Spatiotemporal Spillover in Lawn-to-Garden Program Participation in Long Beach, California

by

Holly Torpey

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)

August 2017

This thesis is dedicated to my husband, John, and my children, Maria, Emmett, and Olivia, without whose patience and support I could not have finished this work.

Table of Contents

List of Figures

List of Tables

List of Equations

Acknowledgements

It is with great appreciation that I acknowledge the contributions of my thesis committee members, Dr. Robert Vos, Dr. Darren Ruddell, and especially my perceptive and supportive advisor, Dr. Elisabeth Sedano. I would also like to thank Dean Wang and Krista Reger of the Long Beach Water Department, who supplied the Lawn-to-Garden participation data and responded to my requests and questions with both efficiency and insight. Finally, I am grateful to the City of Long Beach, Los Angeles County, Zillow, and the United States Census Bureau, for maintaining publicly available, easily accessible, and well-documented geospatial data.

List of Abbreviations

AAR	Adjusted Application Rate
ABCCR	Adjacent-Block Cumulative Completion Rate
BIMBY	Biodiversity in My Back Yard
CCR	Cumulative Completion Rate
4QMAAR	Four-Quarter Mean Adjusted Application Rate
GIS	Geographic information system
IBM	International Business Machines
LBWD	Long Beach Water Department
L2G	Lawn-to-Garden
MATLAB	Matrix Laboratory
MWD	Metropolitan Water District of Southern California
NOAA	National Oceanic and Atmospheric Administration
SFR	Single-family residential
SPSS	Statistical Package for the Social Sciences
ZHVI	Zillow Home Value Index

List of Statistical Abbreviations

ANOVA	Analysis of variance
F	*F* statistic
ln	Natural log
H	Kruskal-Wallis test statistic
Moran's *I*	Moran's global index
N	Sample size
OLS	Ordinary least squares
p	Probability
R^2	R-squared
t	*t* statistic
$\bar{x}$	Sample mean

Abstract

Turf replacement rebate programs are a water-conservation measure promoted by many local and regional government agencies in California. In an effort to reduce outdoor water use, these programs offer financial incentives to homeowners who replace water-intensive lawns with drought-tolerant landscaping and more efficient irrigation systems. Previous studies, however, have found that landscaping choices are based on more than just economic factors; social pressure, neighborhood norms, and property value are also important considerations, and homeowners tend to opt for landscaping similar to that of their neighbors. This study uses GIS, linear regression, binary logistic regression, and a comparison of means to characterize a spatiotemporal spillover effect in turfgrass replacement rebate program participation data for Long Beach, California. The study determines that residents are more likely to participate in a Lawn-to-Garden program when one or more neighbors on the same block have already completed turfgrass replacement projects. In fact, a block with a single project completion is 5.8 times more likely to see a future application submission than blocks where no projects have been completed, and the highest future application rates occur on blocks where more than 8% of households have already completed a Lawn-to-Garden project. Project completions on adjacent blocks were found to be far less influential. These findings indicate that residents are more willing to replace their conventional landscaping with drought-tolerant gardens after an alternative norm has been established on visually adjacent properties, suggesting that local governments should consider focusing their turf replacement program marketing and support efforts on blocks with no prior participation.

Chapter 1 Introduction

As California residents and governments grapple with the severe statewide drought that began in the fall of 2011, turf buyback programs have become an increasingly popular means of curtailing urban water use. Turf buyback programs, often referred to as lawn replacement rebate programs, cash-for-grass programs, or lawn-to-garden programs, are incentive programs wherein local governments offer to pay residents—usually by the square foot of lawn removed—to replace irrigated turfgrass with drought-tolerant landscaping and to replace spray irrigation with more efficient drip systems. The Metropolitan Water District of Southern California (MWD), which sells imported water from the Colorado, Sacramento, and San Joaquin rivers to cities in Southern California, paid out $401.3 million in turf replacement rebates between July 1, 2014 and July 10, 2015, supplementing rebates paid by local water suppliers (Atwater, Schmitt, and Atwater 2015). Turf buyback programs are not without challenges, though. Rates of participation in turf replacement rebate programs can vary dramatically from month to month and from neighborhood to neighborhood, and program attrition is high (Seapy 2015). In fact, Atwater, Schmitt, and Atwater (2015) estimate that only 1–2% of lawns have been replaced in Southern California, including independent (non-incentivized) conversions. On the other hand, results of an ongoing program participation survey by the Irvine Ranch Water District in Southern California indicate that participation is on the rise and that for every three households that replace their lawns in exchange for a rebate, four more households remove their lawns without applying for a rebate, suggesting that turf replacement rebate programs could have a multiplier effect (Johnson 2017; Knickmeyer 2016).

This study examines spatiotemporal patterns of project participation in Long Beach, California, where the Long Beach Water Department (LBWD) has offered its Lawn-to-Garden (L2G) program continuously since 2010. By combining Geographic Information System (GIS) methods with statistical analysis, this study contributes a spatial dimension to the emerging understanding of turf replacement rebate program effectiveness, quantifying the previously unexamined roles of visual adjacency, social contagion, and neighborhood norms in program participation. Results of the study could inform future decisions by local and regional government bodies as they focus marketing and outreach efforts for turf replacement programs.

Specifically, this study aims to determine whether a spatiotemporal spillover effect is at play wherein each L2G project completion increases the likelihood of nearby neighbors participating in the program in the future. Spatial and temporal spillover effects exist when the value of a variable in a given analysis unit and time period affects the values of other variables in neighboring analysis units and future time periods. The presence of a spatial spillover effect would indicate that as the garden-to-turf ratio increases on a residential block, the likelihood of other neighbors replacing their turf with gardens also increases. To that end, more than six years of L2G program participation data provided by the LBWD were geocoded and analyzed for spatial autocorrelation between L2G project completion points in order to determine the extent of that potential spillover effect. Next, city-block-level spatial aggregation units were developed based on that peak autocorrelation distance. Quarterly application and completion rates were calculated and normalized for each city block, and completion rate variables were lagged both temporally and spatially to identify spatial and temporal spillover effects. In addition, several other explanatory variables were derived from Los Angeles County tax assessor parcel data, U.S. Census Bureau household data, and Zillow estimated property values, resulting in a panel dataset of quarterly values for each residential city block in Long Beach. Finally, a linear regression model was fitted to characterize interactions between non-zero city-block-level application and completion rates, comparing annualized future project application rates to rates of previously completed projects on the same and adjacent blocks while controlling for owner occupancy, home value, and rebate amount. In addition, a binary logistic regression model was developed to estimate the likelihood of future application submissions on blocks at a series of project completion rate ranges with the goal of identifying critical completion rate thresholds—both on the same block and on adjacent blocks—after which residents are more likely to participate. Finally, a comparison of means test was used to examine application rates on blocks above and below those same-block and adjacent-block thresholds.

These analyses tested three hypotheses. The first was that future L2G application rates have a significant linear relationship with cumulative completion rates; the second was that the likelihood of application presence on a block is significantly higher when completion rates on the same and adjacent blocks are above zero than when completion rates are zero; and the third hypothesis was that application rates are statistically higher on blocks that have achieved a critical project completion threshold on the same and adjacent blocks.

1.1. Motivation

Homeowners' choices regarding their residential landscaping and methods of irrigation have implications that extend beyond their own property lines and water bills; collectively, residential yards represent the largest land-use class in the Los Angeles Basin (Hevesi and Johnson 2016). Minor et al. (2016, 56) go so far as to claim that "residential yards are perhaps the most underappreciated and understudied ecosystem in the world," citing their collective influence on urban temperatures, water quality and conservation, animal and plant habitat and diversity, and human wellbeing.

Residential irrigation is especially relevant in California, where drought conditions have brought water conservation to the fore. The four years between 2012 and 2015 were the driest on record in California (Hanak et al. 2015), and ten of the last sixteen years—from 2000 to 2016—were drier than average (NOAA 2016). In Southern California, where outdoor irrigation represents about half of urban water use (Hanak et al. 2015), reducing residential outdoor water consumption is a priority for local governments. The City of Long Beach is one of dozens of local governments in California to offer homeowners a financial incentive to replace lawns with drought-tolerant gardens in an effort to achieve that goal (Knickmeyer 2016). The LBWD claims that L2G program participants can reduce their landscape irrigation by 70% (LBWD 2016).

Turf-replacement rebate requirements—and the time and money that must be spent meeting them—can be a barrier to program participation and completion (Hurd, St. Hilaire, and White 2006). In a report to the California Urban Water Council, Seapy (2015) cites several challenges to turf replacement project success. First, many homeowners lack both the skill to produce a successful result and the funds to hire a professional landscape designer. Second, program requirements, project planning, and the length of the conversion process can be overwhelming; Seapy notes that some agencies report a 50% attrition rate between rebate application submission and landscape design approval. Finally, homeowners may hesitate to convert lawns to drought-tolerant landscapes if they feel social pressure to conform to a more conventional landscaping norm. In the same report, Seapy (2015, 12) emphasizes the influence of neighborhood norms: "Powerful in their ability to attract or dissuade customers to a rebate program, social norms can make or break a program's success. For example, agencies have seen that one to two stunning conversions in a neighborhood can catalyze an entire neighborhood's transformation."

The idea that landscaping choices are “contagious” is not new. In a study examining the spatial clustering of easement gardens (gardens planted on the strip of land between street and sidewalk) in Ann Arbor, Michigan, Hunter and Brown (2012) found that properties were more likely to have easement gardens when a visually adjacent property also has an easement garden. Their findings can be explained by an earlier study exploring the role that cultural norms play in homeowner adoption of landscaping innovations. In that study, Nassauer, Brown, and Dayrell (2009) offer two reasons that homeowners may be reluctant to install an innovative landscape even when they appreciate its aesthetic or ecological value. First, homeowners feel social pressure to conform to perceived neighborhood norms dictating more conventional landscaping choices, and second, they fear that unconventional landscaping could reduce their property values. Figure 1 shows a Long Beach neighborhood in which five turf replacements have been completed on a single block, suggesting the establishment of an alternative landscaping norm for that block.

Figure 1. Cluster of L2G projects. In this Long Beach neighborhood, no block has more than one completed project with the exception of a single block with five project completions.

While anecdotal evidence and agency reports support the idea that homeowners are more likely to participate in turf replacement rebate programs after a some of their neighbors have already done so with success, as of 2016, the literature contains scant reference to measurement of a spillover effect in turfgrass replacement participation. By using GIS to conduct a spatial analysis of patterns of turf replacement program participation in both space and time, one can attempt to quantify the social multiplier effect each L2G conversion can have on participation rates among neighboring residents.

1.2. Study Area

Long Beach is a Southern California city situated on just over 50 square miles (130 square kilometers) of urbanized coastal plain between the Los Angeles and San Gabriel rivers.

Located in southern Los Angeles County, immediately adjacent to Orange County, Long Beach is part of the Los Angeles-Long Beach-Santa Ana Metropolitan Statistical Area (Figure 2). The city is bordered on the south by the San Pedro Bay and is surrounded by other urbanized communities to the west, north, and east (City of Long Beach 2009); the city of Signal Hill sits near the center of the city and is bordered on all sides by Long Beach. Signal Hill is not included in this study because its residents are not eligible to participate in the LBWD turf replacement rebate program.

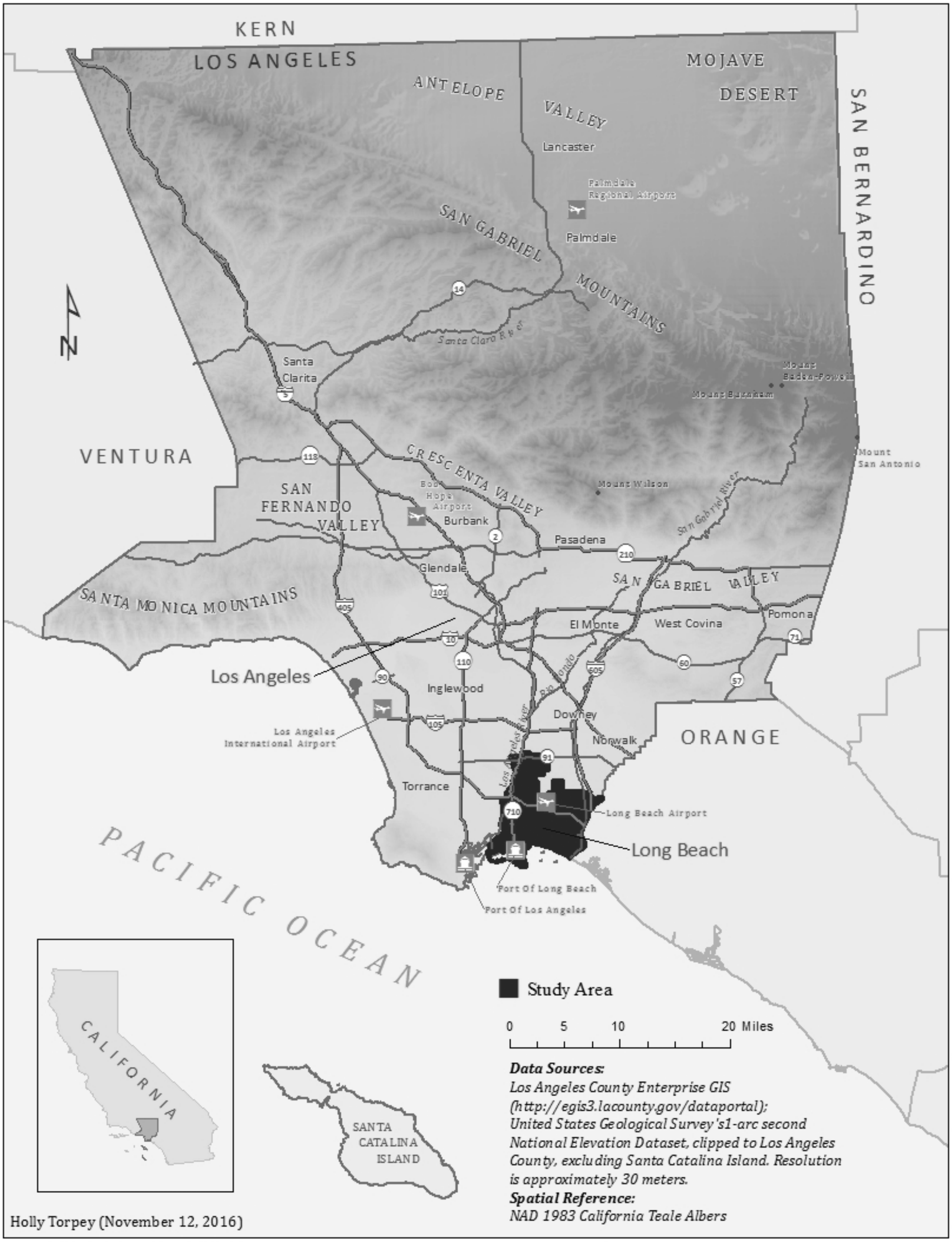


Figure 2. Study area. Long Beach is a coastal city in southern Los Angeles County.

First incorporated in 1888 following the extension of transcontinental railroad lines to Southern California, Long Beach was the fastest growing city in the country in the early 1900s (City of Long Beach 2009). Thousands of eastern and midwestern families moved west, bringing with them their landscaping preference for lush lawns bordered by shrubs and flowers (Mickey 2013); however, in Southern California's Mediterranean climate—characterized by warm, dry summers with little or no rainfall—a green lawn requires regular irrigation. After World War I, oil was discovered in Long Beach and the population jumped from 55,000 to 135,000 in five years. The oil boom spurred development of neighborhoods featuring bungalows and multi-unit residential buildings with small yards and courtyards. The second major wave of residential construction occurred after World War II, as large tracts of post-war bungalows, each with a thirsty front lawn, were built further inland. The average year of construction of single-family homes in Long Beach is 1957 (Los Angeles County Office of the Assessor 2014).

Today, Long Beach is home to about 474,000 people. With over 9,400 residents per square mile (over 3,600 residents per square kilometer), the population density of Long Beach is moderate compared to other cities in the Los Angeles metropolitan area (United States Census Bureau 2015). High- and medium-rise apartment and condominium buildings pepper the downtown area and the coastline; throughout the city, two-, three-, and four-story apartment complexes are common. Single-family residences occupy roughly 63% of all residential parcels. The average single-family house in Long Beach is approximately 1,500 square feet and sits on a 5,830 square-foot parcel (Los Angeles County Office of the Assessor 2014). The lawns and gardens planted on these single-family residential (SFR) parcels are the focus of this study.

1.3. LBWD Lawn-to-Garden Program

The Long Beach L2G program was introduced in 2010. The LBWD initially offered residents $2.50 for every square foot of turf grass they replaced with drought-tolerant landscaping, with a cap of 1,500 square feet and $3,750.00 (Green 2010). In 2013, funding from MWD allowed LBWD to increase the incentive to $3.00 per square foot (Noell 2013) and then to $3.50 per square foot in 2014 (Williams 2014; Machles 2014). By 2015, the MWD subsidy budget had been exhausted, and the incentive was reduced to $2.50 per square foot once again (Smith 2015).

To qualify for the L2G rebate, Long Beach residents must complete a multi-step approval process. They begin by submitting an online application in which they provide the dimensions of the turf they intend to replace; areas of living turf grass up to 1,500 square feet in front yards or parkways (the city-owned strip of land between sidewalk and street) are eligible. Next, property owners complete a short landscape design course online or in person prior to developing and submitting a proposal. The proposal must include a plan to remove and replace turf grass with a drought-tolerant garden in which at least 65% of the former turf area will be covered with vegetation within two years; the remaining area can be covered with mulch or a permeable hardscape. They must also replace conventional sprinkler systems with water-saving irrigation such as drip systems, bubblers, or low-emission rotator nozzles, or no irrigation at all. Finally, LBWD inspects the site upon project completion to verify compliance with program requirements (LBWD 2016). Figure 3 illustrates a successfully completed Long Beach turf replacement project.

Figure 3. Before and after photos of a Lawn-to-Garden project (LBWD 2013).

1.4. Study Design

This study was undertaken with the objective of answering three questions. First, is future application rate dependent on the rate of previously completed projects? Second, does the presence of previously completed projects on a city block increase the likelihood of other residents of that block and neighboring blocks participating in the program as well? And finally, is there a critical completion rate threshold at which the mean block-level L2G application rate increases significantly? This research aims to answer those questions by calculating quarterly block-level L2G project application rates and comparing those to cumulative project completion rates on the same and adjacent blocks using linear and binary logistic regression analyses and a difference of means test, controlling for independent variables such as owner occupancy, housing prices, and rebate amounts.

This project methodology was undertaken in two stages: data acquisition and integration and statistical analysis. In the first stage, data were acquired from the LBWD, the City of Long Beach, Los Angeles County, the U.S. Census Bureau, and Zillow, and processed to achieve uniformity of spatial reference systems and geographical extents. GIS techniques were used to aggregate the data into city-block polygons with attributes describing present and future L2G application rates and same-block and adjacent-block completion rates for 20 quarters from 2011 through 2015 along with explanatory variables like rebate rates, mean real estate value estimates, parcel sizes, and household composition information. Next, block-quarter analysis units were generated by restructuring the tabular polygon attribute data to produce a panel dataset with 20 quarterly records for each residential city block, each containing values for the dependent variables (application rate and four-quarter future application rate) and independent variables. The same-block and adjacent-block cumulative completion rate variables were lagged by one quarter to enable comparison of each quarter's present and future application rate with the previous quarter's completion rates.

In the second stage, the panel dataset was then subjected to three different statistical analyses. First, a linear regression analysis modeled the relationship between non-zero future mean application rates and non-zero same-block and adjacent-block cumulative completion rates, controlling for block-level mean housing value and owner occupancy rate; second, a binary logistic regression model was fitted to quantify the likelihood of a future application being submitted within four quarters for a block-quarter at different categories of same-block and adjacent-block previous-quarter cumulative completion rates (including zero values); and third, a independent samples *t* test was used to compare future application rate means above and below the same-block and adjacent block completion rate thresholds identified by the regression models. All spatial data processing and analysis tools used in this study can be found in the ArcGIS Spatial Statistics Toolbox, while tools for extracting and analyzing descriptive statistics and developing linear and binary logistic regression models are available in IBM's Statistical Package for the Social Sciences (SPSS).

1.5. Organization

The remainder of this thesis will document the results of a literature review of related work followed by a detailed explanation of the research methodology, including data acquisition, integration, and aggregation and statistical analysis. Finally, results will be presented and discussed. The Related Work chapter includes an overview of previous research related to residential outdoor water conservation, neighborhood norms and social contagion, spatiotemporal spillover effects, threshold identification, and similarly structured data. The Data Integration chapter discusses the methods used to acquire and process spatial and tabular data from the City of Long Beach, Los Angeles County, the United States Census Bureau, and Zillow (an online real estate database company) and to spatially and temporally aggregate those data in preparation for statistical analysis. The Statistical Analysis chapter details the statistical methodology employed. The Results chapter will detail the outcomes of the linear regression, binary logistic regression, and comparison of means tests, and the Discussion chapter will delve further into the implications and limitations of those results.

Chapter 2 Related Work

Turf replacement rebate programs are a relatively recent phenomenon; while newspaper articles and white papers assessing program effectiveness and participation abound (Tull, Schmitt, and Atwater 2016; Atwater, Schmitt, and Atwater 2015; Seapy 2015; Green 2010; Addink 2005), few peer-reviewed academic treatments of the topic exist. To date, no research has been published on the topic of spatiotemporal spillover in turf rebate program participation. There is, however, much discussion in the literature of the role of landscaping and irrigation methods in residential water conservation and urban sustainability, as well as of the factors that influence residential landscaping and irrigation choices and behaviors. With regard to analytical methodology, a review of other fields such as econometrics and ecology yields numerous examples of spatiotemporal spillover and threshold identification and provides insight into approaches for statistical analysis of similarly structured data.

2.1. Residential Outdoor Water Conservation

A number of studies address the role of residential landscaping in urban water use and conservation. While there is consensus regarding the importance of landscaping choices and irrigation methods to water conservation efforts, findings are mixed with regard to the effectiveness and appropriateness of turf replacement rebate programs.

Most recently, Tull, Schmitt, and Atwater (2016) investigate monthly water savings for a set of 545 single-family households that participated in residential turf removal programs in three different California water districts. By modeling monthly water usage data at the household level both before and after turf replacement, controlling for household size, the percentage of irrigable area converted, and external influences, their research determines that the households in their study save an average of 24.6 gallons per square foot per year. In a separate study, Atwater, Schmitt, and Atwater (2015) describe their evaluation of participation patterns and outcomes of a turf removal rebate program that began in 2011 in the Moulton Niguel Water District in Orange County, California. The researchers were unable to correlate program participation with other variables including irrigable area, property value, educational attainment, and household size, noting that "the lack of a strong link between participation in rebate programs and the variables measured in this study points to the challenge of developing an effective conservation program" (p. 6). Atwater, Schmitt, and Atwater do not investigate a link between program participation and the presence of nearby completed projects in their study.

Helfand et al. (2006) attempt to quantify homeowners' willingness to pay to replace turfgrass with a more ecologically sound garden, finding that people are willing to pay more for alternative garden landscaping, although their willingness to pay decreases as the monthly costs increase relative to income. This finding is supported by Hurd, St. Hilaire, and White (2006), who evaluate New Mexico homeowners' attitudes and preferences with regard to residential landscaping and water conservation. Their work indicates that homeowners are strongly in support of transitioning from conventional lawns to drought-tolerant native and natural landscaping in that region; however, 25% of study participants cited money as a barrier to completing that transition, and only one of the three New Mexico cities studied offered a rebate for turf removal. The researchers also highlight the importance of evaluating the effectiveness of water conservation programs—including turf rebate programs—since those programs are funded by taxpayer and ratepayer dollars.

Findings differ with regard to the value of replacing turfgrass lawns with drought-tolerant landscaping as a water conservation measure. Sovocool, Morgan, and Bennett (2006) consider the value of low-water use landscaping in Las Vegas as a means of conserving urban water resources in a six-year study that tracks a xeric (low-water use) landscape group, a turf group, and a control group, monitoring their water consumption and landscape-related costs, including conversion and maintenance. Their results show an average decrease of 76% in outdoor irrigation and 30% in total water consumption among the households with a xeric landscape, with a marked reduction in peak summer water use. Participants with xeric gardens also saved $206 each year on landscape expenses and spent over 26 fewer hours maintaining their landscapes. On the other hand, Addink (2005) questions the cost-effectiveness of turf replacement rebate programs in a non-peer-reviewed report for the University of California at Riverside's Turfgrass Research Facility. Having reviewed the outcomes of four turf replacement rebate programs, Addink concludes that irrigation technology and management are more important than vegetation type with regard to water conservation.

Other studies have highlighted the idea that landscape policy makers should consider more than just water conservation. Halper, Scott, and Yool (2012) analyze residential irrigation and patterns of vegetation and thermal comfort in the desert city of Tucson, Arizona, noting the importance of urban vegetation in mitigating heat in cities and suggesting that city governments remain mindful of the risk of creating urban heat islands as they work toward reducing outdoor irrigation. The importance of urban vegetation in mitigating heat is underscored by a Beumer and Martens (2016), who use the BIMBY (Biodiversity in My Back Yard) framework to assess the ecology of various front-yard landscaping in Phoenix, Arizona, and Maastricht, Netherlands. They discuss the potential of residential landscapes to contribute to biodiversity in the urban setting and argue that the higher water use associated with green gardens (as opposed to xeric gardens) is offset by their habitat contributions and climate regulating benefits.

2.2. Neighborhood Norms and Social Contagion

If factors like education and property value are not predictive of turf rebate program participation (Atwater, Schmitt, and Atwater 2015), what factors do motivate homeowners to participate? A study of communication strategies for encouraging water conservation by Seyranian, Sinatra, and Polikoff (2015) suggests that neighborhood norms are powerful motivators of water conservation behaviors. Their study compares four approaches to fostering reduced indoor and outdoor water consumption in an affluent neighborhood in Los Angeles County, and their results indicate that water conservation campaigns emphasizing social norms, social identity, or personal identity are all more effective at influencing homeowner behavior than educational materials alone. Their research suggests that when residents perceive that water-conserving behaviors are linked to desirable social factors, they are more likely to overcome barriers to water conservation. The authors also point out that the role of social factors in household water conservation is worthy of further study, noting that “socially oriented interventions appear to be a promising area of future research in promoting water stewardship and reducing water waste” (2015, 89).

Nassauer, Wang, and Dayrell (2009) lend further support to the idea that neighborhood norms are important factors in homeowners’ landscaping choices. They employ an online questionnaire that asks respondents to first rate their preferences for five different landscape designs and then to rate them again after having been shown pictures of three hypothetical neighbors’ front yard landscaping. The authors report that respondents’ preferences were strongly influenced by the neighbors’ landscape designs: when all the neighbors’ yards had conventional lawns, respondents preferred lawns as well, but when most of the neighbors’ yards had native gardens, the respondents strongly preferred that alternative landscape for their own houses as well. When respondents were shown neighboring yards without a consistent norm, their preferences were more variable. Furthermore, those results were not significantly affected by other demographic variables. Nassauer, Wang, and Dayrell concluded that "...individual homeowners deeply value having a front yard that matches a consistent neighborhood appearance, but that neighborhood appearance does not need to conform to broader cultural conventions" (2009, 290).

Hunter and Brown (2012) further examine the influence of nearby neighbors' landscaping on homeowners' own landscaping preferences by analyzing the spatial distribution of easement (street-side) gardens (as opposed to easements planted with turf) in Ann Arbor, Michigan. The authors find that a property is more than twice as likely to have an easement garden if there is another easement garden within 30 meters, and that clustering is most significant in a neighborhood radius of 91 meters. Similarly, McClintock et al. (2015) find that front yard gardens in Portland, Oregon were spatially clustered rather than randomly dispersed. Further, they identify clusters of front yard gardens occurred in affluent neighborhoods with higher percentages of young, college-educated homeowners, but do not identify garden clustering in a lower-income neighborhood where McClintock speculates that replacing a lawn with a garden "might constitute a transgression of dominant cultural norms" (2015, 12). These results confirm earlier findings that homeowners tend to mimic the landscape elements of their close neighbors (Zmyslony and Gagnon 1998), and Hunter and Brown suggest that local governments could leverage this phenomenon by "seeding" neighborhoods with model gardens for neighbors to imitate (2012, 415).

Promoting the establishment of new social norms to effect behavioral change is further supported by Collier et al. (2013), who consider the role of community participation in urban transitions to sustainable design. The authors emphasize the importance of shifting paradigms, concluding that environmentally sustainable practices must be normalized in order to achieve resilience. Hayden et al. (2015) also call for a need to cultivate new aesthetic norms incorporating water conservation practices, citing an unwillingness to sacrifice aesthetics to reduce water consumption, even among people who recognize the need for conservation.

Finally, Fowler and Christakis (2010) show that altruistic behavior spreads in a cascading fashion through social networks. Their work indicates that when individuals witness others contributing to the greater good, they are more likely to contribute as well; this mimicry causes contributions to propagate through a network. While participation in a turf removal rebate program may personally benefit the participant in the form of the rebate received and reduced monthly water bills, it might also be perceived by others as a contribution to the greater good, particularly in times of drought. When residents see others in their neighborhood network contributing to the greater good by replacing a lawn with a drought-tolerant garden, the desire to mimic that altruistic behavior may also be a powerful motivator for participation.

2.3. Spatiotemporal Spillover Effects

A *spatial spillover* effect occurs when a variable describing one analysis unit indirectly affects outcomes in neighboring units in an endogenous (reciprocal) relationship; a *temporal spillover* occurs when a variable describing an analysis unit affects a future outcome in the same unit. *Spatiotemporal spillover* occurs when both spatial and temporal spillovers are present; in other words, the value of a variable in a given analysis unit and time period affects outcomes in neighboring analysis units and future time periods. The current study attempts to identify and quantify a spatiotemporal spillover wherein the L2G application rate on a city block is influenced by the same-block and adjacent-block cumulative completion rates of the previous quarter.

Various types of spillover effects have been identified in the literature, including spillovers of knowledge, industry, and growth (Zubek and Henning 2016; Capello 2009). Many treatments of spatial and temporal spillover effects have arisen from the field of spatial econometrics (Anselin, Le Gallo, and Jayet 2008; Elhorst 2012; Elhorst 2014), which incorporates spatial effects—namely spatial dependence and spatial heterogeneity—into regression analysis in order to model interactions between variables; this effort to incorporate the effects of space into regression analysis has been used with increasing frequency by researchers in a variety of fields including economics, crime analysis, public health, and ecology (Anselin 2010).

A spatially lagged variable is created by aggregating the values of neighboring analysis units in order to model the interactions between variables in neighboring units. The study herein incorporates a spatially lagged independent variable describing the average rate of L2G project completions on adjacent blocks in order to measure their effect on the application rate of a subject block. While spatial regression models more commonly employ a spatially lagged dependent variable (Elhorst 2010; Anselin 2002), Grubesic and Rosso (2014) and Rosso et al. (2013) argue that by using a spatially lagged independent variable in regression analysis, researchers can measure the effect of neighboring individuals or aggregation units on a dependent variable without having to employ the specialized estimation models often required for spatially lagged dependent variables. Grubesic and Rosso (2014) explain that spatially lagged independent variables can be incorporated into linear models, and highlight their utility for analyzing epidemiological and socioeconomic data related to neighborhood. Researchers often calculate spatially lagged variables by calculating the sum of weighted neighborhood values within a given distance (Grubesic and Rosso 2014; Franzese and Hays 2007); other studies employ the average of weighted neighborhood values instead (Zubek and Henning 2016). The weights are usually reflective of each neighbor's relative distance from the subject case so that closer neighbors are weighted more heavily in accordance with Tobler's first law of geography: "Everything is related to everything else, but near things are more related than distant things" (Tobler 1970).

