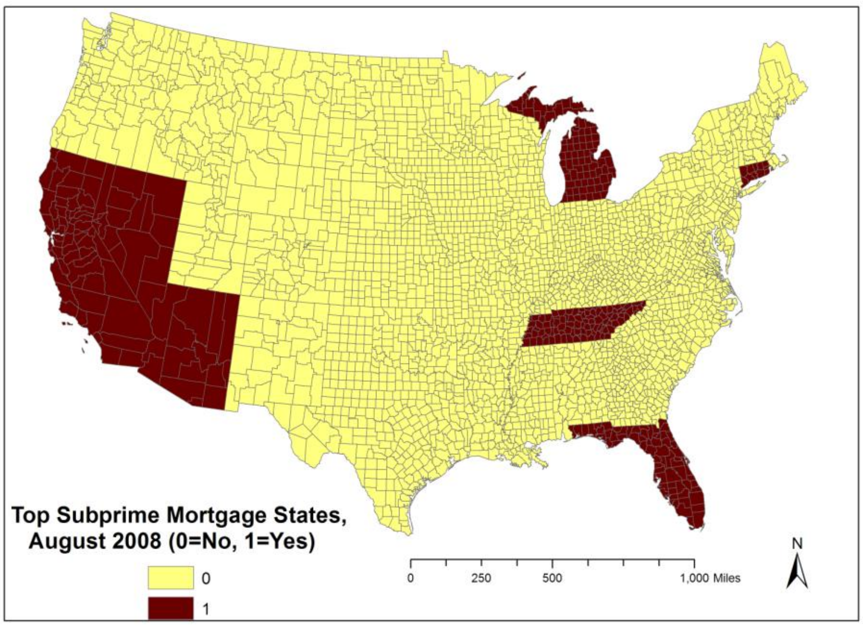

Figure 22. Top Subprime Mortgage States per 1,000 Mortgages, August 2008.

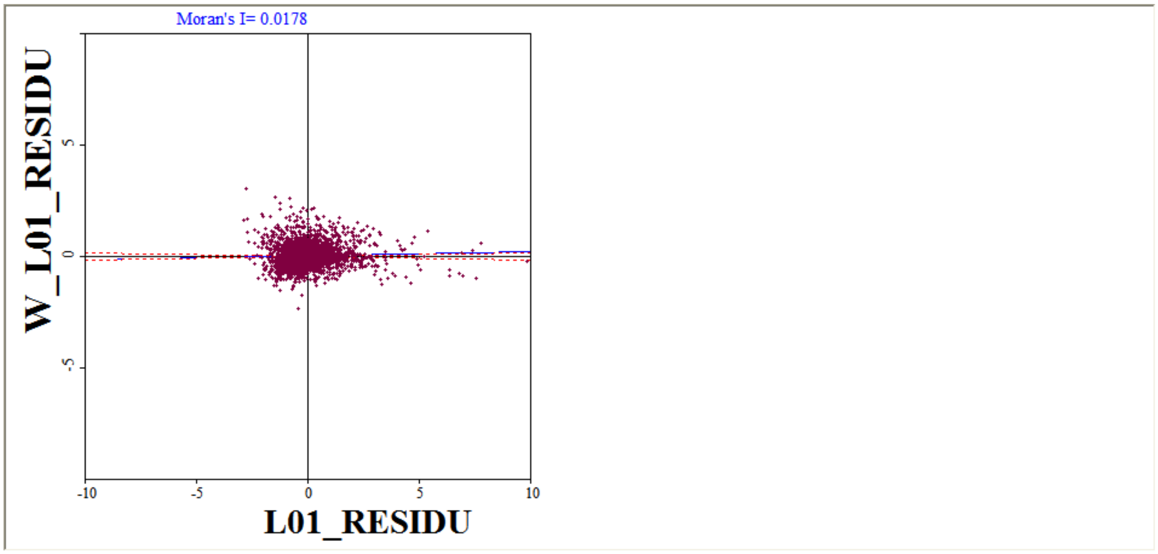

Figure 23. Moran’s I for Spatial Lag Model, November 2001.

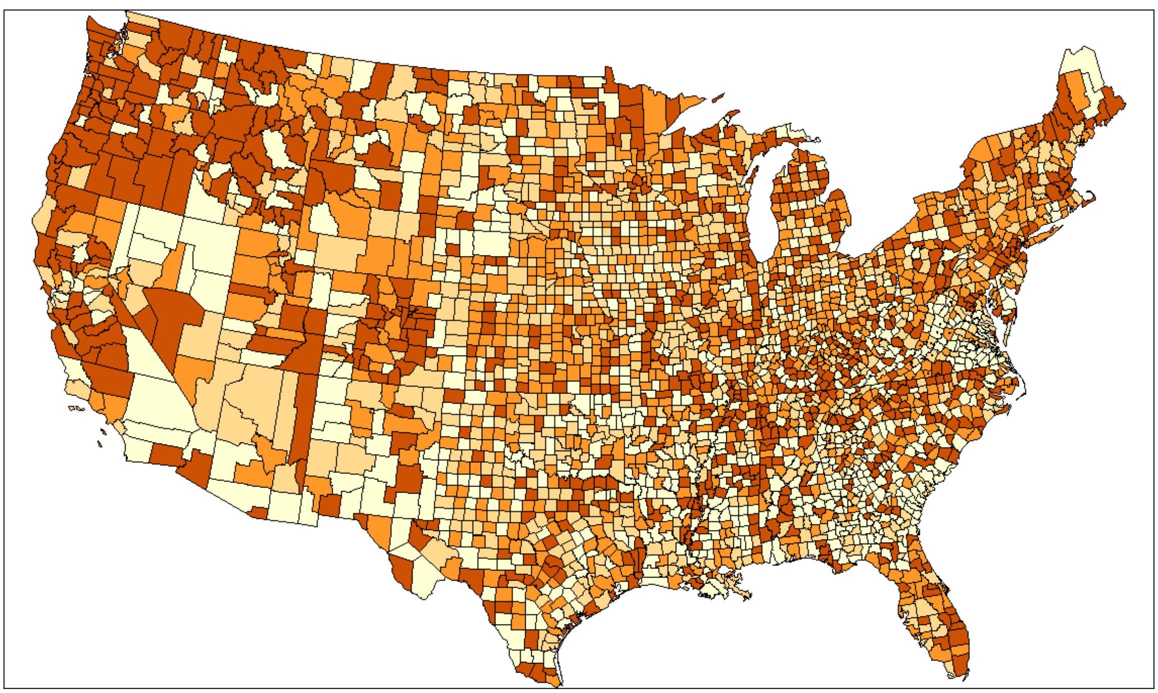

Figure 24. Residual Map for Spatial Log Model, November 2001.

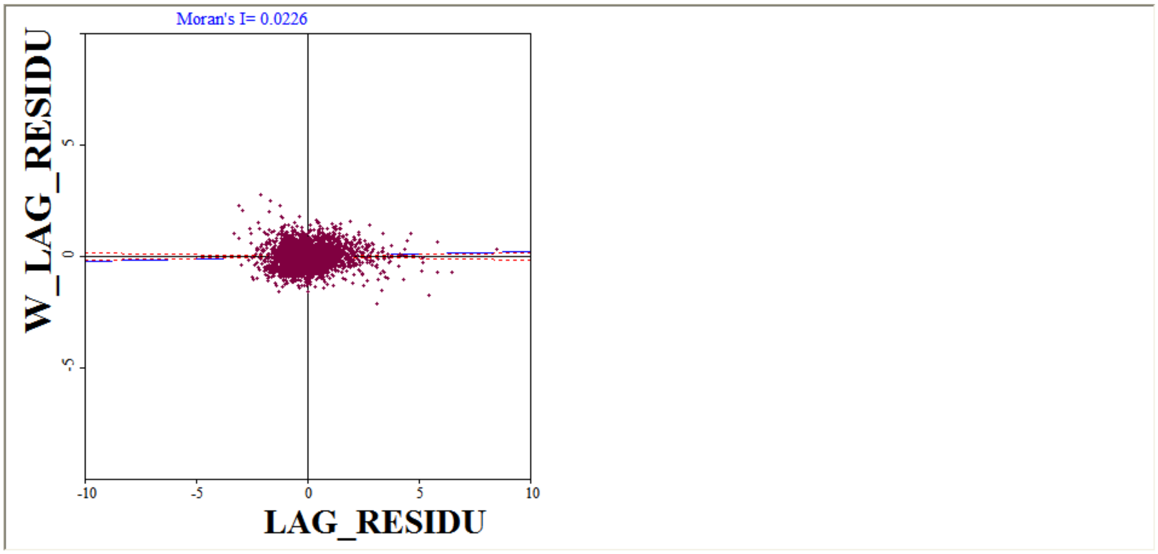

Figure 25. Moran's I for Spatial Lag Model, June 2009.

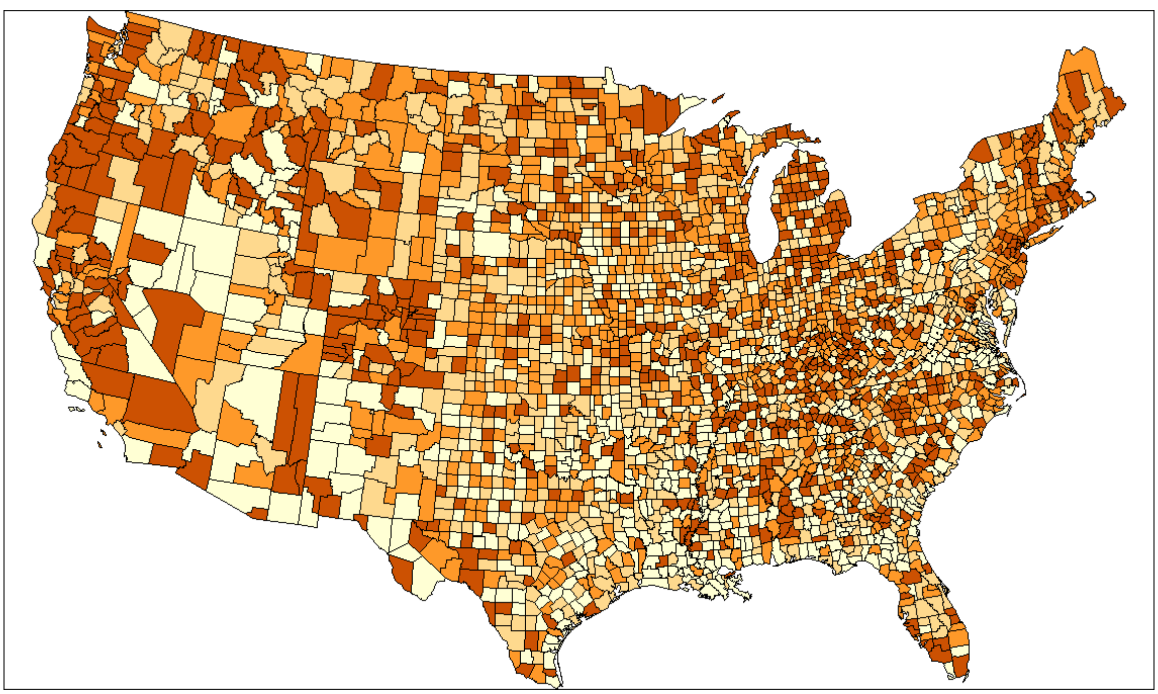

Figure 26. Residual Map for Spatial Lag Model, June 2009.

