PREPARING FOR IMMIGRATION REFORM:

A SPATIAL ANALYSIS OF UNAUTHORIZED IMMIGRANTS

by

Anna Jane Fischer

A Thesis Presented to the
FACULTY OF THE USC GRADUATE SCHOOL
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
MASTER OF SCIENCE
(GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY)

December 2014

DEDICATION

I would like to dedicate this document to my parents and brother for always encouraging me to pursue my academic goals and to Evan Colby, for putting up with my lack of availability on seemingly endless weekends. Thank you, Evan, for being forever helpful and making the day-to-day a little bit easier so that I could focus on accomplishing this goal.

ACKNOWLEDGMENTS

I would like to acknowledge my thesis advisor and mentor, Dr. Karen Kemp, for guiding me through the thesis process. Her feedback as well as her words of encouragement were a major guiding force through the process of research and writing. I would also like to acknowledge Dr. Robert Vos for his direction in the initial formulation of my thesis topic. Lastly, I would like to acknowledge Professor Roberto Suro, Director of the Tomas Rivera Policy Institute, who has taught me much of what I know about immigration policy and who was one of the initial sources of motivation for conducting research on this topic. Thank you also to my family and friends, without whom I could not have made it this far.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACS	American Community Survey
AICc	Akaike's Information Criterion
CBO	Congressional Budget Office
CIR	Comprehensive Immigration Reform
CSII	Center for the Study of Immigrant Integration (CSII)
CPS	Current Population Survey
GDP	Gross domestic product
GNP	Gross national product
DACA	Deferred Action for Childhood Arrivals
DHS	Department of Homeland Security
DOJ	Department of Justice (U.S.)
GWR	Geographically Weighted Regression
INS	Immigration and Naturalization Service
ITIN	Individual Taxpayer Identification Number
KMO	Kaiser-Meyer-Olkin
LAC-MILSS	Los Angeles County Mexican Immigrant Legal Status survey
LPR	Legal Permanent Resident
MPI	Migration Policy Institute
OIS	Office of Immigration Statistics
OLS	Ordinary Least Squares
PCA	Principal Component Analysis
PPIC	Public Policy Institute of California

PUMA	Public Use Microdata Areas
RPI	Registered Provisional Immigrant
TRPI	Tomas Rivera Policy Institute
USC	University of Southern California
USCIS	United States Citizen and Immigration Services
VIF	Variance Inflation Factor

ABSTRACT

An estimated 11.7 million unauthorized immigrants resided in the United States in 2012 according to the Pew Hispanic Center (Passel, Cohn, and Gonzalez-Barrera 2013). Reforming the U.S. immigration system is a clear policy priority for President Barack Obama, and an agenda item for the 113th Congress (U.S. Congressional Research Service 2013). Based on prior legislation, processing of immigrants for legalization is likely to be a complex and time consuming task, necessitating the involvement of nonprofit and public infrastructure. The goal of this study was to design a research methodology for estimating the unauthorized population at the census tract level, as a means for visually representing the relative densities of the unauthorized population in a way that would be useful for planning where to provide services for the unauthorized populations within a community. Using statistical methods, the relationships between the dependent and independent variables was defined at the state level. The state level relationships were then applied to census tract level data in order to make census tract estimates. The results of the analysis were displayed as relative densities using the dot density renderer in ArcGIS Desktop. The performance of this model was verified by comparing the results generated in this study to those of other studies. Based on this verification method, the performance of the model varied by geography, with the western states, in particular, California seeming to have performed the best. The states that appear to have performed the worst are primarily located in northeastern United States and include six out of the eight states with the lowest number of unauthorized persons (<3,000). Within California, between a 0.02 (Orange County) and 3.4 (Bay Area) percentage point difference was found when comparing the regional distribution estimated in this study with those of other studies.

CHAPTER 1: INTRODUCTION

An estimated 11.7 million unauthorized immigrants resided in the United States in 2012 according to the Pew Hispanic Center (Passel, Cohn, and Gonzalez-Barrera 2013). Reforming the federal immigration system of the United States, a stated second-term policy priority for President Barack Obama, and a clear agenda item for the 113th Congress, has garnered a great deal of attention from all sides of the political spectrum (U.S. Congressional Research Service 2013). The 113th Congress has been marked by heated bipartisan debate around proposed immigration related legislation. In the following section, three approaches to immigration reform are introduced with examples describing how these approaches have played out over Obama's presidency. This chapter continues by outlining the research objectives, making a case for utilizing spatial analysis methods in planning for immigration reform, and concludes with an outline of the thesis structure

1.1 Approaches to Immigration Reform

Three leading approaches to reform have presented themselves during Obama's presidency and the 113th Congress, including: (1) comprehensive immigration reform (CIR), where wide-ranging reforms are enacted in one "mega-bill," (2) the piecemeal approach, where rather than floating one bill, several immigration related bills are introduced, and (3) administrative or executive action, unilateral action undertaken by Obama.

On June 27th 2013, the Senate passed a comprehensive immigration reform (CIR) bill: *Border Security, Economic Opportunity, and Immigration Modernization Act (S. 744).* Although this bill garnered a great deal of attention, as of August 2014, John Boehner, Speaker of the House, has not brought S. 744 for a vote on the House floor. Additionally, reports have surfaced claiming that Boehner does not plan to act on the Senate bill this year (Myers 2014). Although the House has not gone to a vote on S. 744, they have continued to be active on the subject of immigration, but in what could be described as a piecemeal approach. As of the end of March 2014, over one dozen immigration related bills, addressing facets of the immigration system were pending in the House (U.S. Congressional Research Service 2013; What's on the Menu? 2014).

As of June 2014, Obama has announced a plan to move forward on immigration reform through unilateral action using his executive powers (Marshall and Garcia 2014). Although as of the second week of August 2014, Obama has not announced a path to legalization, it is speculated that a path to legalization may be announced before the fast approaching end of the summer (Nakamura 2014). Obama employed executive action in 2012 with Deferred Action for Childhood Arrivals (DACA), which offered young unauthorized immigrants that arrived in the United States as children and met certain other criteria, reprieve from deportation and authorization to work.

1.2 Issues Addressed in Immigration Reform Legislation

The bills acted on by the House in the 113th Congress have addressed a number of aspects of the U.S. immigration system including: interior enforcement, employment eligibility verification, worksite enforcement, border security, nonimmigrant visas, and immigrant visas (U.S. Congressional Research Service 2013). Similarly, S.744 addressed many of the same facets through various provisions in the bill. In contrast, S.744 also included provisions for the legalization of unauthorized immigrants as well as humanitarian admissions (U.S. Congressional Research Service 2013). The legalization of unauthorized immigrants was a controversial element in S.744, which would have allowed for most unauthorized immigrants in the United

States to gain legal status. Legal status would first be granted through a new status, Registered Provisional Immigrant Status (RPI). After a period of time, immigrants with RPI status would have been able to apply to adjust to Legal Permanent Resident (LPR) status (U.S. Congressional Research Service 2013).

Due to the provision that would have allowed most unauthorized immigrants to gain legalization, S.744 was projected to have grown the U.S. labor force (U.S. Congressional Budget Office 2013b). The U.S. Congressional Budget Office (CBO) projected that S.744 would boost economic output and increase real gross domestic product (GDP). While per capita gross national product (GNP) as well as average wages would initially fall slightly, they would increase by 2033 (U.S. Congressional Budget Office 2013b). Although the average GNP and wages were projected to have initially fallen, these averages would have included all those newly authorized to live and work in the United States and would not have necessarily indicated a decrease for those already legally present in the United States under current law (U.S. Congressional Budget Office 2013b).

1.3 Immigrant Processing Requirements

Should a path to legalization for unauthorized immigrants be introduced, that targets anywhere near the numbers of those that would have potentially been eligible under S.744, upwards of 8 million unauthorized immigrants may be in need of processing in the United States (U.S. Congressional Budget Office 2013a). Based on past legislation, processing of immigrants for legalization is likely to be a complex and time consuming task, necessitating the involvement of nonprofit and public infrastructure, such as community groups, nonprofits, and legal service providers. S.744 would have required unauthorized immigrants to supply proof of presence in the United States on and after December 31, 2011, proof of immigration status, proof of identity, as well as undergo a background check in order to obtain Registered Provisional Immigrant (RPI) status (U.S. Senate 2013).

Similar detailed and thorough documentation was required to apply for DACA. Preliminary findings from a study conducted by the Tomas Rivera Policy Institute (TRPI) in Los Angeles County, estimates an average of 3 hours of assistance would be required per low-need applicant, those who have a majority of required documents, to process applications for RPI status under S. 774 (Chan, Kabat, and Reyes 2013). Moderate need applicants, those missing required documents, may require between 6–20 hours of assistance (Chan, Kabat, and Reyes 2013). High need applicants, those with criminal records or previous interactions with U.S. Citizen and Immigration Services (USCIS), are likely to require the greatest amount of resources and time. However, no reliable estimate exists for this population because they are generally not served within the network of non-profit service organizations but instead are referred out to attorneys for legal advice (Chan, Kabat, and Reyes 2013). The estimates produced by TRPI only include the time required to help applicants prepare their legalization application. They do not include the time required to process the application once received by the Department of Homeland Security (DHS).

Given these numbers, in Los Angeles County alone, the estimated 900,000 unauthorized immigrants would require a minimum of 2.7 million hours of assistance (Chan, Kabat, and Reyes 2013). If the registration period is limited to one-year, a full-time workforce of 2,700 individuals would be required to process RPI applications alone. This assumes 1,700 work hours per person per year spending 100 percent of their time processing applications. This estimate does not include time that would surely be needed for administrative duties such as set-up, supervision, or training. Regardless of the final form that immigration reform may take, whether through S.774, a piecemeal approach, or executive action taken by Obama, preparing to process a substantial number of applicants will not only require a large enough workforce, but outreach and services in locations that are accessible to the eligible unauthorized population.

1.4 Research Objectives

The unauthorized population is neither limited to discrete locations nor evenly spread out. Additionally, there is no large-scale survey that directly asks about legal status, no reliable estimates at the sub-state level for a majority of the nation, not to mention the lack of estimates at the neighborhood level. In fact, no existing estimates of the unauthorized population at the census tract level were uncovered during the course of this research.

That being said, the goal of this analysis is to design a research methodology for estimating the unauthorized population at the census tract level, as a means for visually representing the relative densities of the unauthorized population in a way that would be useful for planning where to provide services for the unauthorized population within a community.

1.5 Thesis Structure

Chapter two contains a thorough investigation of the current state of the field of research around estimating the unauthorized population, examining several leading estimation methods and then presenting the results of previous research and analysis, including estimates of the number and likely characteristics of unauthorized population. Chapter three follows with a detailed account of the study design and methodology utilized in this study. Chapter three begins with a section on determining the variable inputs for the analysis and continues with defining the relationships between the independent and dependent variables at the state level. The state level relationships are then applied to the census tract level data in order to make census tract estimates of the unauthorized population. Chapter three concludes with an overview of the rendering scheme for mapping the results.

Chapter four presents the results of the analysis through maps of various scales and extents that visualize the relative density of the unauthorized population using dot density renderer in ArcGIS Desktop. Although only four maps are presented, a map could be produced for virtually any location of interest within the study area (forty-eight contiguous U.S. states and Washington, DC). Conclusions on the implication of the analysis and the viability and performance of the analysis method are presented in chapter five. This report concludes with an overview of the challenges, weaknesses, and limitations of the analysis and suggests next steps to carry the research and methodology forward.

CHAPTER 2: BACKGROUND

This chapter presents an overview of the state of the field of research on estimating the unauthorized population by presenting the leading estimation methods as well as the findings of recent studies, which includes both existing estimates of the total numbers as well as characteristics of the unauthorized population. This chapter concludes with an overview of research on immigrant settlement patterns in the United States.

The material presented is the basis for many of the methodological decisions made throughout this analysis. Specifically, the characteristics of the unauthorized population and their settlement patterns in the United States, as determined from prior research and analysis, guided the decisions on what independent variables to include in the analyses. The data generated from previous estimates of the unauthorized were used as the dependent variable as well as the primary method of verifying the results. Not to mention, knowledge of the existing estimation methods influenced the overall study design.

2.1 Methods for Estimating the Unauthorized Population

The following section covers the residual method, community-based probability method, and other statistical methods that have been used to calculate estimates of the unauthorized population.

2.1.1 Residual Method for National and State Estimates

The "residual method" is the leading method for estimating the unauthorized population, used to produce the estimates released by the Department of Homeland Security (DHS) Office of Immigration Statistics (OIS) and the Pew Hispanic Center (henceforth referred to as Pew) (Passel 2013; Baker and Rytina 2013). Simply put, the residual method subtracts the legal foreign-born (legal nonimmigrants, refugees, asylees, and legal permanent residents) from the total number of foreign-born residing in the United States. What remains, after making certain adjustments for factors such as undercounting and mortality, is an estimate of U.S. foreign-born that are not legally present in the United States, the unauthorized population, as they are referred to in this report (Hill and Johnson 2011; Judson and Swanson 2011; Passel 2013; Pastor and Marcelli 2013; Warren and Warren 2013;). A simplified equation for estimating the unauthorized population using the residual method follows. In addition to the equation below, adjustments are made to account for mortality and emigration rates.

total unauthorized population
equals (=)
Total foreign-born population
minus (-)
legal permanent residents (LPRs)
nonimmigrant resident population
refugees admitted
removals of unauthorized population
plus (+)
the undercount

The following data sources are commonly incorporated into the residual method to estimate the number of unauthorized immigrants in the United States:

Table 1 **Residual Method: Common Data Sources**

ORGANIZATION	**ESTIMATE/COUNT**
ACS	Total foreign-born
U.S. Census Bureau	Total foreign-born
CPS	Total foreign-born
Department of Homeland Security (DHS)	Authorized immigrant population
Department of State	Refugee characteristics
DHS and U.S. Citizenship and Immigration Services (USCIS)	Legal permanent residents (LPR) characteristics
USCIS	Asylums granted affirmatively
Executive Office for Immigration Review of the Department of Justice (DOJ)	Asylums granted defensively in removal proceedings
U.S. Customs and Border Protection*	Nonimmigrant admissions
National Center for Health Statistics	Life expectancy tables

*TECS system capturing I-94 arrival-departure records

The residual method alone is restricted to estimating the unauthorized population at the national or state level because of the lack of granularity of required data. Using estimates made from the residual method as a baseline, combined with additional methods, such as survey and statistical methods as well as the use of administrative data, have been employed to estimate the distribution and demographic characteristics of the unauthorized population at the sub-state level.

2.1.2 Residual Method Combined With Other Methods for Sub-state Estimates

Two examples of studies that produced sub-state estimates for California include Pastor and Marcelli (2013) and Hill and Johnson (2011). Pastor and Marcelli (2013) use a "community-based probability method," a combined survey and statistical method, to generate estimates of the unauthorized by sub-counties, or PUMAs. Using this method, of the individuals captured in the ACS as non-citizen foreign-born (excluding those born in Cuba), the probability of being unauthorized is calculated by using legal status predictors generated from Marcelli's 2001 Los Angeles County Mexican Immigrant Legal Status survey (LAC-MILSS). Those with the highest calculated probabilities of being unauthorized are flagged until the total number of those flagged equals the OIS estimates (derived from the residual method) of the total number of unauthorized adults for the top ten countries of origin (Pastor and Marcelli 2013). The characteristics of those flagged as unauthorized are then analyzed and presented as the characteristics of the unauthorized (Pastor and Marcelli 2013).

Hill and Johnson (2011) use statistical methods that include administrative data, Individual Taxpayer Identification Number (ITIN) filer counts, to estimate the total number of unauthorized by zip codes and counties in California. The final zip code estimates are ultimately scaled so that when summed, they equal the total number of unauthorized in California, as derived from the residual method. Hill and Johnson use ITIN filers, excluding those that file from abroad, as a proxy for the unauthorized because they have found that, "the vast majority of ITIN filers do appear to be unauthorized (2011, 11)." Although not all of the unauthorized pay taxes or pay taxes using an ITIN—some do not pay taxes at all or pay taxes using other methods like a false or fraudulent Social Security number— it is unlikely that persons legally in the United States would use an ITIN because they would use a Social Security number or other federal tax ID number instead (Hill and Johnson 2011).