Zubek and Henning (2016) use a spatial econometric approach to study whether spatial spillovers in knowledge and financial means play a role in the allocation of European Union (EU) regional funding in Slovakia. Not unlike the L2G program in Long Beach, the EU regional funding program has a complicated application and approval process; a region's likelihood of applying for and receiving this funding depends on its level of political knowledge and administrative capacity. Zubek and Henning use a spatial Durbin model (a model that uses spatially lagged dependent and independent variables) to test for the existence of a spatial and temporal spillover in a panel dataset. The researchers note that in the case of knowledge spillovers, effects may not be immediately apparent; for that reason, developing a panel dataset—a dataset in which each aggregation unit has multiple records containing data collected at specific time intervals—is essential to capturing the temporal as well as the spatial nature of a knowledge spillover (2016, 372).

An especially pertinent example of spillover analysis is the spatial analysis methodology employed in a study examining water use patterns in Kelowna, British Colombia, which uses a spatial econometrics approach to document a spatial spillover effect wherein water-saving innovations on one property are propagated to neighboring properties (Janmaat 2013). The author Janmaat (2013, 12) explains that this effect is "analogous to the propagation of impacts from a shock." In this study, the author runs a series of spatial structure tests at three distances

(50 m, 100 m, and 200 m) to establish that the water usage data are spatially structured and to create a spatial weights matrix to be used in the estimation of a set of regression models. The regression models provide evidence for the presence of a spatial spillover in the data, suggesting a diffusion process. Finally, the author develops a model that predicts the strength and extent of the spillover effect depending on the locations of and spatial relationships between water-saving innovations. Janmaat (2013, 16) directly relates his work to the topic of this thesis when he notes that his results indicate that requests for turf replacement rebates will tend to be spatially clustered and "...neighborhoods where no one has taken a chance on water saving landscaping are less likely to avail themselves of the subsidy." Janmaat suggests that communities should allocate dollars by spatial divisions to maximize patterns of spillover mimicry.

Similarly, Zhang and Wang (2016) use regression modeling to analyze the effects of low-cost airlines on airports' average ticket prices, as well as the relationship between average fares at an airport and the current and past average fares of neighboring airports. Their results describe a temporal spillover effect wherein an airport's average airfare price drops after a low-cost airline's market share increases, and a spatial spillover effect wherein neighboring airport prices also drop in proportion to their proximity to the airport serving the budget airline. For this study, the researchers used MATLAB, MathWorks' proprietary programming language, to incorporate a spatial weights matrix, a one-year temporal lag, and a one-year spatiotemporal lag into a dynamic spatiotemporal regression model.

Osorio (2015) also investigates a spillover effect by employing spatial econometric methods in a recent study of drug-related violence and police action, finding that escalating drug-related violence in a given location tends to be followed by increased violence in nearby areas. In that study, Osorio develops a first-order autoregressive model with spatial autoregressive disturbances for spatial panel data—data collected for the same geographical units at multiple time intervals—to analyze patterns of spatial diffusion of drug violence in Mexico in response to five categories of violent and non-violent police actions. Caetano and Maheshri (2013) also analyze spatial panel data on crime and crime reduction policies in Dallas, Texas. Their two-stage analysis first estimates equations describing the causal relationships between current and previous levels of each type of crime, and then uses those equations to simulate the effects that reducing levels of a given crime might have on future levels of other crimes.

2.4. Threshold Identification

Xie et al. (2011) find that in various types of social networks, when approximately 10% of a network becomes inflexibly committed to a minority opinion, there is a decrease in the time required for the majority to shift their opinion to that of the initial minority. In other words, adoption of an alternative idea or behavior begins gradually, but then accelerates once a critical percentage of randomly distributed committed proponents is reached. The threshold model (Granovetter and Soong 1983) is one well-established means of modeling social diffusion of innovation; it assumes that the behavior of an individual is dependent on the proportion of other individuals already exhibiting the behavior. The current study attempts to determine whether such a threshold exists in L2G participation data.

Qian and Cuffney (2012) evaluate the appropriateness of multiple threshold models to analyze stream ecological response to urbanization with special focus on means used to identify a threshold response. They conclude that experimenting with multiple models is of paramount importance in identification of thresholds because different models can produce different thresholds, or even identify a threshold where none exists. Andersen et al. (2009) examine several possible methods of detecting thresholds within the ecology domain, specifically with regard to describing ecological regime shifts. Ecological systems often react in a nonlinear manner in response to external pressures, and predicting shifts before they occur is of interest to ecologists. Andersen et al. point out that indicators such as rising variance and growing skewness in data often precede a regime shift. Methods of detecting these change points in a dataset include chronological clustering, cumulative sum control charts, sequential *t* tests, F-tests, and nonlinear diffusion filtering, threshold autoregressive models, and dynamical linear models.

2.5. Zero-Inflated Proportion Data

Many commonly used inferential statistical methods, including *t* tests, analysis of variance (ANOVA) tests, and linear regression, assume normality of data. A normally-distributed dataset is one in which about two-thirds of the data fall within one standard deviation of the mean, and 95% of the values fall within two standard deviations of the mean (Jongman, ter Braak, and van Tongeren 1987). The study herein involves a panel dataset in which the key variables are proportions (expressed as percentages) and zero values predominate; both of these characteristics violate the assumption of normality.

A panel dataset comprises repeated measurements or observations collected at specific time intervals for the same individuals. Here, the “individuals” are residential city blocks, the time intervals are three-month quarters spanning a five-year study period, and the measurements are application and completion rates. Because even a block with a high level of L2G participation might only have eight application submissions and four completed projects by the end of the study period, the application and completion rate variables are positively skewed (values are clustered close to zero), zero inflated (zero values are overrepresented), and do not exhibit a normal distribution of errors (a plot of the differences between predicted and observed values does not resemble a symmetrical bell-shaped curve). These characteristics present unique challenges with regard to statistical analysis.

Studies in the water management and conservation domains have used straightforward methods like linear regression or geographically weighted regression to analyze panel data for water consumption (Chang, Parandvash, and Shandash 2010; Franczyk and Chang 2009), but their data were aggregated at the census block group and county levels where numerous zero values are unlikely. While their choices of regression models cannot be applied to the current dataset in its entirety because of its abundance of zero values, their methods of inferring socioeconomic data from building structure data are relevant to the current study. Chang, Parandvash, and Shandash found that variations in residential water consumption correlate strongly to building structure characteristics, which in turn correlate to socioeconomic variables like income and education. Since income and education data are difficult to obtain at the household or parcel level, the authors suggest that building structure data from publicly available tax assessors' records can be used as proxies for socioeconomic variables and recommend further study to evaluate the effectiveness of that substitution.

For insight into the statistical analysis of zero-inflated panel datasets like the one developed for this study, looking beyond water management and conservation to the field of ecology yields numerous relevant examples. Species presence-absence and abundance data often exhibit positively skewed distributions with large proportions of zero values (Fletcher, MacKenzie, and Villouta 2005; Pearce and Boyce 2005; Dobbie and Welsh 2001). Because this type of data does not meet the assumptions of normality of error distribution built into linear regression models, ecology researchers have developed other methods of analysis. Fletcher, MacKenzie, and Villouta (2005) recommend modeling binary presence-absence data (zeros included) and positive abundance data (zeros excluded) separately. For zero-inflated presence-absence data, a binary logistic regression is suggested, as that nonlinear model does not assume normal distribution of errors; this is not a novel approach, as employing logistic regression to analyze presence-absence data is well documented in the literature (Dobbie and Welsh, 2001; Jongman, ter Braak, and van Tongeren 1987). For the abundance data, Fletcher, MacKenzie, and Villouta compare a continuous, positive dependent variable to one or more explanatory values using a linear model. Because the abundance data—like city-block-level L2G rate data—typically remain positively skewed even after exclusion of zero values, the authors recommend a log transformation of the dependent variable to satisfy the assumption of normal distribution of residual errors. A data transformation is an operation applied to each measured value, replacing it with another value in order to achieve a more normal distribution or facilitate comparison with other variables (Montello and Sutton 2013, 196; Jongman, ter Braak, and van Tongeren 1987, 20). Fletcher, MacKenzie, and Villouta (2005, 46) explain that the advantage of modeling presence-absence and abundance data separately is that researchers can examine the influence of explanatory variables on the two aspects of the data individually without developing a more complicated mixed model approach in which parameters would have to be estimated for both aspects of the data concurrently. This two-stage approach is supported by earlier work; Dobbie and Welsh (2001) adopt a similar methodology in an analysis of repeated counts of species, taking into account the possibility that repeated observations of an individual are often correlated. Their study also employs a logistic regression model to analyze presence-absence data. For their non-zero count data, Dobbie and Welsh used a truncated discrete model, a linear model in which observations are incomplete due to the systematic exclusion of some portion of the data (zero values, in this case). Because the data in the current study are not discrete counts, but rather proportions, a discrete model was ruled out for the current study.

Analysis of proportion data poses additional challenges even when zero values are excluded. Proportion data are unique in that values range between zero and one (or zero and 100 when expressed as a percentage) and can never be less than zero or greater than one (or 100). When relationships to independent variables are plotted, they tend to look more like a curve than a straight line because the difference between observed and predicted outcomes is often greater as values approach those zero and one limits; for this reason, linear regression is only an option if data can be brought closer to normality through a transformation. In addition, linear regression can predict invalid values smaller than zero or greater than one (Crawley 2007, 248-9).

Warton and Hui (2011) discuss the role of data transformation in the evaluation of proportion data, noting that over one-third of papers published by the journal *Ecology* discuss the analysis of that type of data. Because proportion data often exhibit a non-normal distribution, transformations are frequently applied prior to regression analysis. A transformation often mentioned in the papers published in *Ecology* is the arcsine transformation, but Warton and Hui argue that the arcsine transformation is outdated. Instead, they suggest that better options exist, such as using logistic regression for binomial data and applying other transformations—such as the log transformation—for non-binomial data to satisfy linearity assumptions. They suggest a logit transformation, as follows: log(y/[1-y]). A logit transformation is easier to interpret than an arcsine transformation, but cannot be applied to values of 0 or 1, effectively excluding those observations from the dataset. A solution to that problem in situations where researchers must include zero values is to add a small value such as the minimum non-zero proportion to both the numerator and denominator of the logit function. Since this value is added to all data, their relative proportions remain unchanged. In datasets where zero values predominate, however, the data will remain skewed and an approach that excludes zero values, such as those discussed by

Fletcher, MacKenzie and Villouta (2005) or Dobbie and Welsh (2001), might be more appropriate.

Pham et al. (2012) discuss statistical analysis of proportion data in their study of the spatial distribution of vegetation in Montreal in relation to concentrations of minority and low-income populations. Their statistical analysis used descriptive statistics and ordinary least squares (OLS) regression and spatial regression to measure vegetation equity among demographic groups. While that study did not utilize a panel dataset and did not present issues of zero inflation, it did bear similarities to the current study in its aggregation of proportion data at the city-block level. The authors calculated the vegetated proportions of city-block-level aggregation units and then compared them to demographic explanatory variables that were disaggregated from the Canadian equivalent of census block groups to the city block level. The authors chose the city block as their aggregation unit, noting that the city block is relatively homogenous compared to the overall heterogeneity of the urban built environment. Other studies also stress the importance of fine-scale aggregation within the patchy mosaic of the urban ecosystem (Landry and Pu 2010; Grove, Burch, and Pickett 2005; Grimm et al. 2000). The authors of a study of crime hot spots in Seattle, Washington analyze their data at the street segment level since previous studies (Taylor 1997; Wicker 1987) have shown that micro-geographic units such as street segments can function as “social units with specific routines” (Weisburd, Groff, and Yang 2013).

Finally, Lumley et al. (2002) point out that assumptions of normality are not required for comparing means with a two-sample *t* test when samples are sufficiently large. Since the data used in this study represent the entire population of block-quarters during the study period rather than a sample drawn from them, a *t* test can be considered a valid evaluation of the difference of mean values of a variable under different conditions.

Chapter 3 Data Acquisition and Integration

In order to create block-quarter analysis units representing street-segment-based city blocks at a given three-month quarter, each characterized by the same set of dependent and independent variables, a great deal of data integration work was required. This chapter details the steps undertaken to acquire, integrate, aggregate, normalize, restructure, lag, and transform the data using both spatial and non-spatial data processing techniques in preparation for statistical analysis. The statistical analysis methodology is discussed in Chapter 4.

3.1. Data Sources

Data for the study herein were acquired from five sources: the LBWD, the City of Long Beach, Los Angeles County, the United States Census Bureau, and Zillow. All spatial data were imported into an Esri file geodatabase (the "L2G database") and projected to the Lambert Conformal Conic projection with the California State Plane Coordinate System of 1983 (zone 5), as these are the projection and coordinate system used by Los Angeles County and the City of Long Beach. The specific acquisition and processing methodologies for each of these datasets are detailed in the following paragraphs.

3.1.1. Lawn-to-Garden Data

A tabular dataset provided by the LBWD forms the core of this study design. The dataset comprised 5,394 L2G project applications submitted from January 2010 to September 2016, including a street address and zip code for each project application. These addresses were geocoded to yield a point feature class of property locations indicated on turf replacement project applications. Each point has an associated application date as an attribute, and the 2,884 projects that were successfully completed also have a completion date attribute; these date attributes form the basis of the temporal component of the analysis.

3.1.2. City of Long Beach Data

The Long Beach city boundary and Long Beach street centerline datasets were downloaded from the City of Long Beach’s online GIS Data Catalog in shapefile format and imported into the L2G geodatabase. Each street centerline segment represents an unbroken stretch of road between intersections and includes attributes describing its length, street name, and address ranges. These street centerlines were the basis of the block polygons that served as aggregation units for the project application and completion points.

3.1.3. Los Angeles County Office of the Assessor Data

A geodatabase containing 2015 tax assessor parcel data for all of Los Angeles County was downloaded from the Los Angeles County GIS Data Portal. The assessor parcel feature class includes attributes describing property type, building square footage, number of bedrooms and bathrooms, and lot size for every parcel in Los Angeles County.

3.1.4. United States Census Bureau Data

Next, household composition data were acquired from the US Census Bureau in a two-part process. First, TIGER/Line census block polygons for Los Angeles County were downloaded from the Census Bureau’s Maps and Data website, clipped to the Long Beach city boundary. Then, the block-level 2010 Household Type by Tenure dataset was downloaded from the US Census Bureau’s American FactFinder website as a table of comma separated values. This file was imported into Microsoft Excel, where it was reformatted according to Census Bureau instructions in preparation for joining the tabular data with the TIGER/Line census block polygons in ArcGIS (U.S. Census Bureau 2014).

3.1.5. Zillow Home Value Index Data

Finally, tabular neighborhood-level Zillow Home Value Index (ZHVI) single-family-home time series housing value data were acquired from real estate website Zillow for December of each year from 2010 to 2015. The ZHVI is a hedonic home-price index developed to overcome the bias inherent in median sales price due to fluctuations in the composition of homes sold in a given time period. In other words, the same set of homes is not sold in each time period, so median sold prices based on observed home sales can fluctuate in a way that does not reflect actual market trends. The ZHVI incorporates the estimated home valuations (called "Zestimates") of all houses in a neighborhood, so it is not affected by differences in the mix of properties sold in a given period (Bruce 2014). These valuations are based on multiple property characteristics available in public records, and Zillow reports a median error rate for its estimated valuations of 4.2% for the Los Angeles-Long Beach-Anaheim metro area. This means that the actual sales price of 56.1% of all homes sold are within 5% of their estimated valuation, 78.6% are within 10%, and 91.4% are within 20% (Zillow 2017). These data were downloaded as an Excel spreadsheet and included monthly median ZHVI data calculated for single-family homes in 42 Long Beach neighborhoods. In addition, a polygon shapefile representing the neighborhood boundaries used by Zillow was downloaded.

3.1.6. Basemap

To provide visual context for the vector data and a reference for spot-checking data, the ArcGIS Online World Imagery layer was used as a basemap. This high-resolution raster layer enabled visual inspection of the data and provided a means for spot-checking features during the subsequent processing steps.

3.2. Data Integration and Aggregation

The newly acquired datasets required considerable processing and some preliminary analysis to aggregate and integrate the data to produce the panel dataset of block-quarter analysis units that would be used in the statistical analysis stage of the study.

3.2.1. Selection of Spatial and Temporal Aggregation Units

Because multivariate analysis results are extremely sensitive to adjustments to the scale and segmentation (placement of divisions between units) of both spatial and temporal aggregation units—a phenomenon known jointly as the modifiable areal unit problem (Fotheringham and Wong 1991) and the modifiable temporal unit problem (de Jong and de Bruin 2012; Cheng and Adepeju 2014)—spatial and temporal aggregation and segmentation of data must be undertaken with caution. The street-segment-based city block, rather than larger spatial aggregation units like census blocks or tracts, neighborhoods, or zip codes, was selected as the aggregation unit for this study because the literature review had suggested that residents' landscaping choices are most influenced by the choices of their visually adjacent neighbors (Hunter and Brown 2012) and that street-segment-based city blocks function as a relatively homogenous social unit (Weisburd, Groff, and Yang 2013; Pham et al. 2012). Census blocks were also considered because they are a similarly fine aggregation unit in an urban context and their use would obviate the necessity of re-aggregating census-based household data; however, this aggregation option was ruled out because census blocks would have failed to capture the effect of visual adjacency. Because census blocks generally describe a square city block, that areal unit would have included visually non-adjacent neighbors (those on the other three street segments that define a square block) and excluded visually adjacent neighbors on the opposite side of the street. To test the validity of this decision in the context of L2G participation data, a preliminary analysis of spatial autocorrelation of project application dates (expressed as months elapsed since L2G program inception) was performed at this point on the L2G participation data. The Incremental Spatial Autocorrelation tool in ArcGIS identified a peak autocorrelation distance of 300 feet ($z = 4.137$, 0.000), which is less than the mean length of a Long Beach residential city block (1,182 feet). This result indicates that temporal clustering of project application submissions is most intense at a distance less than the length of a typical block, which supports the decision to aggregate data at the city block level in order to measure the effect of cumulative L2G project completions on application rates. A peak autocorrelation distance greater than the length of an average city block would have indicated that visual adjacency was less important than expected; in that case a larger aggregation unit such as census block groups or tracts might have been more appropriate.

With regard to temporal aggregation and segmentation, Silvestrini and Veredas (2008) note that the same model will produce different results for different frequencies of data, but that in general, smaller temporal units provide more information because they represent a larger number of observations. Balancing the benefits of higher frequency data with practical constraints related to the increased time required to aggregate the data to a larger number of smaller units resulted in the selection of three-month quarters as the temporal units for this study. Finer temporal aggregation was deemed unnecessary because even at the quarter temporal aggregation unit, over 97% of block-quarters have quarterly application rates of zero. The three-month temporal units were segmented with the first quarter beginning on the first day of January 2011 and the last quarter ending on the last day in December 2015 since there was no clear periodicity in the data that would compel a less intuitive segmentation scheme. Seasonality was evident in quarterly application data, with the highest rates of applications generally occurring in the spring; this was corrected by averaging each quarter's application rate with the subsequent three quarters to create a four-quarter future application rate for each block-quarter, representing a "moving window" of annualized data to compare to the previous quarter's cumulative project completion rate on the same and adjacent blocks. Creation of these variables is detailed in section 3.2.4.

3.2.2. Lawn-to-Garden Program Participation Data

The L2G project application and completion dataset was furnished in the form of an Excel spreadsheet containing five fields: address, city, zip code, application date, and completion date. The spreadsheet was imported into the L2G geodatabase in ArcGIS, and the tabular data were geocoded using the ArcGIS Online World Geocoding Service. The geocoding process produced a point feature class of 5,397 points corresponding to all L2G project applications submitted since the program was initiated in 2010. Upon examination, three duplicate records were found and removed, leaving 5,394 unique points.

Further examination of the L2G project participation data revealed that some completed projects were for commercial properties, churches, government-owned buildings, and multi-unit residential buildings. Although this study seeks to characterize the likelihood of L2G participation by residents of single-family houses, a decision was made to include these non-residential project completions because the presence of those completed projects is expected to contribute to the landscaping norms of a block. Final counts of L2G applications and completions are shown by year in Table 1, and Figure 4 shows the spatial distribution of applications submitted for L2G projects on single-family residential parcels from 2011 to 2015.

Table 1. Single-family residential (SFR) L2G applications and completions by year

Year	SFR Applications	All Applications	SFR Completions	All Completions
2010 (April–December only)	532	574	112	122
2011	441	508	346	378
2012	448	493	243	263
2013	651	732	310	347
2014	866	961	470	504
2015	1,337	1,507	873	952
2016 (January–September only)	589	619	253	318
Totals	**4,864**	**5,394**	**2,607**	**2,884**

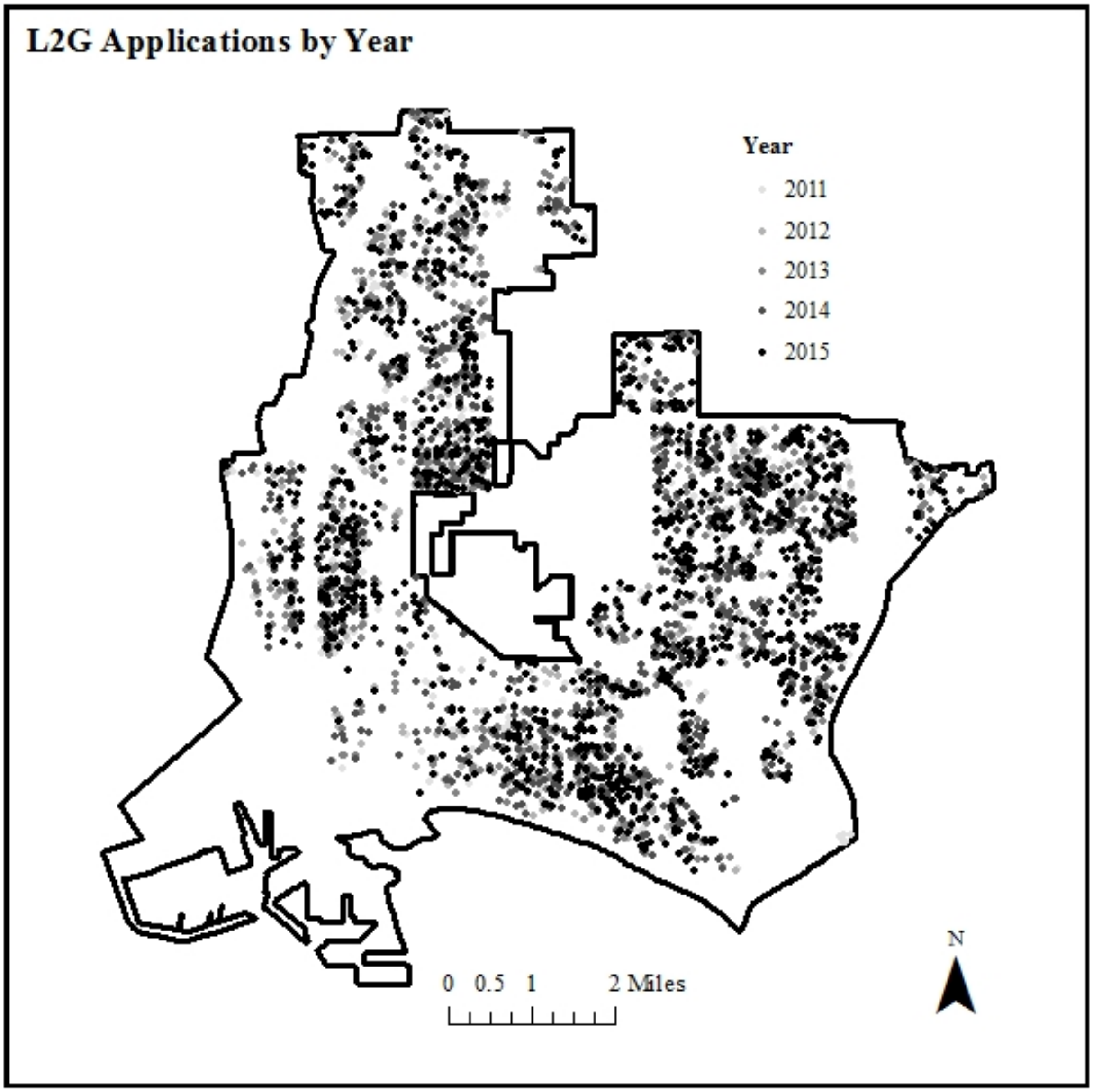


Figure 4. Single-family residential L2G applications by year

Last, a rebate rate attribute was added to each point based on its L2G application date. Rebate rates have changed three times since the inception of the L2G program. The rebate was \$2.50 per square foot of turf removed from the beginning of the program through April 2013; from April 2013 through July 2014 it was \$3.00 per square foot; and from July 2014 to July 2015 the rate reached its peak of \$3.50 per square foot. From July 2015 through the end of 2016, the rate was \$2.50 per square foot.

Figure 5 offers further support for the idea that application rates are influenced by nearby cumulative completion rates. This figure, created in the course of an initial exploration of the L2G participation data, shows the relative density of completed project points by year compared to the density of points representing new application submitted in the following year. In these kernel-density maps, the darkest colors indicate areas with the greatest relative density of cumulative completions or new applications each year. A comparison of the previous year's cumulative completions with the following year's applications reveals that the areas with the greatest application density appear to coincide with or abut areas where the density of cumulative completions is greatest, suggesting the presence of a spatial spillover effect.

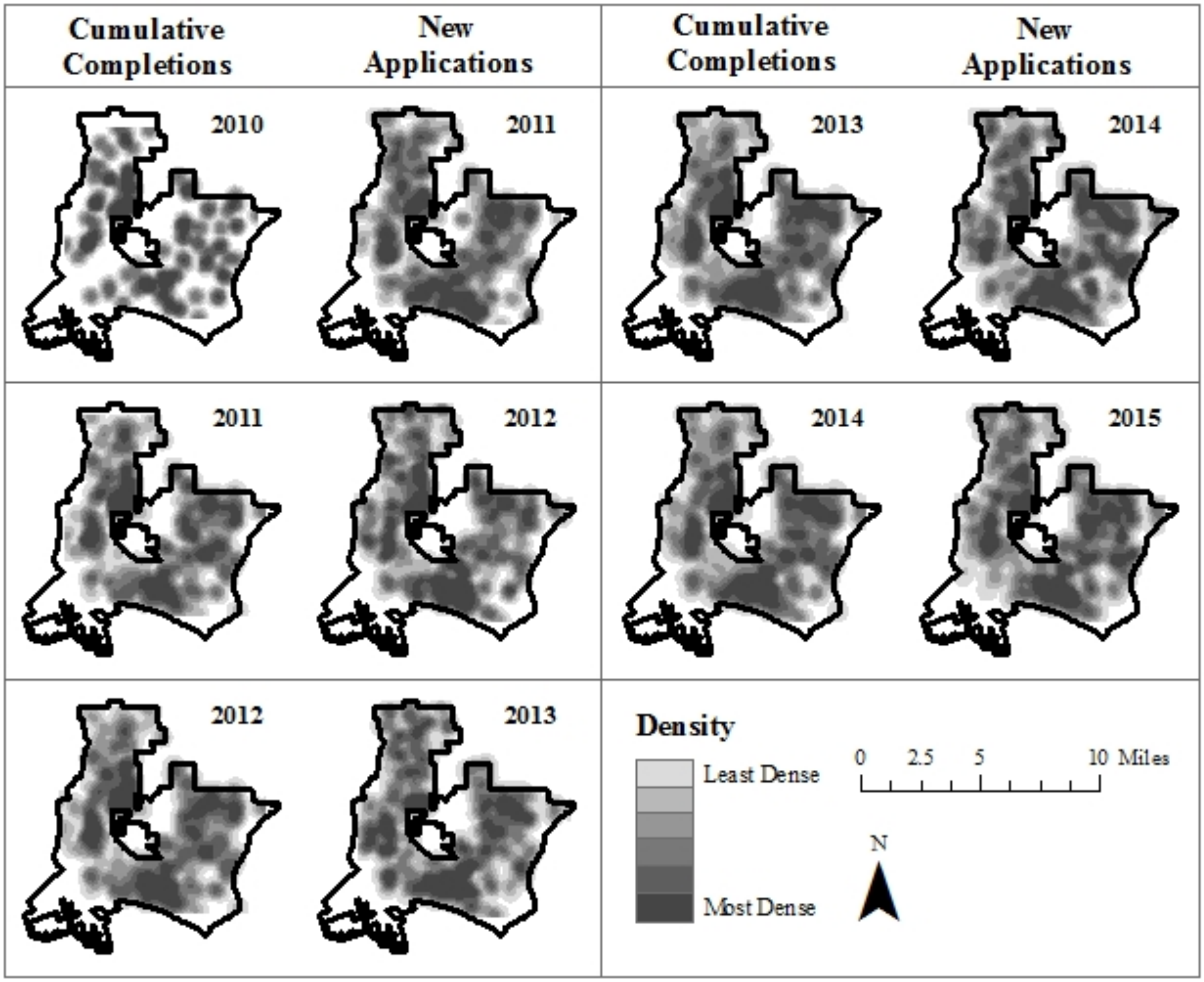


Figure 5. Comparison of kernel density maps of each year's cumulative Lawn-to-Garden project completions alongside the following year's new applications in Long Beach.

3.2.3. Residential Parcel Polygon Data

The next step in the data integration and aggregation stage of the study was to incorporate parcel data. Tax assessor parcels are an essential component of this methodology for two reasons: first, parcel attributes were used to identify applications submitted for single-family housing units, a key piece of information since this study seeks to characterize the likelihood of participation by residents of single-family houses. Second, counts of single-family parcels and total parcels per block were used to normalize application and completion counts, a step deemed necessary because of considerable variation in block length.

Since the Los Angeles County parcel dataset included all parcels in Los Angeles County, the first step was to reduce that dataset to a more manageable one containing only parcels served by the Long Beach Water Department. To that end, parcels whose centroids fell outside the Long Beach city boundary polygon were selected and removed from the Long Beach parcels feature class.

Next, the parcel use type fields were used to select all single-family residential parcels, and these were exported as a separate Long Beach single-family residential parcel feature class. Creating separate feature classes for SFR parcels and all parcels facilitated the calculation of quarterly block application and completion rates discussed later in this chapter; application rates are a ratio of SFR applications to eligible SFR parcels, while cumulative completion rates are a ratio of all project completions to all parcels. The parcels dataset identifies eight types of residential parcels in Long Beach (Table 2). Unlike most single-family residences, many of the housing types listed in Table 2 have no landscaped outdoor space or share a single landscaped space used by multiple residents; consequently, L2G application points for those properties were excluded from the analysis.

Table 2. Specific Residential Parcel Use Types in Long Beach (2015)

<table>
<tr><th colspan="2">Specific Residential Use Type</th><th>Count</th></tr>
<tr><td rowspan="2">Single-Family Residence</td><td>Condominiums: 18,462</td><td rowspan="2">80,531</td></tr>
<tr><td>Houses: 62,069</td></tr>
<tr><td colspan="2">Double, Duplex, or Two Units</td><td>7,549</td></tr>
<tr><td colspan="2">Three Units</td><td>2,114</td></tr>
<tr><td colspan="2">Four Units</td><td>2,933</td></tr>
<tr><td colspan="2">Five or More Units or Apartments</td><td>4,682</td></tr>
<tr><td colspan="2">Rooming or Boarding House</td><td>12</td></tr>
<tr><td colspan="2">Manufactured Home Park</td><td>14</td></tr>
<tr><td colspan="2">Manufactured Home</td><td>61</td></tr>
<tr><td colspan="2">Total Parcels</td><td>97,896</td></tr>
</table>

Further study is needed to determine whether the factors that influence landscaping decisions made by residents of single-family homes are the same as those that influence owners or residents of apartment buildings, multi-unit houses, and mobile homes. Figure 6 and Figure 7, which depict the distribution of single-family residential parcels in two Long Beach neighborhoods, illustrate that the concentration of single-family residential parcels varies considerably from one block to another and from one neighborhood to another. This disparity called for extra care in calculating block-level SFR application rates, as comparing application counts to total parcel counts would have resulted in artificially low rates on mixed-use blocks.