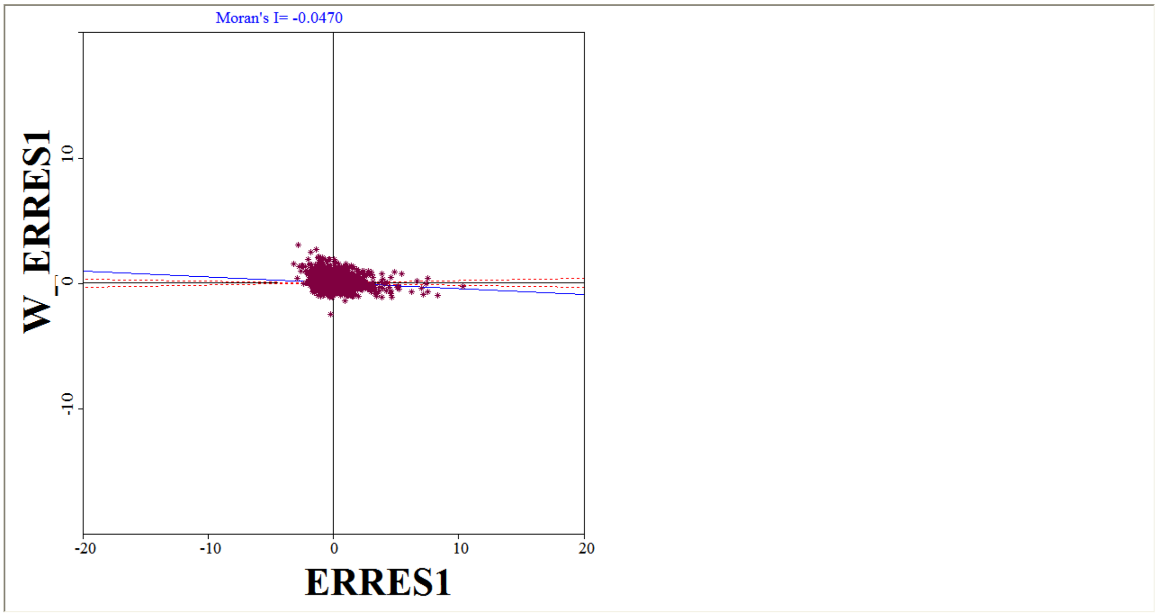

Figure 27. Moran’s I for Spatial Error Model, November 2001.

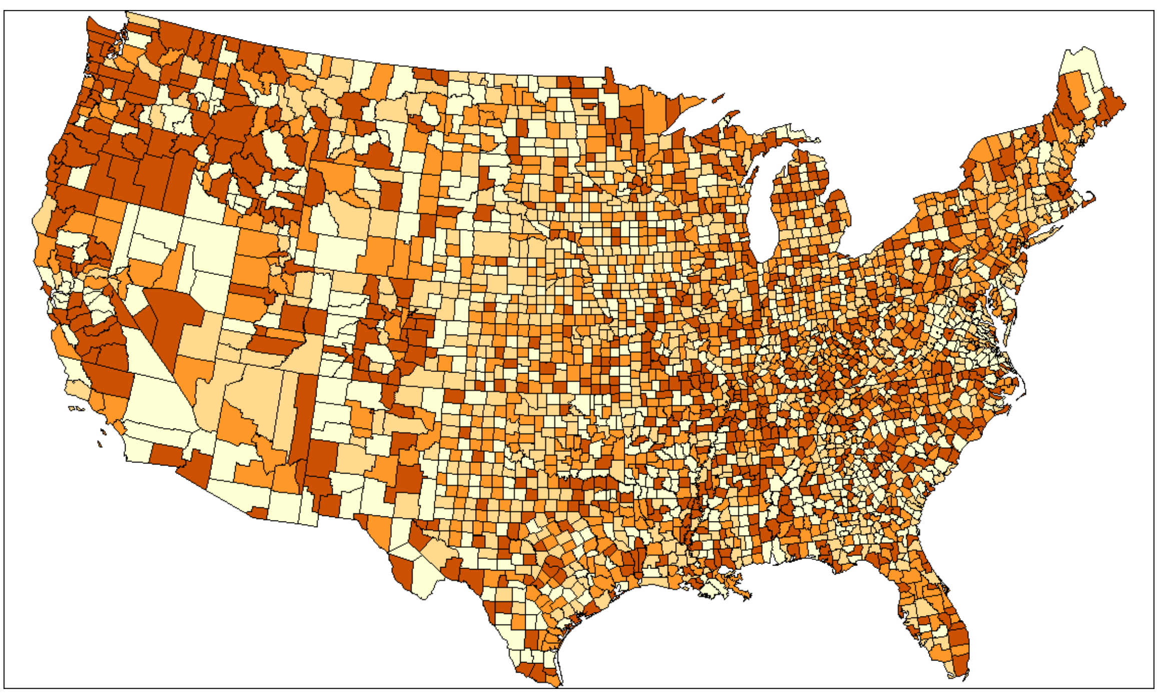

Figure 28. Residual Map for Spatial Error Model, November 2001.

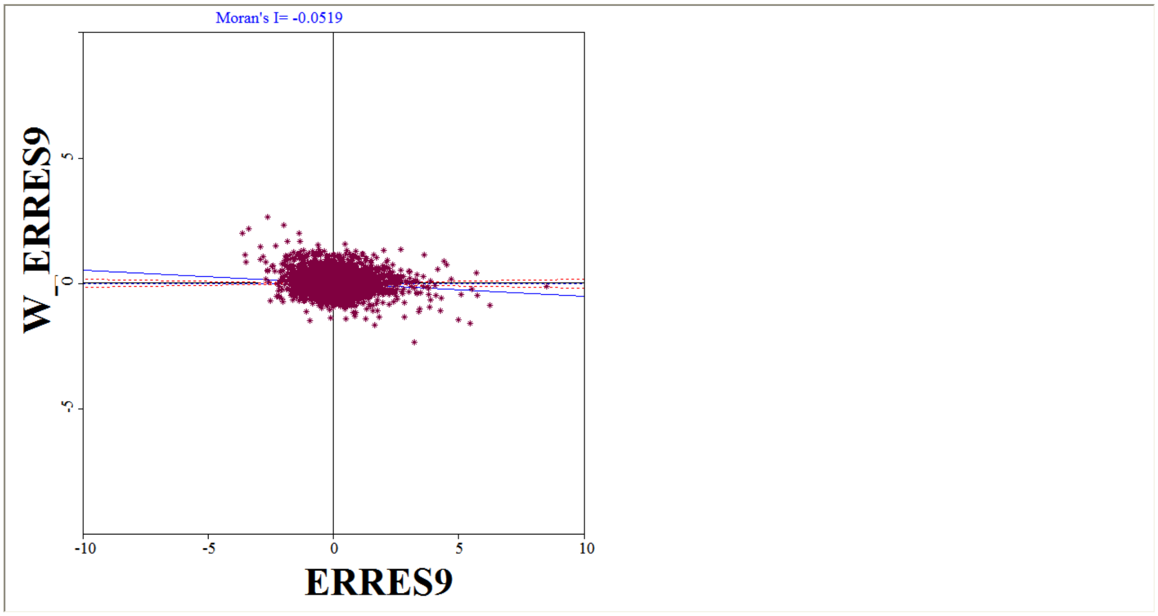

Figure 29. Moran’s I for Spatial Error Model, June 2009.

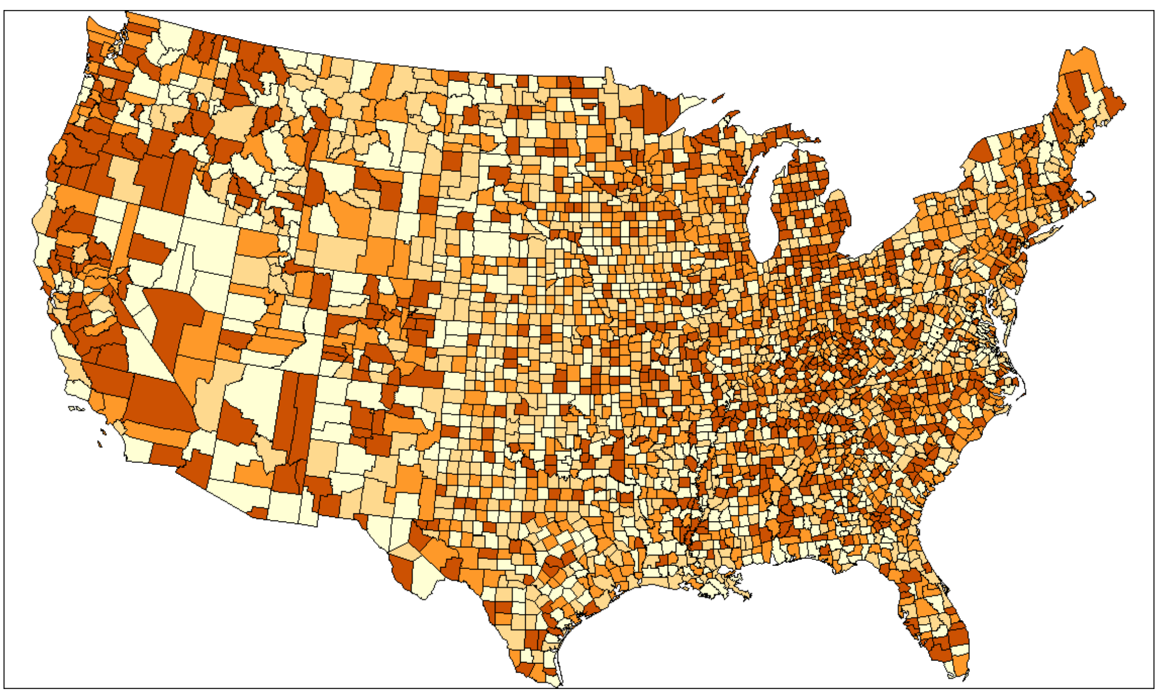

Figure 30. Residual Map for Spatial Error Model, June 2009.

Paste to the top of page 99

Step Three Results: Multilevel Modeling

Because of the limitations imposed by regression methods regarding the requirement of independent random errors and non-constant variance (heteroskedasticity), this step of the research design uses multilevel modeling to further examine the determinants of county unemployment rate variation. The two-level model structure used in multilevel modeling avoids these pitfalls since the models are not estimated with ordinary least squares, but rather with an iterative maximum likelihood procedure as detailed in Raudenbush and Bryk (2002). The advantage of using multilevel modeling here is twofold: 1) it may (or may not) confirm some of the spatial regression results and 2) it provides a deeper insight into the June 2009 study period by examining the effect of a state-level variable, namely, the top states for subprime mortgages (LP1K08D), into the analysis. Data for subprime mortgages were only available at the state level for this study, hence it is used here as a level-2 predictor variable. In the multilevel models presented here, estimation is performed with HLM 6 software using full maximum likelihood, (rather than restricted maximum likelihood).

The county data (level-1) used here have a hierarchical relationship with state level groupings (level-2). The aim of the HLM modeling presented here is to examine how individual county-level variables measuring race (% non-White: NONWHT07), self-employment (% self-employed: SELEMP00), educational attainment (% with a college degree: EDUC00), and industry mix (a binary variable indicating whether the county is manufacturing-dependent or not, 1 = yes, 2 = no: USDAMANF) affect county unemployment rate levels after controlling for the state level effects. Also, a group-level dichotomous variable indicating whether a state contained a high-level of subprime mortgages in 2008 (top 8 states in subprime mortgages/1000 residents, 1 = yes, 0 = no: LP1K08D) is used to exploe the effect of subprime mortgages upon unemployment rates for June 2009.[10] Three HLM models are specified: 1) an *empty null* model, 2) a *random coefficients* model, and 3) an *intercept and slopes as outcomes* model.

