# THE EFFECT OF ENVIRONMENTAL AMENITIES ON HOME VALUES IN THE UPPER SANTA CRUZ BASIN: A HEDONIC ANALYSIS USING CENSUS DATA

**BY**

**GAURAV ARORA**

---

A Thesis Submitted to the Faculty of the

DEPARTMENT OF AGRICULTURAL AND RESOURCE ECONOMICS

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

In the Graduate College

THE UNIVERSITY OF ARIZONA

**2012**

**STATEMENT BY AUTHOR**

This thesis has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this thesis are allowable without special permission, provided that accurate acknowledgement of the source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In other instances, however, permission must be obtained from the author.

SIGNED: ______________________

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

______________________ 5/11/2012

Dr. George Frisvold
Professor of Agricultural and Resource Economics

Date

**ACKNOWLEDGEMENTS**

I would first like to thank my advisor, Professor George Frisvold for allowing me to pursue this study with great supervision, encouragement and a lot of independent thinking in past months. His valuable time, guidance, insight and day-to-day comments have been immensely helpful in the successful completion of this project.

I am thankful to my committee members, Dr. Bonnie Colby and Dr. Sandy Dall'erba for all their help and suggestions. I am grateful to Dr. Laura Norman for providing me with a quality dataset and for her diligent assistance, and Dr. Gary Thompson for his insightful comments on econometric issues, which has greatly helped in refining this research.

Last but not the least; I take this opportunity to express my gratitude towards the Department of Agricultural and Resource Economics, for supporting my higher education for the period of two years.

# DEDICATION

**To my PARENTS**

## TABLE OF CONTENTS

## LIST OF FIGURES

# LIST OF TABLES

## ABSTRACT

For this study, hedonic pricing method is used to measure the effects of natural amenities on home prices in the U.S-side of the Santa Cruz Watershed. Further, multivariate spatial regression techniques are utilized to estimate different factors that affect median home values in 613 census block groups of the 2000 Census, accounting for spatial autocorrelation, spatial lags and spatial heterogeneity in the data. Diagnostic tests suggest that failure to account for spatial effects would bias inferences made from the statistical analysis. The explanatory factors for the hedonic model can be classified as (1) physical attributes of the housing stock, (2) neighborhood characteristics, and (3) environmental attributes. Census data merged with GIS (land-sat) data is used for vegetation and land cover, land administration, measures of species richness and open space, and proximity to amenities and disamenities. IV 2SLS – robust Spatial lag models were estimated to control for heteroskedasticity and variable endogeneity. Results suggest that policies to maintain rich biodiversity provide economic benefits to homeowners, reflected in higher home values. Marginal effects of these explanatory variables are calculated which are interpreted as a continuous rate of change in the home values, given a marginal change in the continuous explanatory variable or a discrete change in the dummy variables from 0 to 1. The land cover variables seem to dominate in terms of the '*localized or direct*' marginal effects while public lands, in terms of '*spillover or indirect*' marginal effects. Further, estimated marginal effects of the discrete variables reveal that there is a definite discount in home values when on the border (<5 miles), near Tucson International Airport & Davis Monthan Airbase, whereas there lies a high premium for being in the Catalina Foothills and Tanque Verde School Districts.

# CHAPTER 1

## INTRODUCTION:

### Background:

The Upper Santa Cruz River supports and sustains the social, economic and environmental wealth of its watershed. That it not only is a source of perennial surface water flows but also supports the groundwater recharge by filtering river water through soil layers before water enters the aquifer bed. The river provides for various ecosystem services via supporting the aquatic and the riparian habitat. The term - aquatic habitat refers to the flora and fauna living in the river whereas riparian habitat refers to the life sustained due to the river flow that lies in a belt around the river. This study concentrates on the economic benefits of the natural habitat supported by the river watershed through surface flows and groundwater flows. What makes the river watershed a primary focus is its importance in providing for a green belt comprising of riparian forests; habitat to native, migratory and endangered species; smooth and sustainable municipal water supplies to households and a unique cultural identity to this semi-arid region in Arizona[1]. However, given the fact that the river flows in the semi-arid state of Arizona, its perennial flow is supported by both - precipitation and treated water effluent streams. The ever-growing population of the region and the increasing urbanization has caused groundwater over-draft and indiscriminate use of surface water to meet peoples' needs, thereby making the river more and more dependent on treated effluents. The treated water supports the perennial flow and replenishes the groundwater in an efficient way. However, at the same time refilling

[1] "A Living River- Charting the Health of the Upper Santa Cruz River 2010 Water Year"- A report published by The Sonoran Institute. Downloaded from http://sonoraninstitute.org/library/reports.html

the river by effluents gives rise to concerns of degraded surface water and groundwater quality, which may affect the long sustained riparian health, and increases hygiene-related concerns for people consuming water for various purposes, including drinking. Hence, there has been a large number of surveys conducted by the Arizona Department of Water Resources (ADWR), Arizona Department of Environmental Quality (ADEQ), Friends Of the Santa Cruz River (FOSCR), National Park Service (NPS), Sonoran Institute, Tucson Audubon Society and U.S. Geological Survey (USGS). These surveys are conducted to monitor the water quality, surface stream flow, groundwater levels, precipitation, weather and sometimes migration of birds to better understand how the river performs (through various research techniques) with an ultimate aim of ensuring sustained flow and good health of the watershed.

**The Issue of Conservation and Restoration of Riparian Habitat in the River Basin:**

The Santa Cruz River starts at the San Rafael Valley, Arizona, enters Mexico and loops back into Nogales, Arizona from Nogales, Sonora (Mexico). Such a bi-national contrast of the river's structure adds complexity to the management of the river water. Sprouse (2005) notes that the US-Mexican border represents the meeting of two cultures and with that are shared power and responsibility by the local, state and federal governments and agencies of the two countries. Due to such typical structural setup, resolutions with respect to shared river water problems move slowly and thus are more complex than water issues for any river located exclusively in the mainland United States.

The perennial flow of the Santa Cruz River is essentially a finite source of water. Most of *it* depends upon the treated effluent stream from the Nogales International Wastewater Treatment Plant (NIWTP), located nine miles north of the US-Mexico border at Rio Rico, Arizona. The wastewater that is received from Nogales, Sonora is important because not only does it support the rich riparian habitat north of NITWP but also recharges the Santa Cruz River Aquifer and provides for irrigation and potable water to the population residing in the watershed (Sprouse 2005). This peculiar importance of the Mexican effluent is a matter of water 'security' to Arizona because Mexico maintains rights to retain this water within its own county (IBWC[2] 1967, Sprouse 2005). Moreover, the Sonoran sewer system does not capture some of the effluent sources raising the concerns regarding quality of this water.

Water that is received and treated by NITWP, Arizona further competes within Santa Cruz AMA for industrial, institutional, irrigation, household and recreational use *versus* conservation and restoration of riparian habitat. Policy-makers face challenges of appropriately allocating and rationing these resources optimally fulfill the demands of each of the classified consumers above.

Along with allocating water resources, arises the issue of allocating limited land resources in the region. The issue of exploding population and increasing migration to cities gives rise to the not-so-environmental-friendly phenomenon of urban sprawl. Thus, allocation of land - a limited resource is competitive for its use for conservation and restoration of riparian habitat and for other uses like housing, agriculture, industries, institutions etc. Due to increasing awareness of environmental issues like decreasing forest cover and increasing developed spaces, Tucson AMA passed the Sonoran Desert Conservation Plan (SDCP). In view of this environmental-friendly policy, over 20,000 acres of land has been purchased by the year 2008 to ensure that a stock of natural space is preserved as a public good (Bark et al. 2008).

[2] International Boundary and Water Commission

**Primary Objectives and Motivation for the Study:**

This project aims at laying down specific homebuyer preferences for environmental amenities in the Upper Santa Cruz Basin to serve as a decision support for critical policy-making issues in the region. For this, geo-referenced data on homes, collected at the census block-group level is used. Vegetation-species'-specific preferences are drawn by hedonic price analysis on this data. Remote sensing data has provided researchers to create various vegetation and environmental indices to demarcate various land-use types and thus homebuyer preferences towards them. Here, I attempt a biodiversity preference models, testing whether biodiversity-rich are particularly paid premiums for when buying homes in the area. Bark et al. (2006) note that differentiating between species while conducting hedonic analysis can lead to better and more comprehensive results. Use of spatial econometrics techniques, control for spatial auto-correlation and spatial heterogeneity, which correct for omitted variable bias in hedonic models largely. The significance of spatial econometrics is well known but is still under-utilized in hedonic models (Brasington and Hite 2005).

*Motivation:*

The above laid structure of the Santa Cruz River derives a tension of acquiring water, providing it in potable form and then allocating it in an 'optimal' manner. This signifies the importance of quantifying consumer preferences within the study area. The term *consumer* here refers to the residents of the Upper Santa Cruz Basin. Since the study area comprises of both the urban and the rural areas, 'develop-able land' is also a scarce resource for which there is a competitive demand. Allocation of land and water resources, involvement of numerous institutions for monitoring and conserving riparian health of the watershed, significant amounts of funds and services directed towards conservation and restoration, and moreover, issues related to U.S. and Mexico's share in the Santa Cruz River water signifies the importance and need for this project.

## LITERATURE REVIEW:

### *Hedonic Price Modeling:*

Hedonic price modeling or hedonic pricing method has extensive applications to derive implicit prices of individual characteristics of goods using observable market expenditures on them. There exists a vast literature on hedonic pricing ranging from appraisal of the policy-driving housing market to even smallest of goods like wine. Majority of literature refers to Court (1939) as the pioneer of hedonic pricing, who created a price index for automobiles using a log-linear specification. However, Goodman (1998), Colwel and Dilmore (1999) and Malpezzi(2002) do point out to the classic works of H.A. Wallace (1926) and G.C. Haas (1922) as the originators of popular hedonic pricing method.

Two oft-cited classic references that present formal hedonic models are that of Lancaster (1966) and Rosen (1974). While Lancaster's work is focused on estimating the 'utility' of individual characteristics of house, Rosen has emphasized on price determination by using 'bid' and 'offer' functions in a perfect completion scenario.

A housing market is typically composed of: a) a house (a heterogeneous) good with peculiar characteristics like size, quality and location, and b) consumers with unique utility functions placing different values on different products (Sirmans, Macpherson and Zeitz, 2005). A hedonic model deals with this heterogeneity by breaking down total housing expenditure into the values (utility) of individual components.

*Fundamental hedonic equation:*

$Price_i = f(S_i, N_i, L_i, C_i)$ where,

$Price_i$ = observed market price for $i^{th}$ housing unit

$S_i$ = structural characteristics,

$N_i$ = neighborhood characteristics,

$L_i$ = locational characteristics, and

$C_i$ = contract type, i.e. whether the housing unit is inhabited by its owner or it is rented (Malpezzi 2002).

Literature refers the above (hedonic) equation as "the first stage" of a two-stage model, which attempts to recover supply and demand parameters of housing characteristics (Malpezzi 2002, Follain and Jimenez 1985b, Witte, Sumka and Erekson 1979). This study uses the above equation to relate observed consumer expenditures with utility-generated by individual housing characteristics via regression analysis and estimate the value/price of this utility.

Although hedonic models are extensively used to assess the real estate markets, there is one caveat pointed out in the literature, that hedonic models are location-specific and thus cannot be generalized over different geographical areas (Sirmans, Macpherson and Zeitz, 2005 - A Meta analysis[3] study published by the National Center for Real Estate Research).

[3] A meta regression analysis involves regressing the regression coefficients (from many econometric models) onto a set of explanatory variables of the important model and data characteristics. (Sirmans et al. 2005)

**Spatial Econometrics & Integration of GIS technologies to Hedonic studies:**

This section explores the literature concerned with a fairly recent and still emerging field of ‘Spatial Econometrics’. Spatial econometrics is different for traditional econometrics in the sense that it focuses around methodological concerns that follow from spatial effects viz.-a-viz. spatial interaction and spatial structure in regression models. Origin of spatial econometrics maps back to 1974, when Professor Jean Paelinck (now at the George Mason University, Virginia) coined the term during his address to the Dutch Statistical Society in 1974. Early work in this field was related to urban and regional science. Transition of spatial econometrics from regional science to mainstream economics happened in 1990’s, exemplified by Case (1991) in demand analysis; Aten (1996) in international economics; Topa (1996) in labor economics; Case, Risen and Hines (1993), Murdoch, Rahmatian and Thayer (1993) in public economics and local public finance; and Benirschka and Binkley (1994) in agricultural and environmental economics (Anselin 1999, Bateman, Jones, Lovett, Lake and Day 2002).

Integration of Geographic Information system or GIS in to environmental and resource economics is a very recent phenomenon. GIS technology has had its critiques before its integration with applied economics work was possible (Bateman et. al. 2002). Smith (1996) argued that from an econometrician viewpoint- ‘spatial dependence is a constraint or an exogenous factor that has to be control for’, whereas it should be viewed as an explanatory dimension for which GIS is well suited. In addition, utilizing GIS services into econometric evaluations allows for using the data “correctly” and “creatively” (Irwin and Geoghegan 2001) because a geo-referenced data set can be very useful to decode the factors explaining the spatial processes.

The housing market may reflect a spatial structures or patterns because of its direct relevance to natural environment. Therefore, there is a noticeable shift in the literature in hedonic models' estimation technique from traditional econometric models to newly developed spatial econometric models. Spatial arrangement of factors like habitat, land-cover or effluent discharges have an effect on species diversity, natural assimilative capacity and nutrient cycle (Turner et. al. 1993, Bateman et.al. 2002). GIS can provide information on location and the ambience of housing units, which together may drive choices and therefore, prices. Some of the examples of drivers of spatial processes can be urban sprawl causing land-use changes (agricultural to residential), distance to amenity/dis-amenity, proximity to pollution sources or national forests can significantly affect market prices (Bateman et.al.2002).

Most of published studies in the past referring to such spatial dependence among housing units include combination of property price information with census data and the locations of amenities – Metz and Clark 1997; Orford 1999; Waddell and Berry 1993; Doss and Taff 1996; Powe et.al. 1997.

**Variable Choice and Functional Form:**

*Variable Choice*:

Specifying the hedonic models correctly demand appropriate attention due to two major reasons. First, that leaving an explanatory variable out from the hedonic equation may lead to omitted variable bias resulting in estimated coefficients to be under- or over-estimated. This leads to biased and sometimes, inefficient estimates. Second, incorrect specification of hedonic models might lead to multicollinearity. Multicollinearity night cause inflated t-statistic values for estimated coefficients of the corresponding 'highly correlated' explanatory variables. A way to correct for multicollinearity is to choose a subset of these and drop the rest from the specified model but this can in turn lead to omitted variable bias.

Most of the hedonic literature is based on household level data where as this research involves analysis of housing data aggregated at the census block-group level. The choice of variables for this study is in majority a derivative of Schultz and King (2001) since it provides for hedonic price analysis conducted at the census block-group level. Some variables used in household-level data that are relevant for aggregated data (as in this study are exemplified in Sirmans et al. 2005 (*Review Articles – Journal of Real Estate Research).*

Structural variables:

The structural variables used in this study include: median number of rooms, % housing units without phone, % housing units without complete plumbing, and distance from the US-Mexico international border. The explanatory variables - median number of rooms is expected to have a positive relationship with median housing prices while % housing units without phone, % housing units without complete plumbing, and 'in-close-proximity' to the border is expected to be negatively related to the median housing prices.