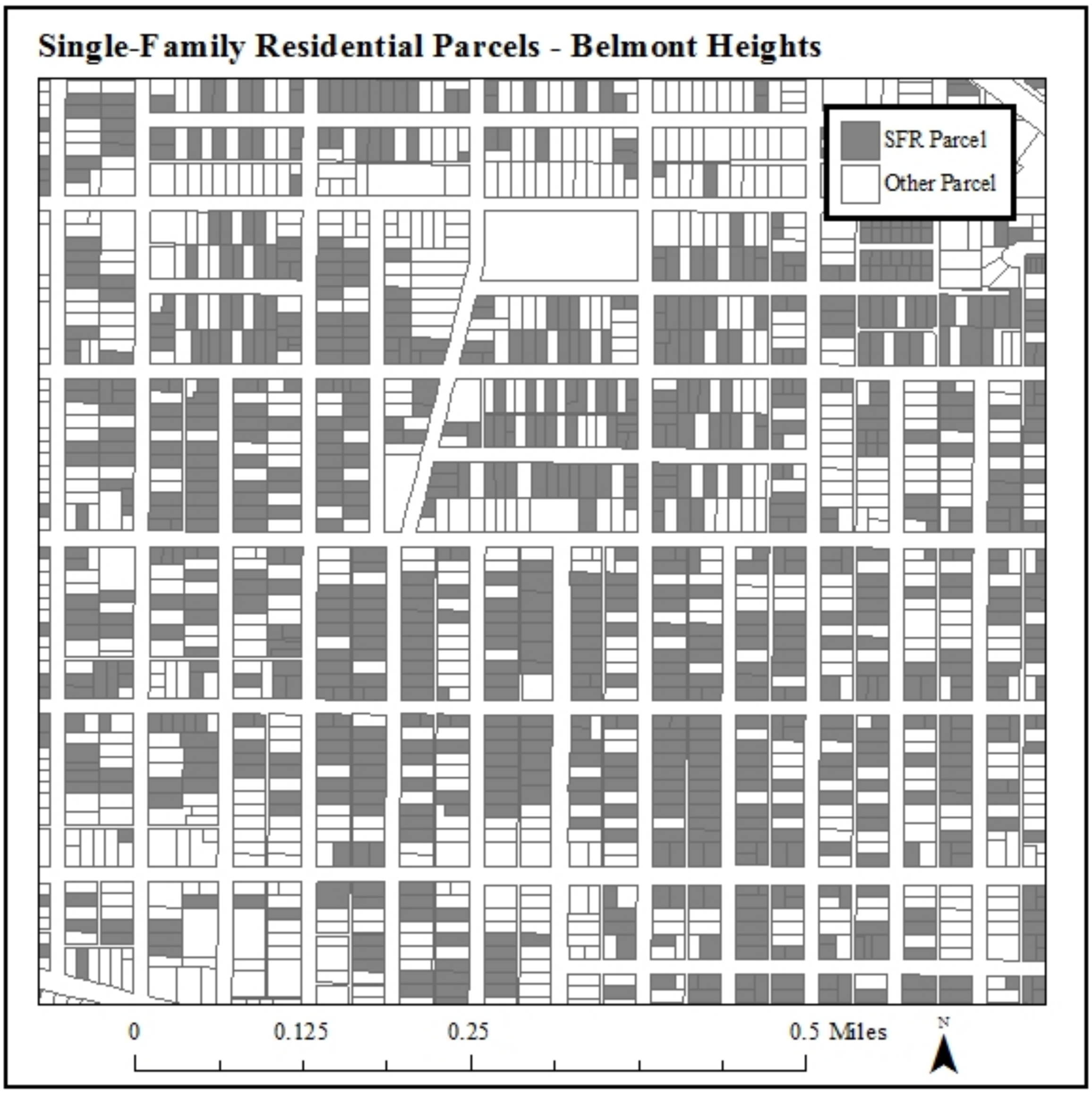


Figure 6. Single-family residential parcels in the Belmont Heights neighborhood

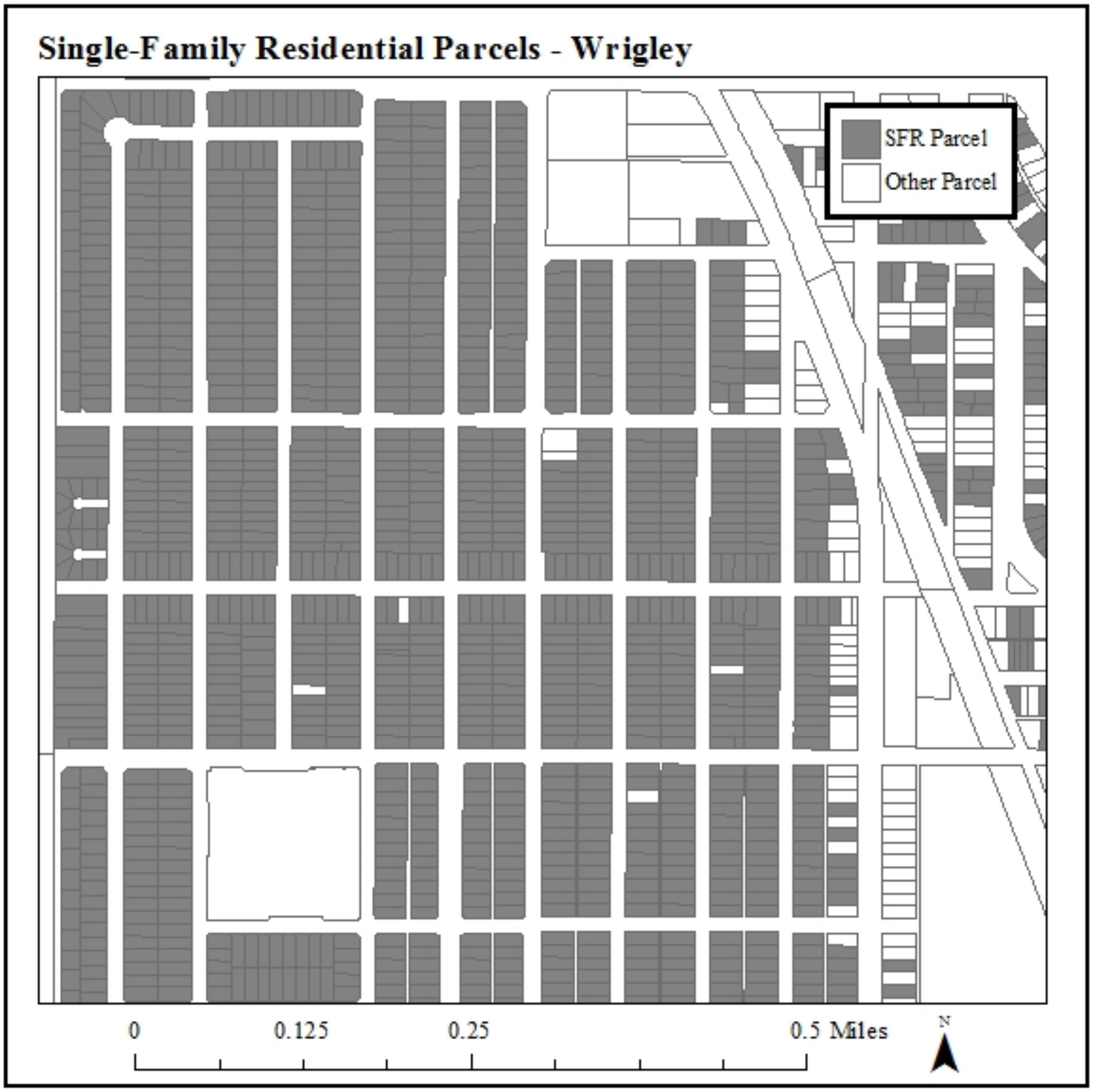


Figure 7. Single-family residential parcels in the Wrigley neighborhood

An initial examination of the Long Beach parcel data revealed duplicate parcels on some blocks. Further investigation of this duplication identified 18,793 stacked parcels (multiple parcels with identical geometry). Of those, 18,462 were categorized as single-family residences, the vast majority of which were classified as condominiums, a unique case in which multiple property owners share ownership of a single plot of land (Solano and Megerdichian 2009). Another 331 stacked parcels were classified as multi-family residential buildings, retail or offices spaces, boat slips, and parking lots where multiple owners share parcel ownership. Because the parcel count was meant to reflect the number of parcels on a block, not the number of parcel owners, duplicate polygons were removed from both the all-parcels dataset and the SFR parcels dataset. Inclusion of stacked parcels would have skewed the L2G application and completion rates by increasing the parcel counts per block (sometimes by hundreds of parcels, as is the case with high-rise condominium buildings).

Finally, L2G application and completion points were spatial joined to the nearest parcel in the all-parcels dataset and a binary attribute was created for each project application to indicate whether it was for a single-family residence or some other type of parcel. Of 4,775 L2G applications submitted between 2010 and 2015, 4,275 were for single-family residences; most of the others were for condominiums, duplexes, and multi-family residences along with a handful for vacant lots and non-residential properties like churches, government-owned buildings, and commercial buildings.

3.2.4. City Block Data

Next, the Long Beach street centerlines were processed in conjunction with both parcel datasets and the L2G project applications and completions data to produce the city block polygons that would serve as spatial aggregation units. Centerline segments were buffered on both sides by 20 meters to enable a subsequent spatial join with adjacent parcels. The 20-meter buffer distance was chosen to assure adequate overlap with parcel boundaries even on wide streets while not extending beyond the rear lot lines of small lots located on narrow streets, as illustrated by Figure 8. Buffering the centerline segments resulted in 9,708 polygons representing each city block in Long Beach, i.e. each stretch of street between intersections.

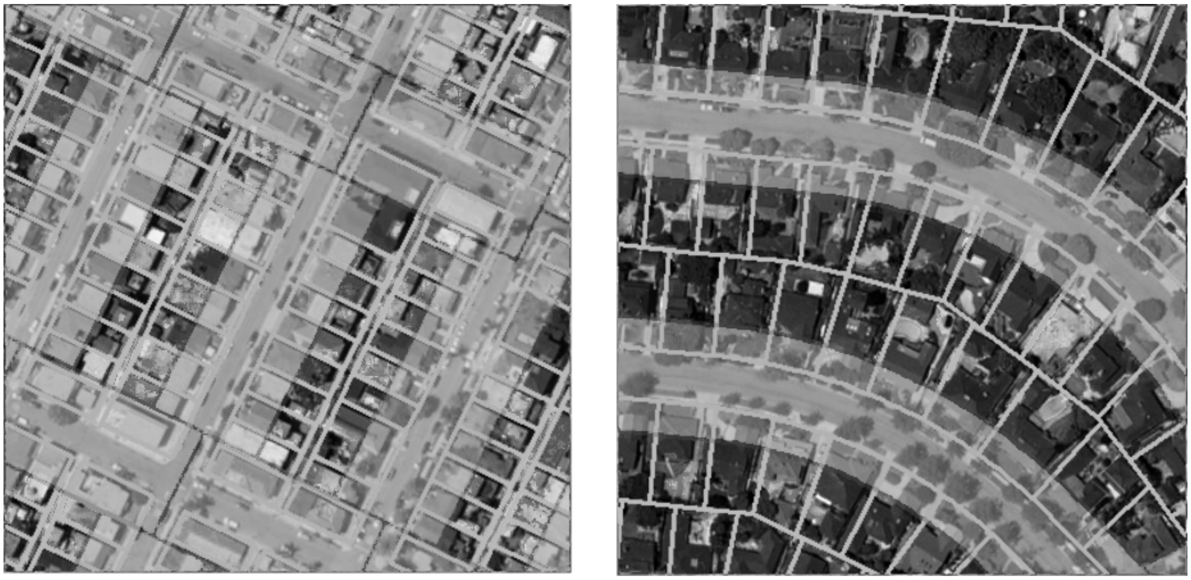


Figure 8. Street centerline buffers. 20-meter centerline buffers overlap the street-side edge of parcel boundaries on both narrow streets with small parcels (left) and wide streets with large parcels (right).

Next, these block polygons were spatially joined to intersecting parcel polygons to calculate counts for both single-family parcels (not including condominiums) per block and total parcels per block. In addition, averages were captured for the following single-family parcel data for each block: parcel size, structure size, and number of bedrooms. Values for parcels classified as condominiums or as property types other than single-family residences were not included in those averages. Finally, 2,187 blocks with no single-family residential parcels were removed from the dataset, ten blocks with null parcel attributes were removed, and two pairs of blocks were merged, leaving a total of 7,509 city blocks with at least one single-family parcel. These remaining block polygons are shown in Figure 9.

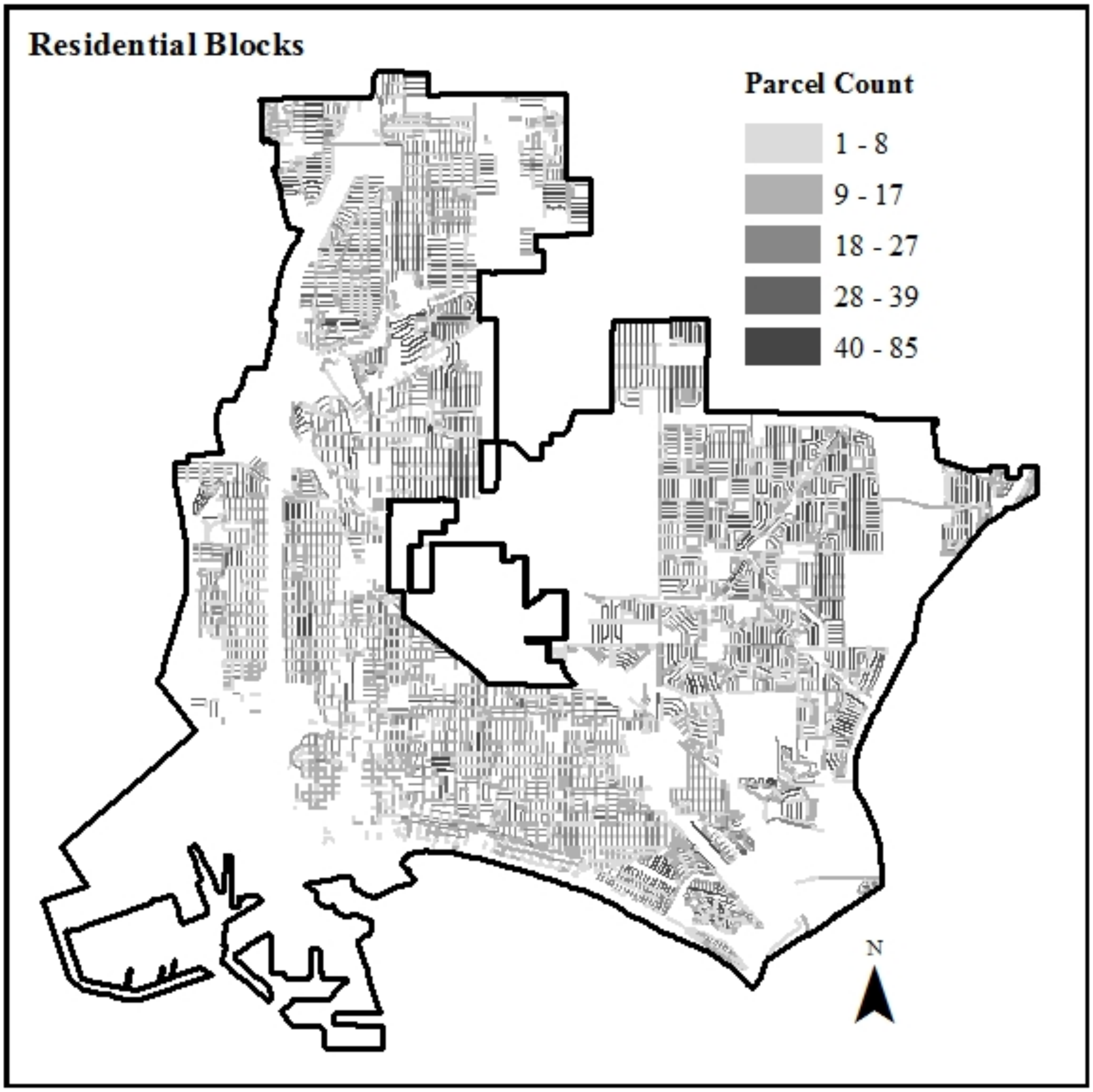


Figure 9. Block polygons with one or more parcels classified as single-family residences.

3.2.3.1. Application and Completion Counts

Because the study aims to characterize patterns of L2G program participation over time, block-level L2G application and completion counts were calculated for all 26 three-month quarters in the dataset (April 2010 through September 2016). These counts were then used to calculate four application and completion rate variables for the 20 quarters in the study period (January 2011 through December 2015), a time span that was selected because each rate variable could be calculated for every quarter. Counts for the three quarters preceding the study period and the three quarters following the study period were included in the calculations because cumulative completion rates incorporated projects completed prior to the first quarter of the study period, and four-quarter future application rates included applications submitted in the three quarters following the end of the study period.

To calculate quarterly application counts for each city block polygon, applications submitted for single-family parcels were iteratively selected in one-quarter selection sets based on application date and spatially joined to the city block polygons, capturing single-family residential application count values for each block in each quarter and storing that information as a block polygon attribute (e.g. Q1 Application Count, Q2 Application Count, etc.). The same iterative process was then repeated for project completions, using completion dates to select projects completed for any parcel type in each quarter and capturing that quarterly block-level count as a block polygon attribute. Then the process was repeated for single-family residential completions only, a value that would be used to adjust the number of eligible SFR parcels on each block for the calculation of quarterly adjusted application rates. Finally, both quarterly cumulative completion counts and quarterly cumulative SFR completion counts were calculated for each block polygon by summing the completion counts of the current quarter and all previous quarters.

3.2.3.2. Same-Block Cumulative Completion Rates

This study hinges on knowing, for each L2G application, how many L2G projects have already been completed on the applicant’s block. Thus, a cumulative completion count is required. However, some block polygons contain many more residential parcels than others due to differences in street segment length or parcel size. For example, a block where three out of ten parcels have completed projects is not equivalent to a block where three out of eighty parcels have completed projects. Accordingly, the cumulative completion rate (CCR) rather than the cumulative completion count was required for analysis to normalize variations in aggregation unit size for the purpose of regression analysis. The CCR value for each block-quarter is equal to each quarter's cumulative completion count divided by the total number of parcels on that block (Equation 1).

Equation 1. Cumulative completion rate

$$\text{Quarterly Cumulative Completion Rate} = \frac{\text{Quarterly Cumulative Completion Count}}{\text{Total Parcel Count}}$$

The end result was a L2G CCR value for each block polygon for each quarter from 2011 to 2015; mean positive CCR values are shown in Figure 10.

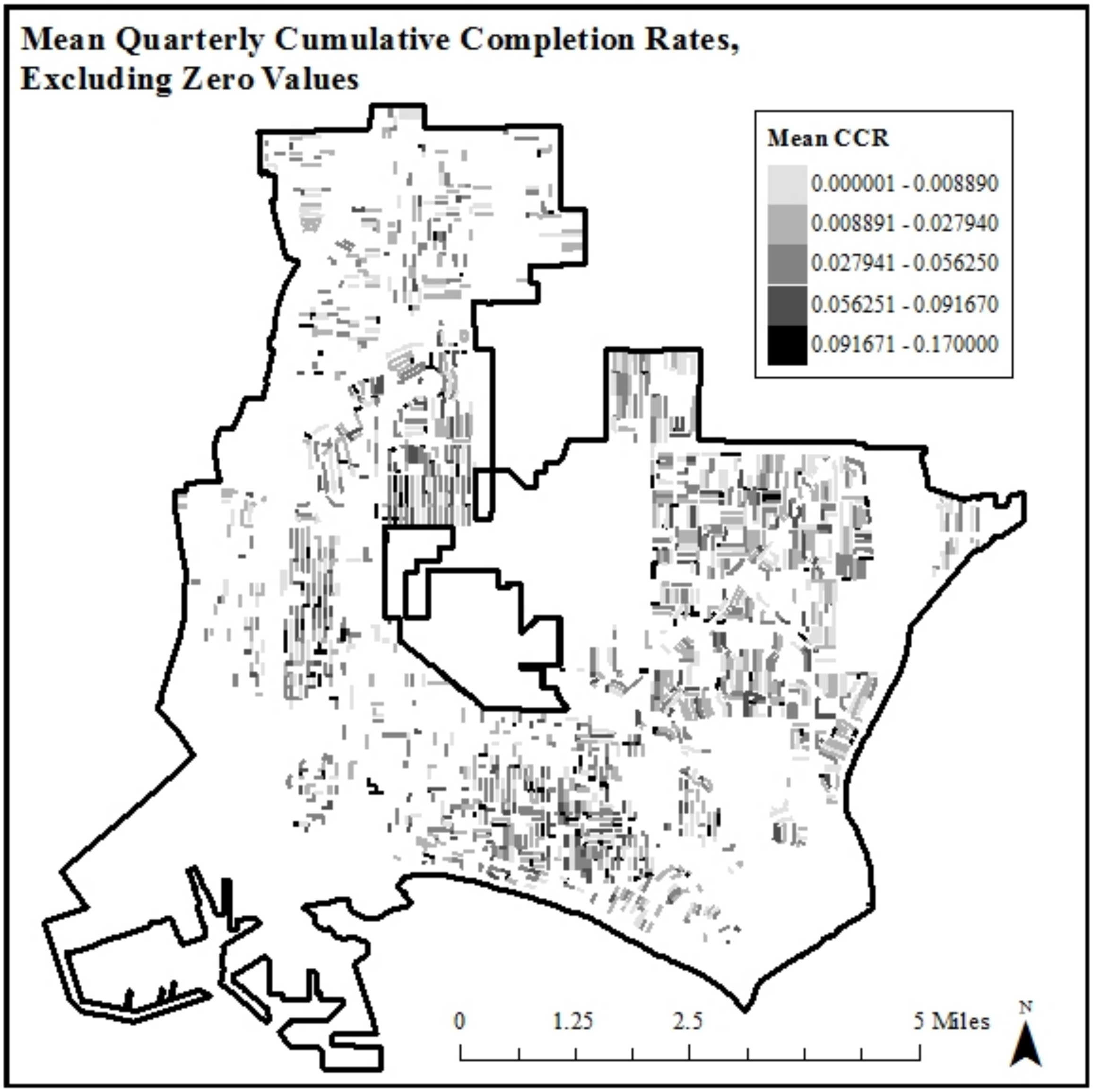


Figure 10. Mean quarterly cumulative completion rates, excluding blocks with no completions.

While this study is focused on the L2G participation of residents and owners of single-family properties, the CCR value includes all completed projects including those for vacant lots, commercial properties, churches, government-owned buildings, condominiums, and multi-family residential buildings. These projects were included in the CCR because they are presumed to contribute to the neighborhood landscaping norms regardless of the nature of the buildings they front; therefore, all completed L2G projects are expected to exert an influence on the future SFR application rate, which reflects the landscaping choices of inhabitants of single-family residences.

Of 7,509 block polygons, 1,826 were found to have a CCR greater than zero at the end of the study period; a CCR of zero would indicate that no L2G projects had been completed on that block. The minimum rate was 0.003 while the maximum rate was 1, but an examination of the data's histogram revealed a non-normal distribution of the rate values in between, with a small number of very high outlier rates skewing the data. Examination of the blocks with the highest L2G conversion rates identified two scenarios most likely to result in an abnormally high conversion rate. First, short "end blocks" like those pictured on the left side of Figure 8 often have a very small number of parcels (usually four, but sometimes only one or two), which is inadequate to produce a rate that accurately reflects a neighborhood trend, since a single project could result in a rate of 0.25 or greater. The second scenario involves blocks with few residential parcels, such as mixed-use blocks where residential parcels were located alongside schools, parks, or retail, office, or industrial parcels. To detect these outliers, the absolute deviation around the median was calculated (Leys 2013) and values that were more than three absolute deviations from the mean were removed. The median value of CCR values greater than zero was 0.083 and the median absolute deviation was 0.124. Therefore, all blocks with L2G conversion rates greater than 0.454 were removed from the dataset, reducing the block count by 22 blocks for a total block count of 7,487. All rates lower than the median fell within a single absolute deviation from the median.

Further examination of the city blocks dataset found 187 blocks for which county assessor data values were null. Satellite imagery revealed that the largest group of these blocks fell within a mobile home park where the parcels had been classified as "Vacant Land" rather than “Manufactured Home” or “Manufactured Home Park” in the assessor data; other blocks with null values included those with condominium or cooperative complexes that were not identified as such in the parcel data, blocks within a small recently built development, and scattered blocks that turned out to be bridges or other uninhabitable stretches of road. The 187 blocks with null values were removed, leaving a total block count of 7,300.

Another issue encountered during block polygon creation was the inclusion of completed L2G projects in the counts and rates of more than one block when projects were located on corner parcels. Because most block polygons overlap the polygons buffering their cross-streets at the corner parcel, projects occurring on corner lots were often included in the CCR calculations of more than one block polygon. This duplication was deemed to be acceptable since a corner-lot project is expected to contribute to the neighborhood norm of both blocks and should thus be represented in both CCR values.

3.2.3.3. Adjacent-Block Cumulative Completion Rates

The spatially lagged dependent variable designed to test for a spatial spillover effect was the mean Adjacent-Block Cumulative Completion Rate (ABCCR). These values were calculated by performing a spatial join between the city block feature class and a copy of itself and calculating the sum of the CCR value for all intersecting polygons. The subject block’s own CCR value was subtracted from that sum and that result was divided by the intersecting block count minus one (the number of adjacent blocks, excluding the subject block) to create a new quarterly variable for each city block representing the mean of all adjacent block CCR values (Equation 2).

Equation 2. Adjacent-block cumulative completion rate

$$\text{Quarterly mean ABCCR} = \frac{\text{Sum of CCR values of intersecting blocks} - \text{same-block CCR}}{\text{Intersecting block count - 1}}$$

The ABCCR variable would allow the regression analyses to test whether residents' willingness to participate in the L2G program is influenced by previous L2G completions on adjacent blocks.

3.2.3.4. Adjusted Application Rates

The dependent variable for each block-quarter was the percentage of eligible residents of single-family residences on that block who submitted applications in a given quarter, referred to herein as the single-family residential adjusted application rate (AAR). The "adjusted" portion of the name refers to the fact that the single-family residential parcel count undergoes an adjustment to remove parcels on which a L2G project has already been completed. Parcels where lawns have already been converted to gardens in previous quarters are generally no longer eligible to participate in the program; summarizing the street address field in the application feature class confirmed that only one application had been submitted for each address.

In order to calculate the AAR for each block-quarter, each quarterly count of single-family residential applications was divided by the difference between the total number of single-family residential parcels on the block and the number of L2G projects previously completed on single-family residential parcels (Equation 3).

Equation 3. Adjusted application rate

$$\text{Quarterly SFR adjusted application rate} = \frac{\text{Quarterly SFR application count}}{\text{Total SFR parcel count - (Sum of all previous SFR completions)}}$$

It is important to note that the AAR is not a cumulative proportion like the CCR, so the set of quarter-block AAR values is composed mostly of zeros interspersed with occasional small positive numbers. Mean positive AAR values are shown in Figure 11.

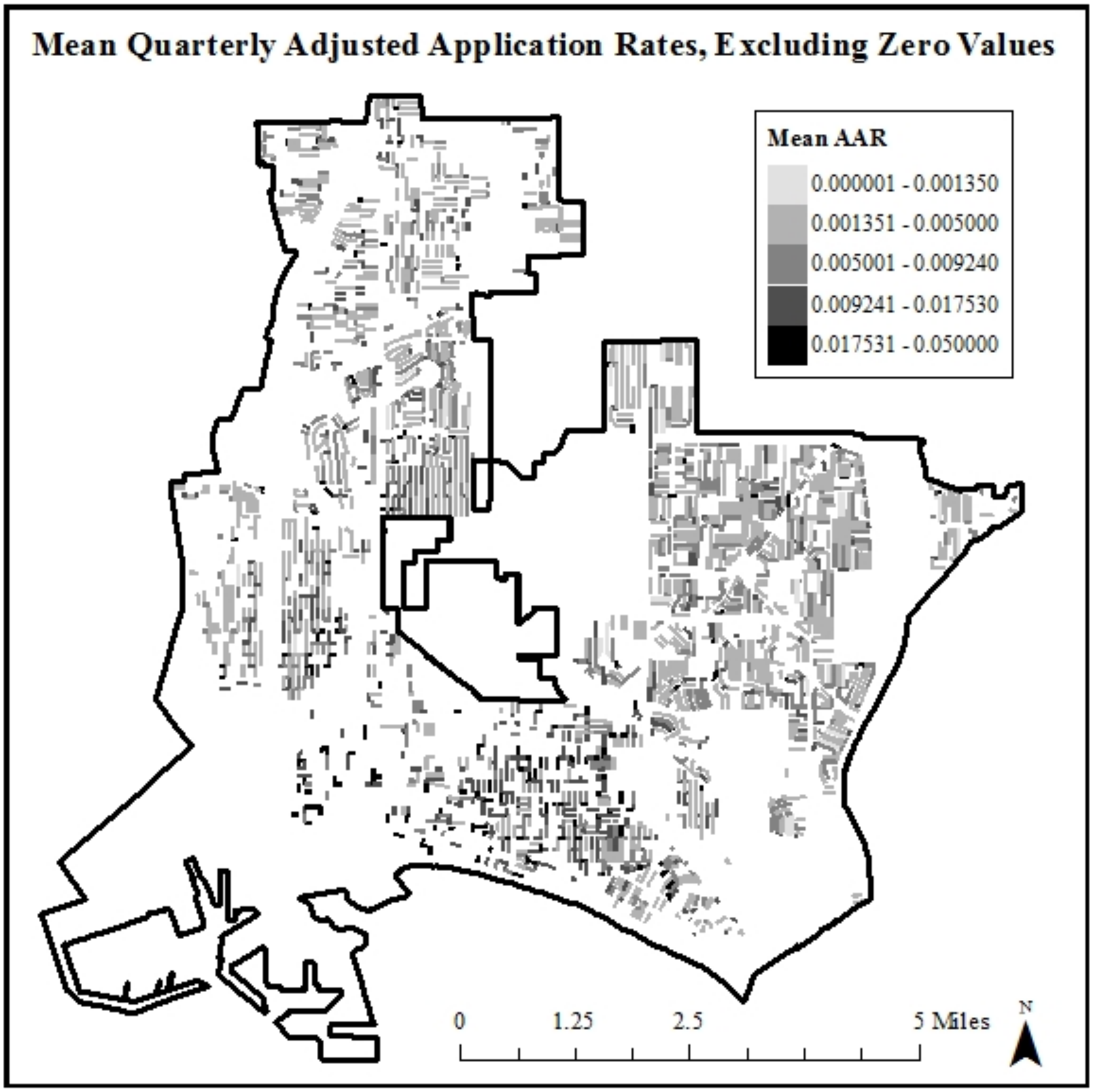


Figure 11. Mean quarterly AAR values, excluding zeros.

3.2.3.5. Four-Quarter Mean Adjusted Application Rates

The mean of the current quarter's AAR rate and that of the three subsequent quarters was also calculated, to provide a measure of long-term L2G project participation outcome. A multi-quarter application rate—rather than the quarterly AAR—was deemed important because while neighbors may be influenced by a new completion on their block or an adjacent block, they might not act on that influence for months or even years because of economic factors, scheduling concerns, or other unknown reasons. The four-quarter mean adjusted application rate (4QMAAR) allowed the effect of an increase of completion rate to be captured anytime during the subsequent year. This four-quarter temporal scale was selected because it ensured that seasonal variations in participation would be equivalent for all block-quarters. Moreover, this temporal scale allowed for analysis of five complete years of data with the currently available dataset; as additional years of data are gathered, it would be interesting to analyze a multi-year future application rate variable as well. The 4QMAAR value was calculated as shown in Equation 4, where AAR_Q represents the AAR value of the current quarter, AAR_{Q+1} represents the AAR value of the following quarter, and so on.