The analysis begins with the empty null model in order to 1) compute the reliability of using multilevel modeling, 2) compute the deviance figure which serves as a model fit comparison, and 3) compute the intraclass correlation coefficient (ICC) which confirms the need for multilevel analysis. The empty null model is expressed with the following equations:

$$\text{Level-1 model: } \mathrm{UR\_JUN09}_{ij} = \beta_{0j} + r_{ij} \tag{9}$$

$$\text{Level-2 model: } \beta_{0j} = \gamma_{00} + u_{0j} \tag{10}$$

$$\text{Combined Model: } \mathrm{UR\_JUN09}_{ij} = \gamma_{00} + u_{0j} + r_{ij} \tag{11}$$

The results for Model 1, the empty null model, are shown in Table 10. The average county unemployment rate for June 2009 is 9.20% (G00), which is significantly different from zero. The ICC value is calculated based on the variance components values and can be computed as: 5.72482/ (5.72482 + 4.92512) = 0.53754. This is a high value indicating that state-level effects account for 54% of the total variance in county

---

[10] The top eight states for subprime mortgages per 1,000 residents, as of August 2000, were: AZ, CA, CT, FL, MI, NV, RI, TN.

unemployment rates for June 2009. In other words, 54% of the total variance in county unemployment rates occurs *between* states. This high value validates the use of multilevel modeling. The reliability estimate measures the adequacy of the data structure, specifically, the reliability of the level 1 intercepts across the 49 states and the differences in outcomes between groups. For this study, the reliability estimate is very high at 96%. The deviance is a measure of model fit and is equivalent to the - 2 log likelihood figure (- 2LL). It is used both to measure model fit and to assess the effects of adding more predictors using a chi-squared test. The HLM 6 software automatically computes the significance of new models with additional predictors to help ensure model parsimony. According to the results, the baseline deviance figure for the null model is 13951.77.

Table 10. Multilevel Model Results for Model 1 (The outcome variable is UR_JUN09)

Final Estimation of Fixed Effects:					
Fixed Effect	Coefficient	Std. Error	T-ratio	d.f.	P-val.
INTRCPT1, B0					
INTRCPT2, G00	9.196407	0.348454	26.392	48	0
Final Estimation of Variance Components:					
Random Effect	Standard Deviation	Variance Component	Chi_Square	d.f.	P-val.
INTRCPT1 (U0)	2.39266	5.72482	3866.4268	48	0
level-1 (R)	2.21926	4.92512			
Note: Deviance = 13951.771; Number of Estimated Parameters = 3					

The next model to be estimated, Model 2, is a *random coefficients* model with four level-1 predictors: SELEMP00, EDUC00, NONWHT07, and USDAMANF. There are no level-2 predictors in the *random coefficients* model. The results are shown in Table 11. The estimate for the variance of the slopes for SELEMP00, EDUC00, and NONWHT07 are 72.33, 39.33 and 9.33, respectively, and they are all highly significant. Therefore, the null hypothesis of no differences in slopes among states for these predictors is rejected.

The "Fixed Effects" output from HLM6 also appears in Table 11. The INTRCPT2 (G00) value of 8.95 shows the average unemployment rate for June 2009 with no predictors. The variables SELEMP00, EDUC00, and NONWHT07 represent percentage values ranging from 0.00 to 1.00 (0% to 100%). As such, their interpretation must be contextualized. The higher the slope, the stronger the relation of county unemployment rate to these variables. The proper interpretation of their values is as follows: The INTRCPT2 (G10) value of - 15.25 indicates the expected decrease in the unemployment rate when SELEMP00 increases by 1 unit. Accounting for the variable structure, the results for SELEMP00 can be interpreted to mean that every percentage point increase in the percentage of self-employed in a county lowers the county unemployment rate by 0.15% (15 basis points). Similarly, the G20 value for EDUC00 of - 11.23 means that every percentage point increase in the percent of college educated in a county lowers the county unemployment rate by 0.11% (11 basis points). Likewise, the G30 value for NONWHT07 of 2.99 means that every percentage point increase in the

Table 11. Multilevel Results for Model 2 (UR_JUN09)

Final Estimation of Fixed Effects: Fixed Effect	Coefficient	Std. Error	T-ratio	d.f.	P-val.
For INTRCPT1, B0					
INTRCPT2, G00	8.956935	0.325321	27.533	48	0.000
For SELEMP00 slope, B1					
INTRCPT2, G10	-15.24904	1.673946	-9.11	48	0.000
For EDUC00 slope, B2					
INTRCPT2, G20	-11.25813	1.068465	-10.537	48	0.000
For NONWHT07 slope, B3					
INTRCPT2, G30	2.988599	0.609238	4.905	48	0.000
For USDAMANF slope, B4					
INTRCPT2, G40	0.841886	0.118065	7.131	48	0.000
Final Estimation of Variance Components: Random Effect	Standard Deviation	Variance Component	Chi_Square	d.f.	P-val.
INTRCPT1, U0	2.24015	5.72482	2976.4295	39	0.000
SELEMP00 slope, U1	8.50458	72.32787	131.13697	39	0.000
EDUC00 slope, U2	6.27113	39.3271	134.91195	39	0.000
NONWHT07 slope, U3	3.05487	9.33221	106.53375	39	0.000
USDAMANF slope, U4	0.49026	0.24035	71.65392	39	0.001
level-1, R	1.76272	3.10717			
Note: Deviance = 12646.665561; Number of Estimated Parameters = 21 Chi-square statistic = 1303.10494; P-value = 0.000					

non-white population increases the county unemployment rate by 0.03%. All of these values are significant. The G40 value for the categorical variable USDAMANF is 0.84, which means that the county unemployment rate is on average 0.84% higher in manufacturing-dependent counties than non manufacturing-dependent counties. This figure is also significant.

The estimated variance components are difficult to interpret in absolute terms though they do provide contextual meaning. The variance components output shows the random effects for the level-2 intercept error term (U0), the error term for the level 2 slopes (U1 through U4), and the error term for the level-1 equation (R). U0 is significant, which simply means that county unemployment rates vary significantly between states. The values U1 through U4 estimate the slopes of the level-1 predictor variables, namely, SELEMP00, EDUC00, NONWHT07 and USDAMANF. These are all significant, so it can be said that there is significant variance in slopes between states for all of these variables and that their relationship to the county unemployment rate differs between states (Figures 31-34).

The variance (R) of the June 2009 county unemployment rate is 3.11, after controlling for SELEMP00, EDUC00, NONWHT07 and USDAMANF. In the intercept-only null model it was 4.93. The difference is 1.82. By dividing this difference by the original variance (1.82/4.93), it can be seen that adding these level-1 predictors to the model reduces the within state variance by 36.9%. Also, adding these predictors to the model dropped the deviance value from 13,952 to 12,647, and this difference was significant, meaning that adding the level-1 predictors improved the model fit significantly.

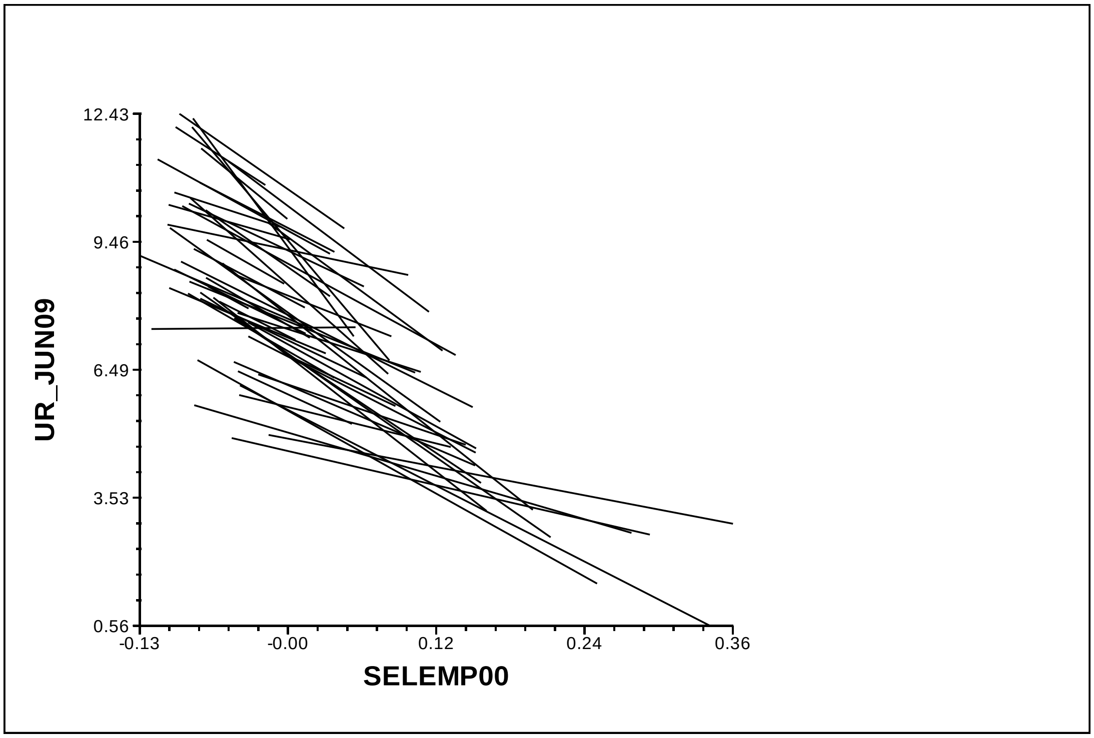

Figure 31. The Slope and Intercept of SELEMP00.

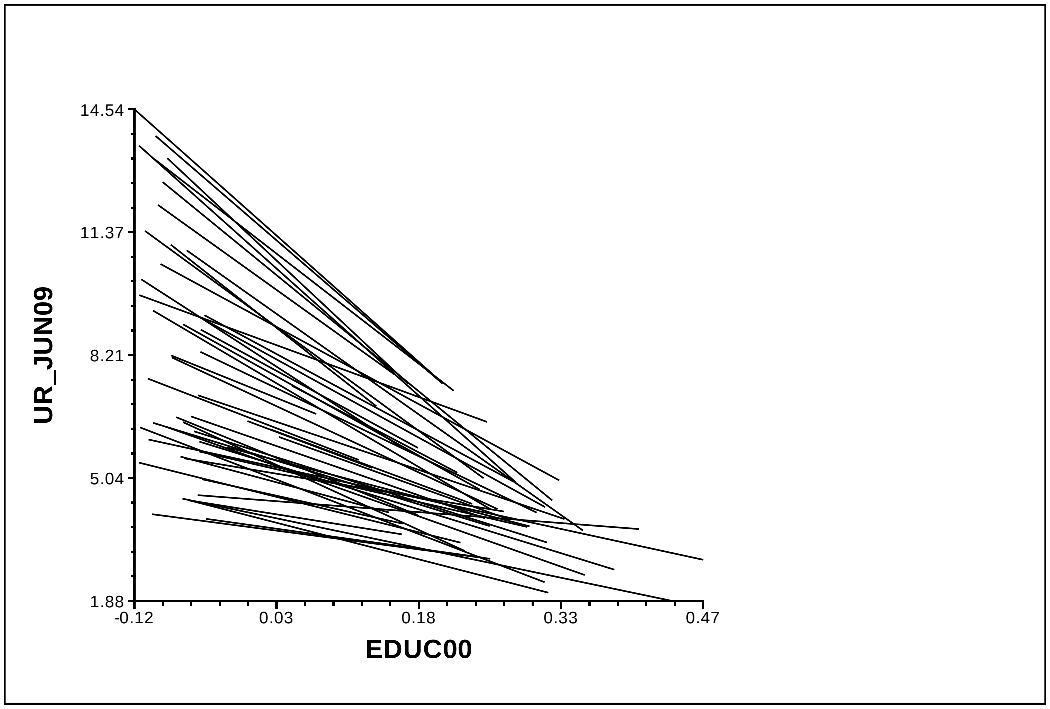

Figure 32. The Slope and Intercept of EDUC00.