2.1.3 Challenges and Weaknesses of Existing Estimation Methods

Estimating the unauthorized is not an exact science, and there are several aspects of the leading methodologies that are subjective. One such aspect is the undercount of the unauthorized population. It is generally understood that a portion of the unauthorized population is missed in the census and other surveys; what is debated is the percentage of the unauthorized population that is not surveyed. OIS uses an undercount of 10 percent, and Pew uses an undercount in the "range of 10-15 percent" (Baker and Rytina 2013; Passel 2013). Warren and Warren (henceforth referred to as Warren), on the other hand, use an undercount of 20 percent (2013). The resulting estimates of the unauthorized are sensitive to the estimated undercount used in the analysis, as shown through sensitivity analysis conducted by OIS (Baker and Rytina 2013).

Another seemingly subjective area is determining "legal status indicators," or characteristics that may indicate an individual as likely to be unauthorized (Pastor and Marcelli 2013). Although, statistical and multivariate regression analysis has been employed to determine these indicators based on the results of smaller scale surveys, the results may be compromised for a variety of factors, including the small numbers of those being surveyed and the known difficulties in eliciting truthful responses when directly inquiring about legal status (Hill and Johnson 2011; Pastor and Marcelli 2013).

2.2 Results of Prior Research and Analysis

2.2.1 Estimates of the Unauthorized

The unauthorized population residing in the United States in 2012 has been estimated at both 11.4 and 11.7 million by the OIS and Pew, respectively (Baker and Rytina 2013; Passel, Cohn, and Gonzalez-Barrera 2013). Pew and OIS offer ongoing yearly reports estimating the unauthorized population in total and by select demographic characteristics (Passel and Cohn 2011; Baker and Rytina 2013; Passel, Cohn, and Gonzalez-Barrera 2013). A third leading source for estimating the number of unauthorized is Robert Warren, Statistics Division, U.S. Immigration and Naturalization Service (INS), and John Robert Warren, Minnesota Population Center, University of Minnesota, who as of January 2010, estimated 11.7 million unauthorized persons were residing in the United States (Warren and Warren 2013).

In California, several studies have attempted to estimate the unauthorized population at a sub-state level, including: county, Public Use Microdata Areas PUMAs (or "sub-county"), and zip code level (Fortuny, Capps, and Passel 2007; Hill and Johnson 2011; Hill and Hayes 2013; Pastor and Marcelli 2013). The finest geographic scale that estimates the unauthorized population for all fifty states is at the congressional district level (Rob Paral and Associates 2006). At the county, sub-county, and zip code level (or for any smaller geography) estimates are only available for select geographic regions.

Estimates of immigrant sub-populations have also been conducted, including estimates of the number of legal immigrants eligible to naturalize and the unauthorized youth eligible for DACA. These analyses have been conducted at various scales and geographies. Estimates of the eligible DACA population have been conducted for the entire United States by metro area and congressional district, and for the state of Illinois by cities/towns, House districts, and Senate districts. For the city of Chicago, these estimates have been made down to the community level. For the state of California, Rob Paral and Associates in collaboration with the USC Center for the Study of Immigrant Integration (CSII) have estimated the number of legal immigrants eligible to naturalize at the California Assembly, Senate and Congressional Districts as well as the census tract level for Napa. (See Rob Paral and Associates "Map Gallery," http://www.robparal.com/gallery/index.html).

While existing studies have increased the overall knowledge of the location of the unauthorized, because the current sub-state estimates are limited to certain regions and there is an overall lack of estimates at a fine geographic scale, the existing estimates are not suitable for planning the outreach and physical infrastructure at the community level for a national initiative.

2.2.2 Characteristics of the Unauthorized Population

Demographic characteristics of the unauthorized population at the national level have been estimated by the OIS and Pew (Passel and Cohn 2009; Baker and Rytina, 2013). Demographic characteristics presented by the OIS include: period of entry, state of residence in the United States, region of birth, country of birth, age range, and sex (Hoefer, Rytina, and Baker 2011; Hoefer, Rytina, and Baker 2012; Baker and Rytina, 2013). A 2009 study from Pew made estimates of the number of unauthorized population by educational attainment, income, and health insurance coverage for the unauthorized population in the U.S (Passel and Cohn 2009). A 2013 study from CSII presents estimates of the characteristics of the unauthorized population in California at the regional (multi-county) level, including race/ethnicity, child population, child poverty, speaks English well, industry, occupation, and labor force participation (Pastor and Marcelli 2013). Based on the findings of these existing analyses, the unauthorized population in the United States, including how they differ from the overall foreign-born and legal immigrant population, can be characterized as follows:

- **Country and region of birth.** Fifty-nine percent of the unauthorized population is from Mexico (Hoefer, Rytina and Baker 2012). And in California the percentage is much higher, with 72 percent of the unauthorized population from Mexico, followed by Central America at 12 percent (Pastor and Marcelli 2013). Of all the immigrants from Mexico (an estimated 11.4 million) residing in the United States in 2008, more than half were unauthorized (Terrazas 2010).
- **Ethnicity.** 76 percent of the unauthorized immigrant population is Hispanic (Passel and Cohn 2009).
- **Age and sex.** The majority of unauthorized immigrants are between 25 and 44 (59 percent). Unauthorized immigrants are less likely to be 65 and older compared to authorized foreign-born and U.S.-born population. Only 1.2 percent of unauthorized immigrants are 65 and older, compared to 16 percent of authorized immigrants, and 12 percent of the U.S.-born (Passel and Cohn 2009). In California, the median age for the unauthorized population is thirty-one compared to forty-four and fifty for authorized and citizen foreign-born population respectively (2009-2011 data) (Pastor and Marcelli 2013). More than half of the total unauthorized population is male (53 percent) (Hoefer, Rytina and Baker 2012).
- **Period of entry.** The vast majority (99 percent) of the unauthorized population currently residing in the United States arrived after 1980 (based on author's calculation of total unauthorized population by year of entry in Hoefer, Rytina, and Baker, 2011). In part, this is likely due to the Immigration Reform and Control Act of

1986 (IRCA), which, allowed immigrants who arrived prior to and had been continually present in the United States since 01 January 1982, to legalize. Of the immigrants that qualified under the "pre-1982" provision, 1.6 million had legalized as of 2009 (Baker 2010).

- **Educational Attainment.** Unauthorized immigrants are less likely to have completed high school or to have attended college than authorized foreign-born. Nearly half (47 percent) of unauthorized immigrants between 25 and 64 did not complete high school compared to around 23.5 percent of legal immigrants. Similarly, 25 percent of the unauthorized population have attended or completed college compared to 54 percent of legal immigrants (Passel and Cohn 2009).
- **Income.** An analysis conducted by Pew found that the 2007 median household income was $14,000 less for the unauthorized than the U.S.-born ($36,000 versus 50,000) (Passel and Cohn 2009). A similar study found even greater income disparities in California, where the median annual income for full time workers was found to be $30,000 less for the unauthorized than the U.S.-born ($20,000 versus $50,000) (Pastor and Marcelli 2013). Additionally, unlike other immigrant groups, unauthorized immigrants do not "make notable gains" corresponding with longer time in the United States (Passel and Cohn 2009).
- **Health Insurance.** Fifty-nine percent of the unauthorized adults did not have health insurance for the entire year of 2007 (Passel and Cohn 2009).
- **Household and home ownership.** Unauthorized immigrants are more likely to live in households with a partner and children (47 percent) than authorized immigrants (35

percent). Unauthorized immigrants are less likely to be homeowners than authorized immigrants (Passel and Cohn 2009).

- **Residency.** In California, the median number of years in the country for unauthorized is 9 compared to 19 for authorized noncitizen immigrants, and 27 for immigrant citizens (2009-2011 data) (Pastor and Marcelli 2013).
- **Language proficiency.** A study conducted using data from 2009-2011 found that of immigrants in California, 42 percent of unauthorized speak English well compared to 61 percent of authorized noncitizen (Pastor and Marcelli 2013).

2.3 Immigrant Settlement Patterns in the United States

Immigrant settlement patterns, defined as trends in where immigrant groups choose to reside in the United States, are affected by a variety of factors, including existing family/social ties, demographic make-up of a community, as well as economy and industry (Bohn 2009). One major change in immigrant settlement patterns that started to occur in the 1990s is the dispersal of immigrants from settling primarily in just a few states (or metro areas within these states) to settling across the wider United States. In 1990, nearly 75 percent of immigrants of working age in the United States resided in just six states, with over 30 percent residing in California (Bohn 2009). In the 1990s the proportion of immigrants residing in California began to fall for the first time since the early 1900s and by the late 1990s, the combined proportion of immigrants living in these six traditional immigrant-receiving states began to fall as well (Bohn 2009). In terms of population growth, the states with the highest ratio of immigrants to nonimmigrants saw some of the lowest immigrant growth rates from 2000-2007 (Bohn 2009). A similar analysis of settlement patterns of Mexican immigrants, found that Mexicans had also begun to settle in non-traditional states in the south and Midwest of the country, such as Georgia, North Carolina, Nebraska and

Ohio (Terrazas 2010). Furthermore, the growth rate of Mexican immigrants did not necessarily coincide with the state's overall growth rate. In Louisiana and North Dakota the Mexican immigrant growth rate grew despite the total population shrinking from 2000 to 2008. And in many states, the growth in Mexican immigrants contributed considerably to the overall population growth of the state; In Rhode Island, Mexican immigrants accounted for nearly 60 percent of the total population growth (Terrazas 2010).

Due to lack of data on the unauthorized population, it is difficult to tell how these patterns may have differed, if at all, between the unauthorized and the foreign-born population as a whole. In the case of California, the change in the proportion of immigrants residing in the state, in major part has been due to fewer newly arrived immigrants choosing to settle in California versus established immigrants migrating out of California (Bohn 2009). A similar study of immigrant settlement patterns conducted by the Brookings Institution, found that recently arrived immigrants that are choosing to settle in non-traditional states are likely to be from Asia or Mexico and have lower rates of U.S. citizenship (Singer 2004).

CHAPTER 3: METHODOLOGY

Given that existing methods for estimating the unauthorized are not suitable for making estimates at the census tract level, the goal of this analysis was to design a methodology that may be suitable for estimating the unauthorized population at the census tract level. That being said, this analysis draws on existing methods and their findings as a basis for the methodology outlined in this chapter. Specifically, known characteristics of the unauthorized population and their settlement patterns, established in prior research and analysis, are a basis for determining what variables to include in this analysis. One of the main data sources for this analysis and the source of all of the demographic data (aside from the estimates of the unauthorized population at the state level) is the Census Bureau's American Community Survey (ACS).

To oversimplify the analysis method in an attempt to explain the methodology designed in this study: suppose that majority of the unauthorized population in the United States is from Mexico and speaks English less than "very well." This method would bring those demographic variables into the analysis as independent variables, define their relationship with the dependent variable (the unauthorized population) using regression analysis and then use the resulting equation to make estimates of the unauthorized at the census tract level by "plugging in" census tract level data. While the method used in this study is fundamentally based on the straightforward approach outlined above, there are several crucial ways that this analysis differs:

- The dependent variable is the percent of the unauthorized out of the total foreign-born population. In fact, all demographic variables are transformed to be percentages of the total foreign-born.
- All demographic characteristics are incorporated into one variable using Principal Component Analysis (PCA). Many of the demographic characteristics of the

unauthorized population used in this analysis are highly correlated. In order to avoid the multicollinearity problem that would arise from including all of the variables into a regression analysis, the variables are reduced to one artificial variable, or component score, using PCA.

- The relationship between the dependent and independent variables were defined using a state level regression equation and then "brought down" or applied to the census tract level data in order to make estimates for each census tract. While there are many challenges (including ecological fallacy) with scaling down state level equations to a smaller geography, this method was chosen because the state level estimates of the unauthorized population are the only available and widely accepted as reliable estimates of the unauthorized population.

3.1 Overview of Analysis Steps

An overview of the analysis steps is shown below. Details of the analysis follow in the next sections:

I. Determine input variables:
   1. Define what is being estimated (the dependent variable)
   2. Identify the demographic variables (the independent variables) that correspond to the characteristics of the unauthorized population, more specifically variables with the potential to differentiate the unauthorized from the larger foreign-born population

II. State level analysis: Define relationship between dependent and independent variables
   1. Derive the first principal component to account for joint variation in correlated independent variables using Principal Component Analysis (PCA)
   2. Compute a component score for each state

3. Conduct exploratory regression analysis; Independent variables include the component score (defined in PCA) as well as other state level variables to identify best regression model
4. Run Ordinary Least Squares (OLS) regression analysis based on results of exploratory regression analysis to calculate the percent of unauthorized out of total foreign-born (dependent variable)
5. Run Geographically Weighted Regression (GWR) analysis in order to determine a unique equation for each state included in the analysis

III. Census tract-level analysis: Estimate unauthorized population at the census tract level using previously defined state level equations

1. Compute component scores for each census tract using the coefficient scores defined in the state level analysis
2. Based on GWR equation for each state, substitute the state level component score with each individual census tract's component score in order to calculate an estimated percent of the unauthorized population out of total foreign-born (dependent variable) for each census tract
3. Multiply the estimated percent of the unauthorized out of the total foreign-born (dependent variable) by the total foreign-born population for each census tract in order to come up with an estimate of the total number of the unauthorized for each census tract

IV. Visualize the results of the analysis

V. Verify results of the analysis and draw conclusions of the viability of the method

3.2 Define Variables

This section reviews the variables included in the analysis and the reasoning behind including (or excluding) certain variables. Which variables to include were determined by consulting previous research findings and methodological approaches. Because the goal of this analysis is to make estimates at the census tract level, only demographic data that is available at the census tract level could be incorporated into the equation. Alaska, Hawaii, and Puerto Rico were not included in the analysis because of lack of geographically near neighbors, a requirement for running Geographically Weighted Regression (GWR) analysis. The time period for the analysis is 2006-2010.

3.2.1 The Dependent Variable

The dependent variable in this analysis is the percent of the unauthorized out of total foreign born. The dependent variable was calculated by dividing the number of the unauthorized from the number of foreign born by state. The estimates of the unauthorized by state (the numerator) were generated from Warren using the residual method. The source for the foreign born population estimates (the denominator) is the ACS.

The percent of the unauthorized was estimated out of a base population, rather than estimating the total number of the unauthorized directly. A base population was used as a method for standardizing all of the demographic data. Standardizing the data not only helps to minimize outliers but also ensures that the patterns or correlations are due to underlying demographic differences not differences in the total population numbers between each state. Two variables were considered as the base population: (1) total foreign-born, and (2) total noncitizen foreign-born. For the reasons outlined below, the base population chosen for the analysis was total foreign-born, resulting in the dependent variable being the percent of the unauthorized out of the total foreign-born population in the United States:

- **There is precedence for using the foreign-born population as the base of the estimation.** One of the leading methods for estimating the unauthorized population, the residual method, uses the foreign-born population as the base for estimating the number of unauthorized at the state level. Similarly, PPIC's estimates of the total number of unauthorized by zip code uses the foreign-born population as a base (Hill and Johnson 2011).
- **The ACS estimates for the foreign-born population have a smaller margin of error than those for the noncitizen population.** The noncitizen population is a subset of the foreign-born population, meaning that the total number of noncitizens is smaller than or equal to the total number of foreign-born in any given geography. Because the ACS estimates are derived from surveying a sample of the population, estimates of smaller populations or within small geographies tend to have lower levels of accuracy due to larger margins of error (U.S. Department of Commerce 2008).
- **There is a strong positive correlation between the foreign-born population and unauthorized population.** This is logical because the foreign-born population, as captured by the ACS, invariably includes a portion of the unauthorized population although the exact proportion is unknown. Visual inspection of the scatterplot (Figure 1) shows a strong positive linear relationship between the foreign-born and unauthorized population (as estimated by Warren), meaning that as the total number of foreign-born increases, so does the number of unauthorized. The strength and direction of the relationship is further corroborated by the results of the Spearman's rank-order

correlation (Table 2). The Spearman's correlation found that an increase in the foreign-born population was strongly correlated with an increase in the unauthorized population in the United States at the state level, r_s (47)= .973, $p < .0005$.