Equation 4. Four-quarter mean adjusted application rate

$$\text{Quarterly 4QMAAR} = \frac{AAR_Q + AAR_{Q+1} + AAR_{Q+2} + AAR_{Q+3}}{4}$$

3.2.3.6. Four-Quarter Application Presence

Finally, a binary four-quarter application presence variable was calculated based on the 4QMAAR variable for each block-quarter. Using the field calculator, a 1 was assigned when 4QMAAR was greater than zero, while a 0 was assigned when 4QMAAR was equal to zero. This variable would be used as the binary presence-absence outcome for the binary logistic regression described in section 4.2.

3.2.5. Census Data

Six independent variables were derived from the census block-level “Household Population and Household Type by Tenure” 2010 Census dataset: owner occupation percentage, family percentage, senior householder percentage, female householder percentage, male householder percentage, and husband-wife householder percentage. Calculation and aggregation of these variables involved a number of steps described in detail in the paragraphs that follow.

To summarize, census block polygons were first joined to the tabular household dataset using their common unique identifier (United States Census Bureau 2014), resulting in 5,181 census blocks attributed with household composition data. An assessment of that data revealed that 1,300 blocks had a population of zero; those blocks were removed from the dataset, leaving a total of 3,931 populated census blocks. Next, the six census variables were calculated for each census block and a separate raster layer was generated for each independent variable. Finally the city block polygons were used to extract the mean values of all intersecting raster cells. These mean independent variable values were stored as attributes of each city block polygon.

This method of rasterizing the census data ensured that intersecting census blocks were represented proportionally in the mean values assigned to each city block polygon. In other words, if a city block polygon overlapped one census block by 95% and the other by only 5%, then about 95% of the intersecting raster cells would store the first census block's value and the remaining 5% of the intersecting cells would store the second census block's value, so the mean of all those intersecting cells would reflect those proportions. Conversely, a simple spatial join of census block polygons with city block polygons would have simply added the two values and divided by two, disregarding the proportions of each city block intersected by the census blocks. Furthermore, most census block polygons include the parcels facing all four street segments that constitute a square city block, while the city block polygons created for this study represent only those houses facing each other across a single street segment. This means that there were nearly twice as many residential city blocks (7,300) as census blocks (3,931), and each city block usually intersected at least two census blocks (adjacent census block polygons typically meet in the center of a city block polygon as illustrated in Figure 12). Proportionally averaging the characteristics of the census blocks intersecting each city block provided a general characterization of a location with regard to the ratios of homeowners to renters, seniors to younger residents, families to non-families, and husband-wife co-householders to male and female heads of household.

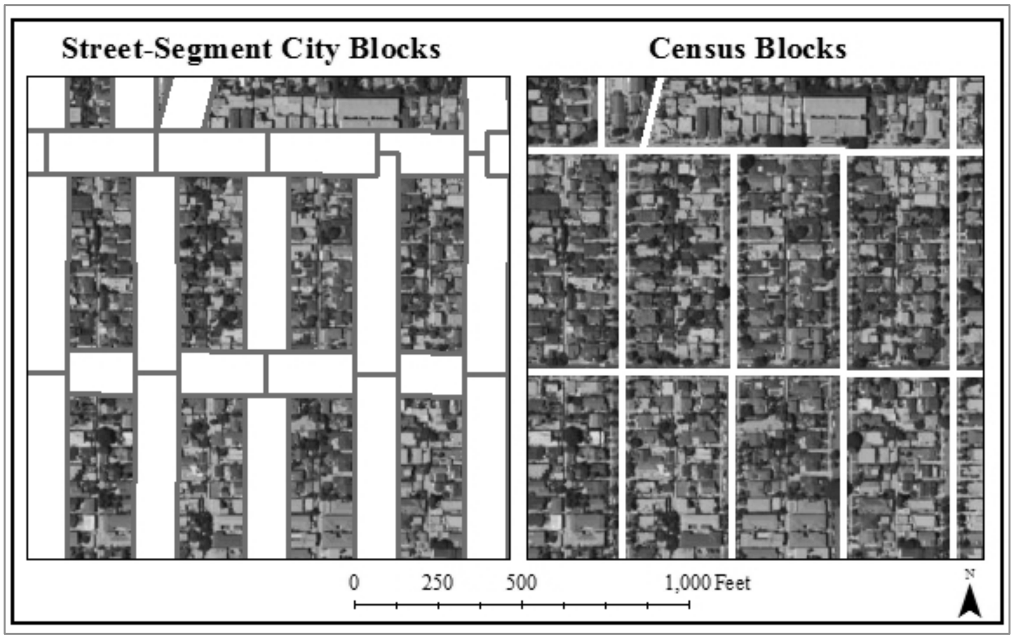


Figure 12. City blocks and census blocks. The white polygons on the left are city-block polygons created by buffering street centerlines, while the white lines on the right outline the larger census block polygons that usually delineate square blocks.

The Household Population and Household Type by Tenure dataset had 86 fields storing household data for 43 different hierarchical categories; each category is described in two fields, one a raw number and the other a percentage. Field names beginning with HD01 are raw numbers, while field names beginning with HD02 are percentages. The 24 fields that were used to create the five independent variables derived from these data are listed in Table 3, along with descriptions of the data contained in each. It should be noted that family households are defined as those in which two or more members are related by birth, adoption, or marriage. Same-sex couples and unmarried couples with no children were not counted in this category in the 2010 Census (Lofquist et al. 2012). Table 4 details how the census data fields were combined to obtain percentages for each of the regression variables. Note that combining the percentage values of each category of variables (occupancy, type, age, and gender) always equals 100.

Table 3. Census data field names in the census-block-level 2010 Household Type by Tenure dataset

Field Name	Description
HD01_S01	Count of occupied housing units
HD01_S02	Count of all units occupied by owner
HD01_S05	Count of all units occupied by renter
HD01_S09	Count of owner-occupied units occupied by families
HD01_S15	Count of owner-occupied units occupied by non-families
HD01_S27	Count of renter-occupied units occupied by families
HD01_S33	Count of renter-occupied units occupied by non-families
HD01_S10	Count of owner-occupied units occupied by families with householder under 65
HD01_S11	Count of owner-occupied units occupied by families with householder 65 or over
HD01_S16	Count of owner-occupied units occupied by non-families with householder under 65
HD01_S17	Count of owner-occupied units occupied by non-families with householder 65 or over
HD01_S28	Count of renter-occupied units occupied by families with householder under 65
HD01_S29	Count of renter-occupied units occupied by families with householder 65 or over
HD01_S34	Count of renter-occupied units occupied by non-families with householder under 65
HD01_S35	Count of renter-occupied units occupied by non-families with householder 65 or over
HD01_S12	Count of owner-occupied units occupied by families with husband-wife householders
HD01_S13	Count of owner-occupied units occupied by families with male householder
HD01_S14	Count of owner-occupied units occupied by families with female householder
HD01_S18	Count of owner-occupied units occupied by non-families with male householder
HD01_S22	Count of owner-occupied units occupied by non-families with female householder
HD01_S30	Count of renter-occupied units occupied by families with husband-wife householders
HD01_S31	Count of renter-occupied units occupied by families with male householder
HD01_S32	Count of renter-occupied units occupied by families with female householder
HD01_S36	Count of renter-occupied units occupied by non-families with male householder
HD01_S40	Count of renter-occupied units occupied by non-families with female householder

Table 4. Census regression variable composition. Each regression variable was calculated by combining the values stored in multiple fields from the Household Type by Tenure dataset and dividing the result by the population value stored in HD01-S01.

Category	Variable (%)	Calculation (where HD01_S01 > 0)
Occupancy	Owner Occupancy	([HD01_S02] / [HD01_S01]) * 100
	Renter Occupancy	([HD01_S05] / [HD01_S01]) * 100
Type	Family Occupancy	(([HD01_S09] + [HD01_S27]) / [HD01_S01]) * 100
	Non-family Occupancy	(([HD01_S15] +[HD01_S33]) / [HD01_S01]) * 100
Age	Senior Householders	(([HD01_S10] + [HD01_S16] + [HD01_S28] + [HD01_S34]) / [HD01_S01]) * 100
	Non-Senior Householders	(([HD01_S11] + [HD01_S17] + [HD01_S29] + [HD01_S35]) / [HD01_S01]) * 100
Gender	Husband-Wife Householders	(([HD01_S12] + [HD01_S30]) / [HD01_S01]) * 100
	Male Householder	(([HD01_S13] + [HD01_S18] + [HD01_S31] + [HD01_S36]) / [HD01_S01]) * 100
	Female Householder	(([HD01_S14] + [HD01_S22] + [HD01_S32] + [HD01_S40]) / [HD01_S01]) * 100

For each calculated regression variable, the Polygon to Raster tool was used to generate a raster layer with cell values set to the value of the variable (Figure 13 through Figure 18). A raster cell size of 50 feet was chosen because it was less than both the peak autocorrelation distance demonstrated in the L2G project completion data (50 feet vs. 100 feet) and the average area of a residential parcel in Long Beach (2,500 feet vs. 6,550 feet). Next, the Zonal Statistics by Table tool was used to calculate the mean values of the raster cells that intersected each city block polygon; the output of this operation was a table containing the identifier of each city block, a count of the intersecting raster cells, and their mean value. That table was then joined to the city block polygons attribute table and the mean variable value was copied into a corresponding field in the city blocks attribute table before removing the join. This process was repeated for each of the six independent variables derived from census data.

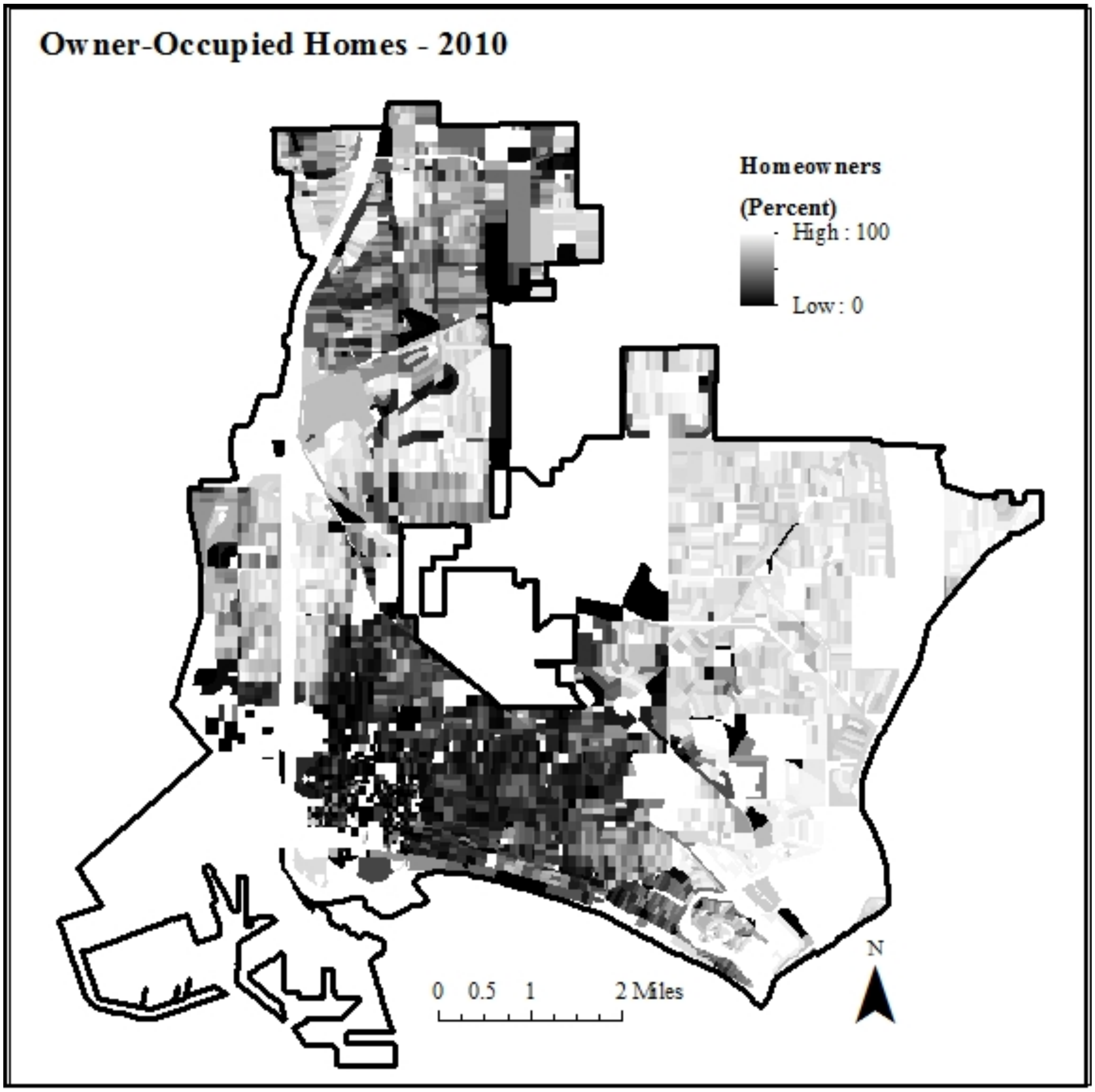


Figure 13. Owner occupancy

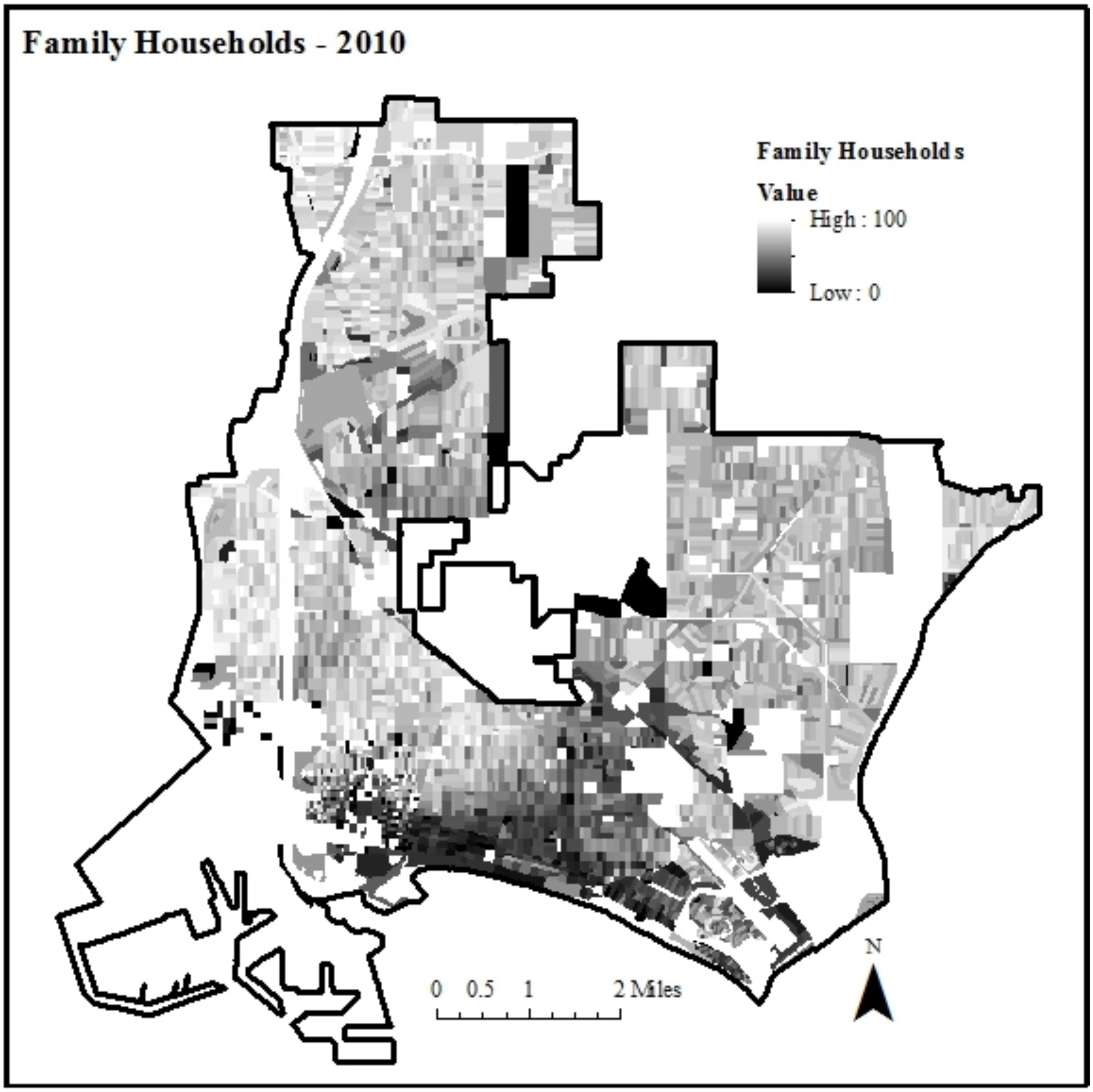


Figure 14. Percentage of family households

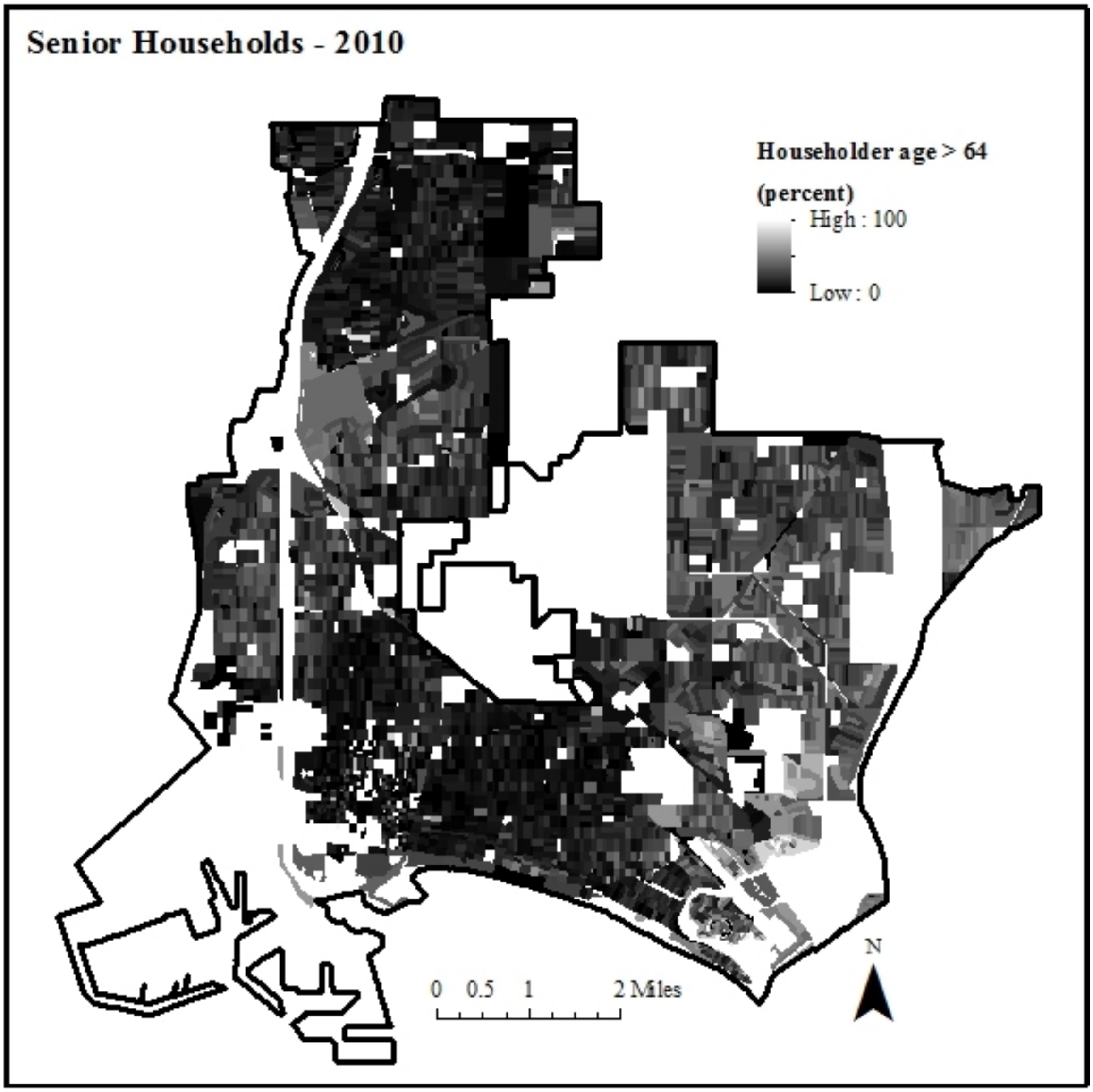


Figure 15. Percentage of heads of household over 65 years old

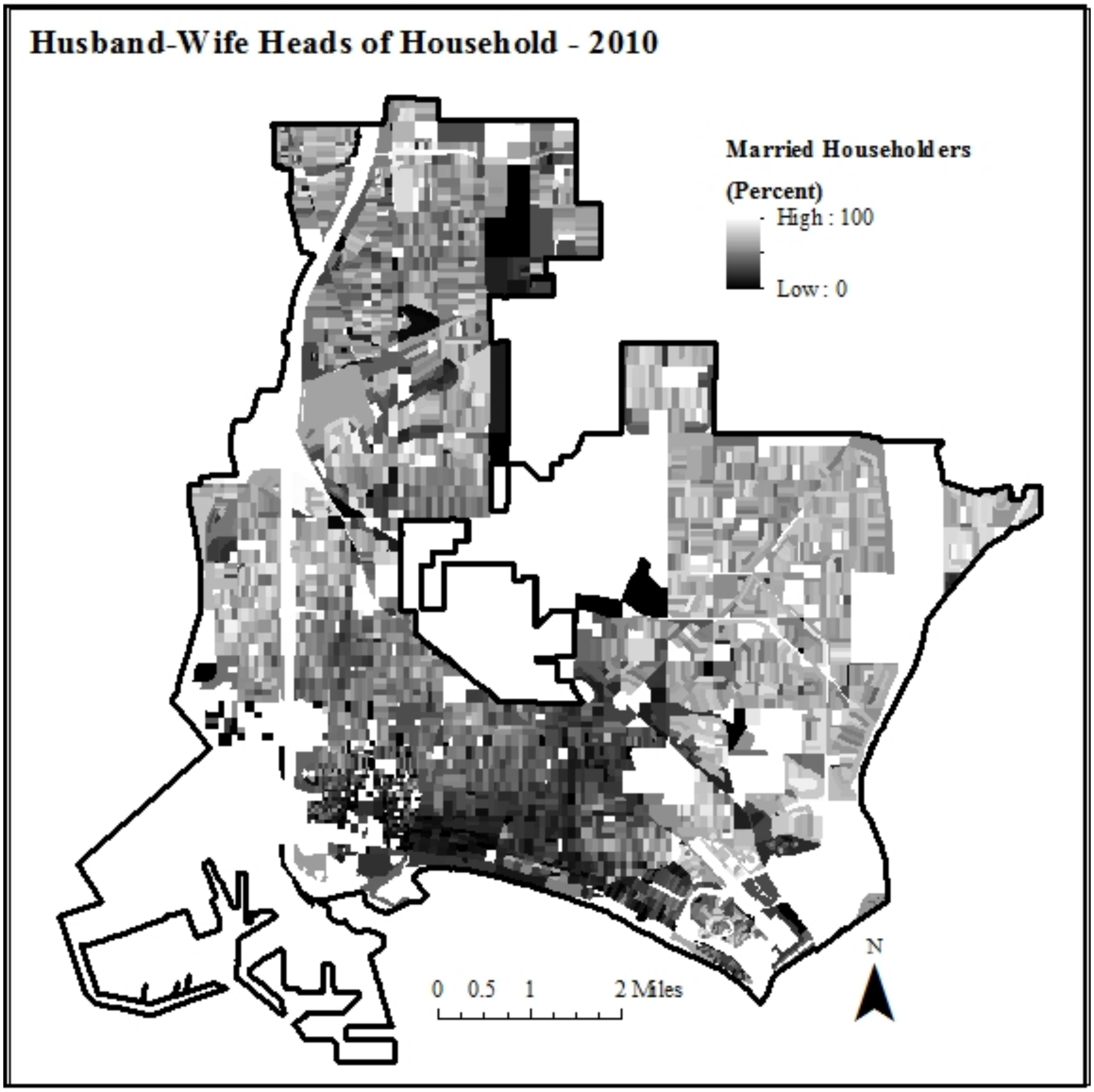


Figure 16. Percentage of households headed by a husband and wife

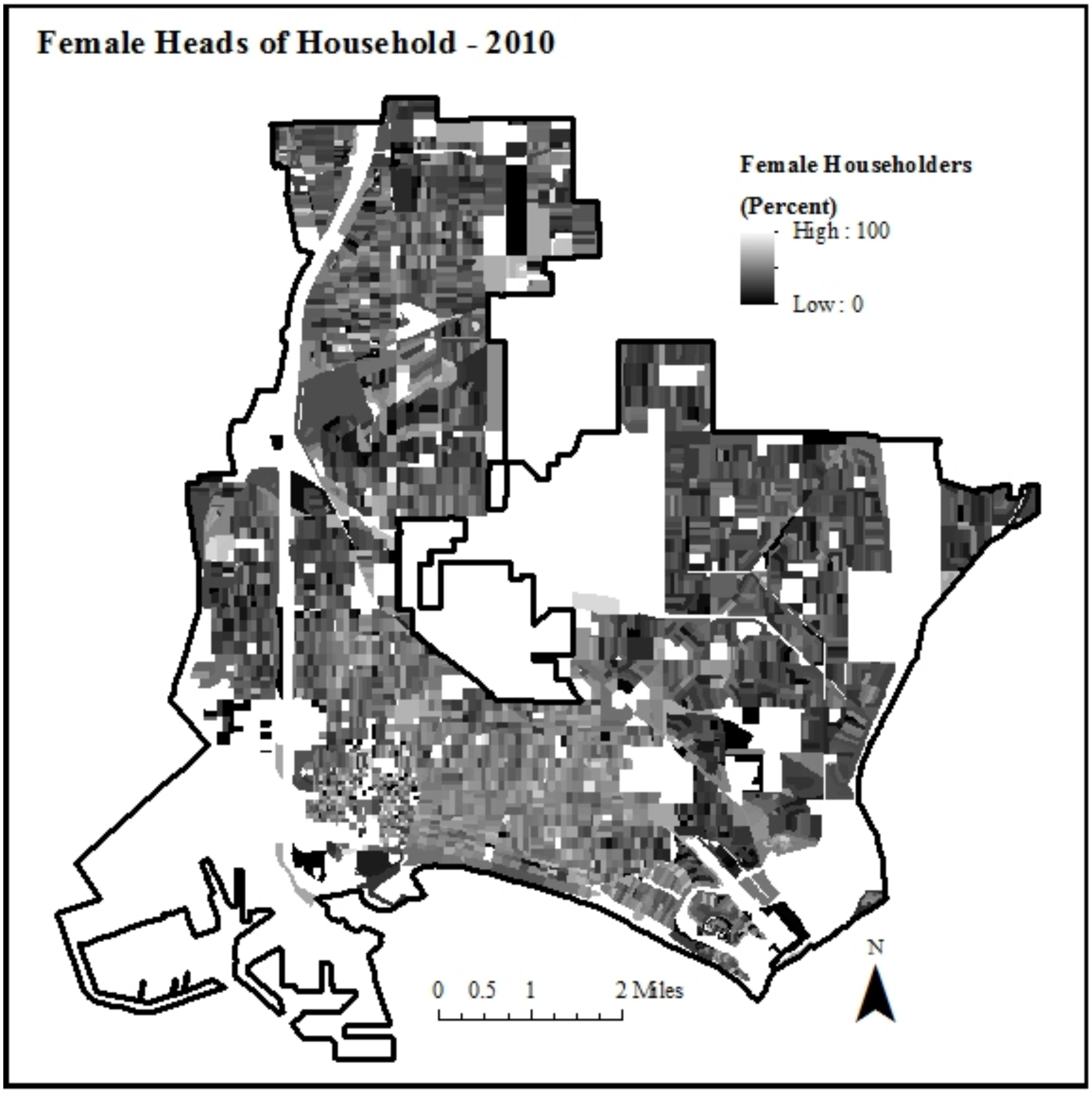


Figure 17. Percentage of households headed by a woman

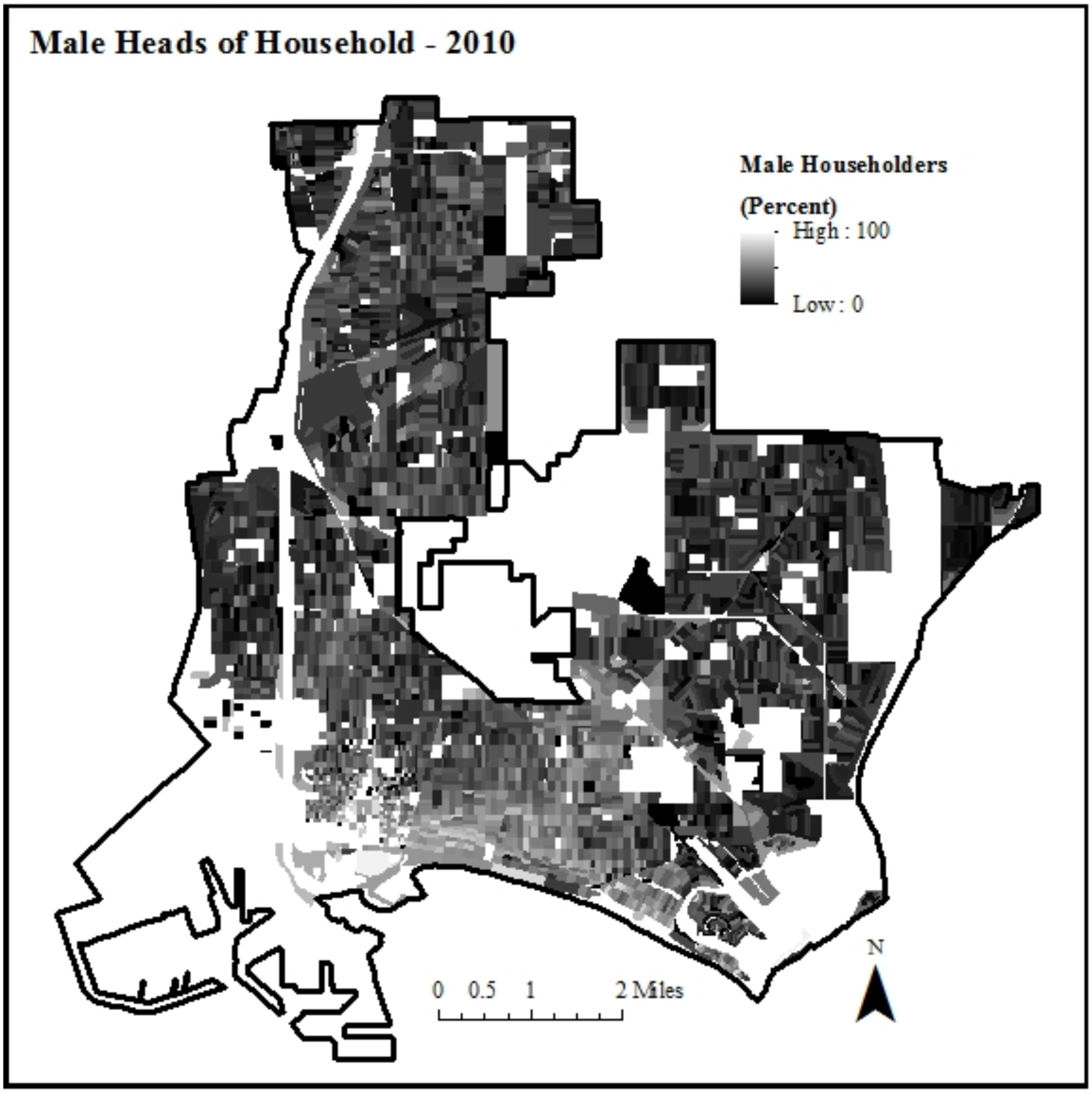


Figure 18. Percentage of households headed by a man

3.2.6. Zillow Data

The methodology for capturing the mean median single-family home ZHVI independent variable for each city block polygon for each year from 2011 to 2015 was similar to that used to capture the census-derived variables, because the Zillow neighborhood polygons were much larger than the city block polygons. First, the Zillow neighborhood polygons for the state of California were limited to those within Long Beach, resulting in a smaller dataset of 51 neighborhoods. These 51 polygons were then joined to the tabular median ZHVI data on the RegionID field. Fifteen of the 51 neighborhood polygons contained no joined median ZHVI values; examination of those polygons revealed that nine of them represented unpopulated areas such as the Port of Long Beach, the man-made offshore oil islands, large public parks, and the Long Beach Airport, so those polygons were removed from the dataset. An additional six neighborhood polygons with no ZHVI values contained small pockets of residential blocks in the downtown area, as well as Carroll Park in the Bluff Heights area and the Peninsula between Alamitos Bay and the Pacific Ocean, adjacent to Belmont Shore. These six polygons were each merged with an adjacent neighborhood polygon based on geometry (longest shared edge, locations of intersecting city block polygons), and local knowledge of shared neighborhood characteristics. Next, for each year from 2011 to 2015, the Polygon to Raster tool was used to generate a median ZHVI raster layer from the Zillow neighborhood polygons with cell values set to store that year's December median ZHVI value (Figure 19). Like the census-derived variable rasters, the rasters were assigned a cell size of 50 feet. Finally, the Zonal Statistics by Table tool was used to calculate the mean median ZHVI values of the raster cells that intersected each city block polygon; this operation produced a table containing the identifier of each city block, a count of the intersecting raster cells, and their mean median ZHVI value. That table was then joined to the city block polygons attribute table and the mean median ZHVI value was copied into a corresponding field in the city blocks attribute table before removing the join. This process was repeated for each year of the ZHVI data. As was described with regard to the census-derived variables, rasterizing the census data ensured that each intersecting neighborhood polygon was represented proportionally in the mean values assigned to each city block polygon.