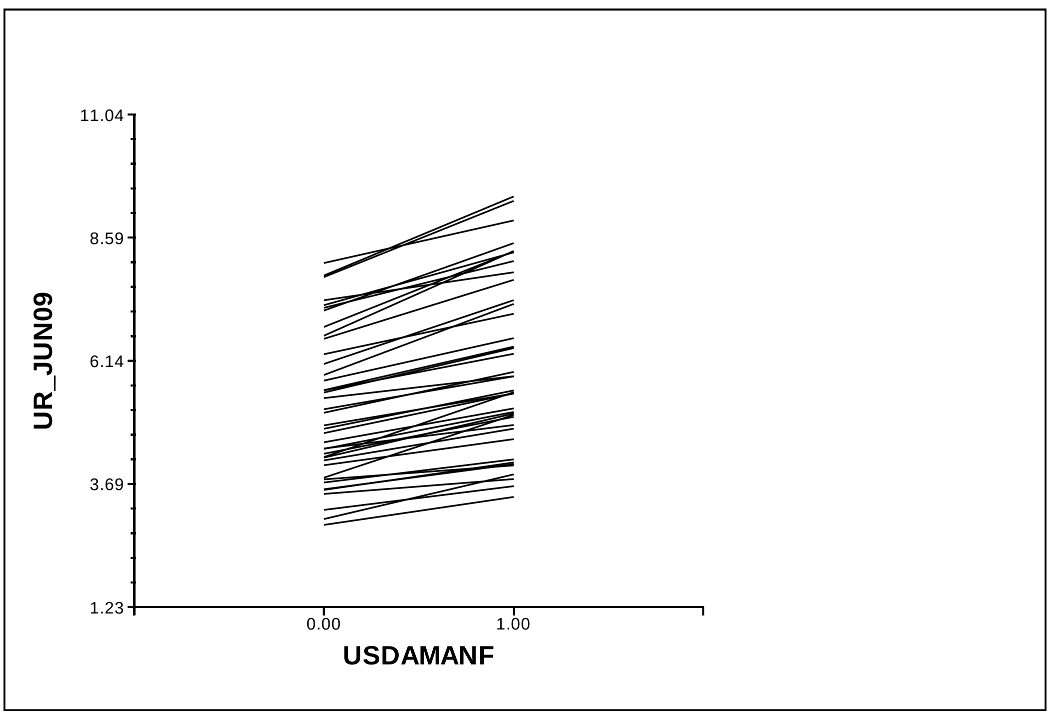

Figure 33. Bar Graph for the Binary Variable, USDAMANF.

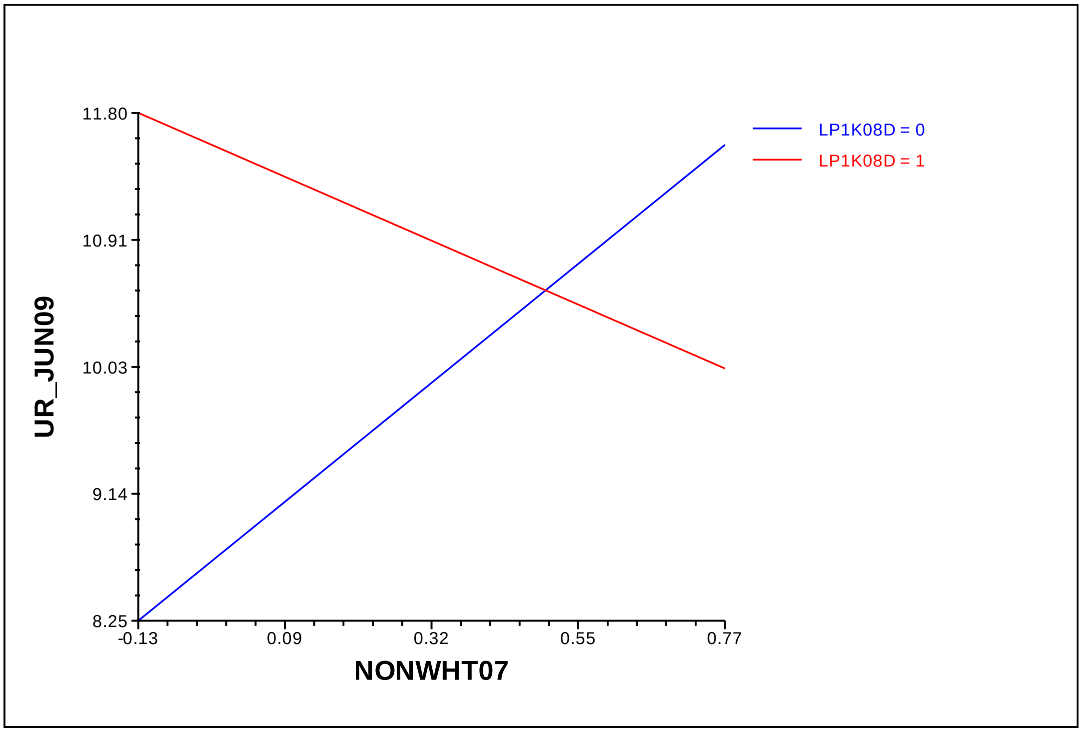

Figure 34. The Slope and Intercept of NONWHT07

The results of Model 3, the *intercept and slopes as outcomes* model appears in Table 12. The “Fixed Effects" sections contains five sets of coefficients. The first set shows the coefficients associated with the level 1 intercepts of county unemployment rates for June 2009, as predicted by SELEMP00, EDUC00, NONWHT07 and USDAMANF. The G01 coefficient indicates the average unemployment rate of 8.5%. The state-level binary variable, LP1K08D, indicates whether or not a county was located within one of the top states for subprime mortgages during 2008 (1 = yes, 0 = no). The G01 value for this variable is 2.84, which means that when LP1K08D increases by 1 unit, i.e., changes to a top subprime state, the unemployment rate increases 2.84%.

The second set of coefficients is associated with the level-1 slope of SELEMP00 as a predictor of county unemployment rates for June 2009. The G10 coefficient is the estimated mean slope of SELEMP00 across all states when LP1K08D is 0. The G10 coefficient value of - 14.29 means that when SELEMP00 increases by one percent, while holding LP1K08D at 0, the unemployment rate decreases by .14 percent. The G20 coefficient value of - 10.77 means that when EDUC00 increases by one percent, while holding LP1K08D at 0, the unemployment rate decreases by .11 percent. The G30 coefficient value of 3.67 means that when NONWHT07 increases by one percent, while holding LP1K08D at 0, the unemployment rate increases by .037 percent. Finally, the G40 coefficient value of 0.82 for the binary variable USDAMANF means that when USDAMANF increases from 0 to 1 (non manufacturing-dependent to manufacturing-dependent), while holding LP1K08D at 0, the unemployment rate increases by .82 percent. The G11, G21 and G41 coefficients were not significant. However, the G31

Table 12. Multilevel Model Results for Model 3 (UR_JUN09)

Final Estimation of Fixed Effects:					
Fixed Effect	Coefficient	Std. Error	T-ratio	d.f.	P-val.
For INTRCPT1, B0					
INTRCPT2, G00	8.498784	0.32245	26.357	47	0.000
LP1K08D, G01	2.840873	0.697076	4.075	47	0.000
For SELEMP00 slope, B1					
INTRCPT2, G10	-14.28909	1.564514	-9.133	47	0.000
LP1K08D, G11	-8.705074	8.271362	-1.052	47	0.298
For EDUC00 slope, B2					
INTRCPT2, G20	-10.7733	1.063043	-10.134	47	0.000
LP1K08D, G21	-2.436136	4.230893	-0.576	47	0.567
For NONWHT07 slope, B3					
INTRCPT2, G30	3.669435	0.581433	6.311	47	0.000
LP1K08D, G31	-5.640612	1.642373	-3.434	47	0.002
For USDAMANF slope, B4					
INTRCPT2, G40	0.822113	0.131409	6.256	47	0.000
INTRCPT2, G40	-0.132152	0.227607	-0.581	47	0.564
Final Estimation of Variance Components:					
Random Effect	Standard Deviation	Variance Component	Chi_Square	d.f.	P-val.
INTRCPT1, U0	1.97806	3.91274	2324.2675	38	0.000
SELEMP00 slope, U1	7.64015	58.37195	129.70131	38	0.000
EDUC00 slope, U2	6.21445	38.61936	129.11065	38	0.000
NONWHT07 slope, U3	2.63200	6.92742	99.42409	38	0.000
USDAMANF slope, U4	0.50032	0.25032	72.42051	38	0.001
level-1, R	1.76192	3.10438			

Note: Deviance = 12623.315132; Number of Estimated Parameters = 26
Chi-square statistic = 23.35043; P-value = 0.001

coefficient of - 5.64 was significant, which means that as LP1K08D increases 1 unit (changing from a non subprime state (LP1K08D = 0) to a subprime state (LP1K08D = 1)), the slope of NONWHT07 decreases by 5.64 units. This means that, controlling for other variables in the model, the relation of NONWHT07 to the unemployment rate is stronger in non subprime states.

Recall that the deviance value produced by Model 3 was 12, 623, and that this statistic equals - 2 times the log-likelihood value. If the log-likelihood value from the June 2009 spatial regression model (- 6,218) is converted to the deviance statistic, its deviance value becomes 12,436. This is approximate to the Model 3 deviance value of 12, 623. Therefore, multilevel modeling has produced a similarly robust explanatory model for county unemployment rate variation in a more parsimonious fashion by using four less explanatory variables.

In summary, multilevel modeling shows that 54% of the total variance in county unemployment rates occurs *between* states. This indicates that state level factors and policies have a strong influence upon county unemployment rate variation. Furthermore, multilevel modeling has confirmed that as the percentage of self-employed and college educated increases in a county, county unemployment rates decline. It also confirms that the percentage of non-white population has a positive effect upon unemployment rates, and that counties which are manufacturing-dependent have higher unemployment rates.

Step Four: Binary Logistic Regression

Binary logistic regression (BLR) is used to calculate odds ratios (OR) as measures of an association between selected explanatory variables and a binary dependent variable. The dependent variable for the BLR analysis presented here is a dichotomous categorical variable (BLRJUN09) indicating whether a county belongs to the upper quartile of counties as measured by the unemployment rate (0 = no, 1 = yes). The purpose of the BLR analysis presented here is to help answer this study's fourth research question: Can counties with a high risk of belonging to the upper quartile of high-unemployment counties near the end of a recession be statistically predicted using racial, socioeconomic and industry mix variables?

Since the goal of the BLR analysis is to develop a predictive and generalizable model for identifying counties belonging to high-unemployment clusters at the end of a recession, the education variable included in the BLR models is more current and time-specific than those used in the spatial regression and multilevel models. However, it is a state-level variable rather than a county-level variable since county-level data for educational attainment is not available for the 2007-2009 recession across all counties (EDUC07D – Figure 35).[11] Other explanatory variables used in the BLR model include a binary variable indicating whether or not a county is manufacturing-dependent

[11] The EDUC07D states include the District of Columbia, CO, CT, MD, MA, NH, NJ, VT and VA.