Neighborhood Variables:

The neighborhood variables include: % housing units that are vacant, % housing units that are mobile homes, # of persons per home, % homes built in past 2 years, % homes built before 1939, school district (dummy) (Leech and Campos 2003, Cheshire and Sheppard 2004, Chiodo et. al. 2005, Bark et al. 2011), near Tucson International Airport or Davis Monthan Airbase (dummy) and % area under Federal Emergency Management Agency (FEMA) flood zone (dummy) (Bin et al 2008, Bark et al. 2011), % minority population, income levels. Census block groups are chosen as the level of aggregation for above. However, even though literature classifies school districts under geographical characteristics (Black 1999, Downes and Zabel 2002, Bark et al. 2011) but since the dataset used here is aggregated, it accounts for neighborhood-level geography. Here, housing value is expected to be increase by increased occupancy, newer homes built within past 2 years, and heritage homes (built before 1939), and higher income levels. Other variables like flood zone, close proximity to commercial area like airport, more mobile homes, high value for persons/home, and higher % of minority population are expected to have a negative effect on housing prices. High occupancy rate might refer to higher demand in the housing market and therefore should increase housing values.

Contract Variables:

% housing units owned (vs. rented) to classify among the different contract types. More owned homes should be a positive signal with respect to consumer preference and thus yield higher prices.

Environmental variables:

The environmental variables include amenities (as % census block group attributed to open water) as well as disamenities (as if % census block group recently mined or developed as impervious surface). The expected sign on the coefficients of amenities is positive and disamenities are negative. In order to evaluate the overall effect of "greenness" in the area, a new biodiversity index is created weighting the % area under each type of vegetation by the total area of the census block group. Such a variable will reveal preferences for diversity of vegetation and intensity of the same. Expected sign on the coefficient of this variable is positive.

*Functional Form:*

Hedonic analysis literature has used the semi-log or the log-linear functional form most extensively as compared to any other form (Sirmans et al. 2005). There are some major advantages of using this functional form over the linear form and the double-log form. First, a semi-log model allows for variation in dollar value of each characteristic (Follain and Malpezzi 1980, Sirmans et al. 2005). Second, coefficients of a semi-log model have a simple and appealing interpretation (Sirmans et al. 2005). The coefficients represent a percentage change in price given one unit of the characteristic (Follain and Malpezzi 1980). Third, semi-log models minimize the structural problems of heteroskedasticity. Fourth, explanatory variables can be chosen as dummy variables in case of semi-log models whereas double-log models are not flexible enough (Sirmans et al. 2005).

Literature does not provide any consensus on the most appropriate or "correct" functional form for hedonic analysis (Freeman 2003, Bitter et al. 2006). Considering the above advantages, this study uses a semi-log model with natural logarithm of median housing price of a census block-group as the dependent variable and liner form of explanatory variables for econometric analysis.

***Application of Hedonic Pricing Method for economic valuation of environmental amenities on residential property prices:***

Hedonic pricing method is being used from almost over four decades now to estimate the market price of environmental goods, earliest being Ridker and Henning (1967). Ridker and Henning, in their study used census-tract level data to evaluate the effect of air quality on housing prices. Likewise, the foremost studies that evaluated the effect of water quality and undesirable land-use on property values were David (1968) and Blomquist (1974). Strikingly, studies prior to 1988 included only a single environmental good (pertaining to air quality, water quality or undesirable land-use) in their models (Boyle and Kiel 2001). Blomquist et al. (1988) modeled all three environmental quality parameters to evaluate their marginal implicit prices.

The hedonic literature provides for numerous studies that are conducted in various geographical regions. Majority of these are completed in the last decade for observations on housing prices, ranging from household level to census tract level data (Geoghegan et al. 1997, Benson et al. 1998, Leggett and Bockstael 2000, Mahan et al. 2000, Acharya and Bennett 2001, Irwin and Bockstael 2001, Mooney and Eisgruber 2001, Schultz and King 2001, Geoghegan 2002, Smith et al. 2002, Leech and Campos 2003, Paterson and Boyle 2003, Cheshire and Sheppard 2004, Bin and Polasky 2005, Chiodo et al. 2005, Anderson and West 2006, Netusil 2006, Bark et al. 2008, Bark et al. 2011, Bark and Colby 2006). These studies have found out that connectivity to open space, lakefront amenities, urban wetlands, coastal water quality and ecological diversity have a definite premium for residential property values. One other observation is that hedonic analyses have shifted from quantity of environmental goods to their quality, being given primary importance while modeling consumer preferences. As stated above, the hedonic studies are model specific or in simpler words, houses vary in structural, neighborhood and environmental characteristics by region and so do consumer's preferences for them. Therefore, this section focuses on literature that involves variables that measure quality of the environmental amenities (like density and diversity of vegetation, biomass index etc. in case of land-use amenities,) within and around the area of focus. In addition, a few studies have controlled for spatial dependence for census block-group level data that are important to include here.

**Relevant Studies:**

Schultz and King (2001) find that census block-group level is the most appropriate level of census geography for land use aggregation for urban Tucson. Authors point out that most of the studies utilize housing data, which is often expensive and not spatially referenced. From 1990 onwards, census data are spatially referenced ‘block-group level data’ and thus can be integrated with various land-use data using geographic information system (GIS) technologies. This research finds that the reliability of hedonic estimates for alternative levels of census (census block, block-group and tract) is not the same. They found that census block-group level data is most appropriate especially, while estimating the marginal implicit prices of non-residential land-uses on housing values.

Hedonic pricing method to calculate the marginal effects of environmental amenities on housing prices has slowly shifted towards evaluating ‘quality’ from ‘quantity’. Most of the studies in the past have used quantitative variables like size of the amenity (Mahan et al. 2000, Bin and Polasky 2005) and distance from the amenity (Smith et al. 2002), whereas recent studies have accounted for heterogeneity of natural habitat in hedonic analysis.

Bark et al. (2008) and Bark et al. 2011 have asked questions like does the condition of vegetation or the riparian health (in semi-arid region) important (Bark et al. 2008)? The idea of differentiating between different types of riparian vegetation is has a two-way motivation. First, the conservation and restoration of these habitats attract significant funds and support from federal and state governments in the U.S. Second, conservation and restoration of green spaces competes for allocation of scarce water resources with other activities (Bark et al. 2008).

Apart from structural and neighborhood variables, Bark et al. 2008 have included various explanatory variables like ‘wetness’, ‘diversity scores’, ‘biomass’, upland connectivity’ to account for health and vigor of riparian habitat. Authors have found that high quality riparian habitat adds value to the nearby homes and homebuyers do value richer biological riparian vegetation. Authors have found that the fact that natural and constructed habitat are being distinguished by consumers. They also differentiate among various natural vegetation viz-a-viz xero-riparian (require less groundwater for their growth) and hydro-riparian (require shallow groundwater for growth). They have found that there is a premium for homes near hydro-riparian species since they are qualitatively and aesthetically different from xero-riparian species and constructed vegetation. They have found a counter-intuitive negative sign of 'biomass' which is essentially vegetation volume, along a transect. In addition, one caveat of this study is that they have used a buffered sample (0.2 miles from the watershed), therefore they do not account for benefits that accrue to homes outside the buffer area and visitors. In addition, this analysis may suffer from endogeneity as the homebuyers who choose to live within the buffer area may be living there due to peculiar preferences found in this study.

Bark et al. (2011) stands on the same rationale of efficient water allocation to test if there is a premium for some particular kinds of habitat. This study utilized 'greenness indices' which can be describes as high resolution, remotely sensed vegetation indices which classify between different kinds of land –uses. Authors used spatially geo-referenced data of 6,676 households in the Tucson metropolitan area with respective ZIP CODES. They utilized a spatial error model to control for spatial autocorrelation. Authors found that homebuyer's pay premiums for greener lots, greener neighborhoods and greener nearby riparian corridors. These studies along with Bark and Colby (2006) endorse the Sonoran Desert Conservation Plan (SDCP), which prioritizes to preserve the remnant riparian habitat and restore the degraded habitat regions. Premiums generated by home values and property taxes due to greener neighborhoods overcome the costs to build them (Bark and Colby 2006).

Brasington and Hite (2005) exemplify the application of spatial econometric techniques to estimate the demand for environmental 'quality'. The analysis includes six major metropolitan areas in Ohio and incorporates 5,051 observations at the census block-level. The authors utilize spatial Durbin model to control for spatial autocorrelation in both the 'first-stage" and the "second stage" hedonic analysis to estimate demand. This study, conducted at the census block-group level, observably uses the 'usual' explanatory variables, (also pertinent to household-level datasets). Authors do also control for variables like tax rate, climate mildness score and number of art performances. The econometric models used for estimating the implicit prices of housing units and estimating the demand of environmental quality have controlled for spatial effects in both the dependent variable (**Wy)** as well as the explanatory variables (**WX)**. The inclusion of **WX** term explicitly accounts for effects of structural characteristics of neighboring houses on the price of each house. A relatively inelastic demand curve for environmental quality is estimated, suggesting that response to environmental quality is not an individual but a collective action. Further, authors suggest that environmental quality and quality of schools in an area are complementary goods. Structural characteristics of homes viz-a-viz larger lot size can substitute for lower environmental quality. In addition, income levels and education levels of residents are positively related to demand of good quality environment and homes with more number of children have a higher demand for environment than the homes without children. Another study using household level data by Kim et al. 2003 has been conducted in Seoul, Korea for owner-occupied housing units. The study controls for space while statistically determining the best-suited dependence structure. Authors provide an insight towards calculating the marginal benefits of improved air quality (elasticity of hedonic prices) from spatial regression models. They report that these marginal benefits are indeed unbiased and consistent but do not differ much from their OLS estimates (only by 4%).

Another dimension of spatial effects on housing prices is termed as spatial heterogeneity. The application of spatial econometric literature on hedonic price analysis has assumed a constant price structure across the market under study. Bitter et al. (2006) exemplifies a study for urban Tucson (Arizona) where the assumption of constant price structure for entire metropolitan is relaxed. Authors have controlled for this spatial heterogeneity using spatial expansion method and geographically weighted regression (GWR). The reasoning behind expecting spatial heterogeneity in housing prices is attributed to demand and supply imbalances (Schnare and Struyk 1976, Michaels and Smith 1990, Goodman 1981 & 1998, Dubin and Sung 1987, Bitter et al. 2006). Spatial patterns in supply of houses arises from the location and neighborhood of the house like near or far from the center of the metropolitan area, (in) accessibility to riparian forest, national park etc. Similarly, spatial patterns result in the inelastic demand houses due to their location and durability, especially in the shorter periods of time (Schnare and Struyk 1976, Bitter et al. 2006). Bitter et al. have controlled for spatial heterogeneity using 1) spatial expansion models and 2) geographically-weighted regression. Spatial expansion method utilizes the x- and y- coordinates representing location of the housing units in the area. These variables are interacted with the explanatory variables like lot size, dwelling area, etc. in the spatial expansion model. The estimated coefficients of these transformed variables were significant signaling strong evidence of spatial heterogeneity. Another technique used to control for spatial heterogeneity was geographically weighted regression. GWR estimates a separate model for each sales point and weights the observations by distance to this point. This allows estimated marginal implicit price to vary location-by-location. The statistical software GWR 3.0 is used to estimate these GWR models. Author suggests that both spatial expansion and GWR methods improve explanatory power and predictive accuracy of the spatial models. Nevertheless, the GWR outperforms spatial expansion model on both the fronts, given that GWR is better able to represent complex spatial patterns whereas spatial expansion model is only capable of picking up the broad trends these patterns.

Following the literature, this study will aim towards controlling for both spatial autocorrelation and spatial heterogeneity in the housing prices observed at census block-group level for Upper Santa Cruz Basin while attempting at quantifying the economic significance and marginal benefits of ‘size’ and ‘quality’ of natural amenities.

**Thesis Contribution:**

This project expands the literature in the following ways.

First, this study acknowledges the ‘shift’ in literature concerned with hedonic analysis, from only quantifying environmental amenities to also, start analyzing their quality. Researchers these days are trying to map consumer preferences with respect to both quantity and quality of environment. Therefore, this study will incorporate both the ‘quantity’ and ‘quality’ of environmental amenities in the Upper Santa Cruz Basin and test their presence against the preferences of home-buyers.

Second, this study explores and controls for spatial structure within the housing market of Upper Santa Cruz Basin. It is realized that spatial econometrics, a technique to control for spatial dependence and spatial patterns in traditional OLS models is well developed. This technique, though has applications in hedonic analyses but can still be termed as under-utilized after glancing through the recent hedonic literature. This study aims at controlling for spatial effects in hedonic models specified in later sections.

Third, literature offers numerous hedonic studies conducted in different areas. Hedonic studies are constantly noted as being “model-specific” due to heterogeneity in housing market, with respect to both houses and consumers who buy them. The hedonic studies are conducted mostly at the household level, covering a small area of interest, making it impossible to generalize the estimation techniques and results. This study will offer more 'generalize-able' results and models as compared to prior hedonic studies because the area incorporated here is relatively very-large (covers urban housing, sub-urban housing and rural housing) and a dataset which comes from a uniform mechanism used to generate data across United States.

Along with being an econometric exercise, this project has some serious policy implications for the water-allocation and land-allocation issues in the region of interest. Programs like Intergovernmental Agreement and Sonoran Desert Conservation Plan are in place in Arizona due to public interest in conserving the riparian health of the state by reserving treated water and land, respectively. The Santa Cruz River supports some very healthy and aesthetically rich riparian habitat. Water and land being limited resources of the modern era are primary concerns for policy-makers, as they must allocate them efficiently. Water and land compete for agriculture, institutions, industries, household activities etc. versus conservation and restoration of riparian habitat, especially in the semi-arid region of interest. Apart from that, the federal and state governments allocate large sums of money towards the aforementioned and like projects. Therefore, there is a need and motivation for an analysis, which maps consumer preferences for environmental amenities.