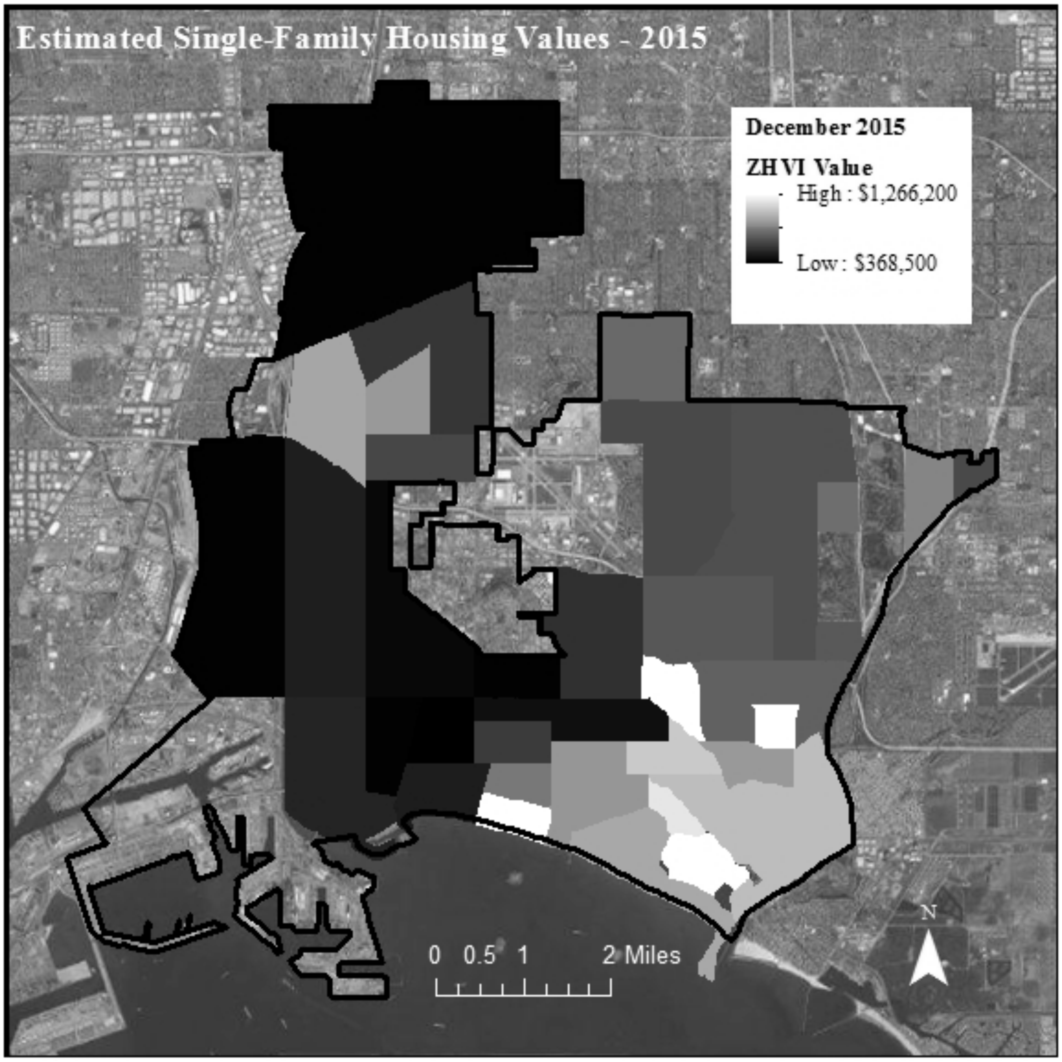


Figure 19. Median ZHVI values for single-family houses by neighborhood, December 2015.

A review of the final dataset found 90 city block polygons (out of 7,300) with null values for the census or ZHVI regression variables. The Zonal Statistics as Table tool stores null values when zone polygons (city blocks, in this case) have overlapping centroids or when the centroids of zone polygons fall beyond the edge of the raster. Of the 90 polygons with null data, 80 were found to be overlapping block polygons created by buffering tiny street segments that occur in complicated intersections, 2 were unpopulated bridges, and 8 were normal block polygons whose centroids fell outside the edge of the census raster layers. The 80 overlapping blocks and the 2 bridge blocks were removed from the dataset, leaving a new total block count of 7,218; census and ZHVI variable values for the remaining 8 normal blocks were manually populated using the field calculator.

3.2.7. Panel Dataset Creation

The next step was to prepare the wide-form data to be restructured to produce a balanced long-form panel dataset. *Wide-form* refers to a data structure in which each aggregation unit is represented by a single record with repeated variables storing values calculated at different time periods (e.g. first-quarter CCR, second-quarter CCR, etc.); repeating each rate variable 20 times results in a very wide table. *Long-form panel data*, by contrast, includes multiple records for each aggregation unit; each record represents a single aggregation unit at a single moment in time, and stores one instance of each variable; this structure results in a table that is long rather than wide. The term *balanced* indicates that every variable was calculated for every block at every time interval; in other words, each aggregation unit was revisited at each time interval in the study. When completed, the panel dataset would have 20 records (one for each quarter of the five years spanning 2011–2015) for each of the 7,218 city blocks and one column for each variable. Each of these records would represent a single block-quarter analysis unit. The 15 dependent and independent variables gathered or calculated for the statistical analyses are outlined in Table 5.

Table 5. Dependent and independent variables for regression analysis

Dependent Variable	#	Independent Variables	Implication or Purpose	Type	
				Time variant	Individual variant
Adjusted application rate (AAR) or Future mean adjusted application rate (4QMAAR) or Binary four-quarter application presence	1	Time-lagged cumulative completion rate (CCR)	Spatial contagion/ spillover effect	✔	✔
	2	Spatially-lagged adjacent-block cumulative completion rate (ABCCR)	Spatial contagion/ spillover effect	✔	✔
	3	L2G Rebate Rate	L2G incentive	✔	
	4	Mean parcel size in square feet	Proxy for income		✔
	5	Mean house size in square feet	Proxy for income		✔
	6	Owner occupancy percentage	Demographics		✔
	7	Family occupancy percentage	Demographics		✔
	8	Senior householder percentage	Demographics		✔
	9	Male householder percentage	Demographics		✔
	10	Female householder percentage	Demographics		✔
	11	Husband-wife householder percentage	Demographics		✔
	12	Mean median single-family home value (ZHVI)	Proxy for income	✔	✔

These variables included six varying regressors, or variables that change both over time and between individuals (AAR, 4QMAAR, binary four-quarter application presence, CCR, ABCCR, and ZHVI), eight time-invariant regressors, or variables that vary between individuals but not over time (owner occupancy, family occupancy, senior householder, husband-wife householder, male householder, female householder, mean parcel size, and mean house size), and one individual-invariant regressor, or variable that varies over time but not between individuals (L2G rebate). The time-invariant variables were not publicly available for each quarter or year of the study, so the same 2015 tax assessor values (mean parcel size, mean house size, mean yard size) and 2010 census data (owner occupation percentage, family percentage, senior householder percentage, and husband-wife householder percentage) were used for every year of the regression analysis. While it is possible that some blocks could have experienced a drastic shift in one or more of these values during the five-year study period relative to other blocks in the study area, it was expected that for most blocks, these data would serve as a valid means of characterizing blocks for comparison purposes since the age of the data is the same for all.

In order for SPSS to restructure the data from wide form to long form, each varying or time-variant variable must be repeated the same number of times. The AAR, 4QMAAR, binary four-quarter application presence, CCR, and ABCCR values had already been tabulated for every quarter, but because the ZHVI values were annual figures, only five ZHVI variables were present. Consequently, each annual ZHVI variable was duplicated three times for a total of four quarterly ZHVI variables for each year of the study, and the resulting 20 ZHVI variables were renamed to reflect quarters rather than years. Similarly, 20 L2G rebate variables were added to the table and populated with the rebate amount values corresponding to each quarter. The assessor-based and census-based variables did not have to be repeated for each quarter because they remained constant for each individual block throughout the five-year study period. The completed long-form dataset was then exported from ArcGIS as a .dbf file and imported into

SPSS, where it was restructured into a long-form panel dataset with 20 cases and 15 variables for each city block.

Next, CCR values were lagged by one quarter to enable comparison of each quarter's application rate or presence values with the previous quarter's CCR values. This was accomplished by populating a new variable (Lagged CCR) with the previous case's CCR value. Because there was no case prior to the first quarter, all first-quarter cases lacked that value and would be excluded from regression analyses.

Finally, four more quarterly variables were added to store log transformations of the positively skewed AAR, 4QMAAR, CCR and ABCCR values in order to bring them closer to a normal distribution. Histograms of the 4QMAAR and CCR values are shown in Figure 20 and Figure 21, with raw rates in the left panel and their more normally distributed natural logs in the right panel. Because log transformations cannot be applied to zero values, this calculation effectively excluded all zero value AAR, 4QMAAR, CCR, and ABCCR values from linear regression analyses. This exclusion was deemed acceptable because a two-stage statistical analysis was used for the study wherein all values (including zeros) would be analyzed using binary logistic regression and only positive values would be analyzed using linear regression. This scheme has been used by multiple studies, especially in the field of ecology (Pearce and Boyce 2005; Fletcher, MacKenzie, and Villouta 2005; Dobbie and Welsh 2001).

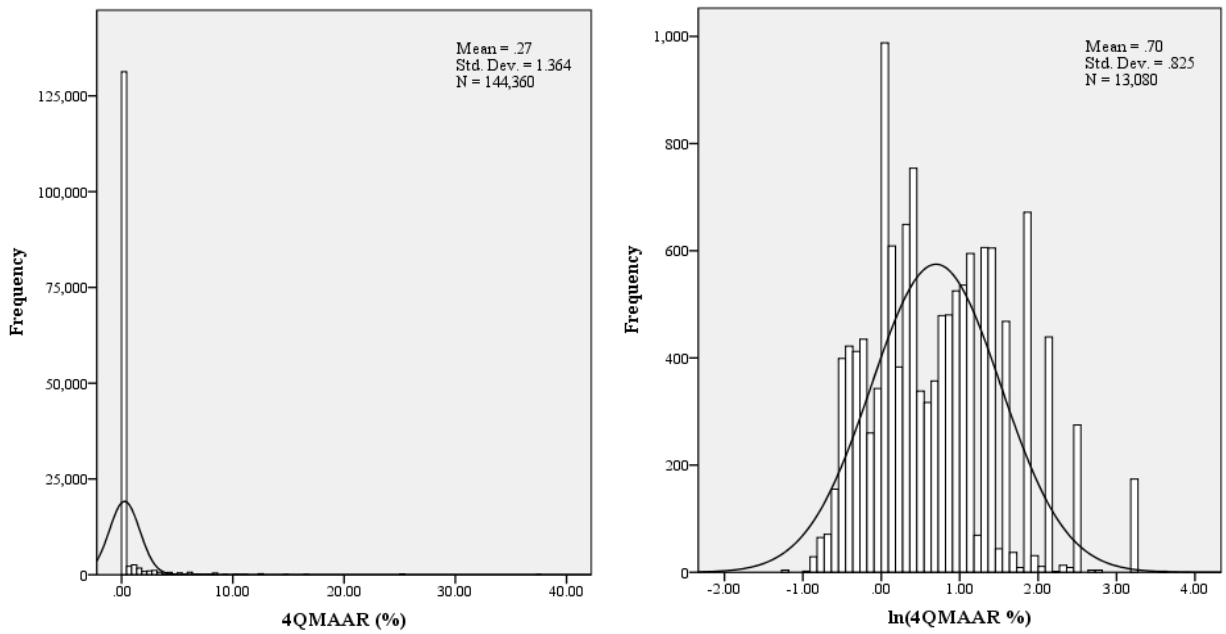


Figure 20. Histograms of raw 4QMAAR values and their natural logs

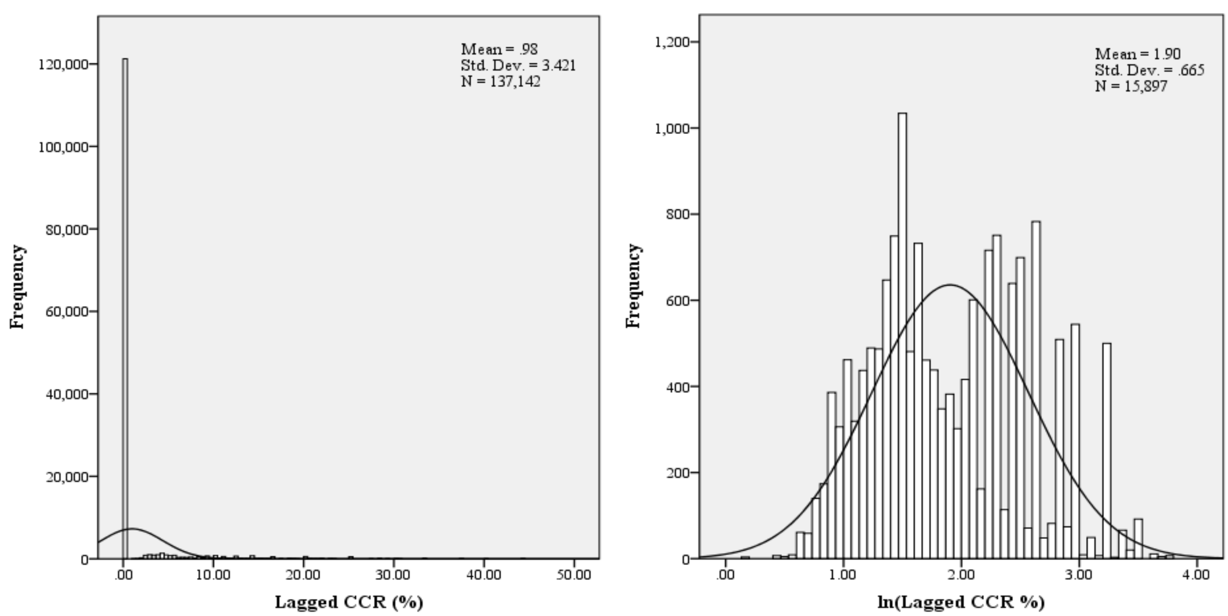


Figure 21. Histograms of raw CCR values and their natural logs.

Chapter 4 Statistical Analysis

The objective of this study was to answer three questions. First, is the four-quarter mean application rate dependent on the rate of previous completions on the same block and adjacent blocks? Second, does an increase in the rate of nearby L2G project completions increases the likelihood of new L2G application submissions in the subsequent four quarters on the same and adjacent blocks? And third, is there a completion rate threshold after which a block's four-quarter mean application rate markedly increases? This chapter describes the statistical analysis approach that was used to address these questions.

Because the total number of L2G applications and completions was quite small compared to the number of parcels multiplied by 20 quarters, the panel dataset was both positively skewed (meaning smaller values were dominant) and heavily zero-inflated. In fact, for the four L2G rate variables, zero values comprised between 64.1% and 97.6% of the dataset (Table 6). This non-normal distribution and abundance of zero values posed special challenges with regard to fitting regression models. The analytical approach described in the following sections draws from earlier treatments of species presence and abundance data (Pearce and Boyce 2005; Fletcher, MacKenzie, and Villouta 2005; Dobbie and Welsh 2001) and proportion data (Warton and Hui 2011, Chao et al. 2005) in the ecology domain, as those datasets exhibit similar distributions and pose similar challenges.

Table 6. Zero values in L2G rate data

Variable[1]	**Zero Values (%)**
Application Rate	97.6
4-Quarter Mean Application Rate	90.9
Cumulative Completion Rate	88.4
Adjacent-Block Cumulative Completion Rate	64.1

[1] Aggregated to block-quarter analysis units.

The analytical methodology employs a two-part regression strategy. First, a linear regression model was used to examine the relationship between positive 4QMAAR values and positive CCR and ABCCR values, excluding cases with zero values for any of these variables. Then, a separate binary logistic regression model was developed to incorporate those zero values into the analysis and determine whether the proportion of existing project completions on the same and adjacent blocks increases the likelihood of project participation on a subject block in the following four quarters. In addition to analyzing the dataset with and without zero values, these two regression models provide different insights: the linear regression model describes how completion rates affect the application *rate* in the future, while the logistic regression model describes how completion rates affect application *presence* in the future. In this way, participation presence and rate data are treated like species presence-absence and abundance data in the ecology literature (Fletcher, MacKenzie, and Villouta 2005; Pearce and Boyce 2005; Dobbie and Welsh 2001). Finally, an independent samples *t* test was performed to determine whether 4QMAAR values differ between blocks above and below the peak CCR and ABCCR thresholds identified by the logistical regression analysis. While common practice limits the use of an independent samples *t* test to normally distributed data, the size and completeness of this dataset justify the test's use, especially since excluding blocks where no applications were submitted would fail to provide a meaningful assessment of the effect of CCR threshold attainment on a block (Lumley et al. 2002; Sullivan and D'Agostino 1992).

The null hypotheses for each of the three analyses are as follows:

1. Four-quarter mean adjusted application rates do not show a statistically significant linear relationship with previous-quarter cumulative completion rates on the same and adjacent blocks

2. The likelihood of application presence within four quarters is not statistically higher at positive same-block and adjacent-block cumulative completion rate thresholds than at same-block and adjacent-block completion rates of zero.
3. There is no statistically significant difference between mean 4QMAAR values above and below peak same-block and adjacent-block completion rate thresholds.

4.1. Linear Regression

First, a linear regression model was fitted to examine the relationship between non-zero four-quarter mean application rates and previous-quarter non-zero completion rates (Figure 22). This analysis examines a subgroup of the dataset limited to block-quarters with positive application and completion rates. While this regression equation is of little real-world predictive value since it excludes block-quarters with application or completion rates of zero, it does provide insight about the degree to which residents' willingness to participate in the L2G program is influenced by the proportion of completed L2G projects on their blocks and adjacent blocks.

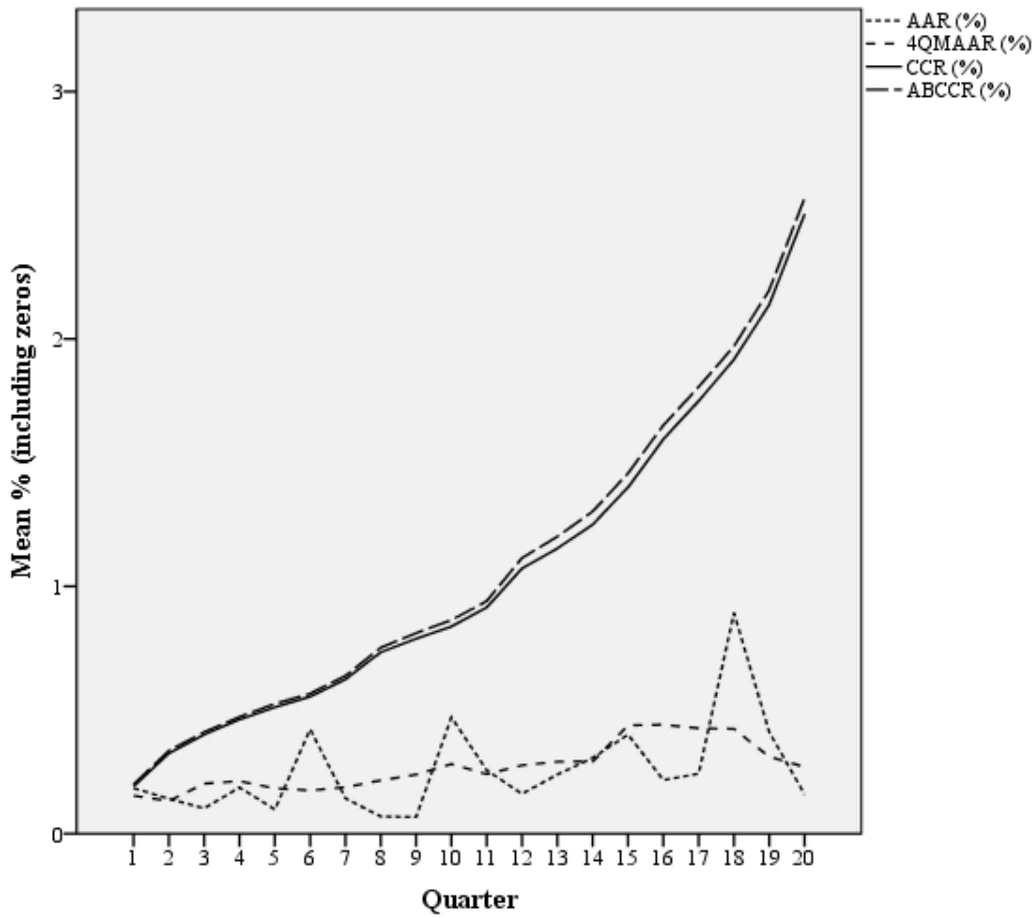


Figure 22. AAR, 4QMAAR, CCR, and ABCCR by quarter

Linear regression equations include coefficients for each variable that can be interpreted as the predicted increase in the dependent variable given a one-unit increase in a single independent variable when all others are held constant. To simplify the interpretation of coefficient results, the AAR, 4QMAAR, CCR, and ABCCR variables were converted to percentages by multiplying each by 100 prior to analysis. This conversion allowed changes to these variables to be described in terms of percentage points without altering relative outcomes. Similarly, ZHVI values were converted from $1 units to $10,000 units because the effect of a single dollar increase in property value is miniscule. Finally, as described in section 3.2.7, histograms of the quarterly participation rate values revealed that the data were strongly positively skewed (Figure 20 and Figure 21). Because linear regression is based on the assumption that input variables are normally distributed, a natural log transformation was applied to all the participation rate variables. Since this transformation is applied to every non-zero value in the dataset, relative proportions were unaffected. The exponentiation operation is the inverse of the natural log operation, so this was applied to predicted natural logs of 4QMAAR to reverse the transformation in order to interpret outcomes.

Two dependent variables, AAR and 4QMAAR, were tested along with all 13 explanatory variables—independently and together in a stepwise model—and only significant ($p < 0.05$) variables that improved model fit and did not exhibit multicollinearity were retained in the final model. The best fit was achieved by modeling the natural log of 4QMAAR with the natural log of CCR (lagged one quarter), owner occupancy percentage (O), ZHVI value (Z), and rebate amount (R).

Using the 4QMAAR variable instead of the AAR variable as the dependent variable in the linear regression equation seemed the logical choice, because as Zubek (2016) points out, effects of knowledge spillovers are not always immediately apparent. A resident influenced by a newly completed L2G project next door might not be ready to apply to the program for several months more; the 4QMAAR variable accounts for all applications submitted within a four-quarter temporal window while AAR only measures applications submitted in a single quarter. The 4QMAAR variable also ensures equivalence of seasonal influences for each block-quarter. The aptness of the 4QMAAR variable was confirmed by an improvement in model fit when AAR was replaced with 4QMAAR, and the bivariate correlation test results shown in Table 7 confirm that while AAR and 4QMAAR are both significantly correlated to the previous quarter's CCR value ($p = 0.000$), only 4QMAAR is significantly correlated to ABCCR.

Table 7. Correlations between L2G participation rate variables.

		ln(AAR%)	ln(4QMAAR%)	ln(Lagged CCR%)	ln(Lagged ABCCR%)
ln(AAR%)[1]	Pearson Correlation	1	0.928**	0.631**	0.040
	Sig. (2-tailed)		0.000	0.000	0.098
	N	3,504	3,504	1,030	1,694
ln(4QMAAR%)[2]	Pearson Correlation	0.928**	1	0.606**	0.070**
	Sig. (2-tailed)	0.000		0.000	0.000
	N	3,504	13,080	3,507	6,107
ln(Lagged CCR%)[3]	Pearson Correlation	0.631**	0.606**	1	0.069**
	Sig. (2-tailed)	0.000	0.000		0.000
	N	1,030	3,507	15,897	9,405
ln(Lagged ABCCR%)[4]	Pearson Correlation	0.040	0.070**	0.069**	1
	Sig. (2-tailed)	0.098	0.000	0.000	
	N	1,694	6,107	9,405	49,218

** Correlation is significant at the 0.01 level (2-tailed).
[1]Adjusted application rate
[2]4-quarter mean adjusted application rate
[3]Same-block cumulative completion rate
[4]Adjacent-block cumulative completion rate

The ABCCR variable was eliminated from the linear regression model despite being a variable of theoretical interest here. Table 7 illustrates that the correlation between 4QMAAR and ABCCR is significant, but quite weak. Consequently, the relationship between these two variables was significant in a linear model that included ABCCR as the only independent variable, but the coefficient was small (0.068, $p = 0.000$) and the adjusted R^2 was only 0.005, indicating that the model accounted for very little variation in the data. When CCR was added to the model, the coefficient of ABCCR decreased to 0.044 ($p = 0.005$) and the adjusted R^2 value jumped to 0.342, indicating that CCR was a much more important explanatory variable than ABCCR. When the owner occupancy rate variable was added to the equation, the effect of ABCCR became insignificant ($p = 0.263$) and the adjusted R^2 value increased to 0.508; therefore ABCCR was excluded from the final regression equation because exploratory linear regression analysis indicated that completion rates on adjacent blocks are not a significant predictor of subject-block application rates in the subsequent four quarters when controlling for same-block completion rate and owner occupancy.

Many of the remaining variables were eliminated because of multicollinearity, which was determined by assessing changes in coefficients and significance as variables were included or excluded from the model as well as by examining Pearson correlation scores in the Bivariate Correlation function in SPSS. 4QMAAR showed significant correlation with owner occupancy, family occupancy, husband-wife householder, male householder, female householder, and senior variables, but these independent variables exhibited varying degrees of multicollinearity when analyzed together. Since owner occupancy had the strongest correlation to 4QMAAR and its inclusion resulted in the largest increase in adjusted R^2 value, it was retained and the rest of the census variables were excluded. Similarly, ZHVI and house size showed evidence of multicollinearity; though neither variable appeared to be strongly correlated with 4QMAAR, ZHVI had the stronger correlation with 4QMAAR and improved the model slightly, so it was retained and house size was excluded. Parcel size was also excluded because it did not have a significant effect on the model when included alongside CCR, owner occupancy, and rebate amount (p = 0.121). Descriptive statistics are summarized for each of the final regression variables in Table 8; quarterly descriptive statistics for the same variables are listed in Appendix A.

Table 8. Descriptive statistics for final regression variables.

Statistic	AAR (%)	4-Quarter Mean AAR (%)	Lagged CCR (%)	Lagged ABCCR (%)[1]	Owner Occupancy (%)	ZHVI ($)
N	144360	144360	137142	137142	144360	144,360
Mean	.2580	0.2692	.9778	.995947	62.4233	496,719.7061
Median	.0000	0.0000	.0000	.000000	72.0577	451,200.0000
Minimum	.00	0.00	.00	.0000	.00	236,600.00
Maximum	100.00	37.50	44.44	22.7564	100.00	1,268,800.00
Std. Deviation	2.64275	1.36438	3.42142	1.8828952	28.73118	206,069.46481
Variance	6.984	1.862	11.706	3.545	825.481	42,464,624,327.809
Skewness	661.912	10.474	27.042	12.151	-1.046	2.185
Kurtosis	21.648	155.345	4.739	2.971	-.569	1.365

[1] ABCCR value is the mean of all adjacent-block CCR values, excluding the subject block.

4.2. Binary Logistic Regression

Second, a binary logistic regression model was used to determine the likelihood of a L2G application being submitted within four quarters at ten different categories of previous-quarter positive CCR and ABCCR threshold values, controlling for owner occupancy and rebate amount, in comparison to blocks with CCR or ABCCR values of zero. The completion-rate thresholds were established by using the SPSS Frequencies procedure to calculate ten percentile-based cut points for both non-zero CCR values and non-zero ABCCR values.

A binary logistic regression employs a dichotomous dependent variable representing future application presence or absence. This binary four-quarter application presence variable is based on the 4QMAAR variable; a 0 indicates no applications will be submitted within four quarters, while a 1 indicates that at least one application will be submitted within four quarters. This analysis incorporates nearly the entire dataset (137,142 cases) with the exception of first-quarter blocks (7,218 cases), as these have missing values for the lagged CCR and ABCCR values because there is no previous quarter value to pull that value from. Completion-rate thresholds with peak likelihoods of four-quarter application presence were used to formulate the independent samples *t* tests described in the following section.

Prior to fitting the equation, a cross-tabulation analysis was run to assess the relationship between the binary four-quarter application presence variable and the categorical same-block and adjacent-block completion rate dummy variables. The chi-square output tables showed that the p-value for the chi-square tests (Pearson chi-square, likelihood ratio, and linear by linear association tests) evaluating the relationship between four-quarter future application presence and CCR and ABCCR threshold categories were all statistically significant ($p = 0.000$). Table 9 contains the chi-square results of the cross-tabulation analysis. Because completion rates on both the same and adjacent blocks significantly improved the predictive value of the binary logistic regression model, both variables were used in the model. In addition, the model controlled for owner occupancy and rebate amount. The effect of ZHVI value was insignificant in the binary logistic regression model, so that variable was excluded from the model.

Table 9. Cross-tabulation results for lagged CCR threshold by four-quarter application presence and ABCCR threshold by four-quarter application presence

Chi-Square Tests: Lagged CCR Thresholds * 4-Quarter Application Presence			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4706.153	10	0.000
Likelihood Ratio	3389.616	10	0.000
Linear-by-Linear Association	1250.048	1	0.000
N of Valid Cases	137142		
Chi-Square Tests: Lagged ABCCR Thresholds * 4-Quarter Application Presence			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	1428.319	10	0.000
Likelihood Ratio	1267.483	10	0.000
Linear-by-Linear Association	1364.368	1	0.000
N of Valid Cases	137142		

4.3. Independent Samples *t* Tests

Independent samples *t* tests were performed to compare the means of 4QMAAR values above and below the peak CCR and ABCCR thresholds identified by the binary logistic regression analysis. The null hypothesis for the first *t* test was that no significant difference exists in mean 4QMAAR values for blocks with previous-quarter CCR values of zero and blocks with previous-quarter CCR values above zero (a cut point of 0.01 was used to define the groups as no CCR greater than zero and less than 0.01 exists). The null hypothesis for the second *t* test was that there is no significant difference between 4QMAAR values for blocks with ABCCR values above and below a cumulative completion rate threshold of 4.28%. For both tests, the 4QMAAR rates were expected to be higher for blocks that exceeded the critical CCR thresholds. Finally, to validate the decision to use the *t* test on skewed data, a non-parametric Independent Samples Kruskal-Wallis test was also performed for 4QMAAR across categories of lagged CCR thresholds.