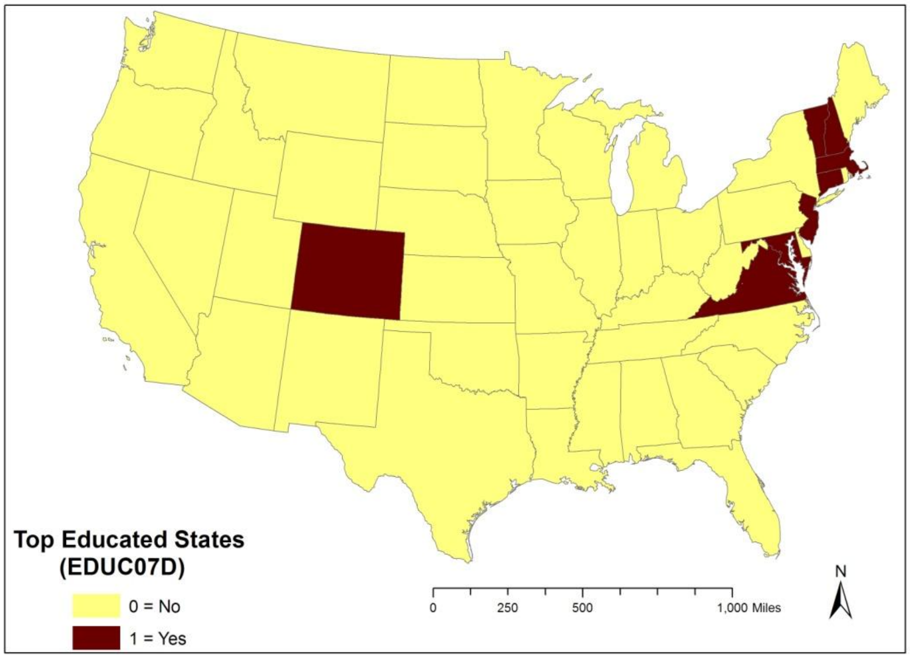

Figure 35. Top Educated States (EDUC07D

(USDAMANF: 1 = yes, 0 = no); a categorical variable indicating which census division a county is located in (CENDVSN – Figure 36); a continuous variable for per capita income (PCPI081K); a continuous variable measuring the percentage of self-employed living within a county multiplied by 10 for interpretative purposes (SELEMPX00_x10); and a path dependent variable used to test the relationship between current county unemployment rate levels and past levels (UR_NOV01). The dependent variable for the model is a binary variable indicating whether or not a county is in the upper quartile of counties as measured by the unemployment rate (BLRJUN09 – Figure 37).

It should be noted that the variable indicating the percentage of subprime mortgages per 1,000 housing units (LP1K08D) was not used for the BLR models since the purpose of these BLR runs is predictive and, thus, its inclusion would prevent the results from being generalized to other recessions since the LP1K08D variable has theoretical applicability only to the recession of 2007-2009. However, it is also noted that, a priori, its inclusion would indeed have increased the predictive power of the June 2009 model (Figure 38).

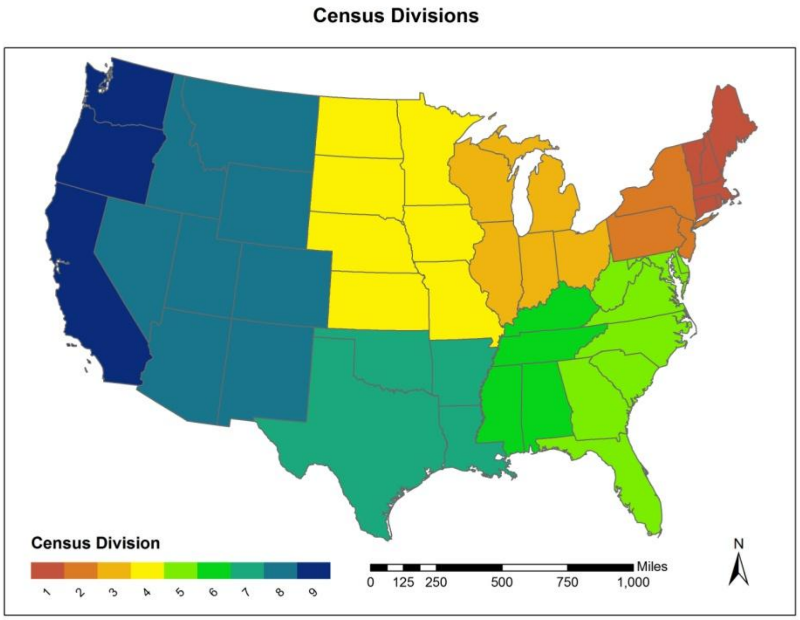

Figure 36. U.S. Census Divisions. Source: U.S. Census Bureau, 2011.

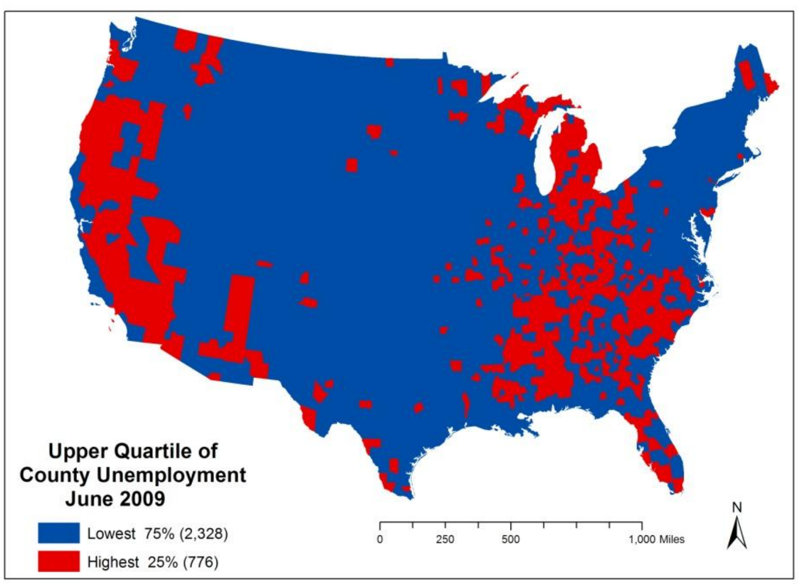

Figure 37. Upper Quartile of County Unemployment, June 2009 (JUN09QTR).

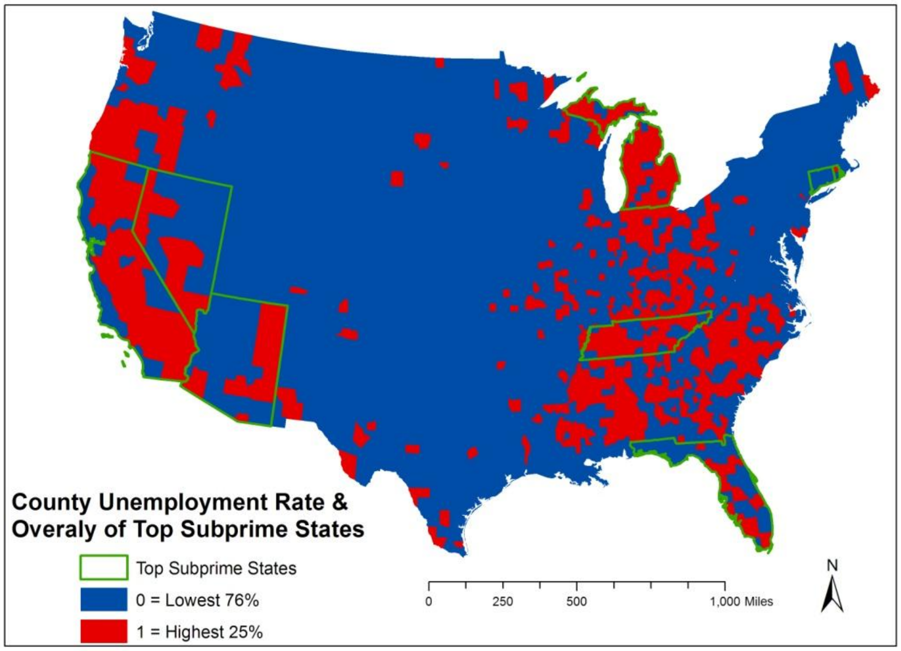

Figure 38. County Unemployment Rate and Top Subprime Mortgage State Overlay.

The results of the BLR model are shown in Table 13. The Hosmer and Lemeshow Test is a goodness-of-fit metric. If it is greater than .05, then the null hypothesis that there is no difference between the observed and predicted values of the outcome variable is rejected. This statistic was not significant for the June 2009 period (0.083), thus the data fit the model at an acceptable and significant level. However, the Hosmer and Lemeshow test does not help us to understand the overall explanatory value of the model. For this, the Nagelkerke R^2 value is used as it approximates the R^2 in a multiple regression. The Nagelkerke value for June 2009 is 0.586, which indicates that the BLR model has a moderately strong level of explanatory power.

The classification table, shown at the bottom of Table 13, compares the predicted values for the outcome variable against the actual values from the data set. Of the 3,104 counties modeled for the June 2009 period, 776 counties (25%) belong in the upper quartile as measured by the unemployment rate, while 2,328 (75%) are not. For comparison purposes, if one were to blindly predict that every county was not in the upper quartile, that prediction would be correct 75% of the time (2,328/3,104). The lower half of the classification table shows the predictive capability of the June 2009 BLR model when the independent variables are included. All but 156 of the 2,328 counties not belonging to the upper-quartile are correctly predicted (93.3%). Of the upper quartile counties, 518 of the 776 counties are correctly predicted (66.8%). The overall percentage of correctly predicted counties is 86.7%, which is a moderately strong improvement over the percentage a blind estimate would yield (75.0%). Thus, it can be said that the BLR model does indeed differentiate between counties which do and do not belong in the upper quartile of high unemployment counties.

The independent variables which help make the BLR model successful are shown in Table 13. Included for each variable are measures for the B coefficient, standard error of B, Wald statistic, significance of the Wald statistic, the exponentiated B coefficient (Exp(B)), and the upper and lower limits of a 95% confidence interval for each odds ratio. The B coefficient represents the relationship between the predictor variables and the dependent variable. The Wald statistic is the ratio of B to the standard error, squared.

Table 13. Results of the Binary Logistic Regression (upper quartile of high unemployment counties, June 2009)

Variables in the Equation					
DV=JUN09QTR	B	S.E.	Wald	Sig.	Exp(B)
EDUC07d(1)	-0.987	0.293	11.348	0.001	0.373
USDAmanf(1)	1.061	0.122	75.888	0.000	2.890
UR_NOV01	0.607	0.044	191.937	0.000	1.836
SELEMP00_x10	-1.190	0.174	46.679	0.000	0.304
PCPI081K	-0.098	0.013	55.709	0.000	0.906
CenDvsn			198.677	0.000	
CenDvsn(1)	-0.686	0.722	0.904	0.342	0.503
CenDvsn(2)	-2.466	0.468	27.818	0.000	0.085
CenDvsn(3)	0.312	0.289	1.165	0.280	1.367
CenDvsn(4)	-0.965	0.331	8.487	0.004	0.381
CenDvsn(5)	-0.415	0.306	1.847	0.174	0.660
CenDvsn(6)	0.203	0.298	0.467	0.495	1.225
CenDvsn(7)	-2.665	0.346	59.169	0.000	0.070
CenDvsn(8)	-0.985	0.357	7.618	0.006	0.374
Intercept	0.222	0.752	0.087	0.768	1.248
Diagnostics					
Hosmer & Lemeshow = 0.083		Nagelkerke R-Sq. = 0.586		N = 3104	
Confidence Intervals (95%)					
DV=JUN09QTR	Lower CI	Upper CI	Sig.		
EDUC07d(1)	0.210	0.663	0.001		
USDAmanf(1)	2.276	3.689	0.000		
UR_NOV01	1.685	2.003	0.000		
SELEMP00_x10	0.216	0.436	0.000		
PCPI081K	0.883	0.93	0.000		
Classification Table		**Predicted** 0 (Not High)	**Predicted** 1 (High Cluster)	**Percent Correct**	
JUN09QTR	0	2328	0	100.0%	
(Blind Model)	1	776	0	0.0%	
	Overall			75.0%	
JUN09QTR	0	2172	156	93.3%	
(Final Model)	1	258	518	66.8%	
	Overall			86.7%	

The corresponding significance level tests the significance of each predictor variable in the model. The Exp(B) measure is also known as the odds ratio (OR), which is defined as "the increase (or decrease if < 1.0) in odds of being in one outcome category when the value of the predictor increase by one unit" (Tabachnick and Fidell, 2001). An OR less than 1 corresponds to decreases in odds, while an OR more than 1 corresponds to increases in odds. An odds ratio near 1 indicates that a unit change in the predictor variable has little or no effect upon the dependent variable. Interpretation of the odds ratio differs slightly between continuous and categorical variables. For continuous variables the OR represents the increase or decrease of the odds for each unit increase in the predictor variable. For dichotomous categorical variables, the OR compares the odds for the two categories. For categorical variables with more than one category, such as the one used here to differentiate the nine census divisions (CENDVN), the OR compares the odds between each category (census division) and the reference category (the West division containing California, Oregon and Washington). The confidence interval for each OR is reported in the results tables. These show an upper and lower value within which we can be 95% sure includes the true value of the OR. The range of a significant variable will not include the value 1.0, which would indicate an equal chance of belonging, or not, to a high-unemployment cluster.