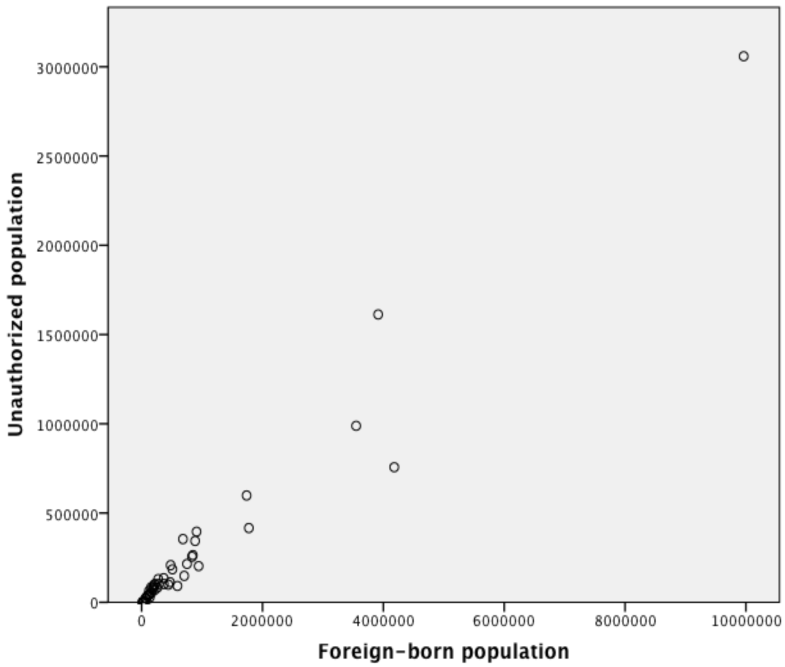

Figure 1 **Scatterplot of the Unauthorized by the Total Foreign-born Population by State**

Table 2 **Spearman's Rank-order Correlation Between the Unauthorized and Foreign-born Population by State**

			Unauthorized	Foreign-born
Spearman's rho	Unauthorized	Correlation Coefficient	1.000	.973**
		Sig. (2-tailed)	.	.000
		N	49	49
	Foreign-born	Correlation Coefficient	.973**	1.000
		Sig. (2-tailed)	.000	.
		N	49	49

**Correlation is significant at the 0.01 level (2-tailed).

Once the dependent variable was determined, there was still the question of time period as well as which source would be used to supply the data for the dependent variable. When considering data options, special consideration was paid to accuracy and recentness of data. The

ACS data was the logical choice for the base population (denominator), offering the most authoritative source for demographic data that is updated regularly and available nationwide at the census tract level. The ACS releases data in 1-year, 3-year, and 5-year estimates. The 5-year estimates were chosen because they are the most reliable and have the largest sample size, particularly important when working with small geographies or when analyzing small populations (U.S. Department of Commerce 2008).

While several state level estimates of the unauthorized population exist, the Warren estimates were chosen as the numerator, because they have been released yearly and for all fifty states (Warren and Warren 2013). A five-year (2006-2010) average of the unauthorized population was taken for the numerator in order to correspond with the 5-year ACS data. This 5-year average was then divided by the 2006-2010 estimates of the foreign-born population released by the ACS in order to come up with the percent of the unauthorized out of the total foreign-born population, the dependent variable.

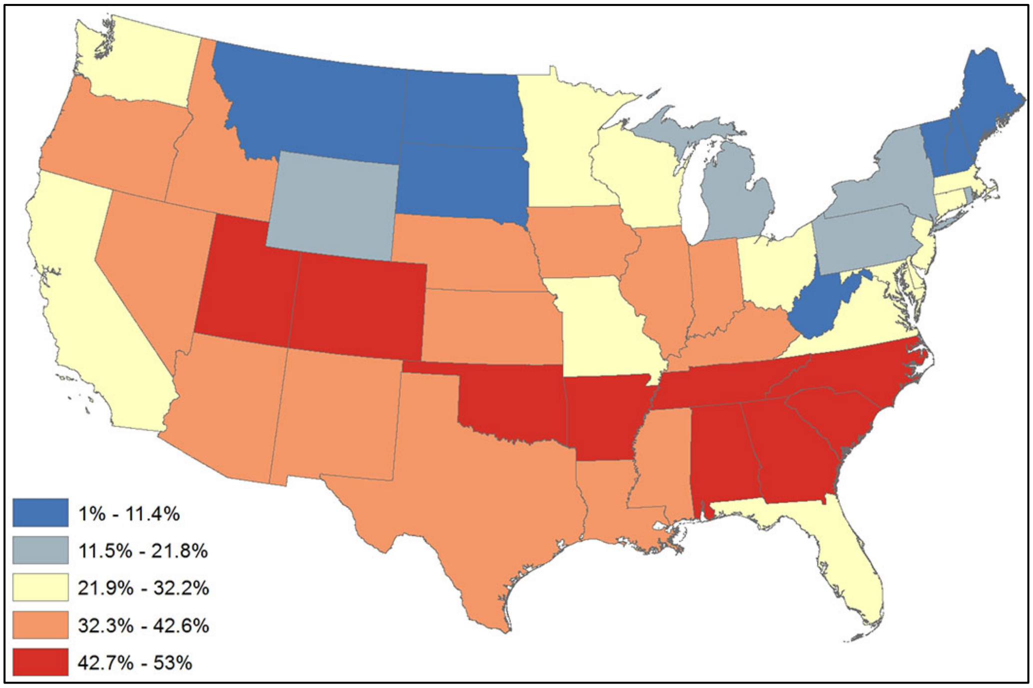

Figure 2 **Observed Dependent Variable: % Unauthorized out of Total Foreign-born Population (2006-2010)**

3.2.2 Corresponding Demographic Variables

Based on findings on the characteristics of the unauthorized population, as determined by prior research and analysis, the demographic variables in Table 3 (as shares of the total foreign-born) were considered for inclusion in the analysis. Several variables were considered in many of the categories (year of entry, language proficiency, educational attainment, income and country of origin) including some "nested variables" where one or more variables make-up another variable. For example, the variable "Income less than 50,000 or no income" includes the variable "no income," which was also initially considered in the study.

Table 3 **Demographic Variables Considered for Inclusion in the Analysis**

Variable	Characteristic	Universe	Spearman's rho: correlation with dependent variable
Entered the U.S. after 2000	Year of entry	Total population born outside of the U.S.	.374**
Entered the U.S. before 1980	Year of entry	Foreign-born	-.708**
Speak a language other than English: Speak English 'very well'	Language proficiency	Foreign-born population 5 years and over	-.402**
Speak a language other than English: Speak English 'not at all'	Language proficiency	Foreign-born population 5 years and over	.765**
Speak a language other than English: Speak English 'less than very well'	Language proficiency	Foreign-born population 5 years and over	.776**
Speak a language other than English: Speak English 'less than well'	Language proficiency	Foreign-born population 5 years and over	.807**
65 years and over	Age	Total foreign-born population	-.777**
Not a U.S. citizen	Citizenship status	Total foreign-born population	.874**
Less than high school graduate	Educational attainment	Foreign-born population 25 years and over	.760**
Graduate or professional degree	Educational attainment	Foreign-born population 25 years and over	-.570**
Bachelor's degree or higher	Educational attainment	Foreign-born population 25 years and over	.609**
No income	Income	Foreign-born population 15 years and over	.673**
Income less than 50,000 or no income	Income	Foreign-born population 15 years and over	.605**
Median income in the last 12 months	Income	Population 15 years and over in the United States with income	.507**
Americas: Latin America: Central America: Mexico	Country or region of origin	Foreign-born population excluding population born at sea	.773**
Americas: Latin America: Other Central America	Country or region of origin	Foreign-born population excluding population born at sea	.375**
Americas: Latin America: Caribbean: Cuba	Country or region of origin	Foreign-born population excluding population born at sea	No clear correlation
Americas: Other Latin America	Country or region of origin	Foreign-born population excluding population born at sea	-.279**
Median Age of Foreign-born	Age	Total population	-.630**
Income in the past 12 months below poverty level: Foreign-born	Income	Total population for which poverty status is determined	.654**
Total Hispanic or Latino foreign-born	Ethnicity	Hispanic or Latino Population	.799**

**Correlation is significant at the 0.01 level (2-tailed).
Data source for all demographic variables is the ACS 2006-10
Red columns indicate demographic variables that were ultimately not retained in the analysis.

Each potential demographic variable's relationship with the dependent variable (percent of the unauthorized out of total foreign-born), and therefore their viability as analysis variables, was examined through visual inspection of scatterplots as well as Spearman's rank-order correlation to test the strength and direction of their relationships with the dependent variable, as shown in Table 3. Where several variables were considered in a particular category, the variable(s) with the strongest relationships with the dependent variable as well as existing theory were considered in determining which variables would be retained for inclusion in the analysis. Ultimately, thirteen demographic variables were retained.

3.3 State Level Analysis: Define Relationship Between Dependent and Independent Variables

3.3.1 Principal Components Analysis (PCA)

Principal component analysis is a statistical method for reducing the number variables in an analysis into a subset of linearly uncorrelated "artificial" variables, called principal components. PCA is a data reduction technique, often utilized as a way of eliminating the redundancy between variables that may be measuring the same or similar construct (O'Rourke and Hatcher 2013). In the case of this analysis, a number of highly correlated demographic variables are reduced to one principal component that represents the maximum variance between the original variables.

The PCA results in a set of actual scores, in this case, one score for each geography (forty-eight contiguous U.S. states and Washington, DC) included in the analysis. These scores were then used in subsequent regression analysis in place of the original variables. So instead of entering all thirteen demographic variables into the regression analysis, only one composite variable (the principal component) was entered into the analysis. PCA was chosen as an analysis method because it eliminates the multicollinearity problems that would have arisen, should all correlated variables have been entered into a regression analysis, without having to eliminate variables altogether.

An initial PCA was run on all retained demographic variables (see Table 3), chosen because of their strength of relationship with the dependent variable (the percent of the unauthorized out of the total foreign-born population). To confirm PCA as an appropriate analysis method, the correlation matrix as well as Bartlett's test of Sphericity and Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy were examined and are explained in the following section.

3.3.1.1 Variable correlation and sampling adequacy

From examining the correlation matrix (See Appendices, Table 19), it is clear that all variables are strongly correlated with at least three other variables at a level of $r \geq 0.3$. The only variable with relatively low level of correlation with other variables is percent of the foreign-born population born in Central America (other than Mexico), which is correlated with three variables right around 0.3, and with all other variables at <0.3. Additionally, Bartlett's test of Sphericity is statistically significant with a p-value <.0005, indicating that overall there are correlations in the variables, suggesting that principal components analysis is an appropriate method for reducing the number of variables in the analysis (Laerd Statistics 2013). While the correlation matrix and Bartlett's test of Sphericity show that there is correlation between variables, there may in fact be too high of correlation between variables. When examining the correlation matrix (Table 19, Appendices), there are three variables with $r \geq 0.9$ which may indicate multicollinearity or singularity with the data.

Kaiser-Meyer-Olkin (KMO) analysis was used to test for sampling adequacy and linear relationship between variables. Sampling adequacy was assessed for the overall equation as well as for the individual variables using KMO analysis. The sampling adequacy for this PCA was found to be .771, which is satisfactory or "middling" on Kaiser's (1974) classification of measure values (Laerd Statistics 2013). This indicates linear relationships between variables and that PCA may be an appropriate analysis method. When assessing KMO measures for individual variables, all variables have strong linear relationships with other variables (KMO >= .5) except for born in Central America (other than Mexico) (KMO = .285).

Table 4 **KMO Measures for Demographic Variables**

Entered 2000 or later*	0.7353
Entered before 1980*	0.7958
Speak a language other than English: Speak English 'less than well'*	0.8598
65 years and over*	0.8335
Not a U.S. citizen*	0.905
Less than high school graduate*	0.777
Bachelor's degree or higher*	0.7348
Median income in the past 12 months	0.6382
Born in Mexico*	0.7704
Born in Central America*	0.2851
Median Age	0.7940
Income in the past 12 months below poverty level*	0.7194
Hispanic or Latino foreign-born*	0.805

*percent of total foreign-born

3.3.1.2 Retaining principal components

From examining the scree plot (Figure 3) and the eigenvalue-one criterion (Table 5) from the initial PCA, it appears that three components could potentially be retained. The first three components have eigenvalues greater than one and each account for over 10 percent of the total variance. That being said, it logically does not make sense to have greater than one component for the purpose of this study, because all demographic variables were included based on their relationship and potential to estimate one variable, the percent of the unauthorized out of the total foreign-born population. The decision to retain only one component is strengthened by examining the component matrix. The component matrix shows that all variables, except "born in Central America," load on the first component at .3 or greater.

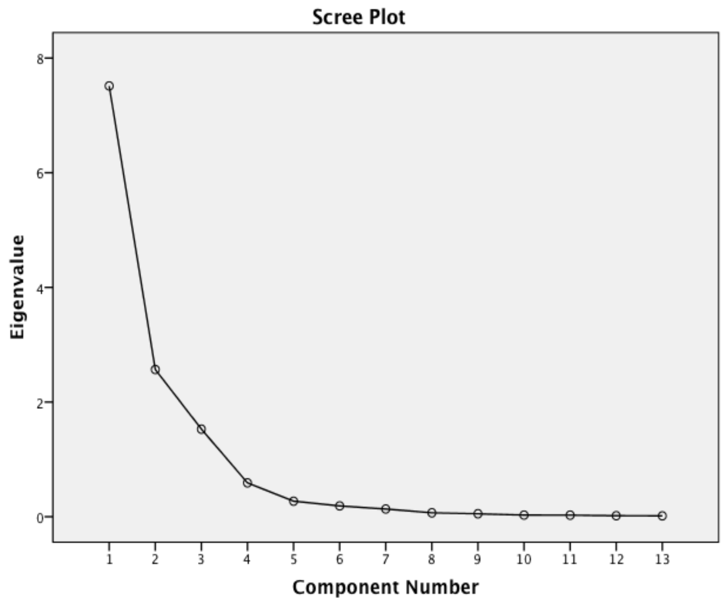

Figure 3 **Scree Plot**

Table 5 **Eigenvalue-one Criterion: Total Variance Explained by Initial PCA**

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.515	57.805	57.805	7.515	57.805	57.805	5.301	40.775	40.775
2	2.568	19.755	77.560	2.568	19.755	77.560	4.263	32.794	73.569
3	1.527	11.750	89.310	1.527	11.750	89.310	2.046	15.741	89.310
4	.591	4.544	93.855						
5	.270	2.081	95.935						
6	0.188	1.447	97.382						
7	0.134	1.031	98.413						
8	0.068	0.523	98.936						
9	0.051	0.394	99.330						
10	0.028	0.218	99.548						
11	0.027	0.207	99.755						
12	0.017	0.131	99.886						
13	0.015	0.114	100.000						

Note: Extraction Method: Principal Component Analysis

The variable “born in Central America” was ultimately removed from the analysis because of the lack of sampling adequacy as measured in the KMO test as well as the relatively low levels of correlation as measured in the correlation matrix. A final PCA was rerun, omitting the “born in Central America” variable and retaining only one component score.

3.3.1.3 PCA Results

Ultimately, one component score was retained with an eigenvalue of 7.5 and which accounts for 62.5 percent of the total variance (Table 7). Twelve demographic variables were incorporated into the principal component, with only one variable having been dropped: “born in Central America.” The output of the final PCA showed improvement from the initial PCA, as reflected in a higher overall KMO measure of .813, which according to Kaiser's (1974) classifications is “meritorious" sampling adequacy (Table 6) (Laerd Statistics 2013). Additionally, the KMO measures for individual variables are now all above .65 (Table 20, Appendices).