Chapter 5 Results

The analytical portion of the study used linear regression to estimate a statistically significant positive relationship between a block's four-quarter non-zero application rate and its previous-quarter non-zero cumulative completion rate on the same block; the linear regression did not find a significant relationship between a block's four-quarter non-zero application rate and the rate of previous completions on adjacent blocks. The binary logistic regression model estimated the likelihood of future application submission on blocks with different cumulative completion rate thresholds on the same and adjacent blocks, including those with zero values; that model predicted a significant increase in likelihood of an application submission within four quarters when either same-block or adjacent-block completion rates exceeded zero, and peak same-block and adjacent-block completion thresholds were identified. Finally, an independent samples *t* test was used to confirm a significant difference between the four-quarter mean application rates of blocks above and below the peak same-block and adjacent-block completion rate thresholds identified by the logistic regression. Detailed results of each of the three analyses are presented in the sections that follow.

5.1. Linear Regression Results

Equation 5 is the linear regression equation calculated to estimate the log-transformed 4QMAAR (Y_{4QMAAR}) for a given block-quarter based on the previous quarter's non-zero same-block CCR ($X_{Lagged\ CCR}$), controlling for owner occupancy rate (X_O), ZHVI value (X_Z), and rebate amount (X_R). Adjacent-block completions are not factored into this equation because non-zero ABCCR did not significantly contribute to the accuracy of the linear regression model.

Equation 5. Linear regression equation

$$Ln(Y_{4QMAAR}) = -0.107 + 0.632(Ln(X_{Lagged\ CCR})) - 0.013(X_o) + 0.002(X_z) + 0.012(X_R)$$

This equation describes a significant relationship between the dependent and independent variables ($F(4, 3502) = 1{,}006.002$, $p = 0.000$) with an adjusted R^2 of 0.534. The coefficient of each independent variable represents the predicted change in ln(4QMAAR) for each one-unit increase in that variable. This means that on blocks with positive non-zero 4QMAAR values and positive non-zero lagged CCR values, ln(4QMAAR) increases 0.632 ($p = 0.000$) following each percentage point increase in CCR, controlling for rebate amount, owner occupancy rate, and home values. Examining the standardized beta coefficients, which represent the change in outcome for a one-standard-deviation increase in a variable (rather than a one-unit increase), allows for comparison of the relative effect of each of the variables and reveals that a positive same-block CCR has a considerably stronger effect than rebate amount on the four-quarter mean application rate. The null hypothesis—that four-quarter mean adjusted application rates would not have a statistically significant linear relationship with previous-quarter cumulative completion rates on the same and adjacent blocks—was partially rejected: 4QMAAR values were significantly related to same-block completion rates, but not to adjacent-block completion rates. Table 10 lists coefficients for each variable, and complete SPSS output from this analysis is included in Appendix B.

Table 10. Linear regression results

Dependent variable: ln(4QMAAR%)	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B	
	B	Std. Error	Beta			Lower Bound	Upper Bound
(Constant)	-0.107	0.073		-1.469	0.142	-0.249	0.36
ln(Lagged CCR %)	0.632	0.015	0.501	41.737	0.000	0.603	0.662
Owner Occupancy (%)	-0.013	0.000	-0.422	-35.020	0.000	-0.014	-0.013
ZHVI in units of $10,000	0.002	0.001	0.034	2.844	0.004	0.001	0.003
Rebate amount in units of $0.10	0.012	0.002	0.072	6.145	0.000	0.008	0.016

Table 11 illustrates the relative effects of each variable by predicting 4QMAAR values after a five-unit increase in each variable compared to a baseline example, holding the other variables constant. Because the linear regression equation predicts the natural log of the 4QMAAR value, the inverse of the natural log was calculated by applying the exponential function, resulting in a predicted 4QMAAR value expressed as a percentage. Increasing rebate amount by five $0.10 units, or $0.50 per square foot, elevates the predicted 4QMAAR value by about 0.17 percentage points. Increasing mean owner occupancy on a block by five percentage points has the opposite effect, lowering the predicted 4QMAAR by 0.17 percentage points, while increasing ZHVI values by five $10,000 units increases the predicted 4QMAAR by only 0.03 percentage points. The largest predicted gains in application rate are seen when the CCR value increases by five percentage points, the equivalent of one newly completed project on a block with 20 eligible parcels. That increase raises the predicted 4QMAAR value from 2.74% to 3.54%, an increase of 0.80 percentage points. In other words, the linear regression equation predicts that an increase of 5 percentage points in same-block completion rate raises a block’s expected application rate over the following four quarters 4.7 times more than a $0.50 hike in rebate amount would. However, it’s important to remember that this equation is based only on blocks with CCR and 4QMAAR values greater than zero.

Table 11. Linear regression model predictions. Each example shows the result of a 5-unit increase in each of the variable values compared to the baseline case.

Case	CCR	Owner Occupancy	ZHVI Value	Rebate Amount	Predicted 4QMAAR	Percentile
Baseline	10%	60%	$700,000	$2.50	2.74%	96.8
5-unit increase in CCR	**15%**	60%	$700,000	$2.50	3.54%	97.6
5-unit increase in Owner Occupancy	10%	**65%**	$700,000	$2.50	2.57%	96.8
5-unit increase in ZHVI	10%	60%	**$750,000**	$2.50	2.77%	96.8
5-unit increase in Rebate Amount	10%	60%	$700,000	**$3.00**	2.91%	97.2

[1]Percentiles refer to non-zero 4QMAAR rates.

5.2. Binary Logistic Regression Results

The binary logistic regression was performed to ascertain the effects of CCR (lagged one quarter and divided into ten percentile-based categories as well as a zero category), ABCCR (also lagged one quarter and divided into ten percentile-based categories and a zero category), and rebate amount (divided into three categories representing the three rebate amounts offered during the study period) on the likelihood of an application submission within four quarters, controlling for owner occupancy. The ZHVI value was excluded, because it did not significantly contribute to the predictive value of the model. This regression model takes into account data for block-quarters with application and completion rates of zero as well for those with positive participation rates. The logistic regression model was statistically significant, $\chi^2(23) = 5{,}661.471$, $p = 0.000$. Results are shown in Table 12.

Table 12. Binary logistic regression results

	B	S.E.	Wald	df	Sig.	Exp(B)	95% Confidence Interval for EXP(B)	
							Lower	Upper
Lagged CCR = 0%			2936.983	10	0.000			
Lagged CCR = 0.01-2.9%	1.760	0.054	1060.306	1	0.000	5.811	5.227	6.461
Lagged CCR = 2.91-3.7%	1.405	0.056	630.981	1	0.000	4.074	3.651	4.546
Lagged CCR = 3.71-4.4%	1.305	0.055	573.080	1	0.000	3.687	3.313	4.103
Lagged CCR = 4.41-5.0%	1.217	0.063	371.730	1	0.000	3.376	2.984	3.821
Lagged CCR = 5.01-6.3%	1.043	0.064	264.506	1	0.000	2.838	2.503	3.219
Lagged CCR = 6.31-8.3%	1.109	0.070	248.585	1	0.000	3.032	2.642	3.481
Lagged CCR = 8.31-10.0%	0.985	0.056	313.787	1	0.000	2.678	2.401	2.986
Lagged CCR = 10.01-12.5%	0.604	0.074	67.384	1	0.000	1.829	1.583	2.112
Lagged CCR = 12.51-16.7%	0.177	0.083	4.511	1	0.034	1.193	1.014	1.405
Lagged CCR > 16.7%	-0.558	0.113	24.549	1	0.000	0.572	0.459	0.714
Lagged ABCCR = 0%			542.375	10	0.000			
Lagged ABCCR > 0-.67%	0.137	0.050	7.399	1	0.007	1.147	1.039	1.266
Lagged ABCCR = .68-.87%	0.117	0.051	5.348	1	0.021	1.125	1.018	1.242
Lagged ABCCR = .88-1.11%	0.132	0.056	5.590	1	0.018	1.141	1.023	1.273
Lagged ABCCR = 1.12-1.52%	0.159	0.048	10.776	1	0.001	1.172	1.066	1.289
Lagged ABCCR = 1.53-1.98%	0.316	0.048	42.555	1	0.000	1.372	1.247	1.508
Lagged ABCCR = 1.99-2.50%	0.351	0.046	58.304	1	0.000	1.421	1.298	1.555
Lagged ABCCR = 2.51-3.33%	0.409	0.048	73.942	1	0.000	1.505	1.371	1.652
Lagged ABCCR = 3.34-4.28%	0.424	0.044	92.892	1	0.000	1.528	1.402	1.665
Lagged ABCCR = 4.29-6.10%	0.694	0.042	278.564	1	0.000	2.002	1.845	2.172
Lagged ABCCR > 6.10%	0.639	0.042	228.133	1	0.000	1.895	1.744	2.059
Rebate = $2.50			959.957	2	0.000			
Rebate = $3.00	0.301	0.024	161.674	1	0.000	1.351	1.290	1.415
Rebate = $3.50	0.713	0.023	959.899	1	0.000	2.040	1.950	2.134
Owner Occupancy (%)	0.009	0.000	595.255	1	0.000	1.009	1.008	1.010
Constant	-3.462	0.030	13210.339	1	0.000	0.031		

The values in the Exp(B) column represent odds ratios; in other words, the Exp(B) value for a categorical variable describes the likelihood of an application being submitted for a block within four quarters, compared with the zero category for that variable, when all other variables are held constant. Figure 23 shows the likelihood of an application submission within four quarters at each same-block CCR threshold category. An examination of these odds ratios indicates that block-quarters with CCR values in the first non-zero category (0.01–2.9%) are 5.811 times more likely to see another application submission within four quarters than block-quarters with completion rates of zero (p = 0.000), pointing to the existence of a temporal spillover effect that is strongest upon completion of the first projects on a block. The second-highest likelihood occurs at the second CCR category, when CCR is between 2.9% and 3.7% (Exp(B) = 4.074, p = 0.000); that trend generally continues as CCR thresholds increase, with the likelihood of an application submission within four quarters gradually diminishing, perhaps as the pool of willing participants is exhausted. Finally, when CCR exceeds 16.7%—a 90th percentile CCR rate—the likelihood of an additional application within four weeks dips below that of a block-quarter with no same-block completions at all.

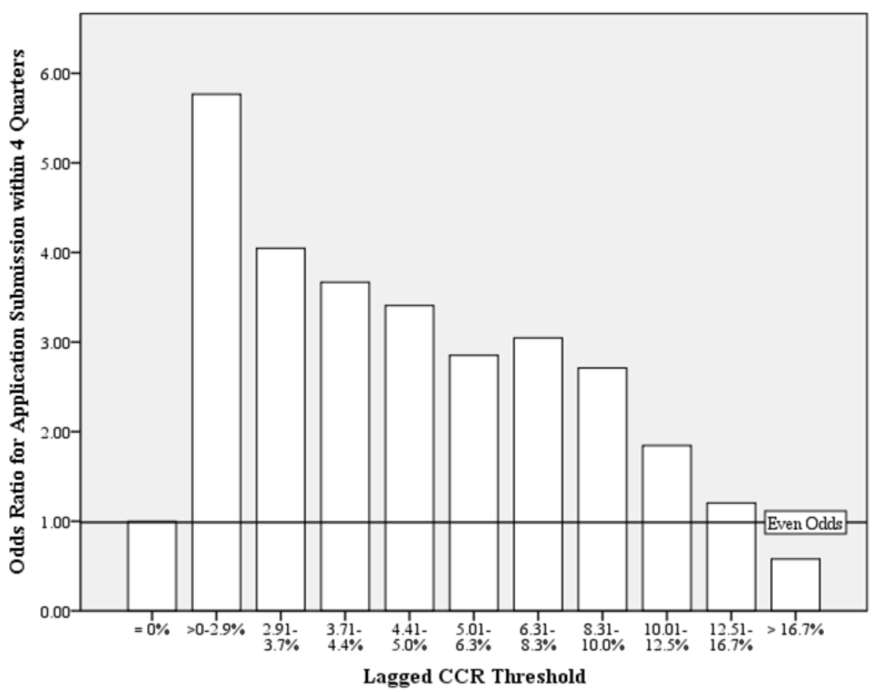


Figure 23. Odds ratios for future application submission relative to CCR = 0.

Figure 24 illustrates the effect of adjacent-block project completions on the likelihood of an application submission within four quarters. A comparison of these two figures reveals that while project completions on adjacent blocks significantly increase the likelihood of an application submission within four quarters, the effect is much weaker than the effect of completions on the same block. Interestingly, odds ratios increase as adjacent-block completion rates increase, while the opposite trend is apparent for same-block completions. The strongest effect of adjacent-block completions was seen at ABCCR values of 4.29% to 6.10%, at which point the likelihood of a future application submission is twice as likely as it is for an ABCCR value of zero (p = 0.000) when all other variables are held constant. No non-zero ABCCR categories resulted in a reduced likelihood of application submission within four quarters compared to an ABCCR value of zero; this result points to the existence of a small spatiotemporal spillover effect.

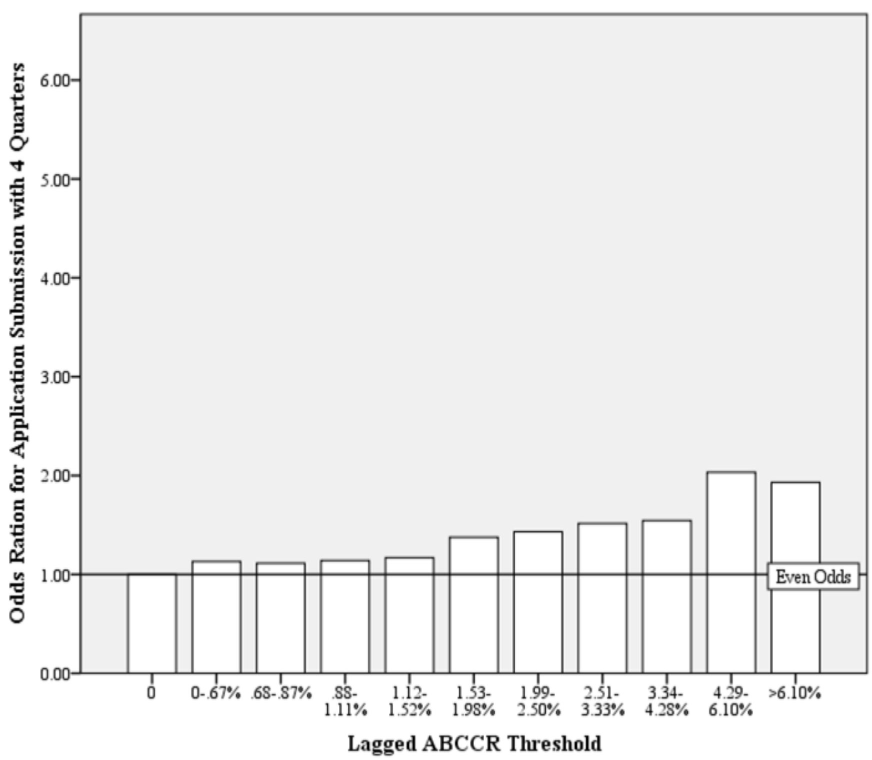


Figure 24. Odds ratios for future application submission relative to ABCCR = 0.

The binary logistic regression model also showed that when all other variables are held constant, a rebate value of $3.50 per square foot doubles the likelihood of an application submission on a block within four quarters compared to a $2.50 rebate ($p$ = 0.000). A rebate of $3.00 increases four-quarter application submission likelihood by 1.351 times (p = 0.000).

Like the linear regression model, the binary logistic regression model does not appear to be a reliable predictor of L2G project participation. In fact, it explained only 8.8% (Nagelkerke R^2) of the variance in application presence and lacked sensitivity, correctly classifying less than one percent of the blocks on which future applications were actually submitted, probably due to the overwhelming preponderance of zero values in the dataset. Despite the poor predictive sensitivity of the model, the null hypothesis—that the likelihood of application presence within four quarters is not statistically higher at positive same-block and adjacent-block cumulative completion rate thresholds than at same-block and adjacent-block completion rates of zero—was rejected. Complete SPSS output from this analysis is included in Appendix C.

5.3. Independent Samples *t*-Test Results

The t tests found that 4QMAAR values were significantly higher for blocks that had attained the peak threshold CCR and ABCCR values identified by the binary logistic regression model. The first independent samples t test showed a significant difference between mean 4QMAAR values on blocks with CCR values above zero ($\bar{x}$ = 0.490%, SD = 1.566%) and blocks with CCR values of zero ($\bar{x}$ = 0.247%, SD = 1.348%); $t(137,140)$ = 18.627, p = 0.000, equal variances not assumed. The second independent t test found that blocks with adjacent-block cumulative completion rates above 4.28% had statistically significantly higher mean 4QMAAR values ($\bar{x}$ = 0.597%, SD = 2.159%) than blocks with cumulative completion rates below 4.28% ($\bar{x}$ = 0.241%, SD = 1.264%); $t(137,140)$ = 17.426, p = 0.000, equal variances not assumed. Taken together, these findings suggest that attaining a critical same-block or adjacent-block CCR value is likely to increase L2G program participation on that block in the future. The null hypothesis—that there is no statistically significant difference between mean 4QMAAR values above and below peak same-block and adjacent-block completion rate thresholds—was rejected. Complete SPSS output from this analysis is included in Appendix D.

To validate the decision to use the *t* test on skewed data, a non-parametric Independent Samples Kruskal-Wallis test was also performed for 4QMAAR values across categories of lagged CCR thresholds. This test confirmed that 4QMAAR values are significantly different across both lagged CCR threshold categories ($H(10) = 4{,}313.117$, $p = 0.000$) and lagged ABCCR threshold categories ($H(10) = 1{,}484.021$, $p = 0.000$) for $N = 137{,}142$.

Chapter 6 Discussion

The results of all three statistical analyses indicate that cumulative completion rates exert a relatively strong temporal spillover effect on four-quarter application presence and rate on the same block; the binary logistic regression and the independent samples *t* test—both of which included participation rate values of zero—also indicate that a spatiotemporal spillover effect is exerted by adjacent-block completion rates. However, the linear regression analysis results suggest that once a block has a same-block completion rate greater than zero, the spatial spillover effect from adjacent blocks is insignificant. At that point, each new project completion on the same block increases both the likelihood of an additional application submission and the predicted rate of application submissions in the subsequent four quarters. To a lesser extent, each new completion also increases the odds of a new project application on any adjacent blocks with no previous completions; however, that new completion is unlikely to affect the application rate on adjacent blocks that already have positive cumulative completion rates of their own, perhaps because any potential participants on those blocks have already been influenced more strongly by their own same-block completion rate as suggested by the linear regression results. These findings support the earlier work of Hunter and Brown (2012), who found that residents were most likely to install innovative landscaping where a visually adjacent neighbor had already done the same.

Specifically, the logistic regression model predicted that blocks with cumulative completion rates between zero and 2.9% were 5.811 times more likely to see another L2G project application on the same block in the future compared to blocks with completion rates of zero. Blocks with a CCR value between 2.9% and 3.7% showed the second-highest likelihood of additional participation in the following four quarters with future applications 4.074 times more likely than on a block with no completions. CCR values from 3% to 16.7% also increased the likelihood of future applications, but to a gradually diminishing extent, and CCR values greater than 16.7% actually decreased the likelihood of additional application submissions on a block in the following four quarters. Project completions on adjacent blocks had a much less striking effect on four-quarter application presence; the peak likelihood of a new application being submitted within four quarters occurred when adjacent-block completion rates surpassed 4.9%, but even then, the likelihood of future application submission was only 2.002 times greater than for an adjacent-block completion rate of zero.

Similar to the binary logistic regression model, the linear regression model also showed that completion rates are strongly predictive of future participation on the same block. Unlike the binary logistic regression model, however, the linear regression model found that rates of project completions on adjacent blocks have no significant effect on a subject block's future L2G participation. It is important to remember, though, that the binary logistic regression model predicted the likelihood of application *presence* within four quarters while the linear regression model predicted application *rate* during that time. In addition, the binary logistic regression included 137,142 block-quarters (95% of the dataset, including a preponderance of zero values), while the linear regression only included the 3,502 block-quarters (2.4% of the dataset) with non-zero 4QMAAR and previous-quarter CCR values. Considered together, these results suggest that on blocks with no same-block completions, residents are more likely to submit their block's first application when adjacent blocks have higher rates of project completions; however, once that block has a completion rate greater than zero, the rate of future applications is more strongly influenced by same-block completion rate; completion rates on adjacent blocks do not appear to predict future application rates on a subject block. Figure 25 shows that future mean application rates reach their peak when 8.3% of households on the same block have completed L2G projects.

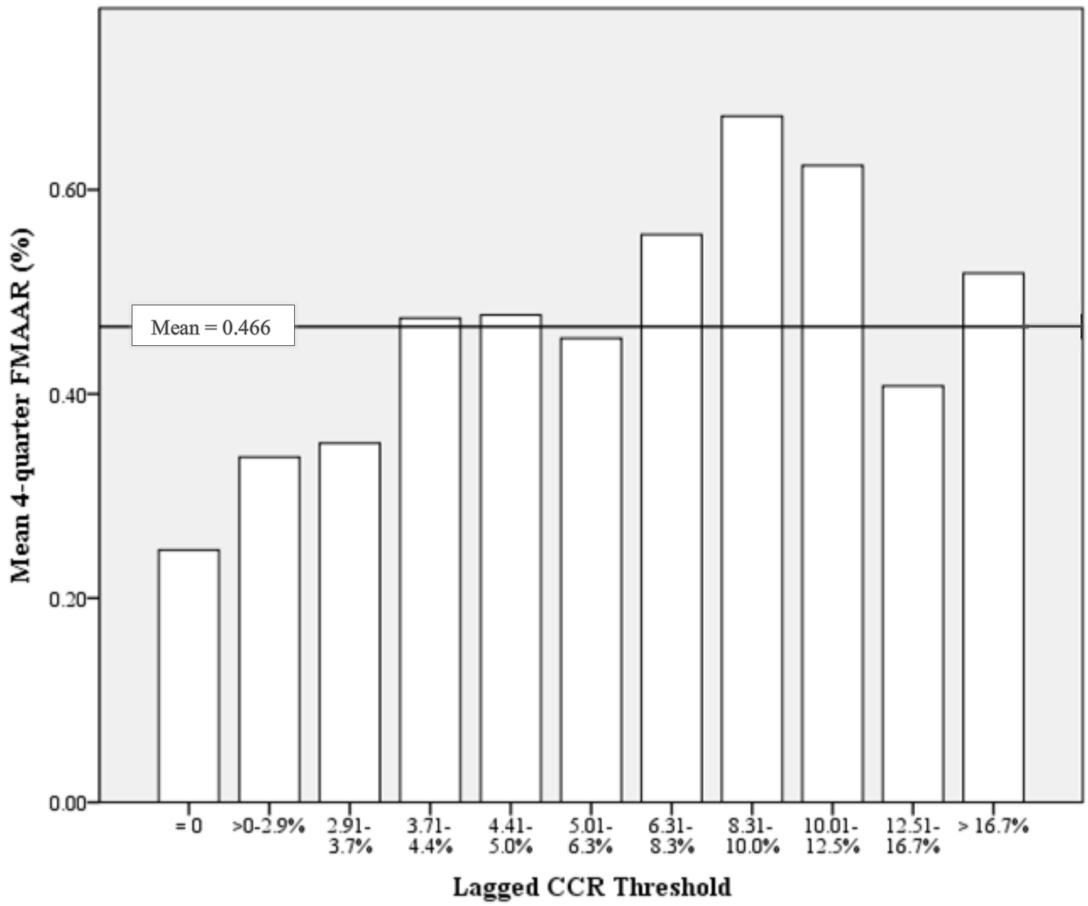


Figure 25. Mean future AAR values at CCR thresholds created from percentile break points. Peak 4QMAAR values occur when cumulative completions achieve a threshold of 8.3%.

The peak application rate seen at completion rates between 8.3% and 10% seems to be further evidence of the “minority influence” described by Xie et al. (2011, 6). Their research found that when approximately 10% of a social network commits to an opinion or innovation, the time until the rest of the network follows suit decreases dramatically. However, while Xie et al. documented an increased rate of adoption leading to a majority consensus, the current study shows a peak in participation rate at 8.3% followed by a generally decreasing rate of participation beginning at 10% completion (Figure 25). Furthermore, consensus—or even majority L2G participation on a block—is still rare; in the five-year study period, only about 10% of blocks achieved a completion rate greater than 16.7%.

This decrease in application rate and apparent barrier to full participation can be explained by the fact that the number of potential participants on a block is often far less than the number of parcels. The CCR value is calculated by dividing the number of completed projects by the total number of parcels, including multi-family, government, and commercial parcels. Many of these parcels do not have lawns to convert, and if they do, commercial property owners may not consider the rebate worth the effort of complying with program requirements. Even among single-family households, the pool of potential participants is probably less than 100%. Some households are ineligible because they do not have lawns to begin with (in the Belmont Shore neighborhood, for example, a courtyard is common), while others may dislike the aesthetic of a drought-tolerant garden more than they care about conserving water. Some residents may be unable to participate because of economic or personal circumstances, and others may have already converted their lawns to gardens without applying for a rebate. The presence of non-single-family households on a block combined with the unknown percentage of residents unlikely to participate may explain why even among the 1,783 blocks with at least one project completion, the mean completion rate at the end of the study period was only 10%.

Both regression models indicated that rebate amount exerts an influence on both four-quarter application presence and rate. While the rate of previous project completions on the same block was the strongest predictor of future participation in both models, the binary logistic regression indicated that increasing rebate amounts boost the odds of a first application submission on a block to a degree similar to that of an increase in adjacent-block completion rate. Specifically, when all other variables are held constant, the model predicts an odds ratio of

1.351 (p = 0.000) for a \$0.50 increase in rebate amount, which is similar to the odds ratio predicted for the fifth ABCCR threshold of 1.53% (1.372, p = 0.000); a \$1.00 increase in rebate amount resulted in an odds ratio of 2.040 (p = 0.000), which is close to the peak ABCCR threshold of 4.28% (2.002, p = 0.000). These findings suggest that on blocks where no projects have been completed on the same or adjacent blocks, an increase in rebate rate is an effective means of encouraging those first projects, which in turn will increase the likelihood of additional project applications on the same and adjacent blocks through the spatiotemporal spillover effect characterized by this study. Rebate values are shown chronologically in conjunction with mean quarterly block-level application rates in Figure 26. It is worth noting that the highest AAR value—by nearly double—occurred in the final quarter of the highest rebate period, immediately before the rebate dropped back to \$2.50 per square foot. The second highest AAR value occurred in the first quarter that rebates were increased to \$3.00 per square foot.

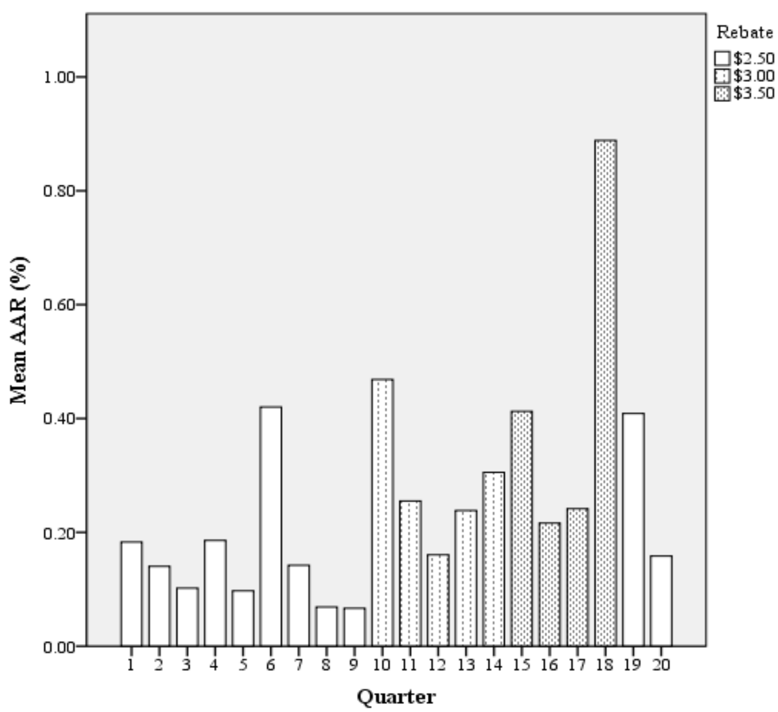


Figure 26. Mean AAR values at different rebate levels over time.

The effect of rebate values is shown in another way in Figure 27, where quarterly mean application rates are grouped by rebate category at different previous-quarter CCR values. Application rates are highest for previous quarter CCR values of 10–12.5% regardless of rebate category, but the magnitude of that peak is clearly greatest for the highest rebate category. In fact, application rates at every CCR category were highest when the top rebate category was in effect, suggesting—not surprisingly—that a combination of high rebates and high completion rates result in the highest rates of future participation.

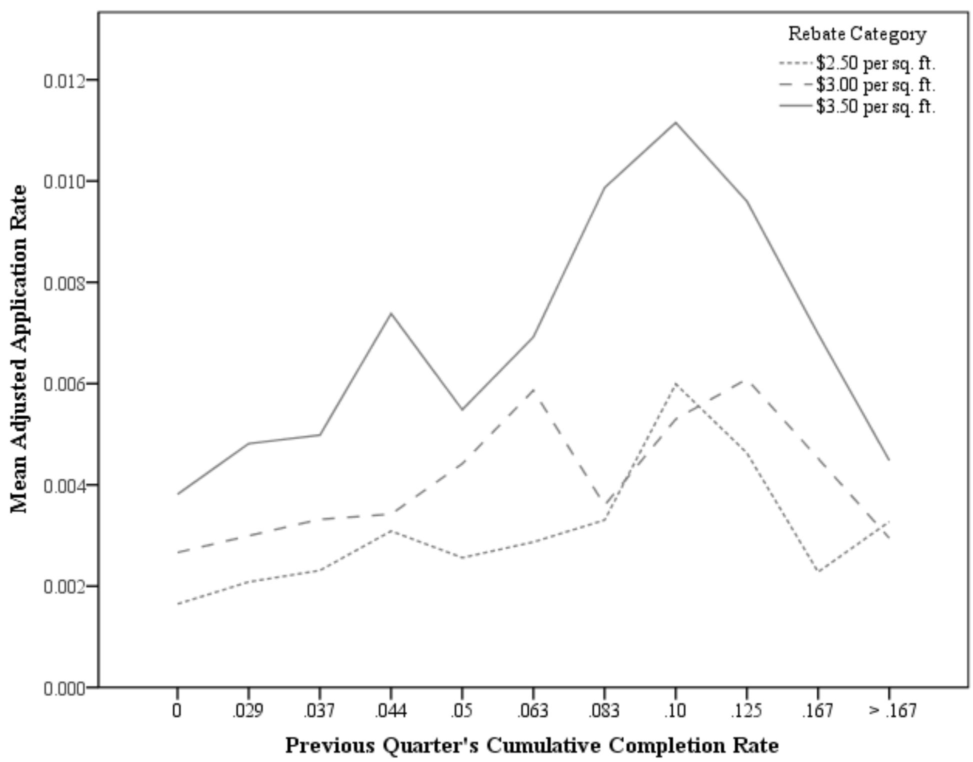


Figure 27. AAR values grouped by rebate amount. Mean application rates were highest on block-quarters with previous-quarter completion rates of about 10% and rebates of $3.50 per square foot.