# CHAPTER 2

## THE UPPER Santa Cruz RIVER BASIN:

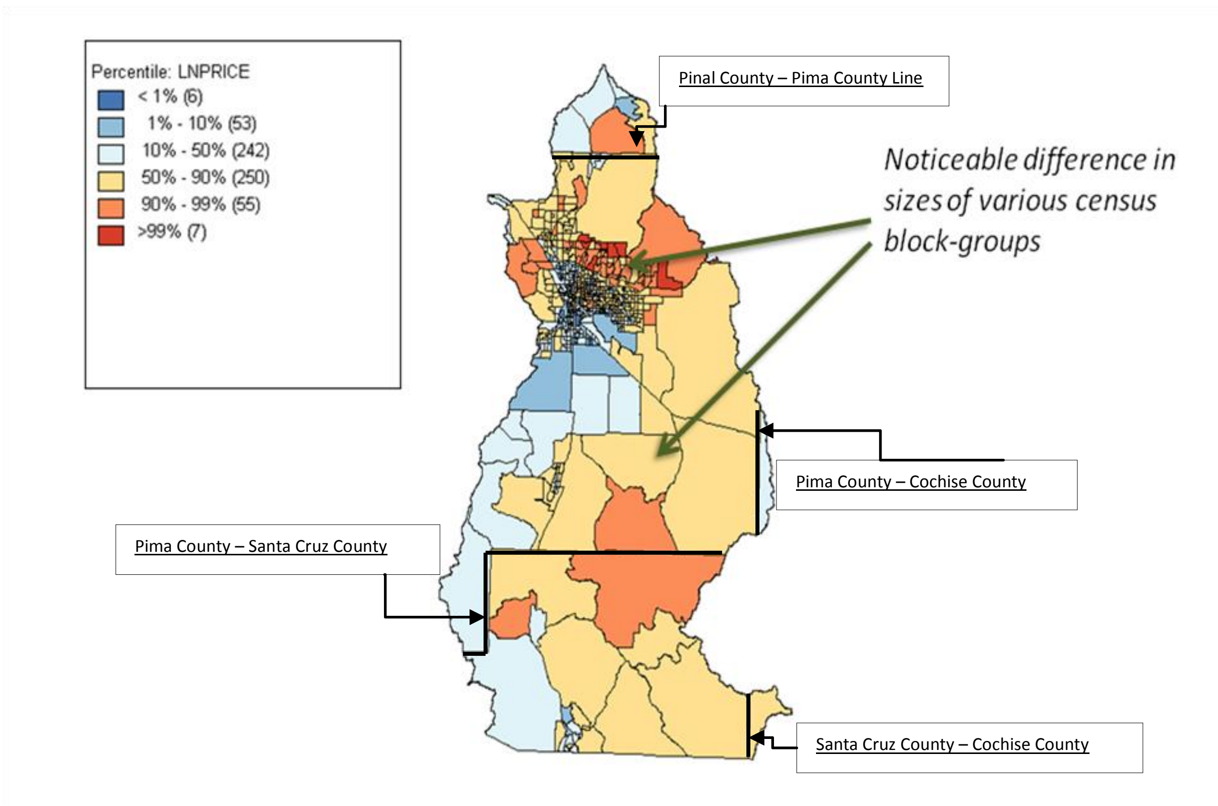
## STUDY AREA:

As noted earlier, this study comprises of observations at the census block-group level allowing a large area for hedonic analysis. The total study area amounts to a total of 8122.63 square kilometers (3136.16 (sq. miles) and includes four (4) counties namely Pinal County, Pima County, Cochise County and Santa Cruz County (in Arizona, USA). It is noteworthy that none of the four (4) counties fall entirely in the upper Santa Cruz river basin, with the basin mostly comprising of Pima County (62%) and Santa Cruz County (32%), while some parts of Pinal County (4%) and Cochise County (2%). Apart from this, the study area includes fifteen (15) school districts, Tucson International Airport, Davis-Monthan airbase, various private and public land parcels (state trust land, bureau of land Management (BLM) land, national park service (NPS) land etc., riparian forests, barren lands, open water parcel as well as recently mined areas. Hence, there is a scope of analysis given amenities and disamenities distributed across the river basin.

Each of the census block-groups can be assumed as *parcels of land or simply as polygons*, which are valued in the market by median price of homes contained within each of them. This "median price" is the observed value of each parcel in the real-estate market, which can be determined, by the unobserved value of each parcel's structural, neighborhood, contractual and environmental attributes. Such an interpretation refers to the importance of understanding in detail, the fragmentation of land-use patterns of each parcel.

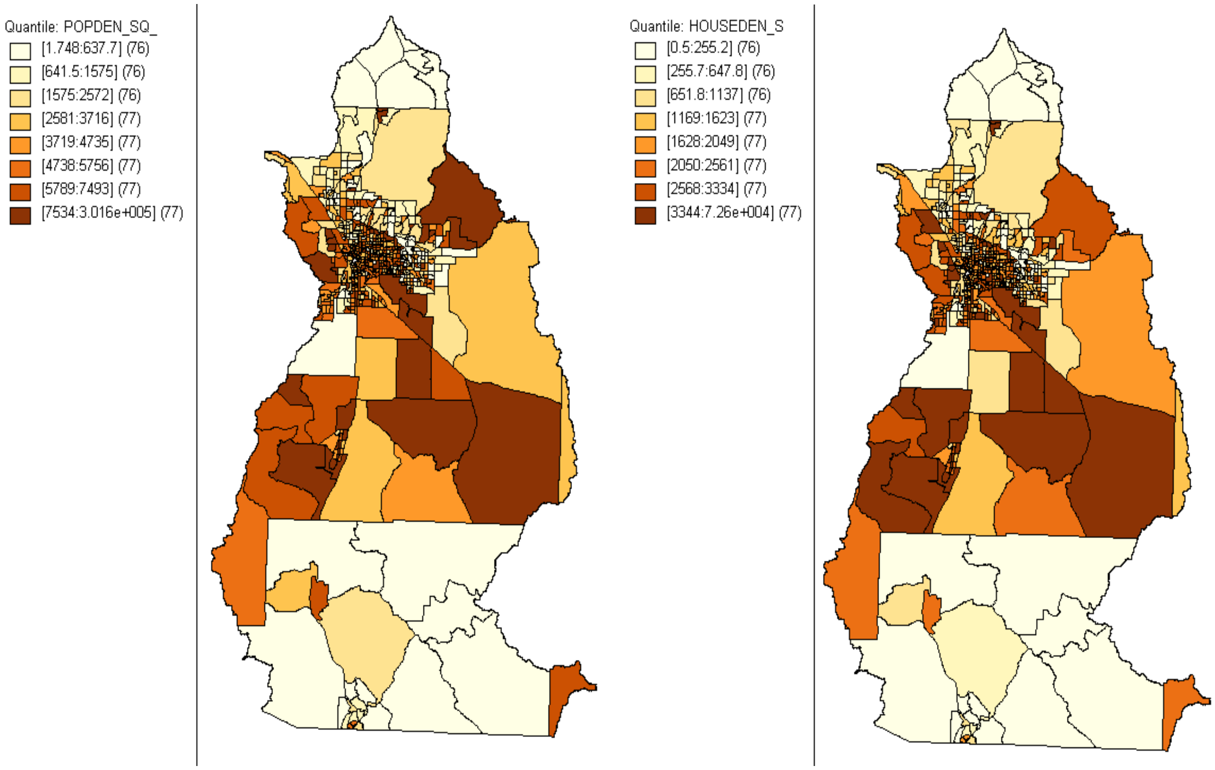
To understand the study area better, Figure 1 below shows a percentile map with respect to natural log of median prices of homes in all the parcels. It can be seen from figure 1 that most of the parcels (almost five-sixths) lie in the $10^{th}$ - $90^{th}$ percentile price range and only a very few (13 out of 613) can be termed as lower or upper outliers.

**Figure 1: Percentile Map for Median Home Price at Census Block-Group Level**




The quantile maps (with eight classifications) for population density and housing density show a stark similarity. The Pima County and Cochise County has in general darker polygons indicating higher population density (per square mile) and housing density (per square mile) as compared to the Santa Cruz County and Pinal County (see figure 2).

## Figure 2: Quantile Map for Population and Housing Density




**THE DATA:**

The dataset used for analysis comes from two major sources. One set comes from the US-Mexico Bi-national census - 2000 values for U.S. side of the Santa Cruz Watershed (SCW) and the other is the derivative of digital maps from Landsat satellite. Given different sources the map scale for them are also different. The map scale for the data from census data is 1: 24,000 and that from landsat images is 1:60,000. The acquired census data and the Land for the year 2000 is georeferenced using GIS applications and pertinent software at the United States Geological Survey (U.S.G.S.), Western Geographic Science Center[4] The georeferenced data set contains

---

[4] All data collection, geo-referencing and compilation are done by: Dr. Laura M. Norman at the U.S.G.S, Western Geographic Science Center, 520 N. Park Avenue, Suite #102K, Tucson, AZ 85719-5035.

each census block group with a unique ID number (Polygon ID). Polygon ID – coded with county name, census tract number and census block-group number – is the key for conducting a spatial econometric analysis. Statistical analysis software – SAS has been used for all data cleaning and data merging purposes and GeoDa –specialized spatial econometrics software is used for *exploratory spatial data analysis and spatial econometric regressions*.

**Home Values and Structural, Neighborhood and Contractual Attributes of the homes:**

Home values are the median prices of housing units in a census block group. Median prices are preferred over mean or average prices because median for home prices is a better guess for aggregation, given very expensive homes and/or very low-prices homes. For the purpose of regression analysis, natural logarithm of median prices [$\log_N$ Price] is the dependent variable. The reasoning for using natural log of median prices rather than median prices per se, is explicitly provided in the above chapters. It is to be noted here that the Census 2000 data used in this study provides for exactly the same median prices (or for that matter natural logarithm of median prices). This feature of the dataset restricts the variability of the dependent variable by a little less than one-sixth of the total observations.

The structural attributes of homes are the primary drivers of the home prices. For the purpose of analysis at the CBG level, not only that there is a focus on structural attributes of homes, but also on the physical attributes of 'home-types' within the designated parcels. The structural attributes are listed in the table below followed by a table compiling descriptive statistics for these variables.

## *Table 1: Variable Names and Descriptions*

| Variables | Description |
|---|---|
| Median Price | Median Home Price. Measured in $. |
| *Dependent Variable* | |
| Log $_N$ Price | Natural log of median home price. |
| *Explanatory Variables* | |
| % Owner | % Housing units that are owned (vs. rented). |
| % Vacant | % Housing units that are vacant. |
| % Mobile Home | % Housing units that are mobile homes. |
| Median Rooms (#) | Median number of rooms. |
| PCNEW | % Housing units constructed in past 2 years. |
| PCOLD39 | % Housing units constructed before 1940. |
| NOPHONE | % Housing units without a phone. |
| INCPLUMB | % Housing units without plumbing. |
| % Flood | % CBG area that fall under FEMA flood zone. |
| TIA_DM_MAS | Dummy, =1 if CBG had Airports - TIA, Davis Monthan in them, and all the blocks that surrounding the real estate they make, otherwise 0. |
| Persons/Home | Population of CBG divided by # of housing units. |
| % Minorities | Percentage non – whites in the CBG. |
| %_earning_lt$25k | Percent of population that earns less than or equal to $ 25,000 p.a. |
| %_earning_gt$60k | Percent of population that earns more than or equal to $ 60,000 p.a. |
| D_ON THE BORDER | Dummy, =1 if distance of centroid of CBG is <= 25miles from US-Mexico Border , otherwise = 0 |
| D_SCHL_DISTRICT | Dummy, =1 if CBG (partly or fully) comes under Catalina Foothills or Tanque Verde School Districts, otherwise =0 |
| D_PIMA_PINAL | Dummy, =1 if CBG lies in Pima County or Pinal County, otherwise =0 |
| Area | Area of the census block group measured in square meters. |

Apart from the above attributes, one of the neighborhood attributes that falls in advertising the real estate market agents for selling homes is ‘school districts’. As noted earlier, there are a total of fifteen (15) school districts in the area under study. Note that one parcel may contain more than one school district or more than one parcel may fall under the same school district. The school district presence is captured by a dummy variable, which equals one (1) for the school district to be present and zero (0) for it to be absent. The following table lists all these fifteen school districts. The presence of a good school district particularly attracts families that have school-going children. The wealthy or ‘the much better-off’ homebuyers can afford good schools even at higher home prices making these homes marketable.

***Table 2: School Districts included in the area under study (15 in no.)***

| | | |
|---|---|---|
| Catalina Foothills School District | Altar Valley School District | Vail School District |
| Sahaurita School District | Continental School District | Santa Cruz Valley Unified School District |
| Tenque Verde School District | Empire School District | Nogales Unified School District |
| Flowing Wells School District | Marana School District | Oracle School District |
| Amphitheater School District | Sunnyside School District | Fort Huachuca Accommodation School District |

***Table 3: Summary Statistics and Expected Signs for pre-described variables.***

| Variables | N | Mean | S.D. | Min | Max | Expected Sign |
|---|---|---|---|---|---|---|
| Median Price | 613 | 109363.94 | 64128.79 | 9999 | 407300 | |
| *Dependent Variable* | | | | | | |
| Log $_N$ Price | 613 | 11.44 | 0.59 | 9.21 | 12.92 | |
| *Explanatory Variables* | | | | | | |
| % Owner | 613 | 0.64 | 0.26 | 0.03 | 1.00 | |
| % Vacant | 613 | 0.08 | 0.07 | 0.00 | 0.49 | |
| % Mobile Home | 613 | 0.11 | 0.22 | 0.00 | 1.00 | |
| Median Rooms (#) | 613 | 4.91 | 1.14 | 2.20 | 9.10 | |
| PCNEW | 613 | 0.02 | 0.05 | 0.00 | 0.48 | |
| PCOLD39 | 613 | 0.05 | 0.12 | 0.00 | 0.83 | |
| NOPHONE | 613 | 0.03 | 0.04 | 0.00 | 0.33 | |
| INCPLUMB | 613 | 0.01 | 0.02 | 0.00 | 0.13 | |
| % Flood | 613 | 8.41 | 14.64 | 0.00 | 100 | |
| Persons/Home | 613 | 2.60 | 4.82 | 0.85 | 118.51 | |
| % Minorities | 613 | 0.24 | 0.17 | 0.01 | 0.99 | |
| %_earning_lt$25k | 613 | 0.35 | 0.19 | 0.00 | 0.87 | |
| %_earning_gt$60k | 613 | 0.25 | 0.19 | 0.00 | 0.83 | |

***Green Space:***

Given that the study area under consideration is a semi-arid region, it comprises of a rather small percentage of land cover under riparian vegetation. The scarcity of this environmental amenity and its usefulness to the residents of the area definitely should have a premium for housing units in their proximity.

The variables green space are considered for three (3) separate "types" of variables viz-a-viz (i) land-ownership variables, (ii) land-use type/ land cover variables or physical attributes of the land area under study, and (iii) a biodiversity index . The first set of variables refers to the 'usual' hedonic variables widely used in the literature, tying home prices to size of the environmental amenity and distance to the amenity. As noted earlier that the recent literature seems to sway away from the usual abstract quantification of environmental attributes towards evaluating the qualitative aspects of environment like vegetation-type, density, diversity and biomass indices. The last sets of variables quantify such details with demarcated land types and their peculiar attributes.

TYPE 1:

**Land-Ownership Variables:**

The dataset divides land ownership into six (6) different sub-divisions, namely – (i) private land; (ii) state trust land; (iii) land under the ownership of Bureau of Land management (BLM); (iv) land under National park Service (NPS); (v) land under U.S. Forest Service (USFS); (vi) local parks.

Size of the amenity (in %age of total land): The reported values of area in square meters within each census block-group are converted to a percentage by using total area in the block-group. Percentage of land under various owners provides a relative measure as compared to the absolute area value, thus providing as a better indicator of the environmental amenity/disamenity.

The land-ownership types differ across the aforementioned types, that is why they have been differentiated and the sub-divisions, which are expected to provide a much more comprehensive analysis. The literature does make a note of difference between private vs. public amenities (Bark et. al. 2011). Further, BLM land is not expected to behave similar to USFS land and NPS land because of the difference in physical attributes – BLM is plains while USFS and NPS lands are hilly areas (generally). In addition, NPS and USFS lands are expected to provide for greater (neighborhood–) scenic and aesthetic views. The state trust land might be a resource as undeveloped (open space) and/or scenic land, but it is supposed to be developed at some point in the future, which might be discounted for in observed real estate values. Another variable by the name of "Percent Scenic" is created to measure the degree of environmental amenity as the sum of -% State Trust land, % National Parks, % Local Parks, % BLM land, % U.S.F.S. land. The descriptive statistics and expected signs for these are followed by two interesting histograms of % private land and % scenic land.