For example, for an independent dichotomous variable having an OR of 2.0, the odds that the variable is associated with a county belonging to the upper quartile of high-unemployment counties would be twice as high for the independent outcome coded as 1 as the outcome coded as 0. For an independent continuous variable having an OR of 2.0, a one-unit increase in the independent variable increases the odds of belonging to a high-unemployment cluster by a factor of 2. The interpretation of the results assumes that the values of the other explanatory variables are held constant.

For the June 2009 model, all of the predictor variables are significant at the $p < .001$ level. The continuous variables in the BLR model are UR_NOV01, SELEMP00_x10, and PCPI081K, while the categorical variables are EDUC07d and USDAMANF. The UR_NOV01 variable has an odds ratio of 1.836, which indicates that as the path dependent unemployment rate as measured from the end of the previous recession increases by one perent, the odds of a county belonging to the upper quartile for the June 2009 period increases by a factor of 1.836. The odds ratio (0.304) for the self-employment variable indicates that as the percentage of a county's self employed increases by a tenth of a percent, the odds of it belonging to the upper quartile decreases by 69.6% (1 - .304). The impact of per capita income upon the model is not as strong since the odds ratio of .906 indicates that as county's per capita income increases by $1,000, the odds of that county belonging to the upper quartile decreases by only 9.4% (1 - .906). The categorical variables in the BLR model are easier to interpret. The odds ratio of 0.373 for EDUC07d indicates that counties located in highly educated states are 62.7% less likely to belong in the upper quartile of counties as measured by the unemployment rate (1 – 0.373). The odds ratio of 2.89 for USDMANF indicates that manufacturing-dependent counties are nearly 3 times more likely to belong in the upper quartile.

The BLR model correctly predicts 86.7% of all counties as to whether or not they belong to the upper quartile of counties, as measured by the unemployment rate. Counties located in highly-educated states are 62.7% less likely to belong in the upper quartile. Similarly, as the percentage of self-employed increases, the chance of belonging to the upper quartile decreases. Conversely, manufacturing-dependent counties are almost three times as likely to belong in the upper quartile. Finally, as the path dependent unemployment rate for each county increase by one percent, the odds of a county belonging to the upper quartile increases by a factor of 1.836.

Conclusion

In summary, the results of the spatial analyses contained herein confirmed most, but not all, of this paper's hypotheses. LISA cluster analysis proves the existence of spatially-dependent clusters throughout the United States for both study periods and, more importantly, many of these clusters persist over time as evidenced by the 278 counties where clusters of high unemployment appear for both study years (H1). The spatial regression models found a positive and significant relationship to unemployment found for the average non-white population of a county, while significant negative relationships were found for the percentage of self-employed and educational attainment (H2). The spatial lag and spatial error models also found that county location plays a significant role as a determinant of unemployment rates (H3). Also, manufacturing-dependent counties were found to have a significantly positive relationship to unemployment (H4), while the percentage of the workforce employed in the government sector was not significant in three out of the four spatial regression models. Furthermore, the relationship between the unemployment rate and farming-dependent counties were mixed: they were not significant in the spatial error models, but were significant and negative for the spatial lag models. The multilevel models found that county 54% of the variance in county unemployment rates occurs across states, indicating that state-level factors exert either upward or downward pressure upon unemployment rates across the entire group of counties within their borders (H5). It also showed that location within one of the top subprime mortgage states had a significant and positive relationship to county unemployment rate variation (H6). Finally, a successful model was developed with binary logistic regression for predicting whether or not counties belonged to the upper quartile of high-unemployment counties. This model successfully predicted 86.7% of all 3,104 counties.

CHAPTER 5

DISCUSSION AND CONCLUSION

Introduction

In a recent paper about the 2007-2009 economic crisis, geographer Ron Martin concluded: "What the paper also suggests is that the geographical study of economic crises and crashes, both past and present, is an area calling for more research" (Martin, 2010). This paper attempts to heed this call by examining the geographical distribution and determinants of county unemployment rates for June 2009 and November 2001. Four main research questions were asked about county unemployment rates, relating to their 1) spatial patterns, 2) determinants, 3) relationship with state-level factors, and 4) predictability. These questions were examined using a four-step research design, the results of which were presented in the previous chapter. This chapter summarizes and contextualizes these findings. The first section of this chapter discusses the spatial patterning of high unemployment county clusters for the study periods June 2009 and November 2001. This is followed by a discussion which compares and contrasts the findings for the two different study periods relating to unemployment rate determinants. Next, the state-level effects upon county unemployment rates for the June 2009 are examined. This is followed with a discussion about the predictability of high-unemployment clusters. Finally, the intellectual merit and broader impacts of this study are assessed and suggestions for future research are proposed.

Major Findings

Using geostatistical techniques applied to a large-scale database, this study has revealed five major findings: 1) geographic clustering of high-unemployment persists to a high degree at the county level in the U.S.; 2) counties that are manufacturing-dependent or that have a high percentage of non-whites show a positive relationship to the unemployment rate, while self-employment and educational attainment are negatively associated; 3) counties located in the top subprime mortgage states show a positive relationship to the unemployment rate for June 2009; 4) 54% of the variance in county unemployment rates occurs *across* states while 46% is due to *within*-state factors; and 5) high risk factors for belonging in the upper quartile of counties, as measured by the unemployment rate, include manufacturing-dependent counties and a high unemployment rate registered at the end of the previous recession. This discussion now turns towards explanations of these findings.

The LISA cluster analysis performed in Step One identified 278 counties in the combined high-high unemployment rate cluster for both June 2009 and November 2001. The average unemployment rate for these persistently high counties in June 2009 was 14.09%, versus the national county average of 9.34%. For analytical purposes, these counties were broken down further into five separate geographical clusters: the West, Michigan, Appalachia, the Carolinas, and the Deep South. Before invoking potential theoretical causes for these persistent clusters, some observations about the underlying data are warranted. As reported in the results, the Moran's I value of .42 for the June 2009 county unemployment rate is indicative of a high degree of spatial clustering. The results of the spatial regression models show that race, per capita income, self employment, manufacturing and, sometimes, farming, had significant relationships to the unemployment rate. Figure 39 shows the Moran's I values for these explanatory variables in 2009 (EDUC00 = .41; NONWHT07 = .69; PCPI_08 = .45; SELEMP00 = .69; USDAMANF = .39; USDAFARM = .31). All of these explanatory variables are themselves highly clustered, particularly race and self-employment. The point to be made is that if the statistically significant independent variables themselves are highly clustered, then it stands to reason that they in some way contribute to the clustering of the unemployment rate. For this reason, assigning a causal force for the clustering of counties with high unemployment is problematic, other than to note that it is the result of a multiple effects. This notwithstanding, the explanatory variables were chosen for theoretical purposes, thus it is possible to posit theoretical causes for the five identified individual clusters of high unemployment found in this study.

A glaringly obvious explanation for the existence of at least some of these clusters is the degree to which their industrial structure is comprised of a high number of manufacturing-dependent counties. Data from the BEA show that while the number of manufacturing jobs in the United States in 2007 was approximately 14.5 million, that figure is nearly 5 million less than manufacturing jobs existing in 1990. Over the past few decades, the forces of globalization have transformed the U.S. labor market from a manufacturing-based economy with a national scope to a service-based economy with a global scope. This global restructuring, accompanied by technological change, has increased global competitiveness with the result that many manufacturing jobs have left the U.S., many of them towards China and India. Four of the five identified unemployment clusters contain large levels of manufacturing-dependent counties, namely, Michigan (35%), Appalachia (32%), the Carolinas (70%) and the South (60%).

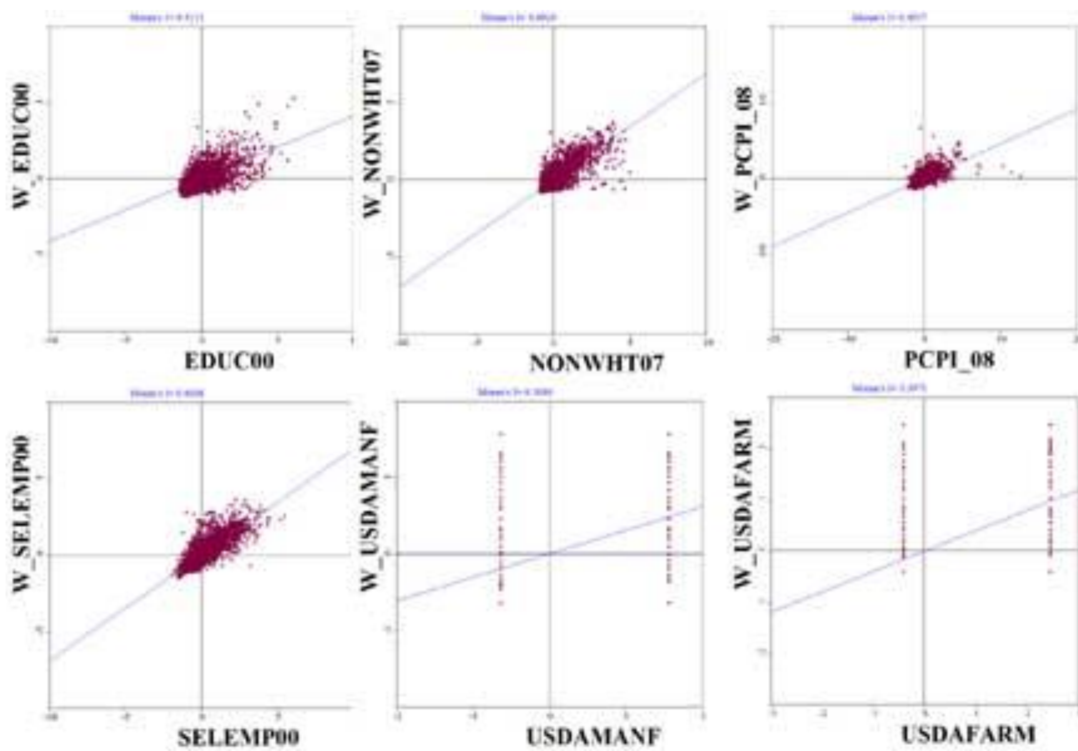

Figure 39. Moran's I for the independent variables: EDUC00 = .41; NONWHT07 = .69; PCPI_08 = .45; SELEMP00 = .69; USDAMANF = .39; USDAFARM = .31.

.