6.1. Implications

These findings highlight the importance of encouraging the completion of a first project on every block; on most blocks, that first completion substantially increases the odds of future project applications on the same block. On an average Long Beach block (about 14 parcels), the first project completion represents a 7% completion rate and the second easily surpasses the 8.3% threshold at which application rates peak. Even on the longest block in the dataset, with 85 parcels, that peak 8.3% completion threshold is nearly met with only seven completed projects. Perhaps, since rebate amounts appear to be a motivator, local governments should consider increasing the rebate for the first project on a block. After that, the reduced financial incentive of a lower rebate would theoretically be supplemented by the more powerful spillover effect of the first completed projects. Alternatively, local governments could focus their marketing and support efforts on blocks with no previous project completions.

6.2. Limitations

Results of this study were limited by several factors. First, the dataset included only turfgrass replacement projects that were carried out through the L2G program. It is likely that many other Long Beach residents have converted their lawns to drought-tolerant landscaping outside of the L2G program or before its inception and that these projects contribute equally to neighborhood norms, but they are not included in the completion rates calculated for this study. It is also unknown if these out-of-program project completions occur at the same rate across Long Beach, or whether they are more likely to occur in some neighborhoods than others. Second, calculated AAR values also may have been lower than the true rate because every single-family residential parcel without a prior L2G project completion was considered eligible. In reality, some of those parcels may not have been eligible for the L2G program because they did not have a lawn. Third, it is possible that inclusion of other explanatory variables in regression models could have improved their predictive value and changed the outcome of the analysis. It is unknown whether educational attainment, political affiliation, or race and ethnicity play a role in L2G project participation in Long Beach. Fourth, the variables derived from census, tax assessor, and Zillow data were not collected for each quarter of the study. All census data were collected in 2010 and all tax assessor data were collected in 2015; Zillow data were acquired on an annual basis though Zillow does provide monthly values. It is likely that the accuracy of these variables decreases as time from collection date increases. Similarly, the census and Zillow data were derived from spatially different aggregation units and re-aggregated to the city block polygons created for this study. Though efforts were made to re-aggregate these data proportionally, the original aggregation units were probably not uniform, so the re-aggregation process likely resulted in some degradation in accuracy. Finally, decisions related to spatial and temporal aggregation and segmentation of data likely influenced the results of the analyses because of both the modifiable areal and temporal unit problems. Further study is needed to discover whether changes to temporal aggregation and segmentation of the L2G participation data would produce different analysis outcomes.

One final potential limitation of this study was the decision to use a mean ABCCR rather than a summed ABCCR as the spatially lagged independent variable. Both versions of the ABCCR variable were tested for use in this study. Exploratory regression models showed slightly better model fit for the summed version (adjusted $R^2 = 0.424$ versus adjusted $R^2 = 0.408$, $p = 0.000$ for both), and Pearson correlation tests showed a slightly higher correlation between 4QMAAR values and the sum ABCCR value (Pearson correlation value of 0.081 versus 0.075 for the mean ABCCR, $p = 0.000$ for both) though both correlations are extremely small. In the end, the mean ABCCR value was chosen because it allowed for more intuitive comparison with the effect of the same-block CCR variable since both were expressed as a percentage. Both methods are supported in the literature—Zubek and Henning (2016) and Beck, Gleditsch, and

Beardsley (2006) create spatially lagged variables by averaging neighboring values, while Grubesic and Rosso (2014) used summed values—but sums were initially preferred for this study because of their capacity to effectively represent not only the trend on surrounding blocks, but also how many blocks were exerting an influence. A typical block has six adjacent blocks (three at each end: left, straight, and right), but some blocks have as few as one adjacent block. It seems likely that a single adjacent block with a completion rate of 10% would exert less influence than six adjacent blocks with completion rates averaging 10%; summing the CCR values of all adjacent blocks rather than averaging them might better capture that expected difference in influence. Additional research is needed to determine if there is strong evidence for selecting one method of spatially lagging variables over the other.

6.3. Future Work

Further research is in order to better understand the variables that affect L2G program participation. An analysis of parcel-level application data taking into account the effect of the number of previously completed projects within a given radius of each parcel (perhaps 300 feet based on the incremental spatial autocorrelation analysis results obtained in this study) calculated for regular time intervals as well as parcel-specific data available in the assessor parcel database might provide additional insight. In addition, it would be informative to conduct a study to determine the duration of the temporal spillover effect of a new L2G project completion; that information could then be used to select a more appropriate temporal scale for future application rate and presence variables in similar studies in the future.

The question of whether L2G participation leads to increased turfgrass removal outside of the rebate program also deserves more study. It remains unknown whether the Long Beach L2G program is experiencing the same multiplicative effect as a similar program in Irvine, California, where early results of a house-to-house survey indicate that for every three households that participate in a turf replacement rebate program, four more replace their lawns without applying for a rebate (Johnson 2017; Knickmeyer 2016). Further study incorporating out-of-program turf replacements—perhaps through ground surveys or analysis of aerial imagery acquired at multiple time intervals—is needed to estimate the full extent of the L2G program's spillover effect and to measure the transformative effect of establishing an alternative neighborhood landscaping norm.

6.4. Conclusion

Previous and ongoing studies examine the roles of household characteristics and neighborhood demographics in turf replacement program participation, and this study is the first to measure spatiotemporal spillover as a driver of program participation. A two-stage analysis of five years of Long Beach L2G program participation data aggregated both spatially and temporally to block-quarter analysis units found that of all the explanatory variables considered, the rate of previous L2G project completions on a block is by far the most effective predictor of future participation on that block. In the absence of same-block completions, completed projects on adjacent blocks also exert an influence, though to a lesser degree. These findings suggest that by encouraging initial projects completions on non-participatory city blocks, local governments and water districts can catalyze a spatiotemporal spillover effect that will increase program participation on the same and adjacent blocks in the future.

References

Addink, Sylvan. 2005. "Cash for Grass" – A Cost-Effective Method to Conserve Landscape Water? A report by the Turfgrass Research Facility at University of California at Riverside. https://agops.ucr.edu/turf/topics/Cash-for-Grass.pdf [accessed September 19, 2016].

Andersen, Tom, Jacob Carstensen, Emilio Hernandez-Garcia, and Carlos M. Duarte. 2009. Ecological thresholds and regime shifts: approaches to identification. *Trends in Ecology & Evo*lution 24, no. 1: 49-57.

Anselin, Luc. 2002. Under the hood: Issues in the specification and interpretation of spatial regression models. *Agricultural Economics* 27: 247–267

———. 2003. Spatial externalities, spatial multipliers, and spatial econometrics. *International Regional Science Review* 26 (2): 153-66.

———. Thirty years of spatial econometrics. 2010. *Papers in Regional Science* 89, no. 1: 3-25.

Anselin, Luc, Elizabeth Griffiths, and George Tita. 2008. Crime mapping and hot spot analysis. In *Environmental Criminology and Crime Analysis*, pp. 97-116. Willan Publishing Cullompton, Devon.

Anselin, Luc, Julie Le Gallo, and Hubert Jayet. 2008. Spatial panel econometrics. In *The Econometrics of Panel Data*, pp. 625-660. Springer Berlin Heidelberg.

Atwater, Patrick, Christopher Tull, Eric Schmitt, Joone Lopez, Drew Atwater, and Varun Adibhatla. 2016. Transforming how water is managed in the West. Conference paper presented at Bloomberg Data for Good Exchange Conference, New York, NY, September 25.

Atwater, Patrick, Eric Schmitt, and Drew Atwater. 2015. Towards California water conservation impact evaluations by default: lessons of a turf removal rebate study in South Orange County. Conference paper presented at Bloomberg Data for Good Exchange Conference, New York, NY, September 28.

Bass, Frank M. 1969. A new product growth for model consumer durables. *Management Science* 15, no. 5: 215-227.

Beck, Nathaniel, Kristian Skrede Gleditsch, and Kyle Beardsley. 2006. Space is more than geography: Using spatial econometrics in the study of political economy. *International Studies Quarterly* 50, no. 1: 27-44.

Beumer, Carijn, and Pim Martens. 2016. BIMBY's first steps: a pilot study on the contribution of residential front-yards in Phoenix and Maastricht to biodiversity, ecosystem services and urban sustainability. *Urban Ecosystems* 19, no. 1: 45-76.

Bousema, Teun, Jamie T. Griffin, Robert W. Sauerwein, David L. Smith, Thomas S. Churcher, Willem Takken, Azra Ghani, Chris Drakeley, and Roly Gosling. 2012. Hitting hotspots: spatial targeting of malaria for control and elimination. *PLoS Med* 9, no. 1: e1001165.

Bruce, Andrew. 2014. Zillow Home Value Index: Methodology. *Zillow Real Estate Research.* https://www.zillow.com/research/zhvi-methodology-6032/ [accessed February 27, 2017].

Caetano, Gregorio, and Vikram Maheshri. 2013. Do Broken Windows Matter? Identifying Dynamic Spillovers in Criminal Behavior. Paper presented at the 5th Transatlantic Workshop on the Economics of Crime, Frankfurt, October 5. No. 2013-252-22.

Capello, Roberta. 2009. Spatial spillovers and regional growth: a cognitive approach. *European Planning Studies* 17, no. 5: 639-658.

Chainey, Spencer, Lisa Tompson, and Sebastian Uhlig. 2008. The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal* 21, no. 1-2: 4-28.

Chang, Heejun, G. Hossein Parandvash, and Vivek Shandas. 2010. Spatial variations of single-family residential water consumption in Portland, Oregon. *Urban Geography* 31, no. 7: 953-972.

Chao, Anne, Robin L. Chazdon, Robert K. Colwell, and Tsung‐Jen Shen. 2005. A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecology Letters* 8, no. 2: 148-159.

Cheng, Tao, and Monsuru Adepeju. 2014. Modifiable temporal unit problem (MTUP) and its effect on space-time cluster detection. *PloS one* 9, no. 6: e100465.

Chiang, Chi-Ting, Ie-Bin Lian, Che-Chun Su, Kuo-Yang Tsai, Yu-Pin Lin, and Tsun-Kuo Chang. 2010. Spatiotemporal trends in oral cancer mortality and potential risks associated with heavy metal content in Taiwan soil. *International Journal of Environmental Research and Public Health* 7, no. 11: 3916-3928.

City of Long Beach Department of Development Services, Office of Historic Preservation. 2009. City of Long Beach Historical Context Statement. http://www.lbds.info/civica/filebank/blobdload.asp?BlobID=3169 [accessed November 16, 2016].

Collier, Marcus J., Zorica Nedović-Budić, Jeroen Aerts, Stuart Connop, Dermot Foley, Karen Foley, Darryl Newport, Siobhán McQuaid, Aleksander Slaev, and Peter Verburg. 2013. Transitioning to resilience and sustainability in urban communities. *Cities* 32: S21-S28.

Crawley, Michael J. 2007. Proportion data. *Statistics: an introduction using R*: 247-262.

Cronk, Brian C. 2012. *How to use SPSS statistics: A step-by-step guide to analysis and interpretation*. Pyrczak Pub.

Cunningham, Ross B., and David B. Lindenmayer. 2005. Modeling count data of rare species: some statistical issues. *Ecology* 86, no. 5: 1135-1142.

de Jong, Rogier, and Sytze de Bruin. 2012. Linear trends in seasonal vegetation time series and the modifiable temporal unit problem. *Biogeosciences* 9, no. 1: 71.

Dobbie, Melissa J., and Alan H. Welsh. 2001. Theory & Methods: Modelling Correlated Zero-inflated Count Data. *Australian & New Zealand Journal of Statistics* 43, no. 4: 431-444.

Dormann, Carsten F. 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. *Global Ecology and Biogeography* 16, no. 2: 129-138.

Elhorst, J. Paul. 2010. Applied spatial econometrics: raising the bar. *Spatial Economic Analysis* 5, no. 1: 9-28.

———. 2012. Dynamic spatial panels: models, methods, and inferences. *Journal of Geographical Systems* 14, no. 1: 5-28.

———. 2014. Spatial panel data models. In *Spatial Econometrics*. Springer Berlin Heidelberg, 37-93.

Fletcher, David, Darryl MacKenzie, and Eduardo Villouta. 2005. Modelling skewed data with many zeros: a simple approach combining ordinary and logistic regression. *Environmental and Ecological Statistics* 12, no. 1: 45-54.

Fotheringham, A. Stewart, and Taylor M. Oshan. 2016. Geographically weighted regression and multicollinearity: dispelling the myth. *Journal of Geographical Systems* 18, no. 4: 303-329.

Fotheringham, A. Stewart, and David WS Wong. 1991. The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning* A 23, no. 7: 1025-1044.

Fowler, James H., and Nicholas A. Christakis. 2010. Cooperative behavior cascades in human social networks. *Proceedings of the National Academy of Sciences* 107, no. 12: 5334-5338.

Franczyk, Jon, and Heejun Chang. 2009. Spatial analysis of water use in Oregon, USA, 1985–2005. *Water Resources Management* 23, no. 4: 755-774.

Franzese Jr, Robert J., and Jude C. Hays. 2007. Spatial econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data. *Political Analysis*: 140-164.

Furth-Matzkin, Meirav, and Cass R. Sunstein. 2016. Social Influences on Policy Preferences: Conformity and Reactance. Preliminary draft available at SSNR: https://ssrn.com/abstract=2816595 [accessed January 24, 2017].

Granovetter, Mark, and Roland Soong. 1983. Threshold models of diffusion and collective behavior. *Journal of Mathematical Sociology* 9, no. 3: 165-179.

Green, Emily. 2010. The Dry Garden: Long Beach lawn rebates prove just how beautiful change can be. *Los Angeles Times*, October 8. http://latimesblogs.latimes.com/home_blog/2010/10/long-beach-drought-tolerant-gardens.html [accessed September 18, 2016].

Grimm, Nancy B., J. Morgan Grove, Steward TA Pickett, and Charles L. Redman. 2000. Integrated Approaches to Long-Term Studies of Urban Ecological Systems. *BioScience* 50, no. 7: 571-584.

Grove, Morgan, William R. Burch, Jr, and S. T. A. Pickett. 2005. Social mosaics and urban forestry in Baltimore, Maryland. In *Communities and Forests: Where People Meet the Land*. Eds. Lee, R.G. and D.R. Field. 250-274. Corvallis, OR: Oregon State University Press.

Grubesic, Tony H. and Andrea L. Rosso. 2014. The Use of Spatially Lagged Explanatory Variables for Modeling Neighborhood Amenities and Mobility in Older Adults. *Cityscape: A Journal of Policy Development and Research* 16, no. 2: 205–214.

Halper, Eve B., Christopher A. Scott, and Stephen R. Yool. Correlating Vegetation, Water Use, and Surface Temperature in a Semiarid City: A Multiscale Analysis of the Impacts of Irrigation by Single-Family Residences. *Geographical Analysis* 44, no. 3 (2012): 235-257.

Hanak, Ellen, Jeffrey Mount, Caitrin Chappelle, Jay Lund, Josué Medellín-Azuara, Peter Moyle, and Nathaniel Seavy. 2015. What If California's Drought Continues? Public Policy Institute of California. http://www.ppic.org/content/pubs/report/R_815EHR.pdf [accessed September 18, 2016].

Harrington, John W., Vu Q. Nguyen, James F. Paulson, Ruth Garland, Lawrence Pasquinelli, and Donald Lewis. 2010. Identifying the "tipping point" age for overweight pediatric patients. *Clinical Pediatrics* 49(7): 638–643.

Hayden, Lillian, Mary L. Cadenasso, Darren Haver, and Lorence R. Oki. 2015. Residential landscape aesthetics and water conservation best management practices: Homeowner perceptions and preferences. *Landscape and Urban Planning* 144: 1-9.

Helfand, Gloria E., Joon Sik Park, Joan I. Nassauer, and Sandra Kosek. 2006. The economics of native plants in residential landscape designs. *Landscape and Urban Planning* 78, no. 3: 229-240.

Hevesi, J.A., and Johnson, T.D. 2016. Estimating spatially and temporally varying recharge and runoff from precipitation and urban irrigation in the Los Angeles Basin, California: U.S. Geological Survey Scientific Investigations Report 2016–5068. http://dx.doi.org/10.3133/sir20165068 [accessed November 5, 2016].

Hill, Alison L., David G. Rand, Martin A. Nowak, and Nicholas A. Christakis. 2010. Infectious disease modeling of social contagion in networks. *PLoS Computational Biology* 6, no. 11: e1000968.

Hunter, Mary Carol R., and Daniel G. Brown. 2012. Spatial contagion: Gardening along the street in residential neighborhoods. *Landscape and Urban Planning* 105, no. 4: 407-416.

Hurd, Brian H., Rolston St. Hilaire, and John M. White. 2006. Residential Landscapes, Homeowner Attitudes, and Water-wise Choices in New Mexico. *HortTechnology* 16(2): 241–246.

Jackson, Donald A. 1997. Compositional data in community ecology: the paradigm or peril of proportions? *Ecology* 78, no. 3: 929-940.

Janmaat, Johannus John. 2013. Spatial patterns and policy implications for residential water use: An example using Kelowna, British Columbia. *Water Resources and Economics* 1: 3-19.

Johnson, Dane. 2017. Multiplier Effect Study for Turf Removal – 2016 Update. Talk presented at American Water Works Association California-Nevada Section Spring Conference. Anaheim, California. April 11.

Jongman, Robert H. G., Cajo J. F. Ter Braak, and Onno F. R. van Tongeren. 1995. *Data Analysis in Community and Landscape Ecology*. Cambridge University Press.

Joshi, Nayan Krishna. 2016. Local house prices and mental health. *International Journal of Health Economics and Management* 16, no. 1 (2016): 89-102.

Katz, Mitchell H. 2011. *Multivariable Analysis: A Practical Guide for Clinicians and Public Health Researchers.* Cambridge University Press.

Knickmeyer, Ellen. 2016. In California, a $350 million social experiment over lawns. Radio interview by Alex Cohen, *Take Two*, 89.3 KPCC, October 31, 2016. http://www.scpr.org/news/2016/10/31/65860/in-california-a-350-million-social-experiment-over/ [accessed January 26, 2017].

Koenig, Walter D. 1999. Spatial autocorrelation of ecological phenomena. *Trends in Ecology & Evolution* 14, no. 1: 22-26.

Landry, Shawn, and Ruiliang Pu. 2010. The impact of land development regulation on residential tree cover: An empirical evaluation using high-resolution IKONOS imagery. *Landscape and Urban Planning* 94.2: 94-104.

Larson, Kelli L., and Jaleila Brumand. Paradoxes in landscape management and water conservation: examining neighborhood norms and institutional forces. *Cities and the Environment (CATE)* 7, no. 1 (2014): 6.

Leys, Christophe, Christophe Ley, Olivier Klein, Philippe Bernard, Laurent Licata. 2013. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology* 49, no. 4:764–766.

Locke, Dexter H., and J. Morgan Grove. Doing the hard work where it's easiest? Examining the relationships between urban greening programs and social and ecological characteristics. *Applied Spatial Analysis and Policy* 9, no. 1 (2016): 77-96.

Lofquist, Daphne, Terry Lugaila, Martin O'Connell, and Sarah Feliz. 2012. Households and Families 2010. A U.S. Census Bureau Brief. https://www.census.gov/prod/cen2010/briefs/c2010br-14.pdf [accessed February 25, 2017].

Long Beach Technology and Innovation Department. Centerlines (shapefile). Posted April 2015. In the Long Beach GIS Data Catalog. http://www.longbeach.gov/ti/gis-maps-and-data/data-catalog/ [accessed October 11, 2016].

———. City Boundary (shapefile). Posted September 2016. In the Long Beach GIS Data Catalog. http://www.longbeach.gov/ti/gis-maps-and-data/data-catalog/ [accessed October 11, 2016].

Long Beach Water Department. Lawn-to-Garden Turf Replacement Program. City of Long Beach, California. http://www.lblawntogarden.com/gardens-open. Undated [accessed September 9, 2016].

———. Project Checklist. City of Long Beach, California. http://www.lblawntogarden.com/Project-Checklist. Undated [accessed September 19, 2016].

Long, J. Scott, and Jeremy Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata*. Stata press.

Los Angeles County Office of the Assessor. 2015. Assessor Parcels – 2015 Tax Roll (file geodatabase). On the Los Angeles County GIS Data Portal. http://egis3.lacounty.gov/dataportal/tag/gis-data/ [accessed November 5, 2016].

———. 2016. FAQ. Los Angeles County Assessor Portal. http://portal.assessor.lacounty.gov/faq [accessed November 5, 2016].

Lumley, Thomas, Paula Diehr, Scott Emerson, and Lu Chen. 2002. The importance of the normality assumption in large public health data sets. *Annual Review of Public Health* 23, no. 1: 151-169.

Machles, Maren. 2014. Long Beach Water Commissioners Increase Turf Removal Rebate. *Long Beach Post*, July 14. https://lbpost.com/news/2000003994-long-beach-water-commissioners-increase-turf-removal-rebate [accessed November 27, 2016].

McClain, Michael E., Elizabeth W. Boyer, C. Lisa Dent, Sarah E. Gergel, Nancy B. Grimm, Peter M. Groffman, Stephen C. Hart et al. 2003. Biogeochemical hot spots and hot moments at the interface of terrestrial and aquatic ecosystems. *Ecosystems* 6, no. 4: 301-312.

McClintock, Nathan, Dillon Mahmoudi, Michael Simpson, and Jacinto Pereira Santos. 2016. Socio-spatial differentiation in the Sustainable City: A mixed-methods assessment of residential gardens in metropolitan Portland, Oregon, USA. *Landscape and Urban Planning* 148: 1-16.

Metropolitan Water District of Southern California, 2013. California Friendly Turf Replacement Incentive Program Southern California, Final Project Report. Bureau of Reclamation website. http://www.usbr.gov/lc/socal/reports/MWDturfreduction.pdf [accessed September 17, 2016].

Mickey, Thomas J. 2013. *America's Romance with the English Garden*. Athens: Ohio University Press.

Minor, Emily, J. Amy Belaire, Amélie Davis, Magaly Franco, and Meimei Lin. 2016. Socioeconomics and neighbor mimicry drive yard and neighborhood vegetation patterns. *Urban Landscape Ecology: Science, Policy and Practice*: 56.

Montello, Daniel R. and Paul C. Sutton. 2013. *An Introduction to Scientific Research Methods in Geography & Environmental Studies.* London: Sage Publications, Ltd.

Mount, Jeffrey and Ellen Hanak. Water use in California. 2016. Public Policy Institute of California. http://www.ppic.org/main/publication_show.asp?i=1108 [accessed September 17, 2016].

Neel, Rebecca, Edward Sadalla, Anna Berlin, Susan Ledlow, and Samantha Neufeld. 2014. The social symbolism of water-conserving landscaping. *Journal of Environmental Psychology* 40: 49-56.

Nassauer, Joan Iverson, Zhifang Wang, and Erik Dayrell. 2009. What will the neighbors think? Cultural norms and ecological design. *Landscape and Urban Planning* 92, no. 3: 282-292.

NOAA National Centers for Environmental Information. 2016. State of the Climate: Drought for July 2016. National Oceanic and Atmospheric Association. http://www.ncdc.noaa.gov/sotc/drought/201607 [accessed September 18, 2016].

Noell, Nicholas. 2013. In Pictures: 2nd Annual Long Beach Lawn-to-Garden Tour. *Long Beach Post*, May 20. https://lbpost.com/life/2000002292-long-beach-lawn-to-garden-tour[accessed September 9, 2016].

Osorio, Javier. 2015. The Contagion of Drug Violence Spatiotemporal Dynamics of the Mexican War on Drugs. *Journal of Conflict Resolution* 59, no. 8: 1403-1432.

Pearce, Jennie L., and Mark S. Boyce. 2006. Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology* 43, no. 3: 405-412.

Pede, Valerien O., R. J. G. M. Florax, and Matthew T. Holt. 2008 Modeling non-linear spatial dynamics: A family of spatial STAR models and an application to US economic growth. Paper prepared for presentation at the American Agricultural Economics Association Annual Meeting, Orlando, Florida, July, pp. 27-29.

Pham, Thi-Thanh-Hien, Philippe Apparicio, Anne-Marie Séguin, Shawn Landry, and Martin Gagnon. 2012. Spatial distribution of vegetation in Montreal: An uneven distribution or environmental inequity? *Landscape and Urban Planning* 107, no. 3: 214-224.

Qian, Song S., and Thomas F. Cuffney. 2012. To threshold or not to threshold? That's the question. *Ecological Indicators* 15, no. 1: 1-9.

Rendell, Luke, Laurel Fogarty, William J. E. Hoppitt, Thomas J. H. Morgan, Mike M. Webster, and Kevin N. Laland. 2011. Cognitive culture: theoretical and empirical insights into social learning strategies. *Trends in Cognitive Sciences* 15, no. 2: 68-76.

Rosso, Andrea L., Tony H. Grubesic, Amy H. Auchincloss, Loni P. Tabb, and Yvonne L. Michael. 2013. Neighborhood amenities and mobility in older adults. *American Journal of Epidemiology* 78, no. 5: 761-769.

Roth, Robert E., Kevin S. Ross, Benjamin G. Finch, Wei Luo, and Alan M. MacEachren. 2013. Spatiotemporal crime analysis in US law enforcement agencies: Current practices and unmet needs. *Government Information Quarterly* 30, no. 3: 226-240.

Ryan, Sandra E., and Laurie S. Porth. 2007. A tutorial on the piecewise regression approach applied to bedload transport data. U.S. Forest Service Technical Report. https://www.fs.fed.us/rm/pubs/rmrs_gtr189.pdf [accessed January 29, 2017].

Scott, Lauren M. and Mark V. Janikas. 2009. Spatial Statistics in ArcGIS. In *Handbook of Applied Spatial Analysis*, eds. Fischer, M.M. and A. Getis, 27–41. Heidelberg, Germany: Springer Berlin Heidelberg.

Seapy, Briana. 2015. Turf Removal & Replacement: Lessons Learned. A report of the California Urban Water Conservation Council. http://cuwcc.org/Resources/Publications-and-Reports [accessed September 22, 2016].

Seyranian, Vivian, Gale M. Sinatra, and Morgan S. Polikoff. 2014. Comparing communication strategies for reducing residential water consumption. *Journal of Environmental Psychology* 41: 81–90.

Silvestrini, Andrea, and David Veredas. 2008. Temporal aggregation of univariate and multivariate time series models: a survey. *Journal of Economic Surveys* 22, no. 3: 458-497.

Smith, Keeley. 2015. LBWD Continues Lawn-to-Garden Program, Despite End of Metropolitan Turf Removal Subsidies. *Long Beach Post*, July 9. https://lbpost.com/news/2000006554-lbwd-to-continue-lawn-to-garden-program-in-smaller-capacity-after-metropolitan-discontinues-turf-removal-subsidies [accessed September 17, 2015].

Solano, Emilio J. and Garo Megerdichian. 2009. *Standards and Procedures for Assessor's Maps.* White paper produced by the Los Angeles County Assessor's Office. http://www.calmapping.org/counties/losangeles/losangeles-mapping-manual.pdf [accessed 23 February, 2017].

Sovocool, Kent A., Mitchell Morgan, and Doug Bennett. 2006. An In-depth Investigation of Xeriscape as a Water Conservation Measure. *Journal (American Water Works Association)* 98, no. 2: 82-93.

Stevens, Matt and Monte Morin. 2015. Southland water district ends popular lawn-removal rebate program. *Los Angeles Times*, July 10. http://www.latimes.com/local/california/la-me-mwd-turf-rebates-20150710-story.html [accessed September 18, 2016].

Sugumaran, Ramanathan, Scott R. Larson, and John P. DeGroote. 2009. Spatio-temporal cluster analysis of county-based human West Nile virus incidence in the continental United States. *International Journal of Health Geographics* 8.1: 43.

Sullivan, L. M., and R. B. d'Agostino. 1992. Robustness of the t-test applied to data distorted from normality by floor effects. *Journal of Dental Research* 71, no. 12: 1938-1943.

Taylor, R. B. 1997. Social order and disorder of street blocks and neighborhoods: Ecology, microecology, and the systemic model of social disorganization. *Journal of Research in Crime and Delinquency*, no. 34: 113–155.

Tobler, Waldo R. 1970. A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46: 234–40

Tull, C., Schmitt, E., Atwater, P. 2016. How Much Water Does Turf Removal Save? Applying Bayesian Structural Time-Series to California Residential Water Demand. Knowledge Discovery and Data Mining 2016 Workshop "Data Science for Food, Energy and Water" (DS-FEW).

UCLA: Statistical Consulting Group. 2017. Logistic Regression SPSS Annotated Output. UCLA Institute for Digital Research and Education. http://stats.idre.ucla.edu/spss/output/logistic-regression [accessed March 12, 2017].

———. 2017. SPSS Annotated Output Regression Analysis. UCLA Institute for Digital Research and Education. http://stats.idre.ucla.edu/spss/output/regression-analysis [accessed March 12, 2017].

United States Census Bureau. American Fact Finder. https://factfinder.census.gov/ [accessed February 2, 2017].

———. Geography: TIGER/Line® Shapefiles and TIGER/Line® Files. https://www.census.gov/geo/maps-data/data/tiger-line.html [accessed February 2, 2017].

———. 2014. *Utilizing Data from American FactFinder with TIGER/Line Shapefiles in ArcGIS.* Instructional document. https://www2.census.gov/geo/pdfs/education/tiger/AFF_TIGERLine_Joining_Presentation.pdf [accessed February 2, 2017].

———. 2015. United States Quick Facts. https://www.census.gov/quickfacts/ [accessed November 12, 2016].

Uren, Hannah V., Peta L. Dzidic, and Brian J. Bishop. Exploring social and cultural norms to promote ecologically sensitive residential garden design. *Landscape and Urban Planning* 137 (2015): 76-84.

Valente, Thomas W. 1996. Social network thresholds in the diffusion of innovations. *Social Networks* 18, no. 1: 69-89.

Warton, David I., and Francis KC Hui. 2011. The arcsine is asinine: the analysis of proportions in ecology. *Ecology* 92, no. 1: 3-10.

Weisburd, David, Elizabeth R. Groff, and Sue-Ming Yang. 2014. Understanding and controlling hot spots of crime: The importance of formal and informal social controls. *Prevention Science* 15, no. 1: 31-43.

Wicker, A. W. 1987. Behavior settings reconsidered: Temporal stages, resources, internal dynamics, context. In *Handbook of Environmental Psychology*, eds. D. Stokels & I. Altman, 613–653.

Williams, Lauren. 2014. Long Beach offering residents incentive for converting lawns. *Orange County Register*, July 21. http://www.ocregister.com/articles/water-629428-long-beach.html [accessed September 18, 2016].

Wilson, Kenneth, and Ian CW Hardy. 2002. *Statistical Analysis of Sex Ratios: An Introduction*. 48-92.

Xie, Jierui, Sameet Sreenivasan, Gyorgy Korniss, Weituo Zhang, Chjan Lim, and Boleslaw K. Szymanski. 2011. Social consensus through the influence of committed minorities. *Physical Review E* 84, no. 1: 011130.

Zhang, Dapeng, and Xiaokun Wang. 2016. Investigating the dynamic spillover effects of low-cost airlines on airport airfare through spatio-temporal regression models. *Networks and Spatial Economics* 16 (3): 821-36.

Zillow. How Accurate is the Zestimate? 2017. https://www.zillow.com/zestimate/#acc [accessed February 27, 2017].

———. Median Home Value – Zillow Home Value Index (ZHVI). https://www.zillow.com/research/data/#bulk [accessed February 1, 2017].

Zmyslony, Jean, and Daniel Gagnon. 1998 Residential management of urban front-yard landscape: A random process? *Landscape and Urban Planning* 40, no. 4: 295-307.

Zubek, Nana, and Christian HCA Henning. 2016. Local Government, Spatial Spillovers and the Absorption of EU Structural Funds. *Journal of Agricultural Economics* 67, no. 2: 368-397.