Table 6 **KMO and Bartlett's Test**

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		.813
Bartlett's Test of Sphericity	Approx. Chi-Square	936.359
	df	66
	Sig.	0.000

The correlation matrix and Bartlett's test of Sphericity (statistically significant with a p-value <.0005), indicate that overall there are correlations between the variables. While these indicators suggest that principal components analysis may be an appropriate method for reducing the number of variables in the analysis, on the other hand, there are indicators that multicollinearity may be a problem. Similarly to the original analysis, three variables continue to be correlated with other variables at $r \geq 0.9$ (Table 21, Appendices). Additionally, the determinant of the correlation matrix is 3.797E-010. A determinant <.00001 indicates that their may be a multicollinearity problem with the data although “strictly speaking,” when conducting PCA, this is not a concern (Field 2013, 21). Although there is concern about model fit with 81 percent (54) of the residuals computed between observed and reproduced correlations are nonredundant residuals with absolute values >0.05 (Table 22, Appendices) (Field 2013).

Table 7 **Eigenvalue-one Criterion: Total Variance Explained by Final PCA**

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.504	62.536	62.536	7.504	62.536	62.536
2	2.376	19.797	82.333			
3	1.236	10.297	92.630			
4	0.271	2.255	94.885			
5	0.197	1.643	96.527			
6	0.161	1.346	97.873			
7	0.094	0.782	98.655			
8	0.061	0.510	99.165			
9	0.033	0.272	99.437			
10	0.028	0.230	99.667			
11	0.023	0.196	99.863			
12	0.016	0.137	100.000			

Note: Extraction Method: Principal Component Analysis

The final output of the PCA, visualized in Figure 4, a unique component score generated for each state, is used in the regression in place of the twelve variables from which it was calculated. The component scores for each state (Table 8) are calculated by multiplying a weight, generated in the course of the PCA, by the original variable and summing the results (Laerd Statistics 2013).

The resulting component scores generated in the final PCA range from -2.05 (Vermont) to 1.54 (Arkansas). By comparing Figure 2 to Figure 4, it appears that in general (and with a few exceptions) those states with high component scores also have high ratios of the unauthorized out of total foreign-born and vise versa. This is one positive indicator of the suitability of using the principal component moving forward in the analysis.

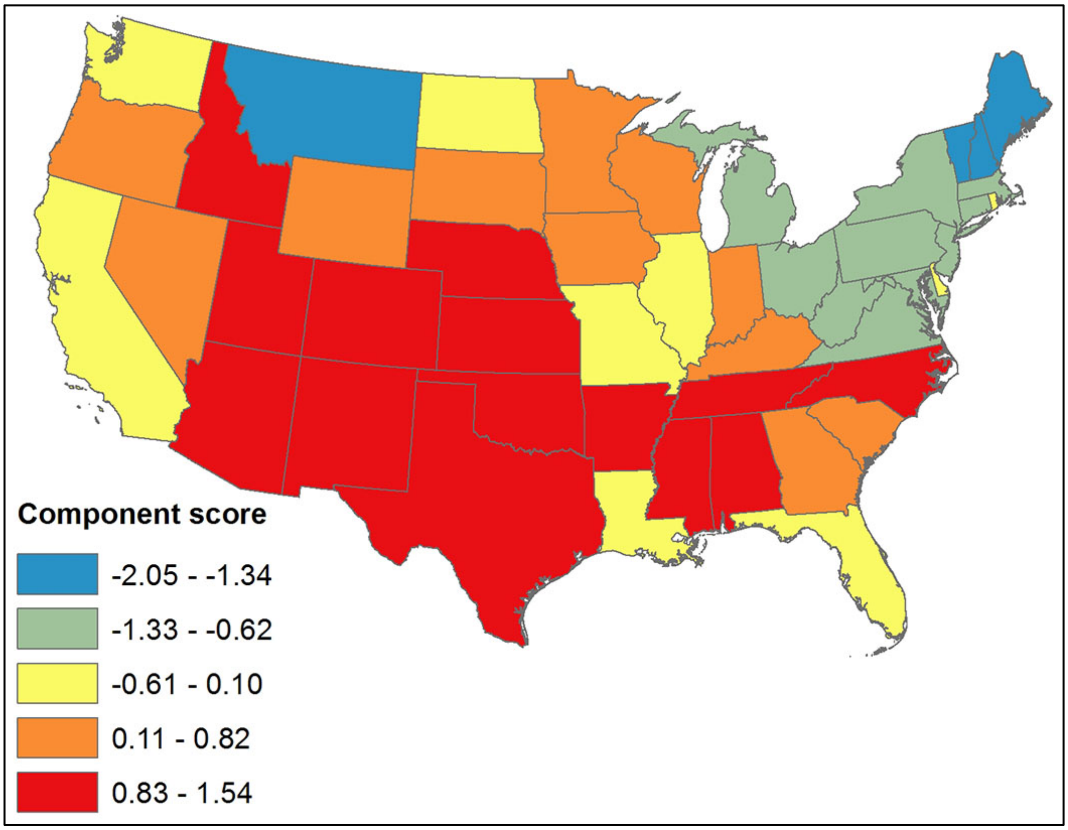

Figure 4 **Component Scores by State**

Table 8 **Component Score Coefficient Matrix**

Entered 2000 or later*	0.062
Entered 1980 or before*	-0.099
Speak a language other than English: Speak English 'less than well'*	0.114
65 years and over*	-0.112
Not a U.S. citizen*	0.125
Less than high school graduate*	0.118
Bachelor's degree or higher*	-0.099
Median income in the past 12 months	-0.080
Born in Mexico*	0.118
Median Age	-0.104
Income in the past 12 months below poverty level*	0.102
Hispanic or Latino*	0.115

*as a percent of the total foreign-born
Note: Extraction Method: Principal Component Analysis;
Rotation Method: Varimax with Kaiser Normalization

3.3.2 Exploratory Regression Analysis

Exploratory regression analysis looks at all possible combinations of independent or explanatory variables and outputs a list of passing models that meet the specified model parameters. Regression analysis was chosen as a method of analysis because of the complexity involved with estimating the percent of the unauthorized out of the total foreign-born population, in particular the challenge of estimating a population (the unauthorized) in which there is an overall lack of reliable data. Therefore, exploratory regression was used as a method to investigate all potential explanatory variables that may be important contributing factors for estimating the unauthorized population. Aside from the principal component, generated in the PCA, several independent variables related to immigrant settlement patterns and changes in settlement patterns were considered for inclusion in the analysis.

3.3.2.1 Independent variables

There is great variance in the percent of the unauthorized out of the total foreign-born population (the dependent variable) by state. In Vermont, the unauthorized make up just 1.2 percent of the total foreign-born compared to 53 percent in Alabama. Because of this wide variance in the percent of the unauthorized out of total foreign-born, and as supported by the literature, it is hypothesized that various state factors may affect immigrant settlement patterns and therefore the make-up of the immigrant population residing in a particular state. The settlement pattern variables are introduced below:

- **Immigrant growth rates**. Rather than focus on the underlying causes of the changes in immigrant settlement patterns, this analysis looks at changes in settlement patterns as reflected by state growth rates of the unauthorized as well as the foreign-born as a whole during three different time periods: (1) 1990-2000, (2) 2000-2010, and (3) the long-term growth rate: 1990-2010. All growth rates were calculated using either the Warren or ACS state level data.

- **Low unauthorized population**. In addition to the component score and growth rate variables, a dummy variable was created for the eight states that had an average of less than 3,000 unauthorized immigrants during the 2006-2010 analysis period. Of these eight states, five were estimated to have fewer than 1,000 persons. These eight states have significantly lower numbers of the total unauthorized, with all other U.S. states included in this study having of greater than 20,000 unauthorized persons (when taking the average of the analysis period). Because these eight states are outliers in many ways, a dummy variable was used as an attempt to account for some of the distinctive characteristics of these states rather than remove the states from the model. Removing these states was not preferable due to the already low number of cases (forty-eight contiguous U.S. states and Washington, DC) included in the analysis. Given the information and theory previously outlined, the following variables were chosen as potential explanatory variables in the exploratory regression analysis:

Table 9 **Independent Variables Included in the Exploratory Analysis**

Description	**Universe**	**Time period**	**Type**	**Spearman's rho: correlation with dependent variable**
Principal component (generated from PCA)	(see Table 3 for universe of input variables)	2006-2010	ordinal	.832**
States with growth rate >100 percent unauthorized	Unauthorized	2000-2010	nominal (dummy)	.361*
States with a decline in number of unauthorized	Unauthorized	2000-2010	nominal (dummy)	-.559**
Growth rate	Unauthorized	2000-2010	ratio	.542**
Growth rate	Foreign-born	2000-2010	ratio	.571**
Growth rate	Foreign-born	1990-2000	ratio	.825**
Growth rate	Foreign-born	1990-2010	ratio	.795**
States with more than double the nations mean immigrant growth rate	Foreign-born	1990-2000	nominal (dummy)	.559**
States with less than 3,000 unauthorized immigrants	Unauthorized	2006-2010	nominal (dummy)	-.636**

**Correlation is significant at the 0.01 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).

Additionally, the strength and direction of the relationships between the dependent variable (the percent of the unauthorized out of total foreign-born) and the potential independent variables (Table 9) were examined through visual inspection of scatterplots as well as Spearman's rank-order correlation before being introduced into the exploratory regression. Although the strength of the relationship varied, Spearman's rank order found that all potential explanatory variables were significantly correlated with the dependent variable and all variables were retained for inclusion in the exploratory regression.

3.3.2.2 Exploratory Regression Analysis Results

After careful considerations of the theory and examination of the data using the exploratory regression method, one model presented itself as most suitable for estimating the rate of the unauthorized population out of the total foreign-born. In order for a model to be considered "passing," it had to meet all of the following criteria:

- Minimum Adjusted R-Squared > 0.50

- Maximum Coefficient p-value < 0.05
- Maximum Variance Inflation Factor (VIF) Value < 7.50
- Minimum Jarque-Bera p-value > 0.10
- Minimum Spatial Autocorrelation p-value > 0.10

After careful consideration of the exploratory regression results, a four variable model met all of the model criteria and all variables were found to be statistically significant at the 0.01 level (see Table 10 and 11).

Table 10 **Passing Model Variables and Direction**

Variable	Time period	Description	Type	direction
Component score	2006-2010	Generated from PCA	interval	positive
Low unauthorized population	2006-2010	States with less than 3,000 unauthorized immigrants	nominal (dummy)	negative
Unauthorized growth rate	2000-2010	Unauthorized growth rate	ratio	positive
Immigrant growth rate	1990-2000	Immigrant growth rate	ratio	positive

Table 11 **Statistics of Passing Model**

Adjusted R-Squared	AICc	Jarque-Bera p-value	Koenker (BP) Statistic p-value	Max VIF	Factor Global Moran's I p-value
0.912280	-162.763921	0.440282	0.738423	2.540353	0.871584

3.3.3 Ordinary Least Squares

Once a suitable model was found using the exploratory regression method, Ordinary Least Squares (OLS) linear regression analysis was performed in order to model the relationship between key variables and the dependent variable. OLS regression analysis results in one set of coefficients that can be multiplied by each state's explanatory variables in order to produce an estimate of the percent of the unauthorized population out of the total foreign-born (dependent variable) for each state.

Regression equation:

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_n X_n + \varepsilon$

OR percent of the unauthorized population out of total foreign-born =

$\beta_0 + \beta_1$(population <3,000) + β_2(component score) +

β_3(unauthorized growth rate) + β_4 (immigrant growth rate) + ε

Where,

Dependent variable (Y)

Explanatory variables (X)

Intercept (β_0)

Coefficients ($\beta_{1 \ldots} \beta_n$)

Residuals (ε)

Table 12 **Retained Variables in OLS Regression**

Explanatory variable (x)	Coefficient (β)	StdError	t-Statistic	Probability	Robust_SE	Robust_t	Robust_Pr	VIF
Intercept	0.2446	0.0182	13.4763	0.000000*	0.0172	14.2366	0.000000*	--------
Low unauthorized population	-0.1511	0.0201	-7.5172	0.000000*	0.0239	-6.3332	0.000000*	1.5292
Unauthorized growth rate	0.0457	0.0133	3.4434	0.001273*	0.0158	2.8974	0.005844*	1.5158
Component Score	0.0527	0.0097	5.4466	0.000002*	0.0077	6.8600	0.000000*	2.5404
Immigrant growth rate	0.0586	0.0147	3.9862	0.000250*	0.0111	5.2668	0.000004*	2.1818

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$).

All signs are expected. All variables are statistically significant. No major aspatial autocorrelation as indicated by the low VIF scores.

Table 13 **OLS Regression Results**

Dependent Variable	% unauthorized out of total foreign-born
Input Features	Contiguous U.S. states
Number of Observations	49
Multiple R-Squared	0.919590
Adjusted R-Squared	0.912280
AICc	-162.7639

Joint F-Statistic	125.799207	Prob(>F), (4, 44) degrees of freedom	0.000000*
Joint Wald Statistic	592.529532	Prob(>chi-squared), (4) degrees of freedom	0.000000*
Koenker (BP) Statistic	1.985517	Prob(>chi-squared), (4) degrees of freedom	0.738423
Jarque-Bera Statistic	1.640680	Prob(>chi-squared), (2) degreesof freedom	0.440282

* An asterisk next to a number indicates a statistically significant p-value ($p < 0.01$).

Adjusted r-squared is .91, indicating that 91 percent of the variance in the dependent variable is explained by the model. The Jarque-Bera Statistic was not statistically significant, indicating that the residuals are normally distributed; a second test, Moran's I, was performed to test whether the residuals exhibit spatial randomness. The results of Moran's I test of spatial autocorrelation, as indicated by a z-score between -1.65 and 1.65 (z = .678) that is not statistically significant (p = .498), implies that the residuals are randomly spatially distributed (see Figure 5 for visual inspection of standard residuals). The default neighborhood search threshold for testing spatial autocorrelation was around 315 miles. To put this in perspective, it is roughly the driving distance from San Diego to Las Vegas or Boston to Philadelphia.

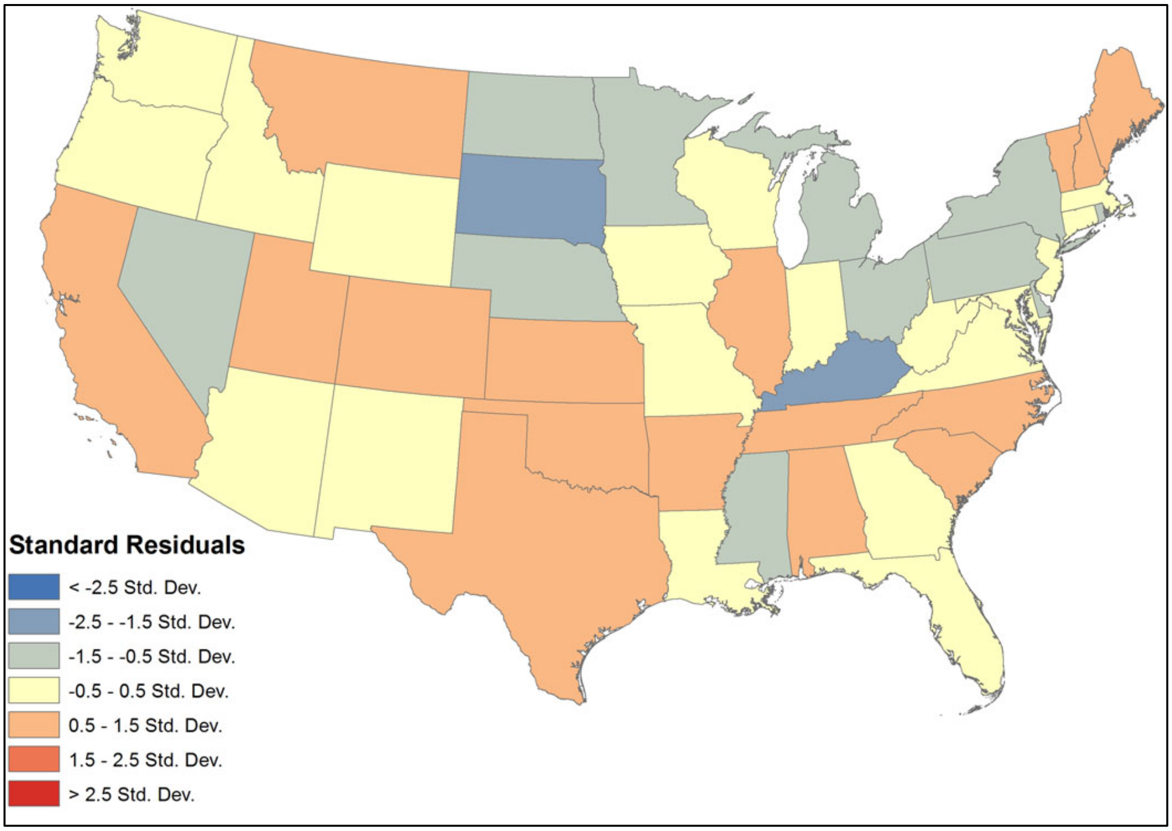

Figure 5 **OLS Standard Residuals**

The Koenker test is not statistically significant, signifying that the relationships between the explanatory variables and the dependent variable are non-stationary, and that the strength of the relationships is likely to stay relatively constant across geographies. Although a non statistically significant Koenker test indicates that the model may not be greatly improved by using geographically weighted regression (GWR), GWR was chosen to be performed regardless due to the known differences in the distribution and characteristics of the unauthorized population across the nation.