The changing global economy has also placed a premium upon education. Manufacturing jobs leave and service sector jobs replace them, but the latter generally require a more educated workforce to meet their staffing needs. Four of the five clusters are below the national county average for college-educated (16%): Michigan (13%), Appalachia (9%), the Carolinas (12% and the Deep South (11%). In the aggregate, only 12% of those residing in the red cluster had a college degree, versus 16% for the U.S. In summary, the change in the U.S. industrial structure brought about by the forces of globalization have caused persistent unemployment problems in most of the identified clusters since they largely remain reliant upon an "old economy" manufacturing base and an uneducated workforce.

Multilevel modeling was used to examine the influence of state-related factors upon county unemployment and to specifically examine the effects of subprime underwriting. The results show that 54% of the total variance in county unemployment rates occurs *between* states. This indicates that state level factors and policies have a strong influence upon county unemployment rate variation. Furthermore, multilevel modeling confirmed that the level of state subprime mortgages per resident had a positive relationship to county unemployment rates in those states. With regard to the subprime mortgage crisis of 2007-2009, geographer Ron Martin remarked this about the 2009 recession, "what became a global financial crisis had distinctly local origins (Martin, 2010). The findings of this paper support this proposition. Having tested the relationship between states having the highest number of subprime mortgages and the unemployment rate of counties in those states, the average county unemployment rate was indeed significantly higher there than for counties in states not heavily burdened by subprime underwritings. Martin further contends that “glocalization” was operationalized during the recent economic crisis. Glocalization theory posits that local factors affect global economic conditions, which in turn reverberate back to the local level. When the global economy grinded to a halt, the U.S. manufacturing sector, already reeling from decades of globalization effects, suffered more job losses. The results of this paper support glocalization theory in that it confirms the positive relationship between manufacturing-dependent counties and unemployment. In other words, the subprime mortgage crisis took root locally in the United States, then spread overseas causing a global economic crisis, the effects of which were in turn felt locally in the United States, particularly in the manufacturing sector.

Finally, this paper posits a predictive binary logistic regression model to identify counties belonging in the upper quartile of high-unemployment counties. This BLR model correctly predicts 86.7% of all such counties. Counties located in highly-educated states are 62.7% less likely to belong to the upper quartile. Similarly, as the percentage of self-employed increases, the chance of belonging to the upper quartile decreases. Conversely, manufacturing-dependent counties are almost three times as likely to belong in the upper quartile. Finally, as the path dependent unemployment rate for each county increases by one percent (as measured at the end of the previous recession), counties are nearly twice as likely to belong to the upper quartile.

Intellectual Merit

The intellectual merit of this thesis lies in its effort to examine economic activity from a geographic perspective. The methods used herein include the advanced techniques of spatial autocorrelation, spatial regression and multilevel modeling; techniques which have been used infrequently within U.S. unemployment rate studies. Their use is intended to broaden and deepen the literature on geography's role in the distribution of county-level unemployment rates by facilitating the understanding of how location-specific factors affect both unemployment rate disparities and spatial patterning. It is hoped that this examination of unemployment rate distribution during the past two recessions will reveal poorly understood aspects of county-level unemployment rates during times of economic duress by identifying the roles that geographic location and local factors play in unemployment rate disparities.

Broader Impact

Macroeconomists and policymakers alike have centered most of their attention on the national unemployment rate as a measure of labor market performance instead of the disparities and geographic patterns found at the local scale (Mollick, 2008). Rather than focusing upon the customary top-down investigative approach, whereby macroeconomic factors are used to explain unemployment rate levels, this thesis uses contemporary statistical methods, developed within the past twenty years (spatial autocorrelation, spatial regression and multilevel modeling), to examine unemployment rate levels geographically from the "bottom-up." The broader impact of this research lies in its potential to explain spatial relationships among U.S. counties relating to unemployment rate differentials and to uncover some of the determinants for these variations. Understanding the effects of geographical and local variables upon county-level unemployment rate variation aids the local economic development process by allowing the actors who play a role in shaping economic development to understand exactly where resources are needed. With such understanding, key actors can identify problem areas so that financial, managerial and governmental resources can be applied towards regional solutions.

Further Research

The nuances and determinants of unemployment rate variation are too complex to be deciphered in a singular paper. This research has focused on its spatial patterns as well as on some elementary explanatory variables. One of the glaring omissions of this work relates to the paucity of industrial mix data analyzed. This is a large-scale investigation whose methods required county contiguity for implementation. Industrial mix NAICS data as provided by the BEA contain too much data suppression, even at the 2-digit classification, to be useful here. The study of unemployment rate variation and its spatial patterning would certainly benefit from micro-scale studies able to process data relating to specific industrial sectors. Smaller-scale investigations focused upon service sector industries and a more extensive mining of the 3, 4 or 5-digit manufacturing classifications seems warranted.

BIBLIOGRAPHY

Anselin, L. (1988). *Spatial econometrics: methods and models*. Dordrecht, Netherlands: Kluwer.

Anselin, L. (1994). Exploratory spatial data analysis and geographic information systems. In M. Painho (Ed.), *New tools for spatial analysis* (pp. 45-54). Luxembourg: Eurostat.

Anselin, L. (1995). Local indicators of spatial association-LISA. *Geographic Analysis, 27,* 93-115.

Anselin, L. (1996). The Moran scatterplot as an ESDA tool to assess local instability in spatial association. In M. Fisher, H. Scholten & D. Unwin (Eds.) *Spatial analytical perspectives on GIS* (pp. 121-138). London: Taylor and Francis.

Anselin, L. (1999). Interactive techniques and exploratory spatial data analysis. In P.A. Longley, M.F. Goodchild, D.J. Maguire, & D.W. Wind (Eds.) *Geographical Information Systems: Principles, techniques, management and applications* (pp. 253-266). New York, NY: Wiley.

Anselin, L. (2003). *An introduction to spatial autocorrelation analysis with GeoDa*. GeoDa Documentation. URL: http://www.csiss.org/clearinghouse/GeoDa/.

Anselin, L. (2005). *Exploring spatial data with GeoDa: A Workbook*. Urbana-Champaign: University of Illinois, Department of Geography and Center for Spatially Integrated Social Science.

Anselin, L., & Florax, R. (1995). Small sample properties for spatial dependence in regression models: Some further results. In L. Anselin & R. Florax (Eds.), *New directions in spatial econometrics* (pp. 21-74). Berlin: Springer-Verlag.

Anselin L., Bera A., Florax R. & Yoon M. (1996) Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics*, *26*(1), 77-104.

Aragon, Y., Haughton, D., Haughton, J., Leconte, E., Malin, A., Ruiz-Gazen, A., & Thomas-Agnan, C. (2003). Explaining the pattern of regional unemployment: The case of the Midi-Pyrénées region. *Papers in Regional Science*, *82*, 155-174.

Audretsch, D.B., Carree, M.A. & Thurik, A. R. (2001), Does entrepreneurship reduce unemployment? *Discussion Paper* TI01-074/3, Tinbergen Institute, Erasmus University Rotterdam.

Baskaya, Y.S. & Mbiti, I.M. (2006). An empirical analysis of black-white employment differences over the business cycles. *White Paper*. Brown University. http://www.econ.brown.edu/students/Isaac_Mbiti/black_white_paper.pdf.

Berman, J. C. (2005). Industry output and employment projections to 2014. *Monthly Labor Review*, *128*(11), 45-69.

Blackaby, D. H., & Manning, D. N. (1992). Regional earnings and unemployment: A Simultaneous approach. *Oxford Bulletin of Economics and Statistics, 54*, 481-501.

Blanchard, O. & Katz, L. (1992). Regional evolutions. *Brookings Papers on Economic Activity, 1*, 1-75.

Blanchard, O. J. & Summers, L. H. (1987). Hysteresis in unemployment. *European Economic Review*, *31*, 288-295.

Blanchard, O. J. (1993). Consumption and the recession of 1990-1991. *American Economic Review, 83*(2): 270-275.

Blau, F., & Ferber, M. A. (1992). *The economics of women, men and work*. Englewood Cliffs, NJ: Prentice-Hall.

Breschi, S., & Lissoni, F. (2001) Knowledge spillovers and local innovation systems: A Critical survey. *Industrial and Corporate Change*, *10*(4), 975-1005.

Bureau of Economic Analysis (BEA). (2011). Local Area Personal Income. Retrieved October 1, 2011, from http://www.bea.gov/regional/reis/.

Bureau of Labor Statistics (BLS). (2009). *Occupational outlook handbook*, 2008-2009. Retrieved March 11, 2011. from http://www.bls.gov/oco/ocos176.htm.

Burridge, P., & Gordon, I. (1981). Unemployment in the British metropolitan labour areas. *Oxford Economic Papers, 33*, 274-297.

Caselli, F. & Coleman W. J. (2001). The U.S. structural transformation and regional convergence: A Reinterpretation. *Journal of Political Economy, 109*, 584-616.

Claessens, S., Kose, M. A. & Terrones, M. E. (2008). *Global financial crisis: How long? How deep?* VoxEU.org (October 7, 2008), Retrieved May 2, 2011, from http://www.voxeu.org/index.php?q=node/2083.

Clemente, J., Lanaspa, L., & Montanés, A. (2005). The unemployment structure of the U.S. states. *The Quarterly Review of Economics and Finance, 45*, 848-868.

Cliff, A. D., & Ord, J. K. (1969). The problem of spatial autocorrelation. In: A. J. Scott (Ed.) *London Papers in Regional Science* (pp. 25-55). London: Pion.

Cliff, A. D., & Ord, J. K. (1973). *Spatial autocorrelation*. London: Pion.

Cliff, A. D. & Ord, J. K. (1981). *Spatial processes: Models and applications*. London: Pion Limited.

Clogg, C. C., & Sullivan, T. A. (1983). Labor force composition and underemployment trends, 1969-1980. *Social Indicators Research, 12*, 117-152.

Cracollici, M. F., Cuffaro, M. & Nijkamp, P. (2007). Geographical distribution of unemployment: An Analysis of provincial differences in Italy. *Growth and Change, 38*(4), 649-670.

Crandall, R. (1988). The Public interest in metals policy. *Mineral Processing and Extractive Metallurgy Review, 3*(1-4), 307-322.

Cronon, W. (1991). *Nature's Metropolis: Chicago and the Great West*. New York, NY: W.W. Norton.

David, P. A. (1997). Path Dependence and the quest for historical economics: One more chorus in the ballad of QWERTY. University of Oxford Discussion Paper in Economic and Social History, Athens, Greece. Retrieved from: http://www.nuff.ox.ac.uk/economics/history/paper20/david3.pdf

Decressin, J. & Fatás, A. (1995). Regional Labor Market Dynamics in Europe, *European Economic Review, 39*(9), 1627-1655.

De Jong, G., & Madamba, A. (2001) A double disadvantage? Minority group, immigrant status and underemployment in the United States. *Social Science Quarterly, 82*(1),117-130.

Delicado, P. & Broner, S. (2008). Distance-based LISA maps for multivariate lattice data. *Technical Report*, Universitat Politµecnica de Catalunya, Preprint. Retrieved March 22, 2010, from http://hdl.handle.net/2117/2421.

Desmet, K., & Fafchamps, M. (2005). Changes in the spatial concentration of employment across US counties: A Sectoral analysis 1972-2000. *Journal of Economic Geography*, *5*(3), 261-284.

Dixon, R. J. & Thirlwall A. P. (1975). *Regional growth and unemployment in the United Kingdom*. London: MacMillan.

Drayse, M. H. (2008). Globalization and regional change in the U.S. furniture industry. *Growth and Change, 39*(2): 252-282.

Elhorst, J. P. (2003). The mystery of regional unemployment differentials: A survey of theoretical and empirical explanations, *Journal of Economic Survey, 17*(5), 709-748.

ERS/USDA (Econmic Research Service/U.S. Department of Agriculture. (2011). Measuring Rurality: 2004 County Typology Codes. Retrieved October 1, 2011, from http://www.ers.usda.gov/briefing/rurality/typology/.