From the histograms below, the most startling facts that come out are the of the total 613 parcels or census block-groups as many as 557 parcels lie under the category of 90% - 100% of land is under private ownership and at the same time 559 parcels have only up to 10 % scenic land. This univariate analysis points out towards the scarcity of scenic lands in the areas that are developed and thus is a strong indication of assured premiums of scenic lands and areas. Although there are parcels with high percentage of private land and high percentage of scenic area at the same time, denying the fact that private land implies scarce scenic land areas.

**Figure 3: Histogram for % Private Land and % Scenic Land**

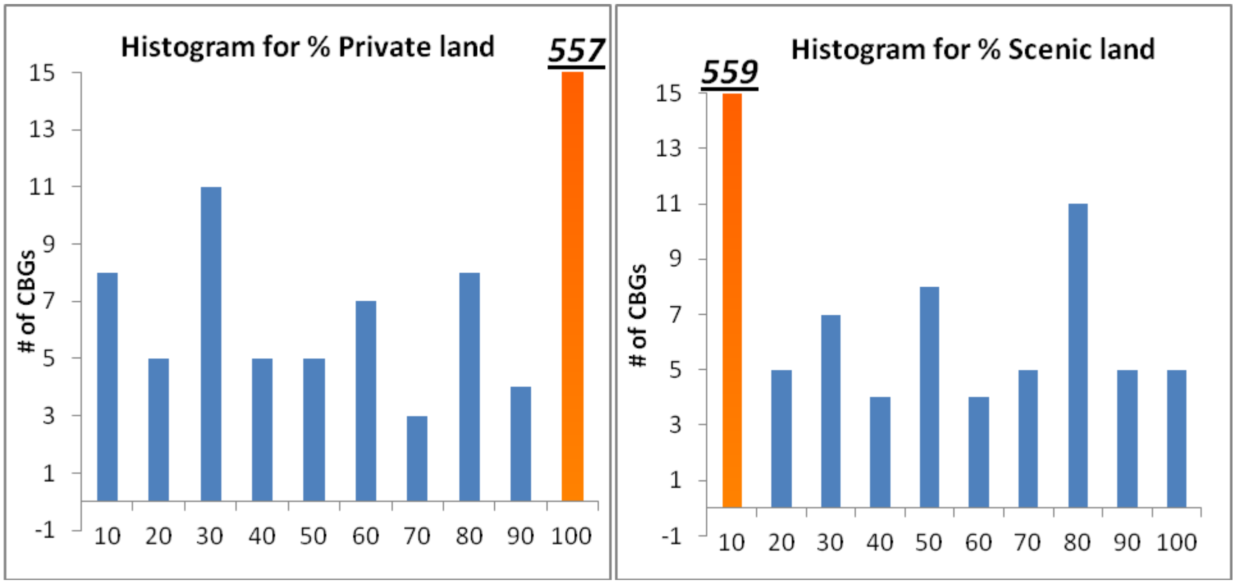



***Table 4: Summary Statistics and Expected Signs for Land-Ownership variables***

| Variables | N | Mean | S.D. | Minimum | Maximum | Expected Signs |
|---|---|---|---|---|---|---|
| %_Private | 613 | 94.32 | 18.73 | 0 | 100 | - |
| %_Trust | 613 | 2.90 | 12.00 | 0 | 97.16 | +/- |
| %_NatPks | 613 | 0.18 | 2.49 | 0 | 39.40 | + |
| %_LclPks | 613 | 0.20 | 2.68 | 0 | 48.80 | + |
| %_BLM | 613 | 0.17 | 1.60 | 0 | 26.56 | + |
| %_Forest | 613 | 1.67 | 10.46 | 0 | 95.26 | + |
| %_Scenic | 613 | 5.13 | 17.53 | 0 | 97.16 | N/A |

TYPE 2:

**Land-use type / Land cover variables:**

As noted above, the physical attributes of the land or (to say -) the different types of vegetation and other surface features might bear an implicit price. This implicit value of the land quality does have an impact on the observed home prices (Bark et al. 2006, 2008, 2011). Data used for this analysis was summarized from Villarreal et al (2011) describing conditions in 1999 using the following descriptions from the NLCD 2001 classification scheme, that were reclassified by Norman et al (2012) for input into a future growth scenario. The data set used for current analysis consists of variables like area of land designated as open water, agricultural land, developed land and different vegetation types like deciduous and evergreen forest, wetlands, pasture, grasslands and also, barren land. Further, there is an explicit differentiation within wetlands as *Palustrine Forested Wetlands* and *Emergent Herbaceous Wetlands* and within developed lands based on the degree of development. The table that follows defines each of the differentiated land cover types as per the land in each parcel attributed to different land-use types.

The 2001 National Land Cover Database (NLCD 2001) provided by U.S. Environmental Protection Agency- describes ten (10) different land cover types – listed in Table 5. On the grounds of these classes, United States Geological Survey (U.S.G.S.) has provided for land-use data having pixels for Sonoran Desert land cover types and arid riverine vegetation in the year 1999 counted by polygon ID, where each pixel = 300 square meters. The U.S.G.S dataset comprises of fourteen (14) classes which can be clumped into ten (10) described by NLCD 2001 (see sub-classes in Table 5). Figure 1 shows how each of these characteristics is distributed in the SCW.

***Table 5: Types of Land cover variables and their respective description.***

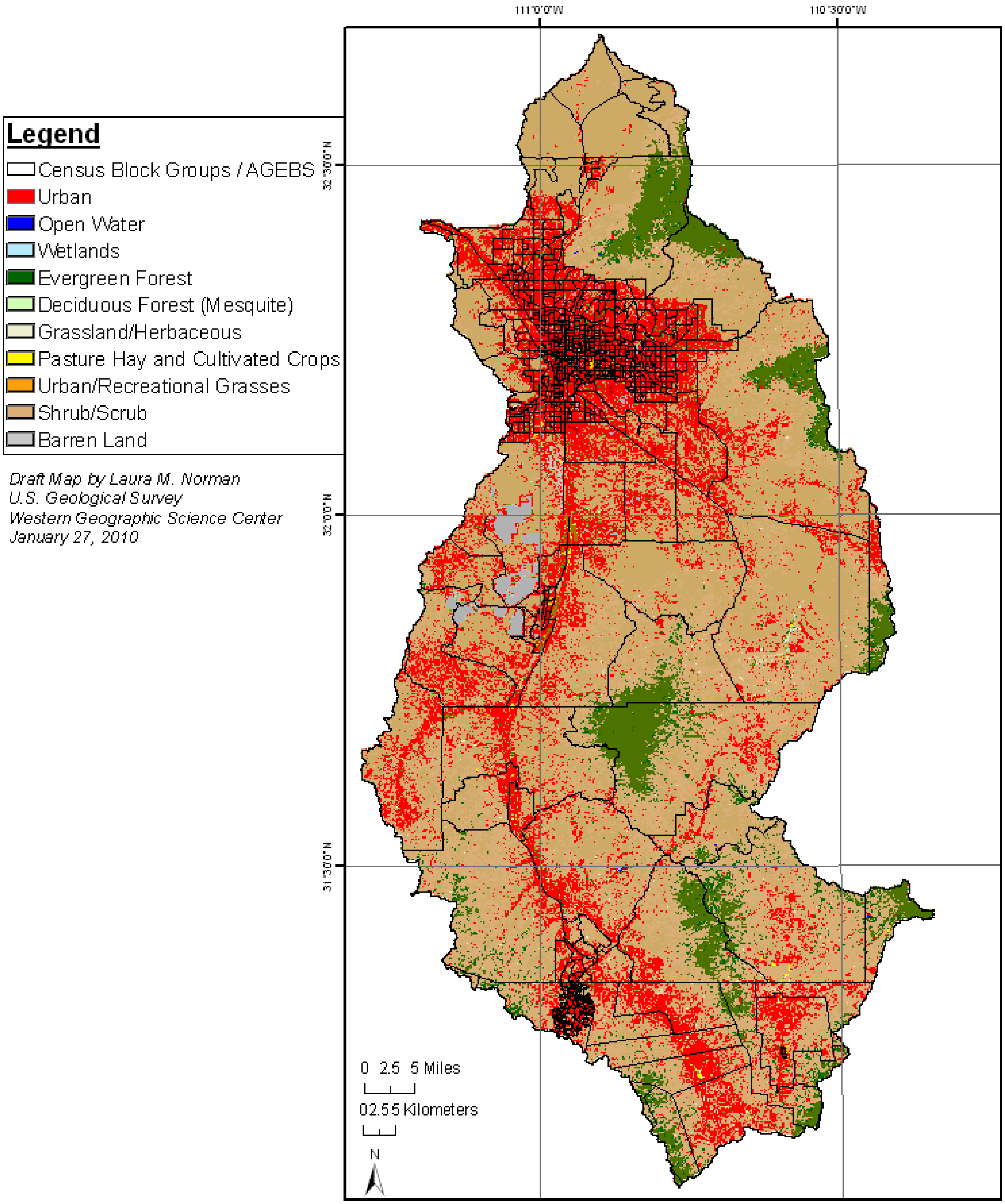
| S.No. | Environmental (dis/) Amenity | Description |
|---|---|---|
| 1. | Open Water | All areas of open water, generally with less than 25 percent cover of vegetation or soil. |
| 2. | **a.** Developed, Open Space | Includes areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for less than 20 percent of total cover. These areas most commonly include large-lot single-family housing units, parks, golf courses, and vegetation planted in developed settings for recreation, erosion control, or aesthetic purposes. |
| | **b.** Developed, Low Intensity | Includes areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 20-49 percent of total cover. These areas most commonly include single-family housing units. |
| | **c.** Developed, Medium Intensity | Includes areas with a mixture of constructed materials and vegetation. Impervious surfaces account for 50-79 percent of the total cover. These areas most commonly include single-family housing units. |
| | **d.** Developed, High Intensity | Includes highly developed areas where people reside or work in high numbers. Examples include apartment complexes, row houses and commercial/industrial. Impervious surfaces account for 80 to 100 percent of the total cover. |
| 3. | Barren Land (Rock/Sand/Clay) | Barren areas of bedrock, desert pavement, scarps, talus, slides, volcanic material, glacial debris, sand dunes, strip mines, gravel pits, and other accumulations of earthen material. Generally, vegetation accounts for less than 15 percent of total cover. |
| 4. | Deciduous Forest | Areas dominated by trees generally greater than 2 m tall, and greater than 20 percent of total vegetation cover. More than 75 percent of the tree species shed foliage simultaneously in response to seasonal change.1 |
| 5. | Evergreen Forest | Areas dominated by trees generally greater than 5 m tall, and greater than 20 percent of total vegetation cover. More than 75 percent of the tree species maintain their leaves all year. Canopy is |

| | | |
|---|---|---|
| | | never without green foliage. |
| 6. | Shrub/Scrub | Areas dominated by shrubs less than 5 m tall with shrub canopy typically greater than 20 percent of total vegetation. This class includes true shrubs, young trees in an early successional stage or trees stunted from environmental conditions. |
| 7. | Grassland/Herbaceous | Areas dominated by graminoid or herbaceous vegetation, generally greater than 80 percent of total vegetation. These areas are not subject to intensive management such as tilling, but can be utilized for grazing. |
| 8. | Pasture/Hay | Areas of grasses, legumes, or grass-legume mixtures planted for livestock grazing or the production of seed or hay crops, typically on a perennial cycle. Pasture/hay vegetation accounts for greater than 20 percent of total vegetation. |
| 9. | Cultivated Crops | Areas used for the production of annual crops, such as corn, soybeans, vegetables, tobacco, and cotton, and also perennial woody crops, such as orchards and vineyards. Crop vegetation accounts for greater than 20 percent of total vegetation. This class also includes all land being actively tilled. |
| 10. | **a.** Palustrine Forested Wetlands | Includes all tidal and nontidal wetlands dominated by woody vegetation greater than or equal to 5 m in height, and all such wetlands that occur in tidal areas in which salinity due to ocean-derived salts is below 0.5 percent. Total vegetation coverage is greater than 20 percent. |
| | **b.** Emergent Herbaceous Wetlands | Areas where perennial herbaceous vegetation accounts for greater than 80 percent of vegetative cover |

***Table 6: Descriptive Statistics and expected signs of Land Cover variables.***

| Variables | N | Mean | S.D. | Minimum | Maximum | Expected Signs |
|---|---|---|---|---|---|---|
| %_Open Water | 613 | 0.0084 | 0.057 | 0 | 0.908 | + |
| %_Developed | 613 | 27.32 | 9.076 | 0 | 35.95 | +/- |
| %_Barren | 613 | 0.78 | 1.659 | 0 | 16.91 | - |
| %_Deciduous | 613 | 0.085 | 0.375 | 0 | 5.24 | + |
| %_Evergreen | 613 | 0.305 | 2.116 | 0 | 34.27 | + |
| %_Shrub | 613 | 3.92 | 7.574 | 0 | 32.91 | +/- |
| %_Grassland | 613 | 0.21 | 0.783 | 0 | 10.58 | + |
| %_Pasture | 613 | 0.63 | 1.611 | 0 | 17.46 | - |
| %_Cultivated | 613 | 0.028 | 0.287 | 0 | 5.69 | - |
| %_Wetlands | 613 | 0.006 | 0.042 | 0 | 0.85 | + |

**Figure 4: U.S.G.S. Map for Land - Use types in the Sanata Cruz River Basin.**



TYPE 3:

**Biodiversity Index:**

Here, the source of data is derived from Villareal et al. (2012). Villarreal et al. (2012) quantified and mapped terrestrial vertebrate species richness using Wildlife Habitat Relation (WHR) models and Wallace et al (2011) vegetation map, and validated the results using with data from local National Park Service biological inventories. 451 potential terrestrial vertebrates were identified as using the habitat available from the vegetation map. A "Species Count" was derived to identify the number of species predicted to habituate in each vegetation type per CBG. In this case, Wallace et al. (2011) identified 34 unique vegetation and land use classes in the Santa Cruz Watershed using imagery from the year 2000. In order to measure the economic significance of biodiversity rich areas (if any), number of species habituated in vegetation 'i' is weighted by its percentage share of the total area in a particular census block-group. This number is aggregated using a summation operator across all vegetation types in a CBG to create the biodiversity index that not only captures the amount of vegetation cover but also specie-diversity among various vegetation types. Note that total number of vegetation types and corresponding share of land for each of these does differ across CBGs.

Mathematically,

$$\textbf{Index} = \sum_{i = 1,\ldots,n} (\textbf{N\_species})_i * (\textbf{\% area under vegetation})_i$$

Thus, higher the Index, the higher will be the biodiversity potential. Such an index is expected to show a positive economic significance as it forms a comprehensive account for green space in a semi-arid region.

# CHAPTER 3

## EXPLORATORY DATA ANALYSIS (EDA) AND EXPLORATORY SPATIAL DATA ANALYSIS (ESDA):

This section is divided into two sections. Section 1 explores different aspects of the geo-referenced dataset used in this study in order to derive various univariate and bivariate statistics and graphs. These will be helpful to explore the distributions of variables of interest, indicate symptoms of spatial interactions 'within' and among them and finally, for appropriately specifying the empirical models. Section 2 determines an appropriate spatial weights matrix (defined to capture the spatial interaction among the polygons or census block-groups), calculate Moran's I statistic (to determine the degree of global spatial autocorrelation) and calculate Local Indicator for Spatial Autocorrelation (to determine if the region/s might be experiencing spatial heterogeneity).