Appendix A: Variable Statistics by Quarter

Statistics by Year and Quarter		AAR (%)	4-Quarter Mean AAR (%)	Lagged CCR (%)	Mean lagged ABCCR %	Owner Occupancy	ZHVI
Y1Q1	N	7218	7218			7218	7218
	Mean	.1831	0.1529			62.4233	415947.0503
	Median	.0000	0.0000			72.0577	376300.0000
	Minimum	.00	0.00			.00	236600.00
	Maximum	100.00	37.50			100.00	1118300.00
	Std. Deviation	2.01990	1.08843			28.73307	184578.20662
	Variance	4.080	1.185			825.589	34069114357.454
	Skewness	975.472	16.746			-1.046	3.278
	Kurtosis	25.260	401.606			-.569	1.625
Y1Q2	N	7218	7218	7218	7218	7218	7218
	Mean	.1404	0.1315	.1892	.194555	62.4233	415947.0503
	Median	.0000	0.0000	.0000	.000000	72.0577	376300.0000
	Minimum	.00	0.00	.00	.0000	.00	236600.00
	Maximum	100.00	37.50	25.00	12.5000	100.00	1118300.00
	Std. Deviation	2.00246	1.00819	1.36298	.6750849	28.73307	184578.20662
	Variance	4.010	1.016	1.858	.456	825.589	34069114357.454
	Skewness	1064.803	18.401	127.886	45.917	-1.046	3.278
	Kurtosis	27.712	490.021	10.230	5.695	-.569	1.625
Y1Q3	N	7218	7218	7218	7218	7218	7218
	Mean	.1021	0.2014	.3234	.330265	62.4233	415947.0503
	Median	.0000	0.0000	.0000	.000000	72.0577	376300.0000
	Minimum	.00	0.00	.00	.0000	.00	236600.00
	Maximum	100.00	25.00	33.33	12.5000	100.00	1118300.00
	Std. Deviation	1.71755	1.29809	1.81684	.8836372	28.73307	184578.20662
	Variance	2.950	1.685	3.301	.781	825.589	34069114357.454
	Skewness	1817.727	12.913	80.448	22.783	-1.046	3.278
	Kurtosis	36.947	215.754	8.000	4.124	-.569	1.625
Y1Q4	N	7218	7218	7218	7218	7218	7218
	Mean	.1860	0.2114	.3988	.405024	62.4233	415947.0503
	Median	.0000	0.0000	.0000	.000000	72.0577	376300.0000
	Minimum	.00	0.00	.00	.0000	.00	236600.00
	Maximum	100.00	25.00	33.33	12.5000	100.00	1118300.00
	Std. Deviation	2.54214	1.34357	2.05252	1.0106670	28.73307	184578.20662
	Variance	6.462	1.805	4.213	1.021	825.589	34069114357.454
	Skewness	1025.660	12.614	63.888	19.114	-1.046	3.278
	Kurtosis	28.452	204.251	7.194	3.789	-.569	1.625
Y2Q1	N	7218	7218	7218	7218	7218	7218
	Mean	.0973	0.1821	.4595	.465671	62.4233	441482.4931
	Median	.0000	0.0000	.0000	.000000	72.0577	398900.0000
	Minimum	.00	0.00	.00	.0000	.00	257000.00
	Maximum	33.33	25.00	33.33	14.0741	100.00	1088300.00
	Std. Deviation	1.16244	1.23793	2.22809	1.0996245	28.73307	185932.97929
	Variance	1.351	1.532	4.964	1.209	825.589	34571072786.681
	Skewness	356.605	13.561	57.881	17.821	-1.046	2.049
	Kurtosis	17.116	238.009	6.833	3.639	-.569	1.391

Statistics by Year and Quarter		AAR (%)	4-Quarter Mean AAR (%)	Lagged CCR (%)	Mean lagged ABCCR %	Owner Occupancy	ZHVI
Y2Q2	N	7218	7218	7218	7218	7218	7218
	Mean	.4201	0.1744	.5092	.517128	62.4233	441482.4931
	Median	.0000	0.0000	.0000	.000000	72.0577	398900.0000
	Minimum	.00	0.00	.00	.0000	.00	257000.00
	Maximum	100.00	25.00	33.33	14.0741	100.00	1088300.00
	Std. Deviation	4.00385	1.23906	2.36385	1.1749722	28.73307	185932.97929
	Variance	16.031	1.535	5.588	1.381	825.589	34571072786.681
	Skewness	393.582	13.663	50.334	16.142	-1.046	2.049
	Kurtosis	17.758	238.698	6.424	3.467	-.569	1.391
Y2Q3	N	7218	7218	7218	7218	7218	7218
	Mean	.1421	0.1865	.5520	.558831	62.4233	441482.4931
	Median	.0000	0.0000	.0000	.000000	72.0577	398900.0000
	Minimum	.00	0.00	.00	.0000	.00	257000.00
	Maximum	100.00	25.00	33.33	14.0741	100.00	1088300.00
	Std. Deviation	2.22516	1.24209	2.44424	1.2129939	28.73307	185932.97929
	Variance	4.951	1.543	5.974	1.471	825.589	34571072786.681
	Skewness	1241.215	13.476	45.597	14.530	-1.046	2.049
	Kurtosis	31.417	235.309	6.109	3.295	-.569	1.391
Y2Q4	N	7218	7218	7218	7218	7218	7218
	Mean	.0689	0.2148	.6230	.631475	62.4233	441482.4931
	Median	.0000	0.0000	.0000	.000000	72.0577	398900.0000
	Minimum	.00	0.00	.00	.0000	.00	257000.00
	Maximum	100.00	25.00	33.33	14.0741	100.00	1088300.00
	Std. Deviation	1.49929	1.28283	2.62126	1.3262073	28.73307	185932.97929
	Variance	2.248	1.646	6.871	1.759	825.589	34571072786.681
	Skewness	2832.788	12.417	40.635	13.601	-1.046	2.049
	Kurtosis	46.665	206.527	5.791	3.211	-.569	1.391
Y3Q1	N	7218	7218	7218	7218	7218	7218
	Mean	.0667	0.2377	.7319	.743065	62.4233	516949.0999
	Median	.0000	0.0000	.0000	.000000	72.0577	468000.0000
	Minimum	.00	0.00	.00	.0000	.00	329800.00
	Maximum	33.33	25.00	33.33	14.0741	100.00	1260200.00
	Std. Deviation	1.05795	1.31560	2.87047	1.4878376	28.73307	200353.67641
	Variance	1.119	1.731	8.240	2.214	825.589	40141595650.280
	Skewness	538.051	11.244	34.798	11.378	-1.046	2.771
	Kurtosis	21.757	173.910	5.374	3.004	-.569	1.518
Y3Q2	N	7218	7218	7218	7218	7218	7218
	Mean	.4686	0.2806	.7861	.799726	62.4233	516949.0999
	Median	.0000	0.0000	.0000	.000000	72.0577	468000.0000
	Minimum	.00	0.00	.00	.0000	.00	329800.00
	Maximum	100.00	25.00	33.33	14.0741	100.00	1260200.00
	Std. Deviation	4.05279	1.37833	2.98110	1.5412877	28.73307	200353.67641
	Variance	16.425	1.900	8.887	2.376	825.589	40141595650.280
	Skewness	375.399	10.106	31.631	9.913	-1.046	2.771
	Kurtosis	17.209	145.139	5.144	2.826	-.569	1.518
Y3Q3	N	7218	7218	7218	7218	7218	7218
	Mean	.2550	0.2398	.8353	.850729	62.4233	516949.0999
	Median	.0000	0.0000	.0000	.000000	72.0577	468000.0000
	Minimum	.00	0.00	.00	.0000	.00	329800.00
	Maximum	100.00	25.00	33.33	14.0741	100.00	1260200.00
	Std. Deviation	2.59498	1.23253	3.09884	1.5971099	28.73307	200353.67641
	Variance	6.734	1.519	9.603	2.551	825.589	40141595650.280
	Skewness	688.302	10.620	30.737	8.849	-1.046	2.771
	Kurtosis	21.948	163.422	5.058	2.704	-.569	1.518

Statistics by Year and Quarter		AAR (%)	4-Quarter Mean AAR (%)	Lagged CCR (%)	Mean lagged ABCCR %	Owner Occupancy	ZHVI
Y3Q4	N	7218	7218	7218	7218	7218	7218
	Mean	.1605	0.2792	.9122	.927079	62.4233	516949.0999
	Median	.0000	0.0000	.0000	.000000	72.0577	468000.0000
	Minimum	.00	0.00	.00	.0000	.00	329800.00
	Maximum	50.00	25.00	33.33	14.0741	100.00	1260200.00
	Std. Deviation	1.82053	1.34851	3.21473	1.6582010	28.73307	200353.67641
	Variance	3.314	1.818	10.334	2.750	825.589	40141595650.280
	Skewness	354.124	9.713	27.233	8.143	-1.046	2.771
	Kurtosis	17.003	134.627	4.768	2.583	-.569	1.518
Y4Q1	N	7218	7218	7218	7218	7218	7218
	Mean	.2384	0.2931	1.0696	1.093161	62.4233	536663.3988
	Median	.0000	0.0000	.0000	.000000	72.0577	489300.0000
	Minimum	.00	0.00	.00	.0000	.00	344300.00
	Maximum	50.00	25.00	37.50	16.6667	100.00	1268800.00
	Std. Deviation	1.99535	1.38510	3.51299	1.8327755	28.73307	205009.67501
	Variance	3.981	1.918	12.341	3.359	825.589	42028966849.198
	Skewness	235.518	9.737	24.281	7.883	-1.046	2.338
	Kurtosis	13.266	134.379	4.486	2.492	-.569	1.445
Y4Q2	N	7218	7218	7218	7218	7218	7218
	Mean	.3054	0.2939	1.1514	1.177877	62.4233	536663.3988
	Median	.0000	0.0000	.0000	.000000	72.0577	489300.0000
	Minimum	.00	0.00	.00	.0000	.00	344300.00
	Maximum	100.00	25.00	44.44	20.0000	100.00	1268800.00
	Std. Deviation	3.18696	1.45354	3.66395	1.9455260	28.73307	205009.67501
	Variance	10.157	2.113	13.424	3.785	825.589	42028966849.198
	Skewness	578.913	9.905	23.411	8.964	-1.046	2.338
	Kurtosis	21.224	133.907	4.368	2.563	-.569	1.445
Y4Q3	N	7218	7218	7218	7218	7218	7218
	Mean	.4126	0.4396	1.2475	1.277484	62.4233	536663.3988
	Median	.0000	0.0000	.0000	.366300	72.0577	489300.0000
	Minimum	.00	0.00	.00	.0000	.00	344300.00
	Maximum	100.00	25.00	44.44	20.0000	100.00	1268800.00
	Std. Deviation	3.37851	1.60128	3.82415	2.0448920	28.73307	205009.67501
	Variance	11.414	2.564	14.624	4.182	825.589	42028966849.198
	Skewness	377.368	7.191	21.069	8.669	-1.046	2.338
	Kurtosis	16.427	77.374	4.169	2.499	-.569	1.445
Y4Q4	N	7218	7218	7218	7218	7218	7218
	Mean	.2161	0.4387	1.3972	1.429590	62.4233	536663.3988
	Median	.0000	0.0000	.0000	.537634	72.0577	489300.0000
	Minimum	.00	0.00	.00	.0000	.00	344300.00
	Maximum	100.00	25.00	44.44	21.5980	100.00	1268800.00
	Std. Deviation	2.21346	1.56447	4.08510	2.2073616	28.73307	205009.67501
	Variance	4.899	2.448	16.688	4.872	825.589	42028966849.198
	Skewness	729.790	7.060	18.762	8.618	-1.046	2.338
	Kurtosis	22.066	75.983	3.951	2.458	-.569	1.445
Y5Q1	N	7218	7218	7218	7218	7218	7218
	Mean	.2414	0.4243	1.5939	1.625923	62.4233	572556.4886
	Median	.0000	0.0000	.0000	.649351	72.0577	533100.0000
	Minimum	.00	0.00	.00	.0000	.00	368500.00
	Maximum	100.00	25.00	44.44	22.2390	100.00	1266200.00
	Std. Deviation	2.70184	1.56655	4.38125	2.3962856	28.73307	210253.66400
	Variance	7.300	2.454	19.195	5.742	825.589	44206603226.695
	Skewness	633.983	7.362	15.386	7.755	-1.046	1.677
	Kurtosis	21.791	82.063	3.620	2.331	-.569	1.352

Statistics by Year and Quarter		AAR (%)	4-Quarter Mean AAR (%)	Lagged CCR (%)	Mean lagged ABCCR %	Homeowners	ZHVI
Y5Q2	N	7218	7218	7218	7218	7218	7218
	Mean	.8884	0.4229	1.7476	1.780415	62.4233	572556.4886
	Median	.0000	0.0000	.0000	.769231	72.0577	533100.0000
	Minimum	.00	0.00	.00	.0000	.00	368500.00
	Maximum	100.00	25.00	44.44	22.2390	100.00	1266200.00
	Std. Deviation	4.22934	1.57332	4.60051	2.5289422	28.73307	210253.66400
	Variance	17.887	2.475	21.165	6.396	825.589	44206603226.695
	Skewness	127.789	7.571	13.951	6.641	-1.046	1.677
	Kurtosis	9.044	87.068	3.458	2.195	-.569	1.352
Y5Q3	N	7218	7218	7218	7218	7218	7218
	Mean	.4088	0.3102	1.9145	1.945730	62.4233	572556.4886
	Median	.0000	0.0000	.0000	.869565	72.0577	533100.0000
	Minimum	.00	0.00	.00	.0000	.00	368500.00
	Maximum	100.00	25.00	44.44	22.2390	100.00	1266200.00
	Std. Deviation	2.91519	1.46965	4.82989	2.6588166	28.73307	210253.66400
	Variance	8.498	2.160	23.328	7.069	825.589	44206603226.695
	Skewness	429.872	9.781	12.279	5.408	-1.046	1.677
	Kurtosis	16.333	133.797	3.265	2.024	-.569	1.352
Y5Q4	N	7218	7218	7218	7218	7218	7218
	Mean	.1584	0.2692	2.1355	2.169262	62.4233	572556.4886
	Median	.0000	0.0000	.0000	1.052632	72.0577	533100.0000
	Minimum	.00	0.00	.00	.0000	.00	368500.00
	Maximum	100.00	25.00	44.44	22.7564	100.00	1266200.00
	Std. Deviation	2.27764	1.41793	5.10558	2.8280353	28.73307	210253.66400
	Variance	5.188	2.011	26.067	7.998	825.589	44206603226.695
	Skewness	1122.416	10.359	10.459	4.315	-1.046	1.677
	Kurtosis	29.468	145.577	3.045	1.853	-.569	1.352
Total	N	144360	144360	137142	137142	144360	144360
	Mean	.2580	0.2692	.9778	.995947	62.4233	496719.7061
	Median	.0000	0.0000	.0000	.000000	72.0577	451200.0000
	Minimum	.00	0.00	.00	.0000	.00	236600.00
	Maximum	100.00	37.50	44.44	22.7564	100.00	1268800.00
	Std. Deviation	2.64275	1.36438	3.42142	1.8828952	28.73118	206069.46481
	Variance	6.984	1.862	11.706	3.545	825.481	42464624327.809
	Skewness	661.912	10.474	27.042	12.151	-1.046	2.185
	Kurtosis	21.648	155.345	4.739	2.971	-.569	1.365

Appendix B: SPSS Linear Regression Output

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Natural log of 4QMAAR (%) Natural log of lagged CCR (%) Owner Occupancy (%) ZHVI in units of $10,000 Rebate in units of $0.10[b]	.	Enter

a. Dependent Variable: Natural log of 4QMAAR

b. All requested variables entered.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.731[a]	.535	.534	.50230

a. Predictors: (Constant), Rebate10 in units of $0.10, ln(Lagged CCR%), ZHVI in units of $10,000, Owner Occupancy (%)

b. Dependent Variable: ln(Future Mean AAR)

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1015.281	4	253.820	1006.002	0.000[b]
	Residual	883.575	3502	.252		
	Total	1898.857	3506			

a. Dependent Variable: Natural log of 4QMAAR (%)

b. Predictors: (Constant), Rebate10 in units of $0.10, ln(lagged CCR%), ZHVI in units of $10,000, Owner Occupancy (%)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-.107	.073		-1.469	.142
	ln(Lagged CCR %)	.632	.015	.501	41.737	.000
	Owner Occupancy (%)	-. 013	.000	-.422	-35.020	.000
	ZHVI in units of $10,000	.002	.001	.034	2.844	.004
	Rebate10	.012	.002	.072	6.145	.000

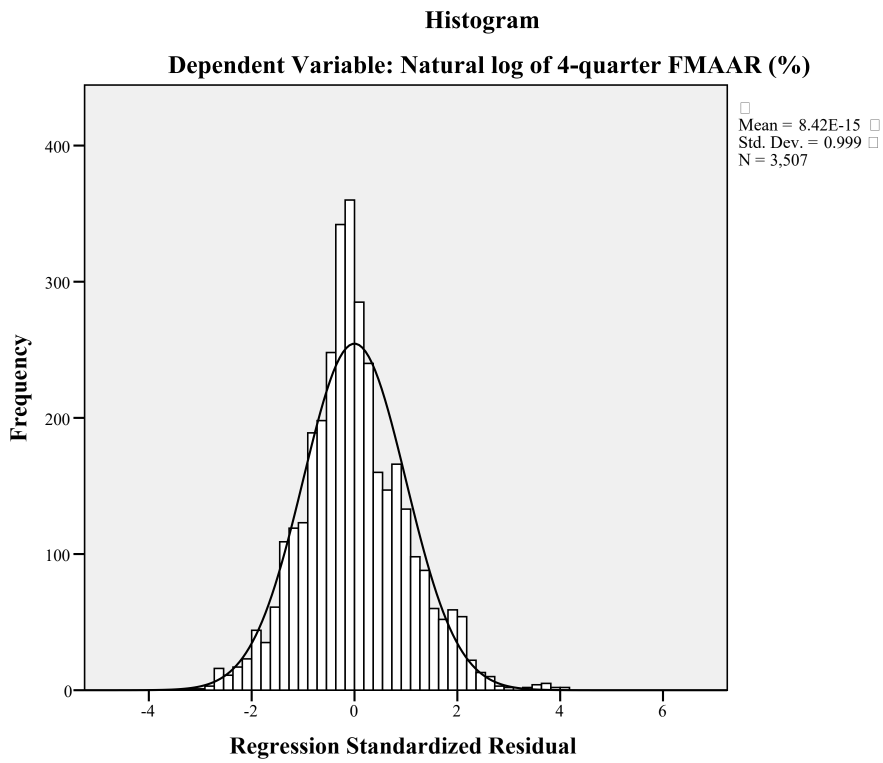
a. Dependent Variable: Natural log of 4QMAAR (%)

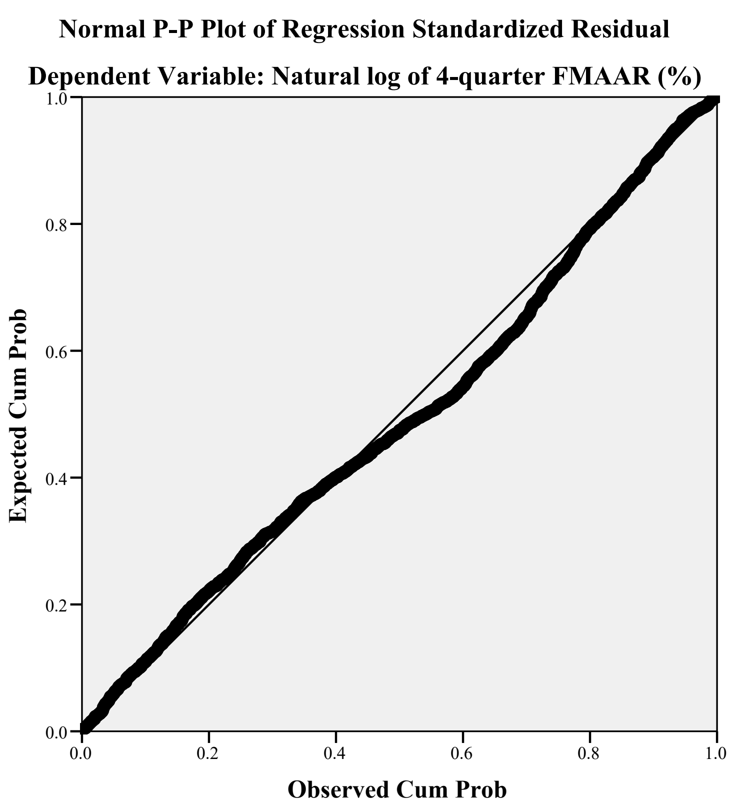
Residuals Statistics[a]

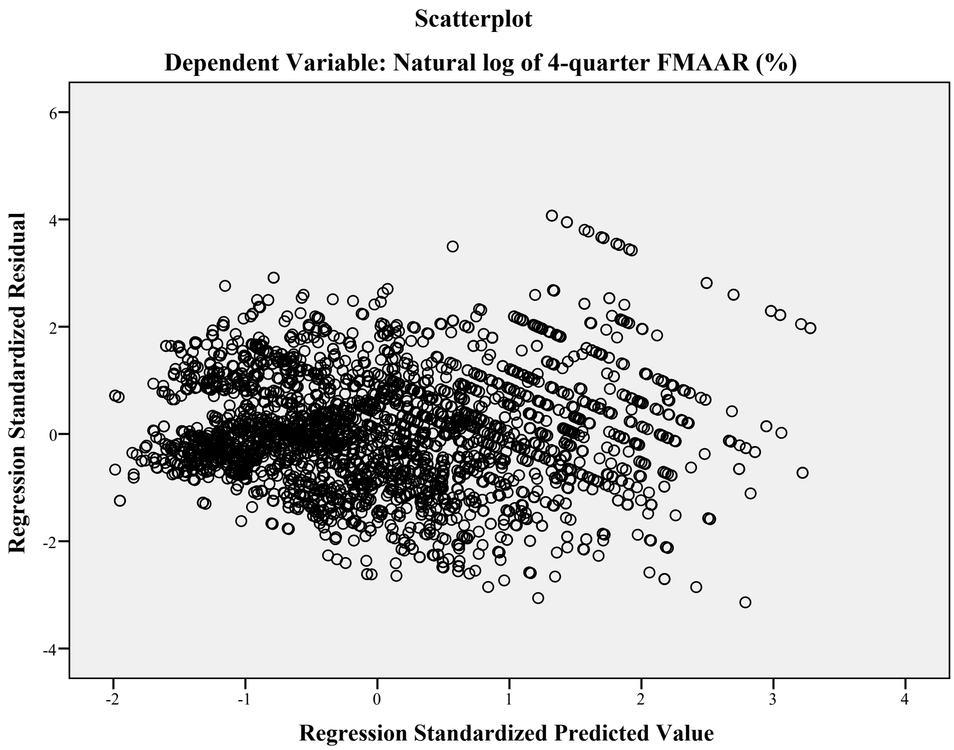
	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-.6062	2.2276	.4630	.53813	3507
Residual	-1.57714	2.504476	.00000	.50201	3507
Std. Predicted Value	-1.987	3.279	.000	1.000	3507
Std. Residual	-3.140	4.071	.000	.999	3507

a. Dependent Variable: Natural log of 4QMAAR (%)

Charts


Histogram
Dependent Variable: Natural log of 4-quarter FMAAR (%)
Mean = 8.42E-15
Std. Dev. = 0.999
N = 3,507
400
300
200
100
0
Frequency
-4
-2
0
2
4
6
Regression Standardized Residual


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: Natural log of 4-quarter FMAAR (%)
1.0
0.8
0.6
0.4
0.2
0.0
Expected Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0
Observed Cum Prob


Scatterplot
Dependent Variable: Natural log of 4-quarter FMAAR (%)
Regression Standardized Residual
6
4
2
0
-2
-4
-2
-1
0
1
2
3
4
Regression Standardized Predicted Value

Appendix C: SPSS Binary Logistic Regression Output

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	137142	95.0
	Missing Cases	7218	5.0
	Total	144360	100.0
Unselected Cases		0	0
Total		144360	144360

a. If weight is in effect, see classification table for the total number of cases.

Dependent Variable Encoding

Original Value	Internal Value
0	0
1	1

Categorical Variables Codings

			Parameter coding									
		Freq.	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Lagged Mean ABCCR Thresholds (subject block excluded)	= 0	87924	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	> 0-.67%	5075	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	.68-.87%	5069	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	.88-1.11%	4226	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
	1.12-1.52%	5435	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1.53-1.98%	4802	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
	1.99-2.50%	5199	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000
	2.51-3.33%	4448	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
	3.34-4.28%	5125	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	4.29-6.10%	4921	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
	> 6.10%	4918	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
Lagged CCR Threshold	= 0%	121245	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	> 0-2.9%	1613	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	2.91-3.7%	1728	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	3.71-4.4%	1902	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000
	4.41-5.0%	1478	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	5.01-6.3%	1503	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000
	6.31-8.3%	1199	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000
	8.31-10.0%	2136	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
	10.01-12.5%	1458	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000
	12.51-16.7%	1492	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000
	> 16.7%	1388	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000
Rebate (in units of $0.10)	25	72180	0.000	0.000								
	30	36090	1.000	0.000								
	35	28872	0.000	1.000								

Block 0: Beginning Block

Classification Table[a,b]

Observed			Predicted: 4-Quarter Application Presence 0	Predicted: 4-Quarter Application Presence 1	Percentage Correct
Step 0	4-Quarter Application Presence variable (1 = presence, 0 = absence)	0	124446	0	100.0
		1	12696	0	0.0
	Overall Percentage				90.7

a. Constant is included in the model.

b. The cut value is .500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-2.283	0.009	60024.874	1	0.000	0.102

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	LagCCRThreshold	4706.153	10	0.000
		LagCCRThreshold(1)	1437.090	1	0.000
		LagCCRThreshold(2)	723.515	1	0.000
		LagCCRThreshold(3)	690.826	1	0.000
		LagCCRThreshold(4)	398.281	1	0.000
		LagCCRThreshold(5)	319.865	1	0.000
		LagCCRThreshold(6)	306.742	1	0.000
		LagCCRThreshold(7)	365.983	1	0.000
		LagCCRThreshold(8)	85.899	1	0.000
		LagCCRThreshold(9)	9.257	1	0.002
		LagCCRThreshold(10)	14.918	1	0.000
		MeanABCCRThresh	1428.319	10	0.000
		MeanABCCRThresh(1)	0.992	1	0.319
		MeanABCCRThresh(2)	0.336	1	0.562
		MeanABCCRThresh(3)	0.000	1	0.990
		MeanABCCRThresh(4)	2.461	1	0.117
		MeanABCCRThresh(5)	28.028	1	0.000
		MeanABCCRThresh(6)	46.446	1	0.000
		MeanABCCRThresh(7)	92.856	1	0.000
		MeanABCCRThresh(8)	150.264	1	0.000
		MeanABCCRThresh(9)	437.209	1	0.000
		MeanABCCRThresh(10)	385.213	1	0.000
		Rebate10	1540.698	2	0.000
		Rebate10 (1)	4.383	1	0.036
		Rebate10 (2)	1342.243	1	0.000
		Owner Occupancy (%)	705.876	1	0.000

Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	5661.471	23	0.000
	Block	5661.471	23	0.000
	Model	5661.471	23	0.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	78943.243[a]	0.040	0.088

3 Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Classification Table[a]			Predicted		
			4-Quarter Application Presence		Percentage
	Observed		0	1	Correct
Step 1	4-Quarter Application Presence variable (1 = presence, 0 = absence)	0	124348	98	99.9
		1	12579	117	0.9
	Overall Percentage				90.8

a. The cut value is .500

Variables in the Equation		B	S.E.	Wald	df	Sig	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	LagCCRThreshold			2936.983	10	0.000			
	LagCCRThreshold(1)	1.760	0.054	1060.306	1	0.000	5.811	5.227	6.461
	LagCCRThreshold(2)	1.405	0.056	630.981	1	0.000	4.074	3.651	4.546
	LagCCRThreshold(3)	1.305	0.055	573.080	1	0.000	3.687	3.313	4.103
	LagCCRThreshold(4)	1.217	0.063	371.730	1	0.000	3.376	2.984	3.821
	LagCCRThreshold(5)	1.043	0.064	264.506	1	0.000	2.838	2.503	3.219
	LagCCRThreshold(6)	1.109	0.070	248.585	1	0.000	3.032	2.642	3.481
	LagCCRThreshold(7)	0.985	0.056	313.787	1	0.000	2.678	2.401	2.986
	LagCCRThreshold(8)	0.604	0.074	67.384	1	0.000	1.829	1.583	2.112
	LagCCRThreshold(9)	0.177	0.083	4.511	1	0.034	1.193	1.014	1.405
	LagCCRThreshold(10)	-0.558	0.113	24.549	1	0.000	0.572	0.459	0.714
	MeanABCCRThresh			542.375	10	0.000			
	MeanABCCRThresh(1)	0.137	0.050	7.399	1	0.007	1.147	1.039	1.266
	MeanABCCRThresh(2)	0.117	0.051	5.348	1	0.021	1.125	1.018	1.242
	MeanABCCRThresh(3)	0.132	0.056	5.590	1	0.018	1.141	1.023	1.273
	MeanABCCRThresh(4)	0.159	0.048	10.776	1	0.001	1.172	1.066	1.289
	MeanABCCRThresh(5)	0.316	0.048	42.555	1	0.000	1.372	1.247	1.508
	MeanABCCRThresh(6)	0.351	0.046	58.304	1	0.000	1.421	1.298	1.555
	MeanABCCRThresh(7)	0.409	0.048	73.942	1	0.000	1.505	1.371	1.652
	MeanABCCRThresh(8)	0.424	0.044	92.892	1	0.000	1.528	1.402	1.665
	MeanABCCRThresh(9)	0.694	0.042	278.564	1	0.000	2.002	1.845	2.172
	MeanABCCRThresh(10)	0.639	0.042	228.133	1	0.000	1.895	1.744	2.059
	Rebate10			959.957	2	0.000			
	Rebate10(1)	0.301	0.024	161.674	1	0.000	1.351	1.290	1.415
	Rebate10(2)	0.713	0.023	959.899	1	0.000	2.040	1.950	2.134
	Owner Occupancy	0.009	0.000	595.255	1	0.000	1.009	1.008	1.010
	Constant	-3.462	0.030	13210.339	1	0.000	0.031		

a. Variable(s) entered on step 1: LagCCRThreshold, MeanABCCRThresh, OOccupation, ZHVI_10K.

Appendix D: SPSS Independent Samples *t*-Test Output

***t* Test – 4QMAAR (%) at Lagged CCR (%) = 0 and Lagged CCR (%) > 0**

Group Statistics

	Lagged CCR (%)	N	Mean	Std. Deviation	Std. Error Mean
4-Quarter Mean AAR (%)	>= .01	15897	0.4896	1.56607	0.01242
	< .01	121245	0.2472	1.34786	0.00387

Independent Samples Test

		Levene's Test for Equality of Variances		*t* test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Future Mean AAR (%)	Equal variances assumed	805.376	0.000	20.896	137,140	0.000	0.24234	0.01160	0.21961	0.26508
	Equal variances not assumed			18.627	19,110.012	0.000	0.24234	0.01301	0.21684	0.26785

***t* Test – 4QMAAR (%) at lagged ABCCR (%) <= 4.28 and ABCCR (%) > 4.28**

Group Statistics

	Mean lagged ABCCR % (not including subject block)	N	Mean	Std. Deviation	Std. Error Mean
4-Quarter Mean AAR (%)	>= 4.28	11474	0.5974	2.15855	0.02015
	< 4.28	132,306	0.2410	1.26386	0.00347

Independent Samples Test

		Levene's Test for Equality of Variances		*t* test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Future Mean AAR (%)	Equal variances assumed	2,006.359	0.000	26.981	143,778	0.000	0.35634	0.01321	0.33046	0.38223
	Equal variances not assumed			17.426	12,164.423	0.000	0.35634	0.02045	0.31626	0.39643