3.3.4 Geographically Weighted Regression

Once the OLS regression equation is properly specified, the same dependent and explanatory variables were included in a Geographically Weighted Regression (GWR) analysis. GWR analysis is a type of linear regression that allows for the strength and direction of relationships of variables to vary across space. Similar to the OLS regression, one of the outputs of GWR analysis is coefficients (β) to be multiplied by the explanatory variables (x) and summed to come up with an estimate of the percent of the unauthorized population out of the total foreign-born (dependent variable) for each state. The primary difference between the two methods is unlike OLS, which outputs one set of coefficient scores for all geographies, GWR outputs unique sets of coefficient scores (β) for each geography. In the case of this analysis, a unique set of coefficient scores (β) is specified for each state, resulting in forty-nine unique regression equations, one for each contiguous states and Washington, DC.

Table 14 **GWR Results**

Input Features	Contiguous U.S. states and D.C.
Number of Observations	49
Dependent Variable	% of unauthorized population out of total foreign-born
Multiple R-Squared	0.9314
Adjusted R-Squared	0.9158
Residual Squares	0.0664
Sigma	0.0412
AICc	-161.2483
Effective number	9.8813

Comparing the results of the GWR to the OLS analysis, the Akaike's Information Criterion (AICc) went up slightly, from -162.76 to -161.25 in the GWR, but the adjusted R-squared also went up slightly, from 0.9123 to 0.9158. Results of Moran’s I, test of spatial autocorrelation on StdResiduals of the GWR, show no spatial autocorrelation:

- Moran’s index: .021
- p-value: .729
- z-score: .346

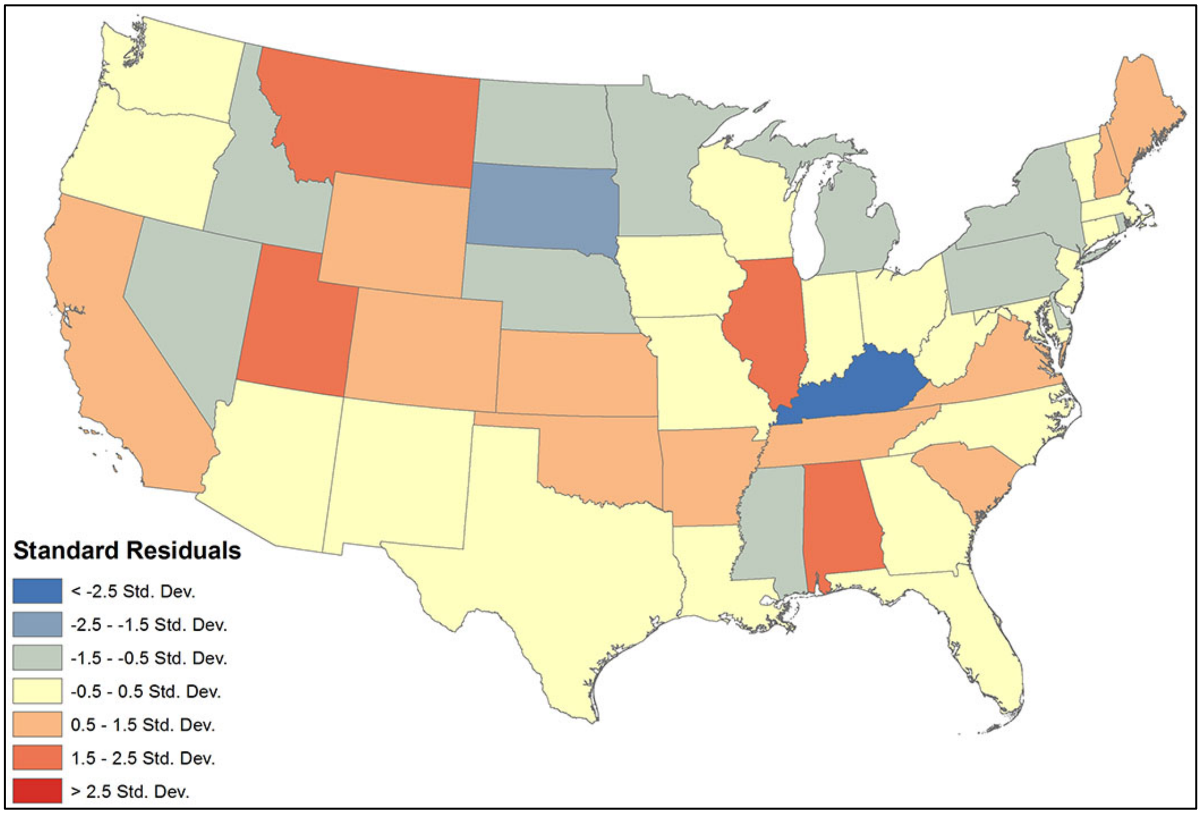

Figure 6 **GWR Standard Residuals**

In addition to outputting a unique regression equation and coefficient scores for each state in the analysis, GWR outputs an adjusted r-squared for each state. The results of the GWR show an adjusted r-squared value between .90-.92, indicating that, depending on the state, between 90 and 92 percent of the variance in the dependent variable is explained by the model.

Additionally, a measure of the influence of each independent variable as a predictor of the percent of the unauthorized immigrant population by state can be explored through each independent variable's coefficients. Figure 7 shows a clear spatial relationship between each variable's strength as a predictor of the dependent variable. In the case of the component score, the strength of this variable as a predictor of the dependent variable is strongest in Texas, Oklahoma, and Kansas and appears to diminish radially from these states. The two growth rate variables appear to have strongest influence on the East coast, with the strength diminishing from

East to West. While the GWR output variable coefficients for all states included in this study for the "<3,000 unauthorized immigrants" variable, because this is a dummy variable, only the eight states that meet this criteria vary in influence on the dependent variables. Additionally, because "<3,000 unauthorized immigrants" variable has a negative relationship with the dependent variable, a lower standard deviation of the variable coefficient indicates a stronger influence on the dependent variable.

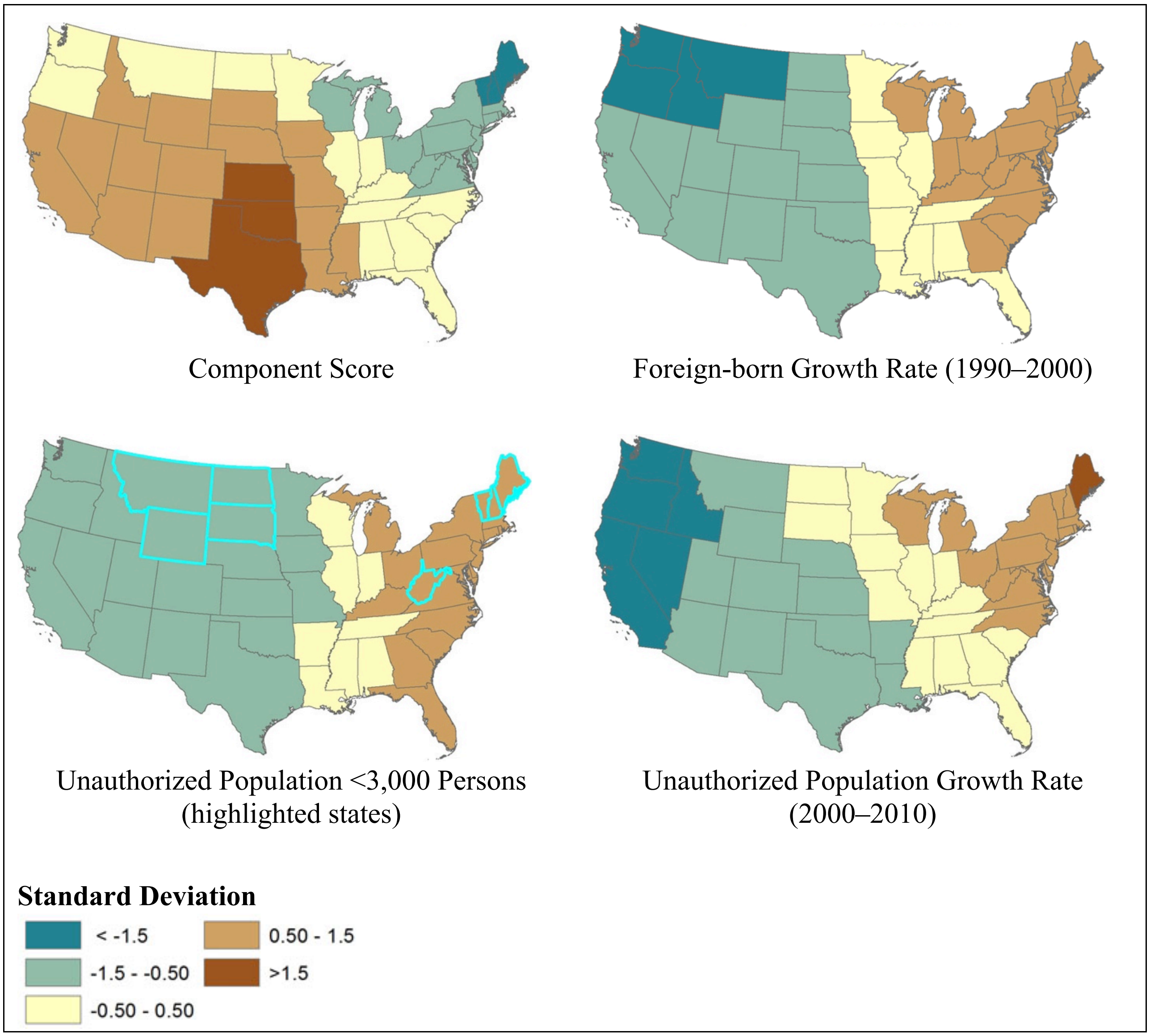

Figure 7 **Strength of Independent Variable Coefficients as Predictors of the % of the Unauthorized Population**

3.4 Census Tract Level Analysis

Once a regression equation was specified for each state using GWR, the next step of this analysis was to apply the state level equations to the corresponding census tract data in each state in order to generate an estimate of the percent of the unauthorized out of the total foreign-born population (dependent variable) for each individual census tract within the geography of the analysis. The key to making an estimate for each census tract was to calculate a unique component score for each census tract.

Although the variables remain the same, the data has changed from state level to census tract level data. Even though it is likely that the relationship between each variable is somewhat different at the census tract level than at the state level, the PCA is not rerun using the census tract data, but the component scores for the census tracts are computed using the coefficient scores generated in the state level analysis (Table 8). The reasoning for not conducting a new PCA using the census tract data is that in order to estimate the unauthorized population, the relationship between the dependent variable and the principal component (generated in the state level PCA) as well as the rest of the independent variables must be defined. This was done using GWR at the state level and is explained in the previous section.

If the PCA were to be rerun using census tract data, the relationship between the component score and the dependent variable would have to be redefined. This is simply not possible because no estimates of the dependent variable exist at the census tract level. Therefore, acknowledging the flaws in this method, the relationships between the independent and the dependent variables were defined at the state level and then applied to the census tract level. The process of using the state level equations to generate census tract level estimates is explained in the following sections.

3.4.1 Calculating a Unique Component Score for each Census Tract

The first step to calculating individual estimates for each census tract is to calculate a unique component score for each census tract. In fact, when applying each state level equation (generated in the GWR) to the census tracts within the state, the only input that changes is the component score variable. Therefore, the key to making unique estimates for each census tract is the component score.

As previously mentioned, the PCA was not rerun using the census tract level data. Instead, component scores were generated for each census tract using the coefficient scores previously generated in the PCA. The component scores for the census tracts were computed manually by multiplying the twelve demographic variable data specific to each census tract by the coefficient scores previously generated in the PCA (Table 8) and then summing the results.

The same coefficient scores were used to calculate every component score generated in this study (for every states and census tract). While the coefficient scores stay the same, the demographic variable inputs are specific to the geography for which the component score is being calculated. The equation, with the coefficient scores, follows:

component score equals (=):

0.062 (Entered 2000 or later)
+ -0.099 (Entered before 1980)
+ 0.114 (Speak English 'less than well')
+ -0.112 (65 years and over)
+ 0.125 (Not a U.S. citizen)
+ 0.118 (Less than high school graduate)
+ -0.099 (Bachelor's degree or higher)
+ -0.080 (Median income in the past 12 months)
+ 0.118 (Born in Mexico)
+ -0.104 (Median Age)

+ 0.102 (Income in the past 12 months below poverty level)
+ 0.115 (Hispanic or Latino)

After calculating a unique component score for each census tract in the United States, the scores were inserted into their respective state level regression equations (depending on which state the census tract was located) in order to come up with an estimate of the rate of the unauthorized out of the total foreign-born for each census tract in the United States.

For example, given the regression equation,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots \beta_n X_n$$

Where,

Estimate of the dependent variable (Y)
Explanatory variables (X)
Intercept (β_0)
Coefficients ($\beta_{1 \ldots} \beta_n$)

From the results of the GWR, the equation for California was found to be:

$$Y = .2697 + -0.1875\, x_1 + 0.0575\, x_2 + 0.0393\, x_3 + 0.0412\, x_4$$

Where,

Y = percent of unauthorized out of total foreign born (dependent variable)
x_1= dummy variable for unauthorized population <3,000 (where "1" indicates <3,000 persons and all other states are "0".)
x_2= component score
x_3= unauthorized growth rate (2000-2010)
x_4= immigrant growth rate (1990-2000)

In order to come up with estimates for each census-tract level (in this case for the state of California) the only variable that would change in the equation would be (X_2), the component score. After calculating the percent of the unauthorized out of total foreign-born, the final step is to multiply the result by the total number of foreign-born per census tract as released by the 2006-2010 ACS.

For example, take two census tracts (A and B):

Census tract	Component score	Total foreign-born
A	**0.0554**	**1,000**
B	**1.13**	**200**

Census tract A

- Calculate the percent of the unauthorized out of total foreign-born using the regression equation for California (generated in GWR):

  0.2697 + -0.1875(0)+ 0.0575(**0.0554**) + 0.0393(0.1283) + 0.0412(0.3724) = **0.2933,**

- Calculate the total unauthorized (percent of the unauthorized out of total foreign-born multiplied by total foreign-born):

  0.2933 * **1,000** = **293**

Results: 29.33 percent of the foreign-born population is unauthorized, an estimated **293** unauthorized out of the **1,000** foreign-born persons.

Census tract B

- Calculate the percent of the unauthorized out of total foreign-born using the regression equation for California (generated in GWR):

  0.2697 + -0.1875(0) + 0.0575(**1.13**) + 0.0393(0.1283) + 0.0412(0.3724) = **0.3551**

- Calculate the total unauthorized (percent of the unauthorized out of total foreign-born multiplied by total foreign-born):

  0.3551 * 200 = 71

Results: 35.51 percent of the foreign-born population is unauthorized, an estimated **71** unauthorized out of the **200** foreign-born persons.