## SECTION 1:

The contents of this section are termed as EDA and are synonymous to descriptive statistics of 'traditional' econometric studies, extensively found in literature. Here, the tools offered by GeoDa – like choropleth maps (percentile map, quantile map and box map), histogram and box plot can be utilized to derive summary statistics. The choropleth maps provide for a visual overview of distribution of the statistical data. The histograms give the actual distribution or spread of respective variables of interest whereas the box plot provides for an overview of the spread of the distribution while identifying the *outliers* or the extreme data values. The lower and upper outliers are the values that lie 1.5 (*hinge*) times below and above the interquartile range (IQR) respectively. IQR is defined as the difference between the values that lie at the 25$^{th}$ and 75$^{th}$ percentiles. In a box plot, the data points are positioned in accordance with their value on the variable where the colored region represents the 25th to 75th percentile values, the red line in the colored region is mean and the green star (or dot) is the median. The difference between upper and the lower bars (or lines) represent the IQR. The data points above and below these bars are termed as outliers. The box map is similar to a percentile map but more informative and strict with respect to identifying outliers on a map.

The figures (box plots, histograms, percentile maps and box maps) that follow this section, give the distribution of the variables of interest while side-by-side comparing the spread on a graph and on a map. First, the dependent variable of this analysis or $\text{Log}_N$ Price is represented using box plot, histogram and a box map. The histogram represents a "normal-like" spread with majority of values concentrated in the middle. However, the box map and box plot also identifies the outliers of this distribution. In all 23 lower outliers and 11 upper outliers are found in price distribution showing the areas with very expensive and very inexpensive homes. Similar analysis for "type - 1" variables or land-ownership variables i.e., percent private land reveals that 476 polygons are upper outliers and 137 lower outliers with a mean of almost 95% land under private ownership. Where the EDA reveals the abundance of private land, it shows scarcity of scenic land with a mean as low as 5% for all the census block-groups. This marks the importance of environmental amenities given scarcity and Santa Cruz River being the sole source of water to water-loving vegetation species in the area.

Among the "type – 2" or the land-use/land cover variables, percent developed land has an average score of 27% for all polygons with only lower outliers (no upper outliers), implying that majority of census block-groups have more than one-fourth of their area as impervious surface due to constructed developments. The percentile map reveals presence of highly developed land cover in and around the Tucson Metropolitan Area. Upon careful inspection of percentile maps for other land-cover types, it is found that areas considered as environmental goods like open water, deciduous forest, evergreen forests, wetlands and grasslands are very limited and are mostly concentrated in less dense areas (with larger polygons). Shrubs and scrubs are found in abundance as compared to the other species. Agricultural and pasture lands which are utilized to produce commercial goods are also not found as much and are concentrated in some census-block groups in west and southwest Tucson. Overall, the distribution of all these variables is highly skewed.

Again, the question of spatial autocorrelation arises from the scenic views and environmental goods provided by the scarce "green areas" of the Santa Cruz Watershed. Either the homes situated in close proximity to or providing for "nicer" views and "fresher" air in the densely populated area are both benefited by the natural amenities in terms of the premiums paid for these. Hence it is important to control for the 'connectedness' of polygons, especially in a semi-arid region like the one under study, where larger trees and green views are not found in abundance and there is a great awareness with respect to conserving and protecting them. Fundamentally, as pointed out by Bitter et al. 2006, the heterogeneity within the housing characteristics as well as home-buyer's preferences leads to "spatial mismatches between supply and demand" causing formation of spatial clusters. Such an argument provides for a technical justification for existence of spatial dependence and/or spatial spillovers within a system.

**Figure 5: Log N Price: - Average = 11.44; Min = 9.21; Max = 12.92**

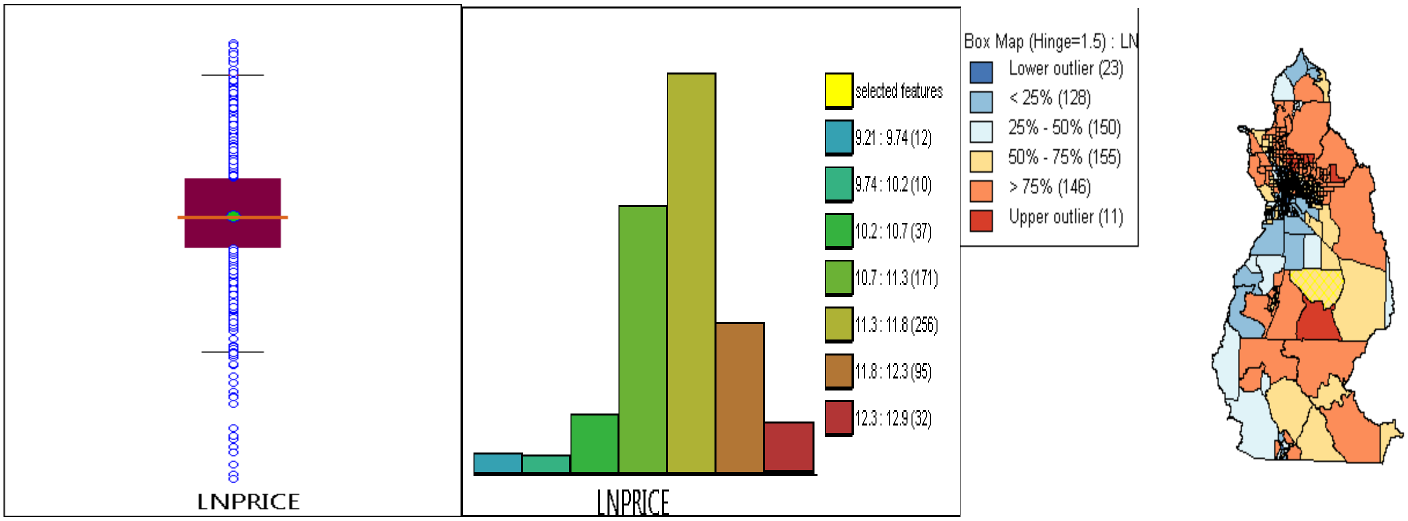



**Figure 6: %_Private Land: - Average = 94.32; Min = 0; Max = 100**

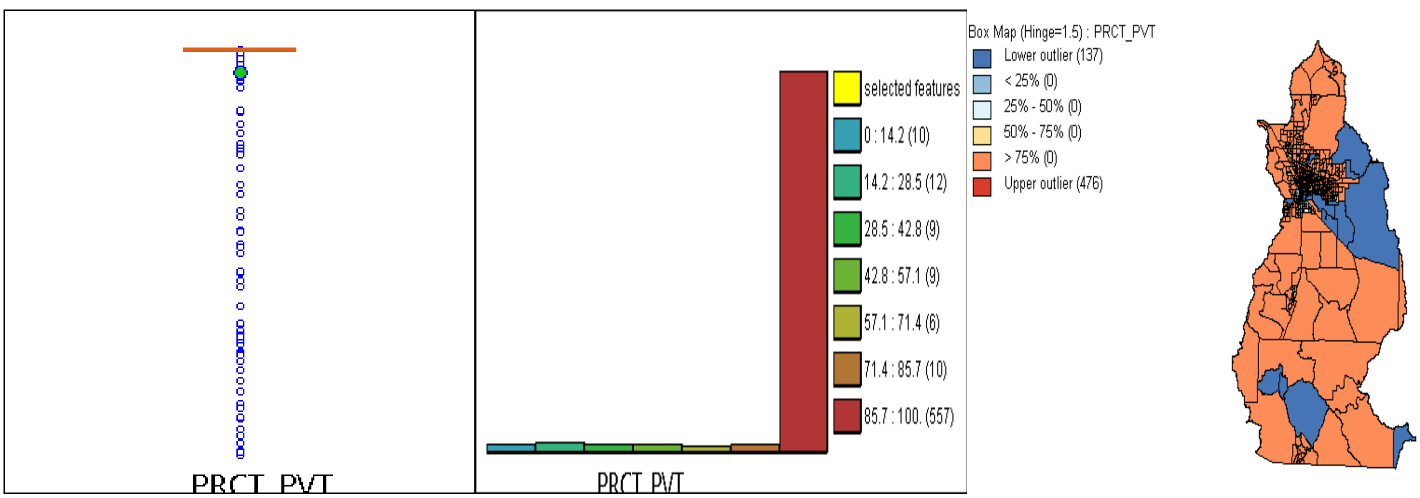


**Figure 7: %_Scenic Land: - Average = 5.13; Min = 0; Max = 97.16**

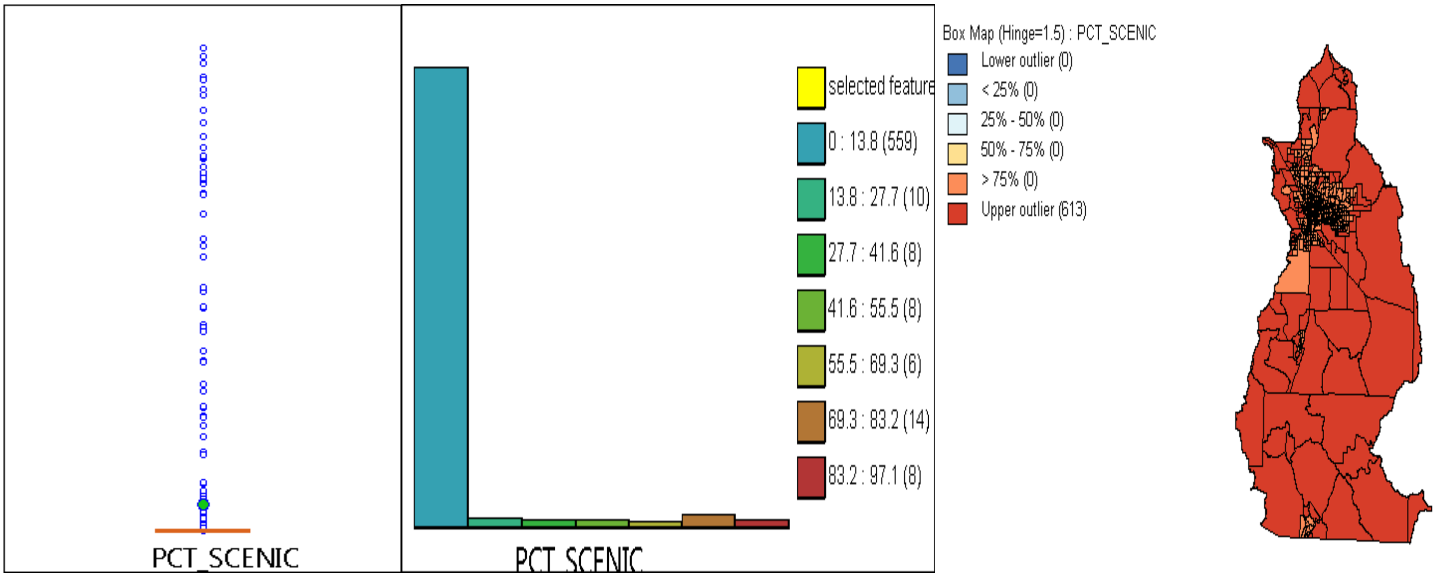



**Figure 8:%_Developed: - Average = 27.32; Min = 0; Max = 35.94**

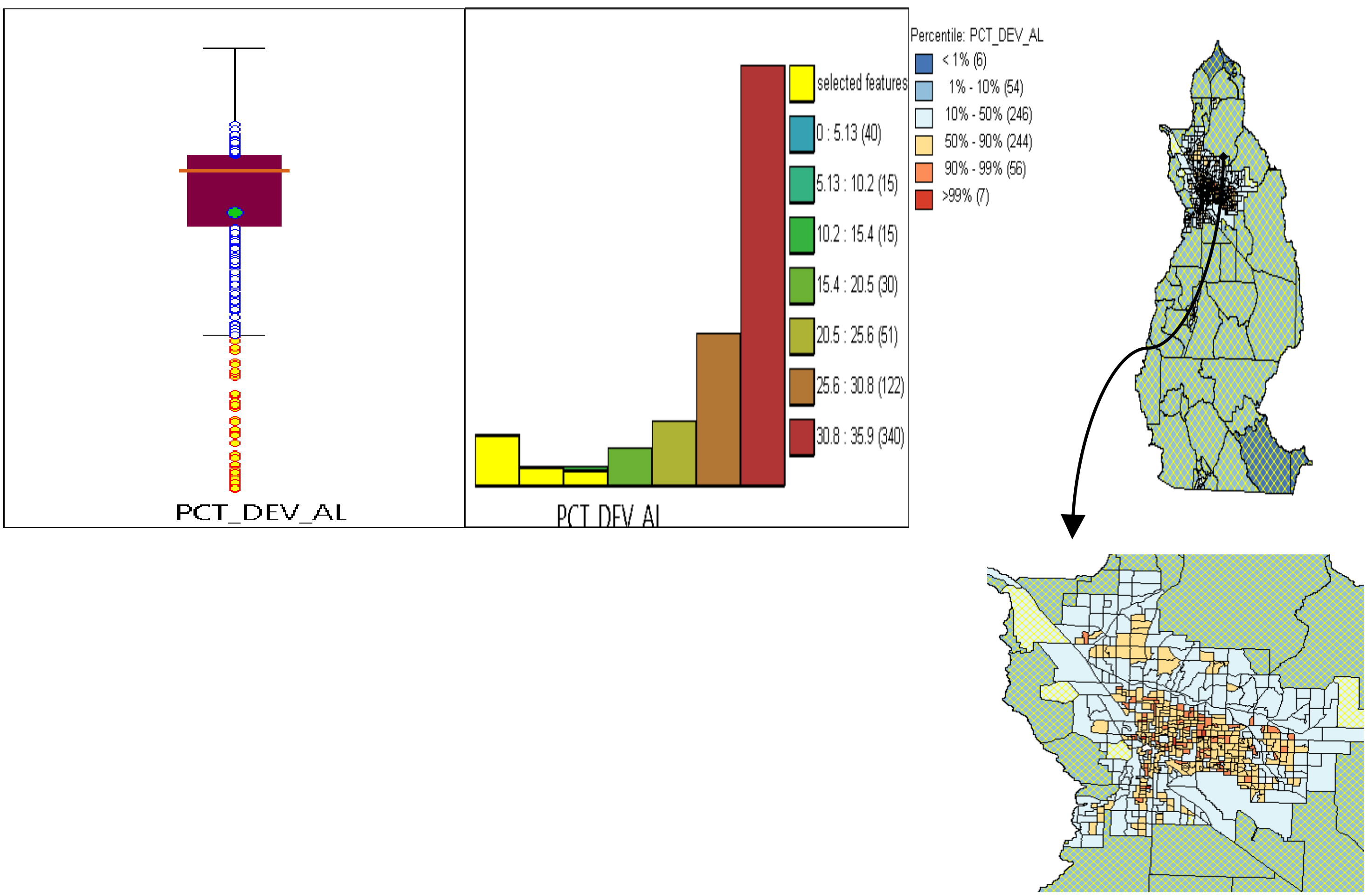
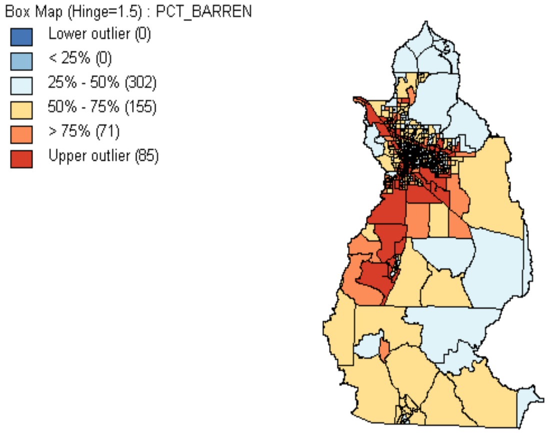
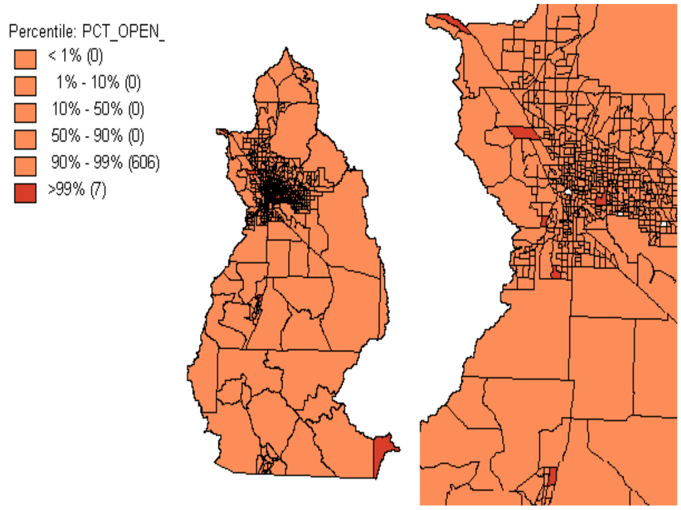