Essletzbichler, J., & Rigby, D. L. (2007). Exploring evolutionary economic geographies. *Journal of Economic Geography*, *7*(5), 549-571.

Ewing, B. T., Levernier, W., & Malik, F. (2005). Modeling unemployment rates by race and gender: A Nonlinear time series approach. *Eastern Economic Journal*, *31*, 333-347.

Fairlie, R. B. & Sundstrom, W. A. (1999). The emergence, persistence, and recent widening of the racial unemployment gap. *Industrial and Labor Relations Review*, *52*, 252-270.

Federal Reserve Bank of St. Louis. (2011, October 11). Federal Reserve Economic Data. Retrieved October 1, 2011, from http://research.stlouisfed.org/fred2/.

Feser, E. J. (1998). Old and new theories of industry clusters. In M. Steiner (Ed.), *Clusters and regional specialisation: On Geography, technology and networks* (pp. 18-40). London: Pion.

Frenken, K., Van Oort, F. & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies, 41*(5*)*, 685-697.

Gardner, J. M. (1994). The 1990-91 recession: How bad was the labor market? *Monthly Labor Review, 117*(6), 3-11.

Getis, A. (2007). Reflections of spatial autocorrelation. *Regional Science and Urban Economics, 37*, 491-496.

Golpe A, & van Stel, A. (2008). Self-employment and unemployment in Spanish regions in the period 1979-2001. In E. Congregado (Ed), *Measuring entrepreneurship. Building a statistical system* (pp. 191-204). New York, NY: Springer.

Green, G. P. & Sanchez, L. (2007). Does manufacturing still matter? *Population Research and Policy Review*, *26*(5-6), 529-551.

Gripaios P., & Wiseman N. (1996). The impact of industrial structure on changes in unemployment in GB travel to work areas 1989-1992. *Applied Economics, 28*, 1263-1267.

Hanson, G. H. (2001). Scale economies and the geographic concentration of industry. *Journal of Economic Geography, 1*, 255-276.

Hartshorne, R. (1936). A new map of the manufacturing belt of North America. *Economic Geography*, *12*(1), 45-53.

Heck, R. H. & Thomas, S. L. (2000). *An introduction to multilevel modeling techniques*. Mahwah, NJ: Lawrence Erlbaum Associates.

Hofler, R. A. & Murphy, K. J. (1989). Using a composed error model to estimate the frictional and excess-supply components of unemployment. *Journal of Regional Science, 29*, 313-228.

Holzer H. J. (1993). Structural/Frictional and demand-deficient unemployment in local labor markets. *Industrial Relations, 32*, 307-328.

Holzer, H. J. (1991). Employment, unemployment and demand shifts in local labour markets. *The Review of Economics and Statistics, 73*, 25-32.

Hoover, E. M. (1937). Location *theory and the shoe and leather industries*. Cambridge, MA: Harvard University Press.

Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam: TT-Publikaties.

Hyclak T., & Johnes G. (1987). On the determinants of full employment unemployment rates in local labour markets. *Applied Economics, 19*, 191-200.

Hyclak, T. (1996). Structural changes in labor demand and unemployment in local labor markets. *Journal of Regional Science, 36*, 653-663.

Hyclak, T. and Stewart, J. B. (1995). Racial differences in the unemployment response to structural changes in local labor markets. *Review of Black Political Economy, Spring 1995*, 29-42.

Ivanović, Zoran; Baldigara, Tea (2006). Multilevel models in educational process Performance Evaluations. *Tourism & Hospitality Management, 12*(2), 25-36.

Kaplan, D. H. (2000). The uneven distribution of employment opportunities: Neighborhood and race in Cleveland, Ohio. *Journal of Urban Affairs, 21*, 189-212.

Kliesen, K. L. (2003). The 2001 Recession: How was it different and what developments may have caused it? *Federal Reserve Bank of St. Louis Review 85*(5), 23-37.

Kolko, J. (1999). Can I get some service here? Mimeo, Harvard University. Retrieved December 10, 2011, from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=985712#

Krmenec, A. J., & Esparza, X. (1999). City systems and industrial market structure. *Annals of the Association of American Geographers, 89*(2), 267-289.

Krugman P (1995) Development, economic geography and economic theory. Cambridge, MA: MIT Press.

Lassman, T. (2005). Beyond the ruins: The Meanings of deindustrialization (review). *Enterprise & Society, 6*(2), 350-353.

López-Bazo, E.T., del Barrio, E. & Artis, M. (2002). The Regional distribution of Spanish unemployment: A Spatial analysis. *Papers in Regional Science 81*(3): 365-389.

Luke, Douglas A. 2004. *Multilevel modeling*. Thousand Oaks, CA: Sage Publications.

Lynch, G. J. & Hyclak, T. (1984). Cyclical and noncyclical unemployment differences among demographic groups. *Growth and Change, January 1984*, 9-17.

Malizia, E. E., & Ke, S. (1993). The Influence of economic diversity on unemployment and stability. *Journal of Regional Studies, 33*, 221-235.

Marshall, A. (1890) *Principles of Economics*. London: McMillan, reprinted 1920.

Martin, R. L. & Sunley, P. J. (2006). Path dependence and regional economic

Evolution. *Journal of Economic Geography*, *6*, 395-438.

Martin, R. L. (2010). The Local geographies of the financial crisis: From the housing bubble to economic recession and beyond. *Journal of Economic Geography, 11*(4), 587-618.

Menard, S. (2010). *Logistic regression: From Introductory to advance concepts and applications.* Los Angeles, CA: Sage Publications, Inc.

Meyer, D. (1983). Emergence of the American manufacturing belt: An Interpretation. *Journal of Historical Geography, 9*, 147-174.

Micceri, Theodore. (2007). *The Advantages of using multilevel modeling to address institutional research questions*. Presentation at the Florida Association for Institutional Research Annual Forum, Cocoa Beach, FL, Feb 7-9, 2007.

Mitchell, W. F. & Bill, A. (2004). Spatial dependence in regional unemployment in Australia. In E. Carlson (Ed.), A Future that works: Economics, employment and the environment, Proceedings of the 6th Path to Full Employment Conference and the 11th National Conference on Unemployment, 312-326.

Molho, I. (1995a). Migrant inertia, accessibility and local unemployment. *Economica, 62*, 123-132.

Molho, I. (1995b). Spatial autocorrelation in British unemployment. *Journal of Regional Science, 35*, 641-658.

Mollick, A.V. (2008). What explains unemployment in US-Mexican border cities? *Annals of Regional Science, 42*, 169-182.

Moran, P. (1948). The interpretation of statistical maps. *Journal of the Royal Statistical Society B, 10*, 243-251.

Murphy, K. J. (1985). Geographic differences in U.S. unemployment rates: A Variance decomposition approach. *Economic Inquiry, 23*, 135-158.

National Bureau of Economic Research (NBER). (2011). US Business Cycle Expansions and Contractions. Retrieved July 15, 2011, from http://www.nber.org/cycles.html.

Neumann, G. R. & Topel, R. H. (1991). Employment risk, diversification, and unemployment. *Quarterly Journal of Economics, 106*, 1341-1365.

Niebuhr , A. (2002). *Spatial dependence of regional unemployment in the European Union*. HWWA discussion paper, 186.

Niebuhr, A. (2003). Spatial interaction and regional unemployment in Europe. *European Journal of Spatial Development, 5*, 1-26.

Nistor, A. (2009). Assessing the effectiveness of human capital investments on the regional unemployment rate in the United States: 1990 and 2000. *International Regional Science Review*. *32*(1), 65-91.

Nordhaus, W. D. (2002): *The mildest recession: Output, profits, and stock prices as the U.S. emerges from the 2001 recession*. NBER Working Paper.

O'Sullivan, D. & Unwin, D. J. (2003). *Geographic Information Analysis*. Hoboken, NJ: Wiley.

Overman, H.G. and Puga, D. (2002). Unemployment clusters across Europe’s regions and countries. *Economics Policy, 34*, 116-147.

Owyang, M., Piger, J. M, & Wall, H. J. (2005). The 2001 Recession and the states of the Eighth Federal Reserve District. *The Federal Reserve Bank of St. Louis Regional Economic Development*, *1*(1), 3-16.

Page, S. (2006). Path Dependence. *Quarterly Journal of Political Science*, 1, 87-115.

Pallant, J. (2007). *SPSS survival manual: A Step-by-step guide to data analysis using SPSS (Version 15)*. New York: Open University Press.

Partridge, M. D., & Rickman D. S. (1995). Differences in state unemployment rates: The Role of labor and product market structural shifts. *Southern Economic Journal, 62*, 89-106.

Patacchini, E. & Zenou, Y. (2008). Spatial dependence in local unemployment rates. *Journal of Economic Geography, 7*(2),169-191.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. London: Sage.

Roberts, J. M., & Morin, N. J. (1999). Is Hysteresis important for U.S. unemployment. *Finance and Economics Discussion Series, 1999*/56, Board of Governors of the Federal Reserve System. doi:10.21.2139/ssrn.205014.

Rosenthal, S. & Strange, W. (2004). Evidence on the nature and sources of agglomeration economies. In J. V. Henderson & J. F. Thisse (Eds.) *Handbook of Urban and Regional Economics* (Vol. 4)*: Cities and Geography* (pp. 2119-2172). Amsterdam: Elsevier.

Scott, A. J. & Storper, M. (2003). Regions, globalization, development. *Regional Studies, 37*(6), 579-593.

Siegers, J. J. (1983). An Economic-Demographic ten-equation model. *De Economist, 131*, 400-443.

Simon, C. J. (1988). Frictional unemployment and the role of industrial diversity. *Quarterly Journal of Economics,* 103, 715-728.

Song, F. & Wu, Y (1997). Hysteresis in unemployment: Evidence from 48 U.S. states. *Economic Inquiry*, *35*, 235-243.

Stoll, M. A., & Raphael, S. (2000). Racial differences in spatial job search patterns: Exploring the causes and consequences. *Economic Geography, 76*, 201-23.

Summers, L. H. (1986). Why is the unemployment rate so very high near full employment? *Brookings Papers on Economic Activity, 2,* 339-383.

Suzen M. L, & Doyuran, V. (2004). A comparison of the GIS based landslide susceptibility assessment methods: Multivariate versus bivariate. *Environmental Geology, 45*, 665-679.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Boston: Allyn and Bacon.

Taylor J., & Bradley S. (1983). Spatial variations in the unemployment rate: A Case study of North West England. *Regional Studies, 17*, 113-124.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46,* 234-40.

U.S. Bureau of Labor Statistics (BLS). (2011). Website. Retrieved March 8, 2011, from http://www.bls.gov.

U. S. Census Bureau. (2011). American Community Survey. Retrieved June 22, 2011, from http://fastfacts.census.gov/servlet/DatasetMainPageServlet?_program=ACS.

U.S. Census Bureau. (2010). Download Center. Retrieved June 15, 2010, from http://www.census.gov/main/www/cen2000.html.

U.S. Census Bureau. (2010). County Business Patterns. Retrieved June 8, 2010, from http://www.census.gov/epcd/cbp/index.html.

U.S. Department of Commerce, Bureau of the Census (2009). County business patterns 2006. Retrieved June 15, 2010, from http://factfinder.census.gov/servlet/IBQTable?_bm=y&-ds_name=CB0600A1.

Vedder, R. & Gallaway, L. (1996). Spatial variations in U.S. unemployment. *Journal of Labor Research*, *17*(3), 445-461.

Walsh, C. (1993). What caused the 1990-1991 recession? *Federal Reserve Board of San Francisco Economic Review*, *2*, 33-48.

Wilsford, D. (1994). Path dependency, or why history makes it difficult but not impossible to reform health care systems in a big way. *Journal of Public Policy*, *14*(3), 251-283.