If census tracts A and B are the exact same size, hypothetically, in a neighborhood that consisted of only these two census tracts, with all other factors the same, more services and/or greater outreach should be provided in census tract A than B, given that there is a higher density unauthorized persons in census tract A.

CHAPTER 4: RESULTS

This chapter reviews the results of the analysis as visually represented using dot density renderer in ArcGIS Desktop. Estimates generated in this analysis are then verified by comparing the results of this analysis with estimates made in prior analyses.

4.1 Relative Densities and Distribution

The estimates of the number of unauthorized by census tract generated in this study were not released. Rather, the estimates were visualized using dot density renderer in ArcGIS as a method for communicating relative densities and concentrations of the unauthorized population. This analysis concludes with relative density maps rather than releasing estimates for each census tract, for two primary reasons:

1. Census tract boundaries do not have much meaning on their own in regards to this analysis, as they are administrative boundaries that do not necessarily correspond with neighborhood or community boundaries nor service areas for providing immigrant services. In fact, in dense areas, several hundred or even thousands of census tracts could be located in a particular service areas. When taken together, on the other hand, the total number of unauthorized per census tract paints a picture of the landscape of the service area.
2. Given that there is not even a consensus as to how many unauthorized people reside in the entire United States, it is unreasonable to believe that the number of unauthorized population can be estimated at as fine a geographic scale as the census tract level with any real accuracy. Rather than try to present these frequencies, the total numbers are used to present the relative density or distribution of the population.

4.1.1.1 Dot Density Renderer

The dot density renderer displays the number estimates of the unauthorized as a random dot pattern within each census tract, where each dot represents a certain number of people. In order to maintain density, as the zoom level increases, the number of people represented by each dot diminishes, while the size of the dot stays the same. Using dot density renderer in Esri ArcGIS Desktop is the preferred method for presenting the results for the following reasons:

- The optimal dot to person ratio can be manually adjusted to best communicate density depending on the particular geography being displayed.
- By mapping the results, the distribution patterns and clusters of high numbers of unauthorized become apparent. This would be difficult to determine looking at a table of estimates alone.
- The results could in the future be combined with other potentially relevant infrastructure information for planning purposes, such as accessibility by public transportation or existing physical office locations of service providers.

4.1.1.2 Maps of Relative Density

In Figure 8, the density of the unauthorized population is displayed for the entire United States. In Figure 8 one dot represents 1,500 people. Because the density of the unauthorized population varies greatly, the dot density map is not particularly informative at this level.

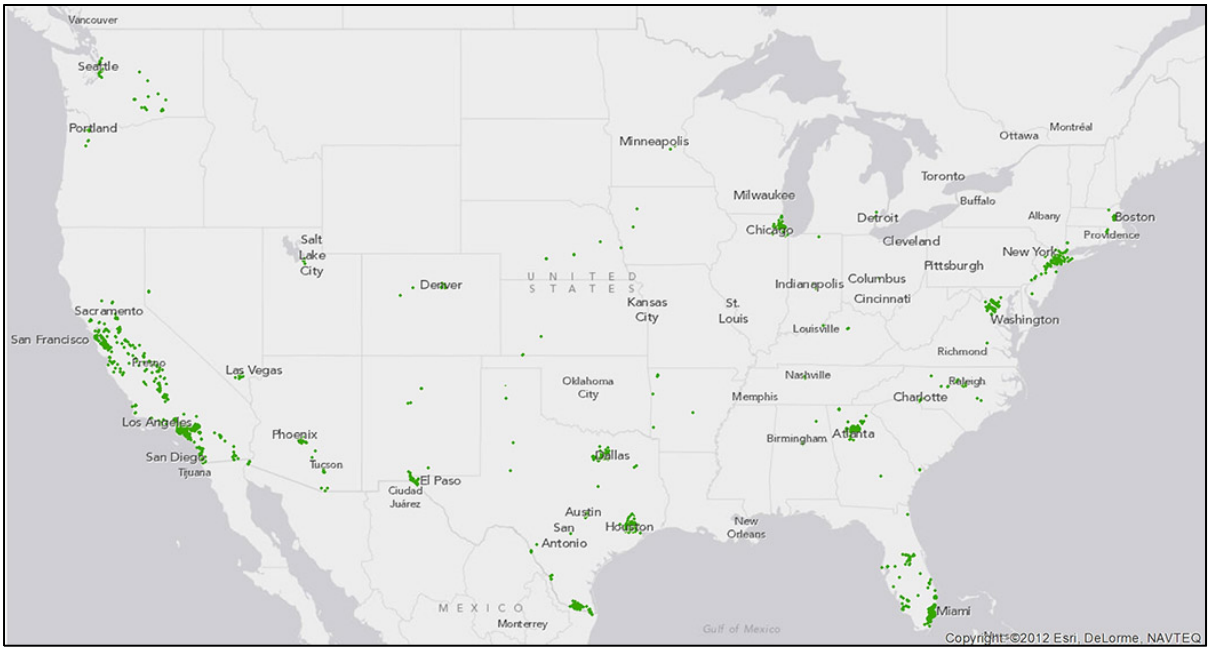

Figure 8 **Unauthorized Population by Census Tract in the United States.** 1 dot = 1,500 people

Figure 9 starts to show areas of density in California. In Figure 9, one dot represents 500 people. While this is potentially useful for state level planning and implementation, that is not the goal of this analysis.

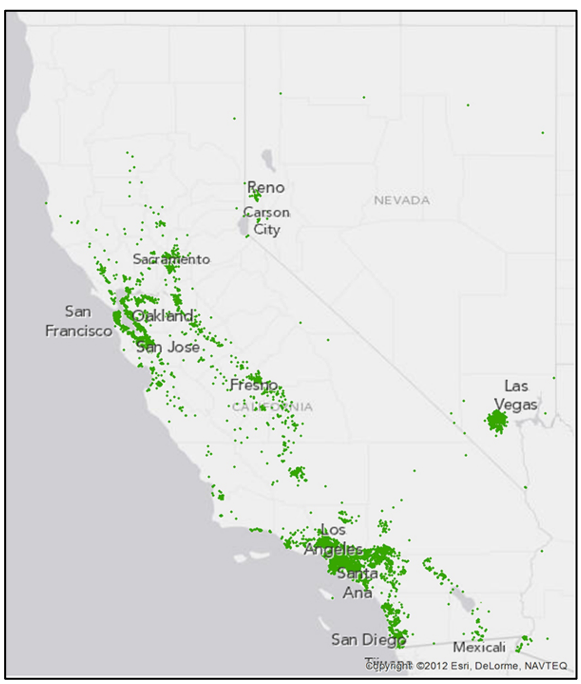

Figure 9 **Unauthorized Population by Census Tract in California.** 1 dot = 500 people

The applicability of the analysis for local level planning starts to become apparent by looking at Figure 10 (one dot represents one-hundred people) and even more so with Figure 11 (one dot represents fifty people). Although only four maps are presented here, using this methodology and the dot density renderer scheme, maps could be made for virtually any geography in the 48 states and Washington, D.C.

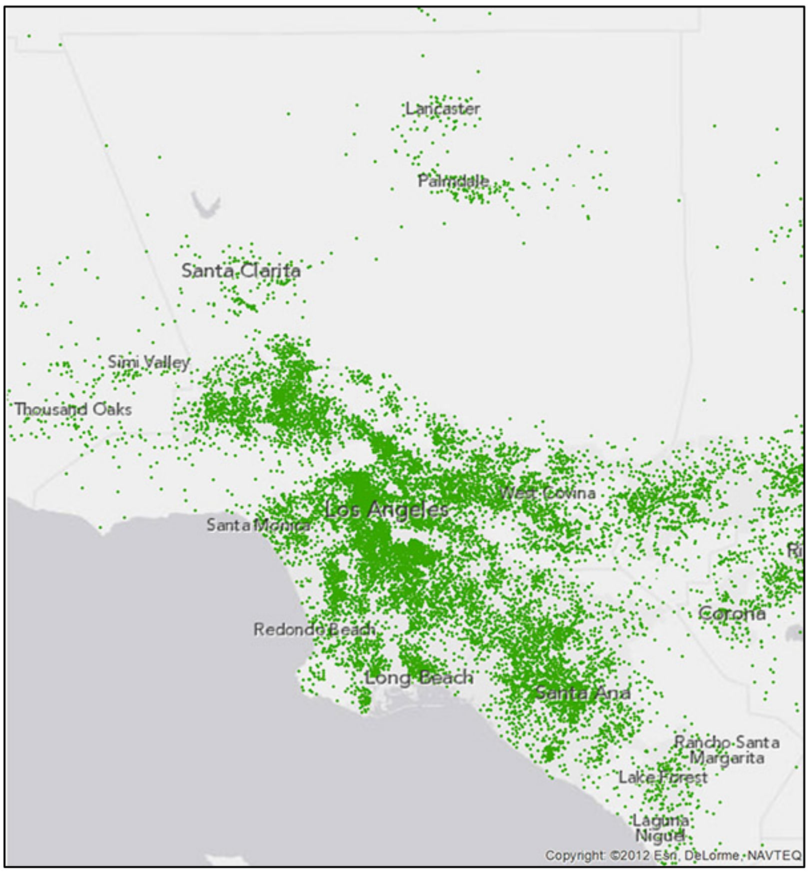

Figure 10 **Unauthorized Population by Census Tract in Los Angeles County.** 1 dot = 100 people

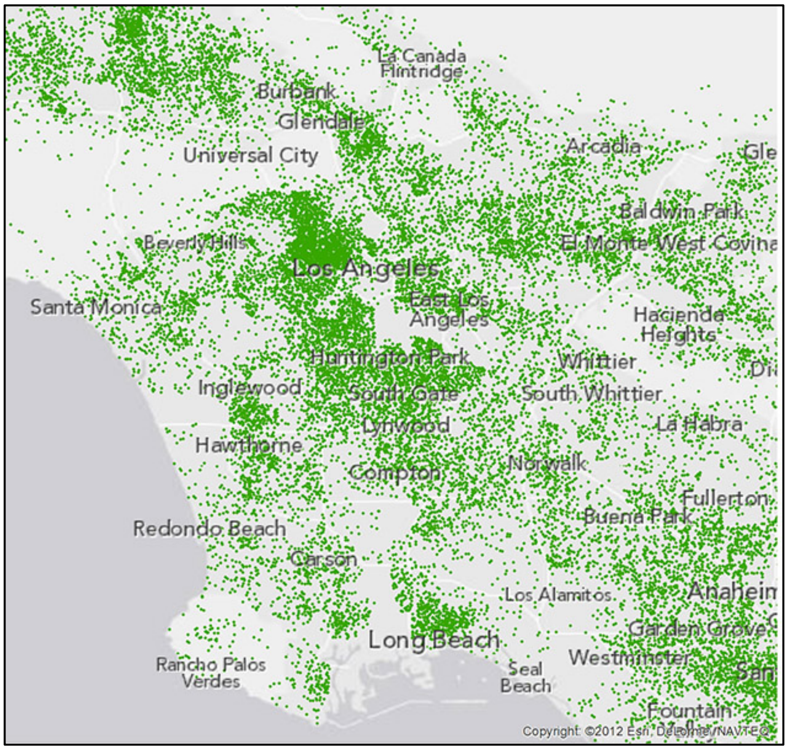

Figure 11 **Unauthorized Population by Census Tract in Los Angeles.** 1 dot = 50 people

4.2 Model Performance and Verification of Results

The primary method for drawing conclusions about the accuracy of the estimates produced in this analysis, was to compare the results of this analysis to those of other studies. Specifically, the census tract level estimates produced in this analysis were summed up to various geographies and compared to estimates made for those geographies by Warren and Warren, the Public Policy Institute of California (PPIC) and the USC Center for the Study of Immigrant Integration (CSII). In the following sections, the estimates generated in this study are referred to as "Fischer" estimates.

4.2.1 Comparison with State Level Estimates Generated by Warren and Warren

The census tract level estimates generated in this analysis were summed by state and compared to those estimates generated by Warren (2013) using the residual method. The state level estimates generated by Warren are the very same estimates upon which this analysis was based. The Warren estimates were the numerator in the dependent variable (percent of the unauthorized out of total foreign born) for the state level regression analyses conducted in this study. The absolute percent differences of estimates generated in this study were compared to those generated by Warren were calculated by taking the absolute value of the following equation: (Warren estimate - Fischer estimate) / Warren estimate. The results are presented below in Table 15 and Figure 12.

Table 15 **Absolute % Difference Between Estimates**

	Unauthorized population estimates by…		absolute % difference of Fischer estimate from Warren estimate
	Warren	Fischer	
California	3,059,069	3,074,782	1
Georgia	395,838	391,721	1
Louisiana	51,467	52,021	1
Mississippi	23,807	24,186	2
Idaho	34,183	33,529	2
Iowa	46,373	45,150	3
Indiana	103,268	100,294	3
Wisconsin	81,988	79,384	3
Nebraska	39,494	40,832	3
Arizona	343,887	327,291	5
Virginia	264,453	277,257	5
Oregon	134,817	127,897	5
Missouri	70,031	73,651	5
New Mexico	80,317	75,570	6
Nevada	184,848	202,035	9
North Carolina	354,355	320,780	9
South Carolina	99,470	88,615	11
Texas	1,612,281	1,429,830	11
Kansas	72,618	64,332	11
Washington	255,464	284,721	11
Tennessee	130,475	115,481	11
Florida	988,384	1,106,241	12
Maryland	215,259	242,902	13
Colorado	207,881	181,054	13
Wyoming*	2,945	2,550	13
Illinois	598,574	518,163	13
Delaware	21,337	24,414	14
Arkansas	64,789	54,787	15
Montana*	793	924	17
Oklahoma	87,584	72,490	17
Utah	102,534	84,840	17
Washington, DC	23,006	18,962	18
Connecticut	112,595	133,308	18
Minnesota	102,516	121,538	19
New Jersey	416,144	494,469	19
Alabama	84,291	67,223	20
Kentucky	43,809	52,973	21
Ohio	98,564	122,933	25
Rhode Island	27,985	35,355	26
Massachusetts	202,790	257,687	27
New York	756,996	1,052,310	39
Pennsylvania	148,215	208,818	41
Michigan	91,766	149,391	63
Maine*	2,024	3,645	80
West Virginia*	818	1,814	122
South Dakota*	953	2,166	127
New Hampshire*	2,047	5,181	153
North Dakota*	425	1,209	185
Vermont*	298	1,693	468

*states with unauthorized population <3,000 persons as estimated by Warren (2013)

The states with the largest absolute percentage difference from the Warren estimates were the six out of eight states with immigrant populations less than 3,000, indicating that this study's analysis method may not be suitable for states with such low unauthorized immigrant populations. While the actual absolute difference between the estimates was between 784 and 3,134 persons for those states with <3,000 unauthorized persons, this equated to an 80 to 468 percent difference from the original Warren estimates. For example, the Warren estimate for Vermont is 298 persons, while this study estimated 1,693 persons. This equates to a difference of 1,395 people or 468 percent of the total Warren estimate. Of these 8 states, Wyoming and Montana performed moderately, with differences from the Warren estimates being 13 and 17 percent respectively. Excluding the six "low-population" states that performed very poorly, of the remaining forty-three states and D.C. included in this analysis the results varied:

> **Very good.** Sixteen states had less than 10 percent difference from the Warren estimates, with California's estimate being the best with a less than 1 percent difference, followed closely by Georgia, Louisiana, Mississippi, and Idaho (all less than 2 percent).
>
> **Good.** Nineteen states performed well with differences from 10-20 percent of the Warren estimates.
>
> **Moderate:** Five states performed moderately with 20-30 percent differences from the Warren estimates.
>
> **Poor**: Four states performed poorly with differences of over 39 percent from the Warren estimates, with Michigan having the largest percent difference, 62.8 percent.