**Figure 9-16: Summary Statistics for land Use Variables**

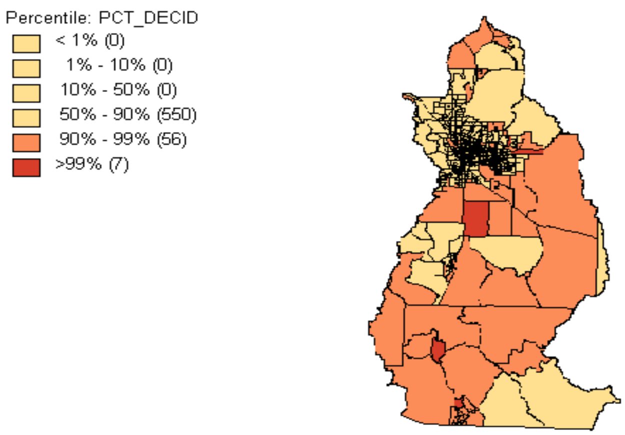
**%_Barren Land: - Average = 0.78; Min = 0; Max = 16.90**



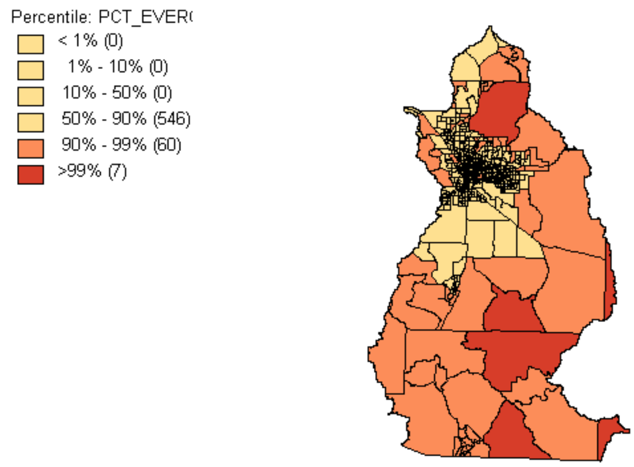

**%_Open Water: - Average = 0.008; Min = 0; Max = 0.91**



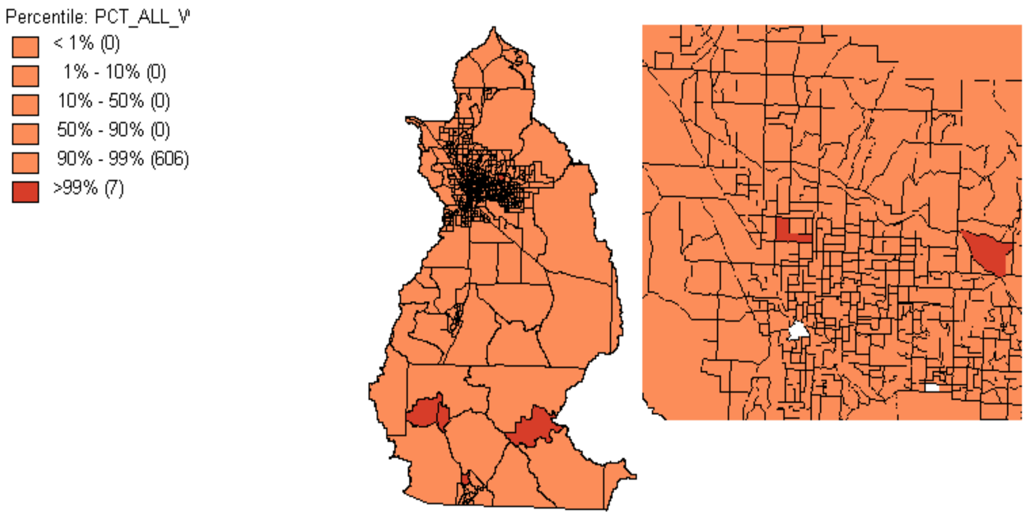

**%_Deciduous Forest: - Average = 0.085; Min = 0; Max = 5.24**



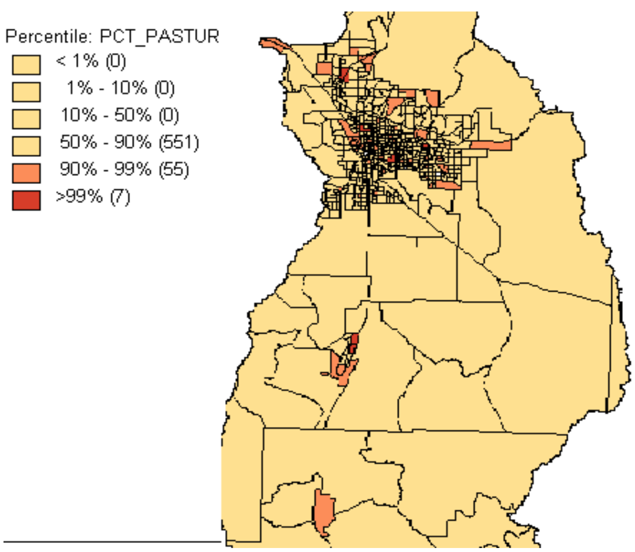

**%_Evergreen Forest: - Average = 0.31; Min = 0; Max = 34.27**



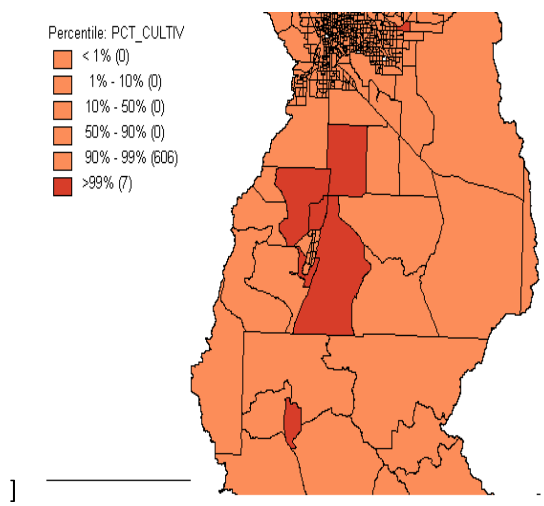
**%_Wetlands: - Average = 0.006 Min = 0; Max = 0.846**



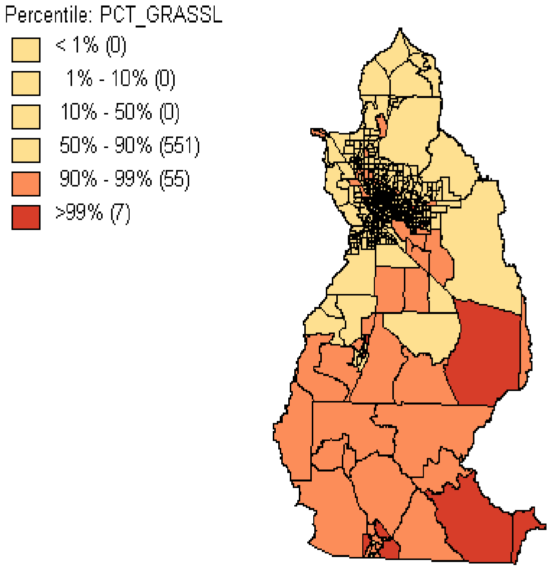

**%_Pasture: - Average = 0.63; Min = 0; Max = 17.46**



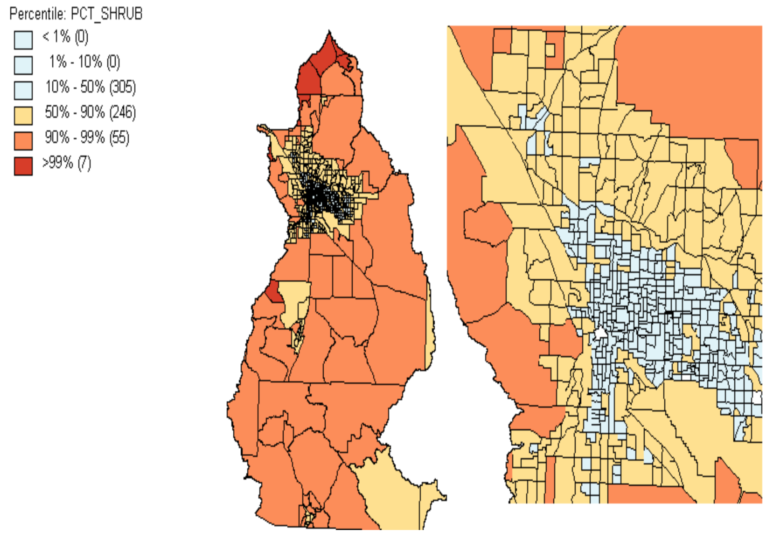

**%_Cultivated: - Average = 0.028; Min = 0; Max = 5.69**



%_Grassland: - Average = 0.21; Min = 0; Max = 10.58




%_Shrubs/Scrubs: - Average = 3.92; Min = 0; Max = 32.91



**SECTION 2:**

Given the above analysis, there arises a need to define the interaction of marketed real estate values over space. For this purpose, we define a weights matrix, which statistically defines the connectedness of each polygon with others in a matrix of numbers. The key is to interact of this weights matrix with the dependent variable to define a spatial lag or the effect of neighborhood polygons on respective polygons and eventually the real estate values per se. The scatter plot between the lagged dependent variable and the variable itself provides for a measure of spatial autocorrelation. The slope of the "best-fit" to this scatter plot is termed as Moran's I test statistic, which determines the extent or degree of spatial auto correlation (*or simply connectedness*).

The first step in this pursuit will be to determine and define an appropriate weights matrix. Literature provides a list of three different types of weights matrices: Contiguity, Great Circle Distance (GCD) and K-nearest neighbors - weights matrix. The connectedness of sides and vertices of a polygon with another polygon define the contiguity weights matrix. There are two types of contiguity weights matrices: Rook and Queen. Rook matrix defines connectedness with respect to polygon sides whereas queen does with both sides and vertices. For example, if a polygon shares common boundary with its neighbor then in case of Rook matrix - a value of 1 is assigned for the corresponding neighbor, otherwise a zero is assigned for all others. In contrast to contiguity matrix, the Great Circle Distance weights matrix counts the neighbors that lie within the assigned radius of the circle drawn from the centroid of a polygon. A k-nearest distance weights matrix counts the number (K) of nearest neighbors and gives a statistical weight to only these.

### *Table 7: Determining the Appropriate Weights Matrix*

| *Weights* | *Log $_N$ Price* | *% Owner* | *% Vacant* | *%_Local Parks* | *%_Evergreen* | *%shrub/scrub* |
|---|---|---|---|---|---|---|
| **Contiguity Weights Matrix** | | | | | | |
| Queen | 0.6063 *(0.001)** | 0.4707 *(0.001)** | 0.2818 *(0.001)** | 0.0816 *(0.020)** | 05410 *(0.001)** | 0.7708 *(0.001)** |
| **Great Circle Distance Weights Matrix** | | | | | | |
| W_16884 | 0.0836 *(0.001)** | 0.0675 *(0.001)** | 0.1174 *(0.001)** | -0.0003 *(0.871)** | 0.5309 *(0.001)** | 0.4461 *(0.001)** |
| W_18386 | 0.0678 *(0.001)** | 0.0492 *(0.001)** | 0.1132 *(0.001)** | -0.0010 *(0.778)** | 0.5185 *(0.001)** | 0.4194 *(0.001)** |
| W_22892 | 0.0255 *(0.001)** | 0.0295 *(0.001)** | 0.1049 *(0.001)** | -0.0000 *(0.969)** | 0.4316 *(0.001)** | 0.3308 *(0.001)** |
| **K-Nearest Neighbors Weights Matrix** | | | | | | |
| K (5) | 0.6099 *(0.001)** | 0.4679 *(0.001)** | 0.3156 *(0.001)** | 0.1321 *(0.008)** | 0.3771 *(0.001)** | 0.6542 *(0.001)** |
| K (10) | 0.5343 *(0.001)** | 0.4253 *(0.001)** | 0.2668 *(0.001)** | 0.0648 *(0.015)** | 0.2549 *(0.001)** | 0.5655 *(0.001)** |
| K (15) | 0.5075 *(0.001)** | 0.4024 *(0.001)** | 0.2352 *(0.001)** | 0.0406 *(0. 018)** | 0.1992 *(0.001)** | 0.5099 *(0.001)** |
| K (25) | 0.4525 *(0.001)** | 0.3721 *(0.001)** | 0.1870 *(0.001)** | 0.0759 *(0.001)** | 0.1805 (0.001)* | 0.2105 *(0.001)** |

***p-value upon 999 permutations (Null hypothesis: I = 0, i.e. No spatial autocorrelation)***

*Decision –rule for determining the most appropriate weights matrix:*

There exists no thumb-rule to determine a right or a wrong weights matrix but a criterion that can be used to determine one is a higher value of the Moran's I statistic. A higher Moran's I would mean higher degree of spatial autocorrelation captured. From the above table it is clear that the GCD matrix is not an appropriate choice as the calculated Moran's I is very low and thus is not a correct measure of the degree of spatial autocorrelation. Similarly, a choice has to be made between the Queen Contiguity weights matrix and K (5) – (with five nearest neighbors) weights matrix. Observation of Moran's I statistic reveal that both of these Weights matrices give an almost same value for the dependent variable whereas the two are found to be competing for other structural and environmental variables.

Queen is chosen over the K(5) weights matrix for two major reasons. One, the queen's weights matrix is sparse and easy to interpret, along with fulfilling the criteria of being exogenous to all the variables involved in the analysis. Second, that due to constraints offered by the latest version of GeoDa, it does not allow for regression analysis with non-symmetric weights matrix.

*Detecting Spatial Autocorrelation:*

*Global Spatial Autocorrelation:*

As mentioned earlier, the Moran's I gives the degree of global spatial autocorrelation. By the term 'global', is meant an average value for all the areas – urban and rural; forests and settlements, etc. for the entire area. A plot for the Moran's I is provided in the figure below. It is the slope of the scatter plot between the standardized values of the dependent variable and its spatial lag. Moran's I here, is found to be 0.60. This implied that a shock of degree – 'x' on the neighbors of a parcel will affect its prices by a degree of 0.60(x) and vice versa.

An important feature of the Moran's I scatter plot is that the each of the four quadrants determines the relationship prices of a parcel in a particular quadrant and its neighbors. Rightmost upper quadrant has data points for parcels, which have higher prices than the average and their neighbors have higher prices too (HIGH – HIGH). Similarly, the lower left quadrant consists of parcels, which have lower median prices (than the overall average) and are surrounded by the ones with lower median prices (LOW – LOW). The LOW – HIGH and HIGH – LOW quadrants are defined in the same way. A feature in GeoDa upon selecting the points on the Moran's I scatter plot highlights the corresponding polygons on the map. This feature given an idea of the clusters of four types of polygons related as HIGH – HIGH, LOW – LOW, HIGH – LOW and LOW – HIGH. The robustness of this clustering is not given by the Moran's I and hence we rely on the LISA statistic discussed as follows.

*Local Spatial Autocorrelation:*

The idea of local spatial autocorrelation is derived from the concept of submarkets or spatial heterogeneity in the spatial statistics literature. Spatial heterogeneity implies a non-constant price structure for the entire area. The Moran's I do hint towards some local spatial autocorrelation or local clusters. The Local Indicator for Spatial Autocorrelation (LISA) provides for a rather robust method to determine spatial clusters and their significance levels. The corresponding figures that explicitly show these clusters and their significance levels follow this section.

**Figure 107: Moran's I Scatterplot**

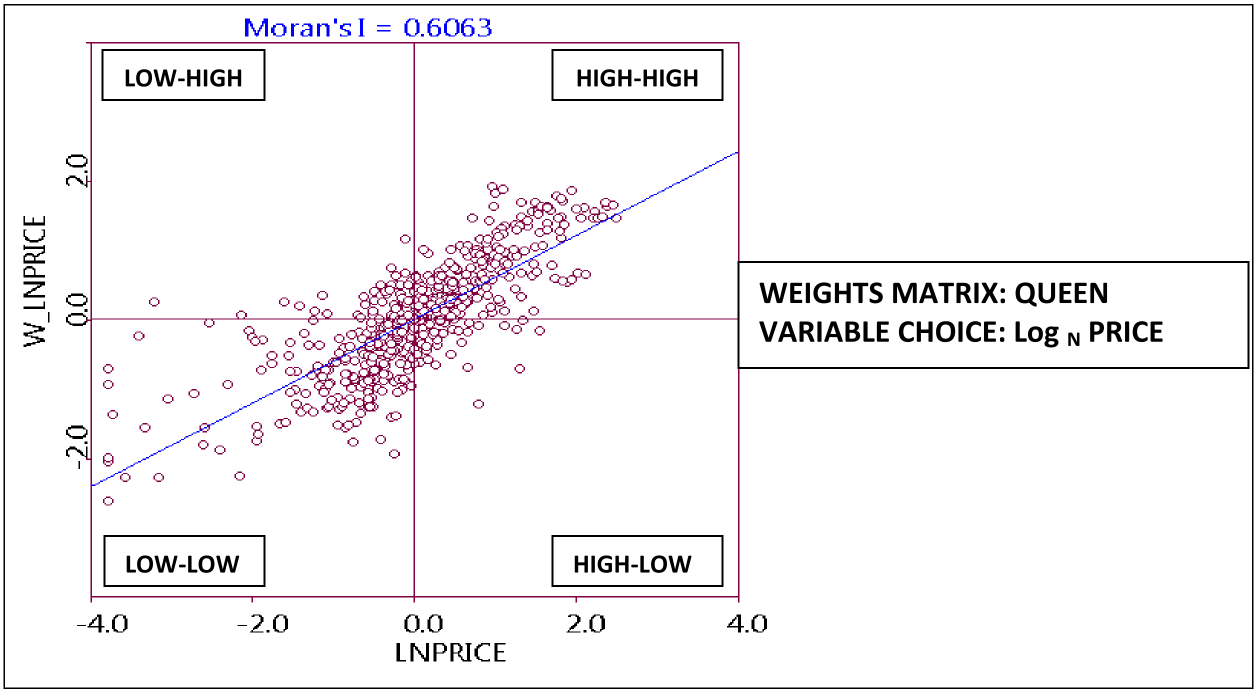



**Figure 118:Four Quadrants of Moran's I scatterplot on a map**

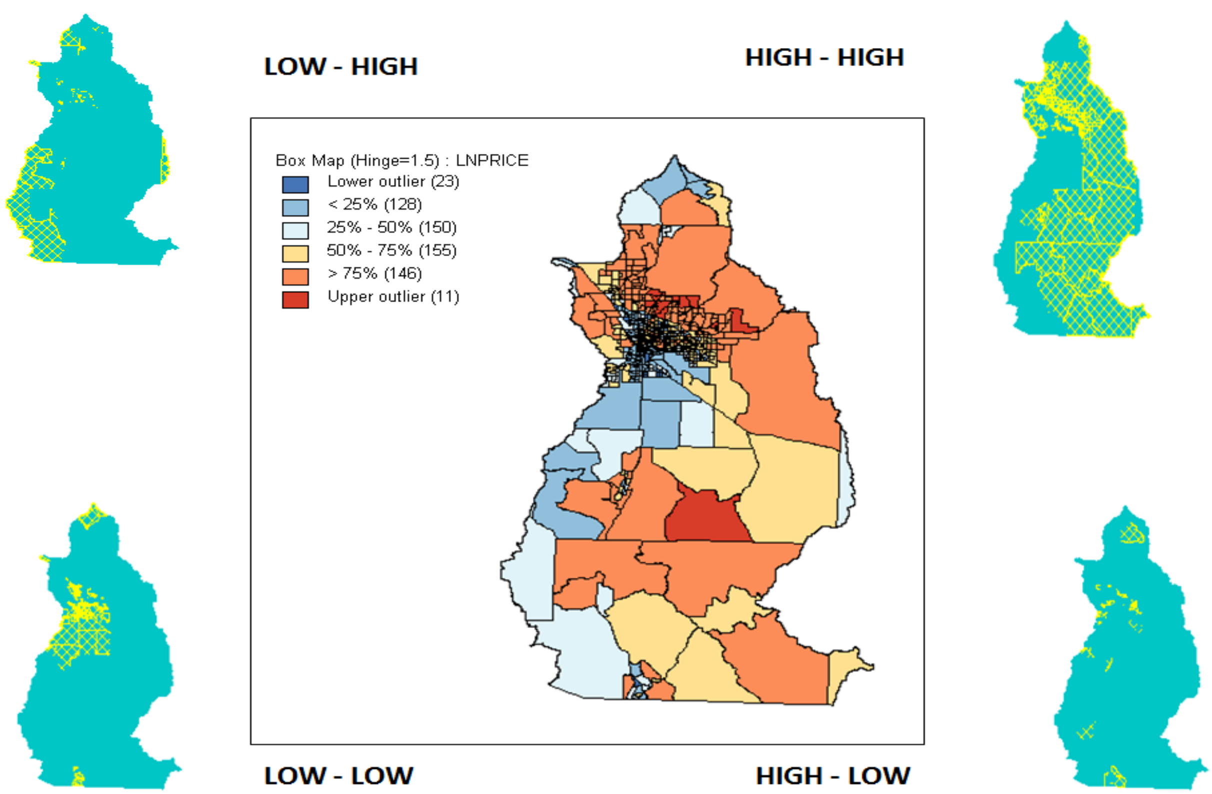
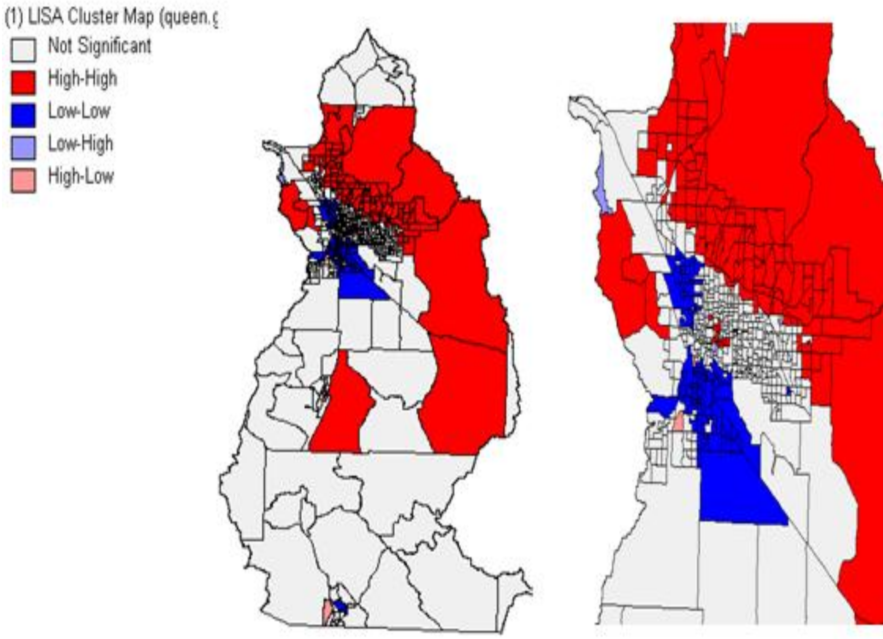
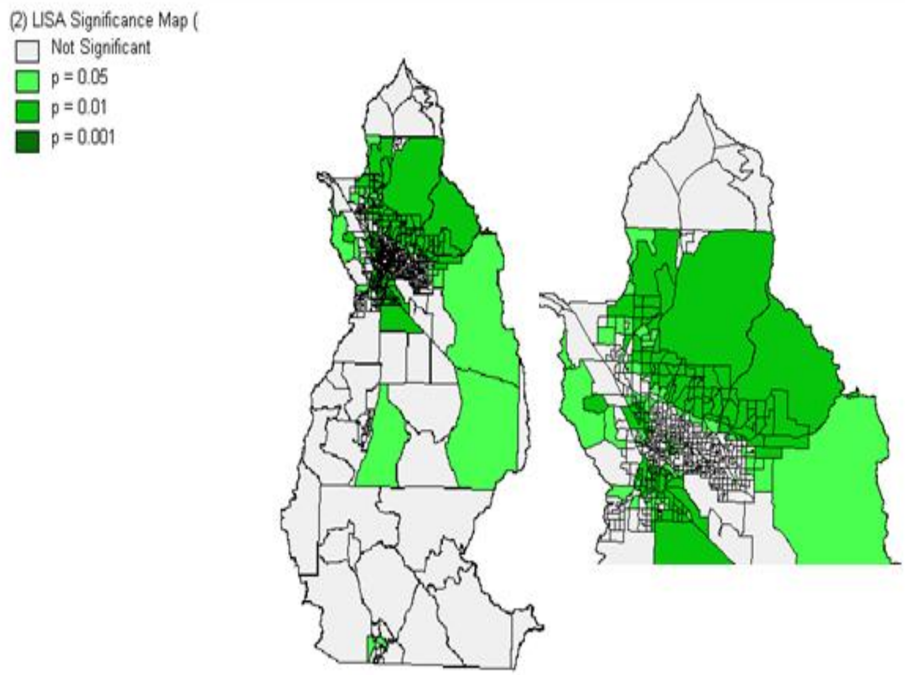


**Figure 12: LISA CLUSTER MAP & LISA SIGNIFICANCE MAPS**




**LISA CLUSTER MAP:**
*(Log $_N$ Price)*




**LISA SIGNIFICANCE MAP:**
*(Log $_N$ Price)*

## CHAPTER 4

### The Econometric Model:

### An insight into the functional form of Spatial Econometric Models:

This section is a derivative of Anselin (1999) which is dedicated to significance and development of theoretical and applied spatial econometric models.

As pointed out above, spatial econometrics deals with the situations where there is an interaction between the decision-making agents or decision-driving units in a heterogeneous system. For example, if one housing unit at the boundary of a national forest or located on foothills has a higher price due to better view and green ambience, then another unit (a little) farther from the boundary also carries the positive externalities to a certain extent, thereby raising its prices. Traditional econometrics deals with a similar effect in the time-series models by introducing a lag-dependent variable with a notion of 'shift' along the time axis whereas spatial effects are multi-dimensional with observations located irregularly in space. To visualize spatial effects in the case of hedonic pricing, the reader can think in terms of housing prices. Price of a particular housing unit – "A" is affected by the attributes all other units 'bordering' it and, the attributes of this unit "A" affects the price of all other units 'bordering' it too. Such an effect is termed as spatial dependence or spatial correlation, which is a result of interaction between different agents in space. Mathematically,

$Cov[y_i, y_j] = E[y_i y_j] – E[y_i].E[y_j] \neq 0$, for $i \neq j$ ------ **(1)** (Anselin 1999)

Where i, j refer to individual locations and $y_i$ , $y_j$ are the random variables of interest at location 'i' and 'j', respectively. The covariance (above) is no longer zero when there is a spatial arrangement, spatial interaction or spatial structure among the variables of interest. Such an effect causes OLS model parameters to be *unstable* (Irwin 2000). To solve this issue of spatial dependence, spatial lags are used. For a spatial lag, first spatial weights are created which encode the connectedness nature for existing data points. The spatial weights for each variable are recorded in a square matrix known as spatial weights matrix, **W**. For N unique observations, the weights matrix, **W** has N rows and N columns with each location "i" as a row, its neighbors as columns and corresponding non-zero elements $w_{ij}$ ($i \neq j$), thus forming a fixed (*non-stochastic*), positive matrix. Since a location "i" can never be a neighbor to it(self), the diagonal elements of **W** (i.e. $w_{ii}$'s) are always zero (Anselin 1988a, Cliff and Ord 1981). In order to correct for spatial auto-correlation exhibited by the dependent variable or the residuals, "spatial smoother" or spatial lag variable is used. A spatial lag variable is a weighted average defined as

Spatial Lag for $y_i$ : $[\mathbf{Wy}]_i = \sum_{j=1,\ldots,N} w_{ij} .y_i$ OR $\mathbf{Wy}$ (in matrix form)

An important point to consider while defining the weights matrix **W** is that a spatial weights matrix must be exogenous to avoid identification problems in econometric models (Manski 1993). In addition, that **W** is row-standardized so that the spatial lag can be interpreted as a weighted average. Essentially, row-standardization process of the weights matrix implies the for each i, $\sum_j w_{ij} = 1$, where $w_{ij} = 1$ if $i = j$ . The structure if weights matrix must be constrained in order to estimate consistent and asymptotically normal estimators.

Another way of incorporating the spatial dependence structure above, (apart from a spatially dependent variable as an additional regressor) is in the model error structure. Such a model incorporating spatial process in the regression dependence term is called as the spatial error model. Here, we focus on the *nuisance dependence* given the fact that,

Cov[**ƐƐ'**] = E [**ƐƐ'**] ≠ 0 or E [$\varepsilon_i \varepsilon_j$] ≠ 0, for i ≠ j -------- **(2)**

Moreover, it arises from the fact that spatial data is being used for analysis, potentially biasing the influence of spatial influence through the MAUP (modifiable aerial unit problem). MAUP refers to the issue of unit of observation chosen for analysis, related to aggregation-levels and choice of map scale (coarse vs. fine resolution maps) for spatial analysis (Getis 2002 UCSB workshop – lecture video).