Each state's absolute percent difference from the Warren estimates was mapped in order to identify spatial patterns. Based on the absolute value measures, Figure 12, illustrates that the model performed well in the western United States. While, seven out of nine states that performed the worst (with an absolute difference of 31 percent or higher) were located in the northeast of the United States.

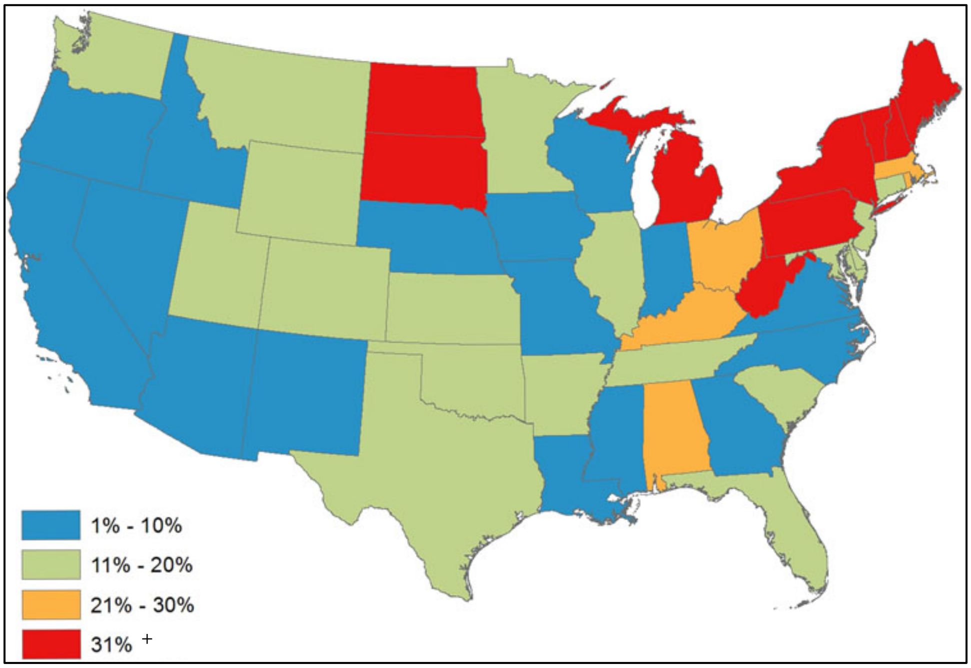

Figure 12 **Absolute % Difference from Warren Estimates**

4.2.2 Comparing Results to Independent Sub-state Estimates for California

The results were further verified at the county level in California, where independent estimates of the unauthorized population have been released by the PPIC and CSII. In order to compare the results of this study with those of PPIC and CSII, the census tract level estimates were summed to correspond with the county areas for which PPIC and/or CSII estimates have been released. Table 16 shows the results of the comparison, with the estimates generated in this study labeled as “Fischer” estimates.

Table 16 **Absolute % Difference Between Estimates of the Unauthorized by Region in CA**

	Unauthorized population estimates			Absolute % difference of Fischer estimate from…	
	Fischer (2006-10)	CSII (2009-11)	PPIC (2008)	CSII	PPIC
EAST BAY (Alameda & Contra Costa Counties)	201,935	153,910	203,000	31	1
INLAND EMPIRE (San Bernardino and Riverside Counties)	290,473	259,130	296,000	12	2
ORANGE COUNTY	274,677	236,569	289,000	16	5
SILICON VALLEY (Santa Clara and San Mateo Counties)	249,168	173,815	235,000	43	6
CENTRAL VALLEY (Fresno, Kern, Kings, Madera, Merced, San Joaquin, Stanislaus, and Tulare Counties)	286,978	331,584	260,000	13	10
BAY AREA (Alameda, Contra Costa, Marin, Napa, Santa Clara, San Mateo, and San Francisco Counties)	552,499	386,947	498,000	43	11
LOS ANGELES COUNTY	1,081,991	892,081	916,000	21	18
SACRAMENTO METRO (El Dorado, Placer, Sacramento, Sutter, Yolo, and Yuba Counties)	118,398	83,480	not available	42	not available

Looking at the regional estimates in Table 16, it appears that the estimates produced in this study are comparable to those produced by PPIC with a 1–18 percent absolute difference. On the other hand, there is a 12–42 percent difference from the estimates of CSII. It is important to note that similar to the methodology of this study, PPIC used the Warren estimates as the basis of their estimates, while CSII does not (For more detailed information about their estimates, see section: Residual Method Combined With Other Methods For Sub-state Estimates) (Pastor and Marcelli 2013; Hill and Johnson 2011). Another reason that the estimates generated in this study may differ from those of other studies is because the study period differs.

Because this analysis is focused on the relative densities or the distribution of the unauthorized population, another test of the validity of the results was conducted by looking at the estimated differences of the distribution of the unauthorized population between studies. The first step was to calculate the distribution or percent, rather than the frequencies, of the unauthorized population by region in the state of California. This was calculated by dividing each regional estimate by the corresponding total state estimate. For example, in the case of CSII, all regional estimates were divided by CSII's estimate for the state of California, 2,654,752**.** The results being the distribution or percent of the total unauthorized population by region across the state of California. The results are compared in Table 17 and 18 with the estimates generated in this study labeled as "Fischer" estimates.

Table 17 **Estimates of the Total Unauthorized Population in CA**

	CSII	**PPIC**	**Fischer**
California	2,654,752	2,876,000	3,074,782

Table 18 **Differences in the Distribution of Unauthorized Population by Region in CA**

	% of unauthorized by region as estimated by…			**Percentage point difference from Fischer results**	
COUNTY	**CSII**	**PPIC**	**Fischer**	**CSII**	**PPIC**
ORANGE County	8.9	10.0	8.9	0.0	-1.1
INLAND EMPIRE (San Bernardino and Riverside Counties)	9.8	10.3	9.4	-0.3	-0.8
EAST BAY (Alameda & Contra Costa Counties)	5.8	7.1	6.6	0.8	-0.5
SILICON VALLEY (Santa Clara and San Mateo Counties)	6.5	8.2	8.1	1.6	-0.1
CENTRAL VALLEY (Fresno, Kern, Kings, Madera, Merced, San Joaquin, Stanislaus, and Tulare Counties)	12.5	9.0	9.3	-3.2	0.3
BAY AREA (Alameda, Contra Costa, Marin, Napa, Santa Clara, San Mateo, and San Francisco Counties)	14.6	17.3	18.0	3.4	0.7
LOS ANGELES County	33.6	31.8	35.2	1.6	3.3
SACRAMENTO METRO (El Dorado, Placer, Sacramento, Sutter, Yolo, and Yuba Counties)	3.1	not available	3.9	0.7	not available

While the estimate of the unauthorized population made in this analysis varied by 12–42 percent from those of CSII when comparing the number estimates of the unauthorized, the differences in distribution of the unauthorized varied by 0.02–3.4 percentage points. These results are encouraging because while the estimates generated in this analysis vary from those of PPIC and CSII, the differences between this study and the other two leading studies in the percent of the unauthorized by regions in California is no more than 3.4 percentage points (Bay Area).

CHAPTER 5: CONCLUSIONS

While the validation method indicates that the methodology generated in this study may be an appropriate analysis method for estimating the unauthorized population at the census tract level, it is worthwhile to discuss some limitations of this analysis. This chapter begins with the weaknesses, challenges and limitations of this analysis method, focusing on limitations around data availability. The chapter continues with ideas for future research, including suggestions on refining the methodology as well as the need for greater verification of results.

5.1 Weaknesses, Challenges, Limitations and Next Steps

There were a number of challenges in the analysis, including lack of available data, missing data and data uncertainty, as well as concerns of model accuracy, and difficulty in verifying reliability of the methodology and overall results of the analysis. Additionally, there are a number of ways the research presented in this report could be continued in order to strengthen and further verify the results. Lastly, the visual display of the results could be refined and presented in a way that allows users to interact and query the results based on their area of interest.

5.1.1 Missing Data and Data Uncertainty

One weakness of the analysis is the number of census tracts with missing demographic data. Estimates could not be calculated for census tracts with missing variables. Due to missing data, no estimate were generated for 4,420 census tracts, roughly 6 percent of the 72,539 census tracts within the geography of this analysis. Fourteen states had over 10 percent of census tracts where no estimates were generated due to at least one missing demographic variable. The missing predictions could greatly change the impression of the visual patterns in the analysis and therefore changing the interpretation of the results. Additionally, the census tracts with missing variables could have greatly changed the interpretation of the verification method, which involved summing all of the census tract estimates. That being said, it is hypothesized that a great many of census tracts with missing demographic variables are missing because of very low numbers of foreign-born (and therefore likely low numbers or no unauthorized), but this may not exclusively be the case. See Appendices, Table 23, for the percentage of census tracts with missing variables by state.

Additionally, all of the variables used in the analysis were estimates rather than known counts. This is unavoidable, given that no known counts exist of the unauthorized population. As previously explained, the Warren estimates of the unauthorized population are made using the residual method. The ACS data used to generate the component scores is also an estimate, albeit statistically sound, based off a survey of a subset of the population (U.S. Department of Commerce 2008).

5.1.2 Ecological Fallacy

The estimates of the rate of the unauthorized out of the total foreign-born for each census tract were based on the relationship of the independent variables at the state level. The state level relationships are assumed to be the same at the census tract level in order to make census tract level estimates. The ecological fallacy being that inferences about the group at the census tract level are deduced from correlations of the variables at the state level. While this is not ideal, the assumptions about the census tract level relationships between variables was necessary given the lack of data available at the census tract level.

5.1.3 Refinement of Independent Variables

While the choice of a dependent variable is quite limited by data available (or lack thereof), there are numerous possibilities on independent input data, particularly demographic data input in the

PCA. In future analysis, it is recommended that greater exploration be conducted to determine which variables to include in the analysis. There are two weaknesses to the PCA analysis that are further discussed in the next sections: (1) a concern of the multicollinearity and singularity of the data, and (2) not all characteristics of the unauthorized were captured in the input data.

5.1.3.1 Multicollinearity and Singularity of PCA Input Data

As previously mentioned, while there were a number of indicators that PCA was an appropriate analysis method, there were other indicators that there may be a problem with the multicollinearity of singularity of the data, namely that some variables were measuring ostensibly the same thing. In a future analysis, including or omitting variables should be considered, particularly those variables that were related to year of entry in the U.S. and age.

5.1.3.2 Differing Characteristics of the Unauthorized

There is also the issue that the unauthorized population is not uniform. While similar demographic characteristics may be used to describe the majority of the unauthorized population, in reality, every person has varying combinations of demographic characteristics. There are certainly characteristics, other than the ones included in this analysis, which would better differentiate different groups of unauthorized.

5.1.4 Improved Method for Verifying the Results

Further verification of the results is necessary to determine the reliability of the methodology outlined in this report. As previously discussed, this report has only verified the results at the sub-state level for the state of California and no verification has been conducted for the accuracy of the estimates below the county level. A similar method of comparing the results of this report to that of prior studies, could be conducted for other states where prior studies have produced sub-state estimates. To verify the results at a finer geography than that of other studies (such as the census tract level, where it is believed that no other estimates exist), a survey that asks about legal status could be conducted.

5.1.5 Sensitivity and Reliability Analysis

Another method for verifying the robustness of the analysis method would be to conduct a sensitivity analysis to see how sensitive the analysis is to changes in the analysis inputs. There are countless ways that the analysis or the input data could be adjusted to conduct a sensitivity analysis. One idea is to fill in a number in place of the missing variables to see how big an influence the missing variables may have on the results. Similarly, reliability analysis for the PCA could be conducted in SPSS.

5.1.6 Refine Display of Results

5.1.6.1 Use of Masking in the Dot Density Renderer

The results of this analysis is an estimate of total unauthorized population for each census tract in the United States, visualized in ArcGIS Desktop using dot density renderer to display relative densities. The dot density rendering method could be improved upon through the use of masking. Through masking, the area for which dots can be rendered is restricted within the polygon boundaries (census tracts in the case of this analysis) to those that may be inhabited. No dots would be rendered in areas that are within the census tract boundaries but are clearly uninhabited, such as bodies of water or national park land. By removing uninhabited lands from the rendering area, the dot density renderer more accurately displays relative density.

5.1.6.2 Interactive Web Application

Creating an interactive web application could increase the accessibility of the results of this research. A web application would allow users to explore the results of the analysis for their geography of interest, without requiring the manual adjustment of density display properties. Ideally, the dot density renderer would automatically adjust the dot size and density display properties to best communicate distribution and relative densities in the selected area.

The optimal density display properties would need to be more nuanced than those that are standard in ArcGIS Desktop that maintain density by making adjustments based solely on zoom level, but would also require consideration of the average number of unauthorized in the geography being viewed. While these settings can be manually adjusted in ArcGIS Desktop, automating this process and making it available online may improve the access and therefore usefulness of this tool as an applied research product.

5.2 Lessons Learned and Potential Impacts

Overall, the results of this analysis indicate that the method designed in this study may be a viable means for estimating the unauthorized at the neighborhood level, at least in certain geographies, such as the West Coast. That being said, this is a first attempt at an entirely new methodology, which will undoubtedly require both refinement of the method and greater verification of the results before being useful for planning purposes. Now is the time to start investigating methods such as this one, so that if and when immigration reform occurs, those on the ground providing services to the unauthorized will have the information needed to effectively and efficiently process potentially upwards of 8 million people.

REFERENCES

Baker, Bryan. 2010. "Naturalization Rates among IRCA Immigrants: A 2009 Update." Washington, D.C.: Office of Immigration Statistics, Department of Homeland Security. Retrieved http://www.dhs.gov/xlibrary/assets/statistics/publications/irca-natz-fs-2009.pdf (last accessed 11 August 2014).

Baker, Bryan, and Nancy Rytina. 2013. "Estimates of the Unauthorized Immigrant Population Residing in the United States: January 2012." Washington, D.C.: Office of Immigration Statistics, Department of Homeland Security. Retrieved http://www.dhs.gov/sites/default/files/publications/ois_ill_pe_2012_2.pdf (last accessed 10 August 2014).

Bohn, Sarah. 2009. "New Patterns of Immigrant Settlement in California." San Francisco, CA: Public Policy Institute of California. Retrieved http://www.ppic.org/content/pubs/report/R_709SBR.pdf (last accessed 10 August 2014).

Chan, Alex, Joanna Kabat, and Jesse Reyes. 2013. "Implementing Immigration Reform in Los Angeles: Lessons from DACA" Los Angeles, CA: Tomas Rivera Policy Institute, *forthcoming*.

Esri. Released 2012. ArcGIS Desktop for Windows, Version 10.1. Redlands, CA: Esri.

Field, Andy P. 2013. "Chapter 17: Exploratory Factor Analysis" In *Discovering Statistics using IBM SPSS Statistics*. London; Thousand Oaks, CA: SAGE Publications. Retrieved http://www.sagepub.com/field4e/study/resources.htm (last accessed 15 August 2014).

Fortuny, Karina, Randy Capps, and Jeffrey S. Passel. 2007. "The Characteristics of Unauthorized Immigrants in California, Los Angeles County, and the United States." Washington, DC: The Urban Institute. Retrieved http://www.urban.org/UploadedPDF/411425_Characteristics_Immigrants.pdf (last accessed 10 August 2014).

Hill, Laura, and Hans Johnson. 2011. "Unauthorized Immigrants in California: Estimates for Counties." San Francisco, CA: Public Policy Institute of California. Retrieved http://www.ppic.org/main/publication.asp?i=986 (last accessed 10 August 2014).

Hill, Laura, and Joseph Hayes. 2013. "Unauthorized Immigrants." San Francisco, CA: Public Policy Institute of California. Retrieved http://www.ppic.org/main/publication_show.asp?i=818 (last accessed 10 August 2014).

Hoefer, Michael, Nancy Rytina, and Bryan Baker. 2011. "Estimates of the Unauthorized Immigrant Population Residing in the United States: January 2010." Washington, D.C.: Office of Immigration Statistics, Department of Homeland Security. Retrieved http://www.dhs.gov/xlibrary/assets/statistics/publications/ois_ill_pe_2010.pdf (last accessed 10 August 2014).

———. 2012. "Estimates of the Unauthorized Immigrant Population Residing in the United States: January 2011." Washington, D.C.: Office of Immigration Statistics, Department of Homeland Security. Retrieved http://www.dhs.gov/xlibrary/assets/statistics/publications/ois_ill_pe_2011.pdf (last accessed 10 August 2014).

IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.

Judson, Dean H., David A. Swanson. 2011. Estimating Characteristics of the Foreign-Born by Legal Status: An Evaluation of Data and Methods. Vol. 2. Dordrecht; New York: Springer Netherlands. doi:10.1007/978-94-007-1272-0.

Laerd Statistics. 2013. "Principal components analysis in SPSS." https://statistics.laerd.com (last accessed 11 August 2013).

Marshall, Serena, and Jon Garcia. 2014. "Obama Announces Unilateral Action on Immigration." *abc News*, June 30. Accessed 11 August 2014. http://abcnews.go.com/Politics/president-obama-announces-unilateral-action-immigration/story?id=24368748

Minnesota Population Center. *National Historical Geographic Information System: Version 2.0*. Minneapolis, MN: University of Minnesota 2011. Retrieved https://www.nhgis.org (last accessed 10 August 2014).

Myers, Laura. 2014. "Reid: No chance of comprehensive immigration reform this summer." *Las Vegas Review-Journal,* July 2. Retrieved http://www.reviewjournal.com/politics/reid-no-chance-comprehensive-immigration-reform-summer (last accessed 10 August 2014).

Nakamura, David. 2014. "Obama readies executive action on immigration." The Washington Post, August 1. Accessed 11 August 2014. http://www.washingtonpost.com/politics/obama-readies-executive-action-to-legalize-millions-of-undocumented-immigrants/2014/08/01/222ae2e8-18f8-11e4-85b6-c1451e622637_story.html

O'Rourke, Norm, and Larry Hatcher. 2013. A step-by-step approach to using SAS for factor analysis and structural equation modeling, second edition. Cary, N.C: SAS Institute Inc.

Passel, Jeffrey S. 2013. "Unauthorized Immigrants: How Pew Research Counts Them and What We Know About Them." Washington, DC: Pew Hispanic Center. Retrieved http://www.pewresearch.org/2013/04/17/unauthorized-immigrants-how-pew-research-counts-them-and-what-we-know-about-them (last accessed 10 August 2014).

Passel, Jeffrey S., and D'Vera Cohn. 2009. "A Portrait of Unauthorized Immigrants in the United States." Washington, DC: Pew Hispanic Center. Retrieved http://www.pewhispanic.org/2009/04/14/iii-demographic-and-family-characteristics/ (last accessed 10 August 2014).

———. 2012. "Unauthorized Immigrants: 11.1 Million in 2011." Washington, DC: Pew Hispanic Center. Retrieved http://www.pewhispanic.org/2012/12/06/unauthorized-immigrants-11-1-million-in-2011 (last accessed 10 August 2014).

Passel, J.S., D'Vera Cohn, and Ana Gonzalez-Barrera, A. 2013. "Population Decline of Unauthorized Immigrants Stalls, May Have Reversed." Washington, DC: Pew Hispanic Center. Retrieved http://www.pewhispanic.org/2013/09/23/population-decline-of-unauthorized-immigrants-stalls-may-have-reversed (last accessed 10 August 2014).

Pastor, Manuel, and Enrico A. Marcelli. 2013. "What's at Stake for the State: Unauthorized Californians, Immigration Reform, and Our Future Together." Los Angeles, CA: Center for the Study of Immigrant Integration. Retrieved http://csii.usc.edu/documents/whats_at_stake_for_the_state.pdf (last accessed 10 August

2014).

Rob Paral and Associates. 2006. “Unauthorized Immigrants in Congressional Districts.” Retrieved www.robparal.com/MapPage.html?map=14&type=G. Accessed May 5, 2011 (last accessed 10 August 2014) (Google Earth plug-in required).

Singer, Audrey. 2004. “The Rise of New Immigrant Gateways.” New Yoek, NY: The Brookings Institution. Retrieved http://www.brookings.edu/~/media/research/files/reports/2004/2/demographics%20singer/20040301_gateways.pdf (last accessed 11 August 2014).

Terrazas, Aaron. 2010. “Mexican Immigrants in the United States,” *http://www.migrationpolicy.org,* last modified 22 February 2010, http://www.migrationpolicy.org/article/mexican-immigrants-united-states-0 (last accessed 11 August 2014).

U.S. Congress, Senate. 2013. *Border Security, Economic Opportunity, and Immigration Modernization Act. S.744.* 113th Cong., 1st sess., June 27.

U.S. Congressional Budget Office. 2013a. *Cost Estimate: S.744 Border Security, Economic Opportunity, and Immigration Modernization Act.* Washington, DC: Congressional Budget Office, 2013. Retrieved http://www.cbo.gov/sites/default/files/cbofiles/attachments/s744.pdf (last accessed 28 August 2014)

———. 2013b. *The Economic Impact of S. 744, the Border Security, Economic Opportunity, and Immigration Modernization Act.* Washington, DC: Congressional Budget Office, 2013. Retrieved http://www.cbo.gov/sites/default/files/cbofiles/attachments/44346-Immigration.pdf

(last accessed 28 August 2014)

U.S. Congressional Research Service. 2013. *Immigration Legislation and Issues in the 113th Congress.* By Andorra Bruno, Michael John Garcia, William A. Kandel, Margaret Mikyung Lee, Marc R. Rosenblum, Alison Siskin, and Ruth Ellen Wasem. CR R43320, Washington DC: Library of Congress, Congressional Research Service, 2013. Retrieved http://fas.org/sgp/crs/homesec/R43320.pdf (last accessed 10 August 2014)

U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau. 2008. *A Compass for Understanding and Using American Community Survey Data What General Data Users Need to Know.* Washington, D.C.: United States Government Printing Office, 2008. Retrieved http://www.census.gov/acs/www/Downloads/handbooks/ACSGeneralHandbook.pdf (last accessed 10 August 2014)

Warren, Robert, and John Robert Warren. 2013. "Unauthorized Immigration to the United States: Annual Estimates and Components of Change, by State, 1990 to 2010." International Migration Review 47 (2): 296-329. doi:10.1111/imre.12022.

"What's on the Menu? Immigration Bills Pending in the House of Representatives in 2014," *immigrationpolicy.org*, last modified 26 March 2014, http://www.immigrationpolicy.org/sites/default/files/docs/summary_of_house_bills_final_4-15-14.pdf (last accessed 10 August 2014)

APPENDICES

Table 19 **Correlation Matrix: First PCA**

		Entered 2000 or later	Entered 1980 or before	Speak English 'less than well'	Age 65+	Not a U.S. Citizen	Less than high school	Bachelor's degree+	Median income	Born in Mexico	Born in Central America	Median Age	Income in the past 12 months below poverty level	Hispanic or Latino
Correlation	Entered 2000 or later	1.000	-.666	.094	-.650	.631	.072	.063	-.265	.149	.223	-.824	.368	.120
	Entered 1980 or before	-.666	1.000	-.594	.948	-.772	-.473	.239	.106	-.456	-.339	.889	-.340	-.524
	Speak English 'less than well'	.094	-.594	1.000	-.646	.701	.889	-.790	-.419	.794	.138	-.469	.539	.888
	Age 65+	-.650	.948	-.646	1.000	-.866	-.605	.374	.210	-.611	-.300	.925	-.448	-.622
	Not a U.S. Citizen	.631	-.772	.701	-.866	1.000	.748	-.555	-.477	.792	.209	-.848	.687	.785
	Less than high school	.072	-.473	.889	-.605	.748	1.000	-.861	-.532	.883	.075	-.478	.676	.887
	Bachelor's degree+	.063	.239	-.790	.374	-.555	-.861	1.000	.642	-.774	.113	.267	-.606	-.781
	Median income	-.265	.106	-.419	.210	-.477	-.532	.642	1.000	-.568	.360	.323	-.886	-.403
	Born in Mexico	.149	-.456	.794	-.611	.792	.883	-.774	-.568	1.000	-.139	-.523	.707	.873
	Born in Central America	.223	-.339	.138	-.300	.209	.075	.113	.360	-.139	1.000	-.165	-.241	.229
	Median Age	-.824	.889	-.469	.925	-.848	-.478	.267	.323	-.523	-.165	1.000	-.514	-.452

Income in the past 12 months below poverty level	.368	-.340	.539	-.448	.687	.676	-.606	-.886	.707	-.241	-.514	1.000	.549
Hispanic or Latino	.120	-.524	.888	-.622	.785	.887	-.781	-.403	.873	.229	-.452	.549	1.000

Table 20 **Anti-image Correlation, Final PCA**

	Entered 2000 or later	Entered 1980 or before	Speak English 'less than well'	Age 65+	Not a U.S. Citizen	Less than high school	Bachelor's degree+	Median income	Born in Mexico	Median Age	Income in the past 12 months below poverty level	Hispanic or Latino
Entered 2000 or later	.682[a]											
Entered 1980 or before		.775[a]										
Speak English 'less than well'			.870[a]									
Age 65+				.902[a]								
Not a U.S. Citizen					.887[a]							
Less than high school						.831[a]						
Bachelor's degree+							.819[a]					
Median income								.645[a]				
Born in Mexico									.851[a]			
Median Age										.777[a]		
Income in the past 12 months below poverty level											.745[a]	
Hispanic or Latino												.845[a]

Measures of Sampling Adequacy (MSA)

Table 21 **Correlation Matrix: Final PCA**

		Born in Mexico	Entered 2000 or later	Entered 1980 or before	Age 65+	Less than high school	Bachelor’s degree +	Median income	Speak English 'less than well'	Not a U.S. citizen	Median age	Income in the past 12 months below poverty level	Hispanic or Latino
Correlation[a]	Born in Mexico	1.000	.149	-.456	-.611	.883	-.774	-.568	.794	.792	-.523	.707	.873
	Entered 2000 or later	.149	1.000	-.666	-.650	.072	.063	-.265	.094	.631	-.824	.368	.120
	Entered 1980 or before	-.456	-.666	1.000	.948	-.473	.239	.106	-.594	-.772	.889	-.340	-.524
	Age 65+	-.611	-.650	.948	1.000	-.605	.374	.210	-.646	-.866	.925	-.448	-.622
	Less than high school	.883	.072	-.473	-.605	1.000	-.861	-.532	.889	.748	-.478	.676	.887
	Bachelor’s degree +	-.774	.063	.239	.374	-.861	1.000	.642	-.790	-.555	.267	-.606	-.781
	Median income	-.568	-.265	.106	.210	-.532	.642	1.000	-.419	-.477	.323	-.886	-.403
	Speak English 'less than well'	.794	.094	-.594	-.646	.889	-.790	-.419	1.000	.701	-.469	.539	.888
	Not a U.S. citizen	.792	.631	-.772	-.866	.748	-.555	-.477	.701	1.000	-.848	.687	.785
	Median age	-.523	-.824	.889	.925	-.478	.267	.323	-.469	-.848	1.000	-.514	-.452
	Income in the past 12 months below poverty level	.707	.368	-.340	-.448	.676	-.606	-.886	.539	.687	-.514	1.000	.549
	Hispanic or Latino	.873	.120	-.524	-.622	.887	-.781	-.403	.888	.785	-.452	.549	1.000

a. Determinant = 3.797E-010

Table 22 **Reproduced Correlations and Residuals: Final PCA**

		Born in Mexico	Entered 2000 or later	Entered 1980 or before	Age 65+	Less than high school	Bachelor's degree +	Median income	Speak English 'less than well'	Not a U.S. citizen	Median age	Income in the past 12 months below poverty level	Hispanic or Latino
Reproduced Correlation	Born in Mexico	.780[a]	.414	-.653	-.741	.781	-.657	-.532	.754	.831	-.690	.677	.762
	Entered 2000 or later	.414	.220[a]	-.347	-.393	.414	-.348	-.282	.400	.441	-.366	.359	.404
	Entered 1980 or before	-.653	-.347	.547[a]	.620	-.654	.550	.445	-.632	-.696	.578	-.567	-.638
	Age 65+	-.741	-.393	.620	.703[a]	-.741	.623	.505	-.716	-.789	.655	-.643	-.724
	Less than high school	.781	.414	-.654	-.741	.782[a]	-.657	-.532	.755	.832	-.691	.678	.763
	Bachelor's degree +	-.657	-.348	.550	.623	-.657	.553[a]	.448	-.635	-.699	.581	-.570	-.642
	Median income	-.532	-.282	.445	.505	-.532	.448	.362[a]	-.514	-.566	.470	-.462	-.520
	Speak English 'less than well'	.754	.400	-.632	-.716	.755	-.635	-.514	.730[a]	.804	-.667	.655	.737
	Not a U.S. citizen	.831	.441	-.696	-.789	.832	-.699	-.566	.804	.885[a]	-.735	.721	.812
	Median age	-.690	-.366	.578	.655	-.691	.581	.470	-.667	-.735	.610[a]	-.599	-.674
	Income in the past 12 months below poverty level	.677	.359	-.567	-.643	.678	-.570	-.462	.655	.721	-.599	.588[a]	.662
	Hispanic or Latino	.762	.404	-.638	-.724	.763	-.642	-.520	.737	.812	-.674	.662	.745[a]
Residual[b]	Born in Mexico		-.265	.198	.129	.102	-.117	-.036	.039	-.038	.167	.030	.110
	Entered 2000 or later	-.265		-.319	-.257	-.342	.412	.017	-.307	.190	-.458	.009	-.284
	Entered 1980 or before	.198	-.319		.328	.181	-.311	-.339	.037	-.077	.312	.227	.114
	Age 65+	.129	-.257	.328		.136	-.249	-.295	.070	-.077	.270	.195	.102

	Less than high school	.102	-.342	.181	.136		-.204	-6.273E-005	.133	-.084	.213	-.002	.124
	Bachelor’s degree +	-.117	.412	-.311	-.249	-.204		.195	-.155	.145	-.313	-.035	-.139
	Median income	-.036	.017	-.339	-.295	-6.273E-005	.195		.095	.089	-.147	-.425	.117
	Speak English 'less than well'	.039	-.307	.037	.070	.133	-.155	.095		-.103	.198	-.116	.150
	Not a U.S. citizen	-.038	.190	-.077	-.077	-.084	.145	.089	-.103		-.114	-.034	-.027
	Median age	.167	-.458	.312	.270	.213	-.313	-.147	.198	-.114		.085	.222
	Income in the past 12 months below poverty level	.030	.009	.227	.195	-.002	-.035	-.425	-.116	-.034	.085		-.113
	Hispanic or Latino	.110	-.284	.114	.102	.124	-.139	.117	.150	-.027	.222	-.113	

Extraction Method: Principal Component Analysis.
a. Reproduced communalities
b. Residuals are computed between observed and reproduced correlations. There are 54 (81.0%) nonredundant residuals with absolute values greater than 0.05.

Table 23 **Percent of Census Tracts with Missing Variables by State**

State	% missing variables
California	0
Oregon	1
Nevada	1
Washington	1
Florida	1
Connecticut	1
New Jersey	1
Rhode Island	1
Massachusetts	1
New Hampshire*	1
Arizona	2
Texas	2
Utah	2
New York	2
Vermont*	2
Idaho	3
New Mexico	3
Maryland	3
Colorado	3
Delaware	3
District of Columbia	3
Minnesota	3
Wisconsin	4
Maine*	4
Virginia	7
North Carolina	7
Illinois	7
Georgia	8
Kansas	8
Wyoming*	8
Michigan	8
Nebraska	9
Pennsylvania	9
Oklahoma	11
Iowa	12
Missouri	12
South Carolina	12
Indiana	13
Montana*	13
Tennessee	14
Arkansas	15
Ohio	15
North Dakota*	16
Louisiana	18
Alabama	18
Kentucky	18
Mississippi	20
South Dakota*	20
West Virginia*	29

* unauthorized population <3,000 (Warren and Warren 2013)