*Spatial Regression Models*:

The two ways to incorporate spatial dependence in simple regression models are discussed as under:

*Case 1: Spatial Lag Model:*

Spatial dependence occurs due to interaction in the dependent variable i.e.

From (1) → Cov[$y_i$, $y_j$] ≠ 0, for i ≠ j

A spatial lag or a mixed regressive, auto-regressive (SAR) model is expressed as:

$$\mathbf{y} = \rho \mathbf{W y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \textbf{(3)}$$

$$\boldsymbol{\varepsilon} \sim \mathbf{N}\,(\mathbf{0}\,,\, \Theta^{2}) \qquad \textbf{(4)}$$

Where **y** is a N X 1 vector of observations for dependent variable; **Wy** is the spatial lag for **y**, again a N X 1 vector and ρ (row) is the associated coefficient; **X** is a N X k matrix, incorporating k explanatory variables and **β** is a k X 1 vector for each of the corresponding explanatory variables; **Ɛ** is the error term (N X 1).

From (3) → $$\mathbf{y} = (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{X}\boldsymbol{\beta} + (\mathbf{I} - \rho \mathbf{W})^{-1} \boldsymbol{\varepsilon} \qquad \textbf{(5)}$$

The $\mathbf{(I - \rho W)^{-1}}$ can be expanded into an infinite series, including both the explanatory variables and the error term. The spatial lag term is, thus treated as an endogenous variable making the OLS estimates biased and inconsistent due to simultaneity of spatial dependence it causes to both explanatory variables and the error term.

*Case 2: Spatial Error Model:*

The spatial dependence structure is reflected by the nuisance term or the disturbance in the linear regression model:

From (2) → $\mathrm{Cov}[\boldsymbol{\varepsilon\varepsilon'}] = E\,[\boldsymbol{\varepsilon\varepsilon'}] \neq 0$

A spatial error model is functionally expressed as:

$$\mathbf{y = X\beta + \varepsilon} \qquad \mathbf{(6)}$$

$$\mathbf{\varepsilon = \lambda W \varepsilon + \mu} \qquad \mathbf{(7)}$$

$$\mathbf{\mu \sim N\,(0\,,\,\Theta^{2})} \qquad \mathbf{(8)}$$

Equations (6) and (7) imply that

$$\mathbf{\varepsilon = (I - \lambda W)^{-1}\mu} \text{ and thus, } \mathbf{y = X\beta + (I - \lambda W)^{-1}\mu}$$

Which on further simplification boils down to the spatial Durbin model which is a more general form of expressing spatial dependence on dependent and the explanatory variables as:

$$\mathbf{y = \lambda W y + X\beta + (I - \lambda W)^{-1}\, X\beta + \varepsilon} \qquad \mathbf{(9)}$$

*Which model specification is best suited - A spatial lag model or spatial error model?*

The above question seems to be rather under-explored in terms of statistical diagnostics for determining spatial error or spatial lag be the best specification for the issue at hand. Anselin et. al (1996) provides for the *Lagrange Multiplier* approach towards the solution to this problem. Estimating the OLS regression models using the dependence structure defined by the weights matrix used gives value of LM (lag) and LM (error) along with their p-values. The more significant of the two provides for the decision rule for the most prominent of the spatial lag and spatial error processes. Kim, Phipps and Anselin (2003) is the only known hedonic analysis study to use this statistical diagnostics to determine the most appropriate spatial dependence criteria.

**Functional Form for the Hedonic analysis at hand:**

This study deals with a semi-log hedonic model, explaining price of homes as a function of its attributes. Following functional form will be followed for ordinary least squares, spatial lag / spatial error regression models:

*(1) OLS:*

$\text{Log}_N \text{ Price}_i = \alpha + \mathbf{S_i} \boldsymbol{\beta_1} + \mathbf{N_i} \boldsymbol{\beta_2} + C_i \beta_3 + \mathbf{E_i} \boldsymbol{\beta_4} + \boldsymbol{\varepsilon_i},$ i = 1 to 613.

Where,

$\text{Log}_N \text{ Price}_i$ = natural logarithm of the median price of polygon i

$\mathbf{S_i}$ is the 1 X $k_1$ vector for the structural housing variables at location i, $\boldsymbol{\beta_1}$ is its coefficient vector ($k_1$ X1).

$\mathbf{N_i}$ is the 1 X $k_2$ vector for the neighborhood variables at location i, $\boldsymbol{\beta_2}$ is its coefficient vector ($k_2$ X1).

$C_i$ is the 1 X 1 vector for the contractual variable (% Owned homes) at location i, $\beta_3$ is its coefficient.

$\mathbf{E_i}$ is the 1 X $k_3$ vector for the structural housing variables at location i, $\boldsymbol{\beta_4}$ is its coefficient vector ($k_3$ X1).

$\boldsymbol{\varepsilon_i}$ is the normally distributed error term for location i and $\alpha$ is the corresponding constant term or the intercept.

*(2) Spatial Lag Model:*

$$\text{Log}_N \text{ Price}_i = \rho.(\sum_{j=1,\ldots,N} w_{ij}. \text{ Log}_N \text{ Price}_j) + \alpha + \mathbf{S_i}\,\boldsymbol{\beta_1} + \mathbf{N_i}\,\boldsymbol{\beta_2} + C_i\,\beta_3 + \mathbf{E_i}\,\boldsymbol{\beta_4} + \boldsymbol{\varepsilon_i}, \quad i = 1 \text{ to } 613,\ i \neq j$$

Where,

$\text{Log}_N$ Price $_i$ = natural logarithm of the median price of polygon i

$\sum_{j=1,\ldots,N} w_{ij}.$ $\text{Log}_N$ Price $_j$ is the spatial lag for the dependent variable and $\rho$ is its coefficient.

$\mathbf{S_i}$ is the 1 X $k_1$ vector for the structural housing variables at location i, $\boldsymbol{\beta_1}$ is its coefficient vector ($k_1$ X1).

$\mathbf{N_i}$ is the 1 X $k_2$ vector for the neighborhood variables at location i, $\boldsymbol{\beta_2}$ is its coefficient vector ($k_2$ X1).

$C_i$ is the 1 X 1 vector for the contractual variable (% Owned homes) at location i, $\beta_3$ is its coefficient.

$\mathbf{E_i}$ is the 1 X $k_3$ vector for the structural housing variables at location i, $\boldsymbol{\beta_4}$ is its coefficient vector ($k_3$ X1).

$\boldsymbol{\varepsilon_i}$ is the normally distributed error term for location I and $\alpha$ is the corresponding constant term or the intercept.

*(3) Spatial Error Model:*

$$\textbf{Log}_{\textbf{N}}\ \textbf{Price} = \alpha + \textbf{S}\,\boldsymbol{\beta_1} + \textbf{N}\,\boldsymbol{\beta_2} + \textbf{C}\,\beta_3 + \textbf{E}\,\boldsymbol{\beta_4} + (\textbf{I} - \lambda\textbf{W})^{-1}\mu, \quad \mathcal{E} = \lambda\textbf{W}\mathcal{E} + \mu$$

Where,

**Log $_N$ Price** = N X 1 vector for natural logarithm of the median price.

**S** is the N X $k_1$ matrix for the structural housing variables; $\boldsymbol{\beta_1}$ is its coefficient vector ($k_1$ X1).

**N** is the N X $k_2$ matrix for the neighborhood variables; $\boldsymbol{\beta_2}$ is its coefficient vector ($k_2$ X1).

**C** is the N X 1 matrix for the contractual variable (% Owned Homes); $\beta_3$ is its coefficient.

**E** is the N X $k_3$ matrix for the structural housing variables; $\boldsymbol{\beta_4}$ is its coefficient vector ($k_3$ X1).

$\mathcal{E}_i$ is the non-normal error matrix exhibiting spatial interaction in the above functional form with $\lambda$ being its coefficient and W as the weights matrix; $\boldsymbol{\mu}$ distributed normally.

## Variable Choice:

The natural logarithm of the median home values in each census block-group is explained as a function of various physical, neighborhoods and environmental attributes of homes, also known as regressors. The regressors in the econometric models to follow have been classified into four types: structural, neighborhood and contractual; land ownership; land cover and biodiversity variables. Almost all the variables have been discussed in the prior section with their descriptive statistics and respective distributions except: "Ratio - %_Dev / %_Pvt" has been created to evaluate the price responsiveness of developed share of the privately owned land area. Apart from the biodiversity variable is an index created by summing - the product of number of species by respective vegetations' percentage cover of land - over all the vegetation types found in that particular parcel. More number of vegetation types and correspondingly larder number of species of each will result in a greater value of the self-created index. Such an index is thus believed to capture the biodiversity of the area. Higher index value implies rich biodiversity and vice versa.

In addition, this study attempts at defining various spatial regimes or submarkets to address the issue of spatial heterogeneity. This allows for a different price structure for each of these pre-defined submarkets rather than a uniform price structure for the entire area under study. The demarcation of these spatial regimes is done usually done by introducing dichotomous and/or polychotomous variables, which further may be interacted with other explanatory variables in the estimated models. First to test for presence of spatial regimes, it has been hypothesized that residents of highly urbanized areas have different preferences for green space than the ones in rural areas. To statistically test for this, first a dummy variable is created for highly urbanized core as a set where %_Developed Land > 10% AND no Public Land in the entire CBG. The chow test (F-test) conducted later, rejects the above hypothesis of structural difference between highly urbanized core versus rural areas. Given the fact that there are no spatial regimes existant in the region in an attempt towards controlling for the issue of spatial heterogeneity, dummy variables have been introduced for census block-groups or polygons to define submarkets that lie on the U.S.-Mexico border (D_ON_THE_BORDER); for polygons that belong to Catalina Foothills and Tenque Verde School Districts (D_SCHL_DISTRICT) and for polygons that lie under Pima and Pinal Counties (D_PIMA_PINAL). The rationale behind controlling for each of the above submarkets pertains to the specificity of the region under study. The value of homes near the border region is expected to be lesser than the value of homes away from it. Similarly, the above two school districts are considered as ‘elite’ areas and are marketed highly by realtors while selling homes. The distinction between Pima and Pinal Counties *versus* Santa Cruz and Cochise Counties follows from the stark distinction in population and house densities between the two. The former are highly populated areas as compared to the latter ones.

**Various Model Specifications:**

There are a total of five econometric models specified in this study. Basis of model specification remains consistent with the literature: ad hoc, considering important issues of multicollinearity and omitted variable bias while maximizing the goodness of fit. Each model has its particular significance. Model-I through Model V move from OLS to spatial econometric estimation. Model I controls for all the structural, neighborhood and contractual variables along with dummies for the three spatial regimes defined earlier. Model II advances to discern the simultaneous effects of different types of land ownership whereas Model III indicate which land cover types provide for economic benefits in terms of premiums in home values. Model IV attempts to figure out the premiums paid for bio-diversity rich areas. Model V encompasses the land ownership and land cover variables together, in order to test consumer preferences for ‘brand’ of the land versus its actual physical attributes in terms of green space. It is noteworthy that all the above models are being specified keeping in consideration the correlation coefficients of each of the regressor, thus minimizing the multicollinearity issue and maximizing the goodness of fit.

**Results:**

**The *best-suited* econometric model:**

Spatial Lag vs. Spatial Error Models:

The estimated models (Model I through V) within the classic regression context reveal that spatial lag models out-run spatial error models. The decision rule for such a specification selection is given by the LM test or the score test. It can be observed from all five (5) OLS models that (Robust) LM - lag is more significant than (Robust) LM - error. Hence, following this decision rule further estimation for all five (5) models is done using spatial structure dependence rather than spatial dependence among the unobserved or omitted variables.

**Other (relevant) Specification Tests:**

About the classic regression models the adjusted – $R^2$ ranges from 0.746 to 0.775, which given estimation on a cross-sectional data (N = 613) can be regarded as a decent goodness of fit. For the spatial lag models estimated with the same cross-sectional data, the log-likelihood value, Akaike info criterion and Schwarz Criterion provide the goodness-of-fit measure. The log-likelihood value should be maximized and the latter two should be minimized in order to attain a better fit of the models. The Schwarz criterion is more stringent and reliable as it has larger penalty for errors than the Akaike info criterion. In accordance with the above decision rule Models II through V perform reasonably equivalently in terms of goodness-of-fit.

The OLS estimation models provides for the Jarque-Bera test statistic to test normality of errors. The hypothesis of the test – normal errors is rejected across the five (5) models. This is the first indication towards spatial component not being controlled for in the models, which leads to non-normality of errors. Further the Lagrange Multiplier for spatial lag and spatial error structures confirm and identify the most appropriate form of spatial dependence.

The classic regression models perform tests for multicollinearity and heteroskedasticity issues. For multicollinearity, the condition number that, by convention should not go over 30 to avoid this problem, here ranges from 50 – 65. This implies that multicollinearity might be an issue but at the same time, a modeler must also avoid omitted variable bias. Model I, which accounts only for the first set of variables has the least goodness-of-fit and still suffers from a high conditional number of 50.11. Therefore, a better fit for the models is compensated by a slight increase in the condition number. Three tests for heteroskedasticity used are Breush-Pagan test; Koenker-Basset Test and White's Test. The decision rule is that if two of the tests reject the null hypothesis of homoskedasticity then there is an issue of heteroskedasticity. The problem of heteroskedasticity remains in all five (5) OLS as well as spatial lag models. The coefficient estimates from the models having this issue are still consistent but they lose on the grounds of efficiency. Another important issue in consider for the analysis at hand, is variable endogeneity especially with regards to regressors controlling for income and demographics of the residents. Therefore, in order to control for heteroskedasticity and variable endogeneity IV-2SLS robust Spatial Lag models are estimated for all five (5) specifications mentioned above.

Further, the likelihood ratio test in all the five (5) spatial lag models suggests that each of them out-perform their counterpart OLS models.

**Estimation:**

This section summarizes the responsiveness of coefficients for explanatory variables across the five (5) IV-2SLS robust spatial lag models. First, it is noticeable that the spatial lag coefficient - ρ (rho) is significant at 99% C.L. across Model I through V and ranges from 0.18 to 0.21.

The coefficient of %_owner is found to be negative and significant at 99% C.I. in Models I through V. This is an interesting find as the extent of ownership in an area is usually, if not provide a premium, is not expected to discount the housing prices. %_Vacant is insignificant across all models while, %_Mobile Homes have a negative and significant effect on $Log_N$ *Price*, both at 99% C.L. NOPHONE (% homes w/o phone) and INCPLUMB (% homes w/o complete plumbing) do not have a statistically significant effect on home values. Another significant premium (@99% C.I.) on home values comes from heritage homes (PCOLD39) built before 1939 whereas newly built homes (within last 2 years) do not have any significant effect. Also, the models reveal that higher percentage of minority and less earning (< $25k) population has a discount on the home values whereas higher earners provide for a significant premium. Other variables like Persons/Home and % FLOOD (under FEMA flood zone) do not have a significant effect on the home values.

The dummy variable for Tucson Intl. Airport and Davis-Monthan Military Airbase reveals a definite disamenity to home buyers, to which they respond by paying lower prices. Other dichotomous variables included discerning the effect of sub-markets reveal that real estate market is definitely different at the U.S-Mexico international border and Catalina Foothills/Tanque Verde from other regions. The coefficient on PIMA/PINAL vs. SANTA

CRUZ/COCHISE counties is not significant in four out of the five models – ruling out the possibility of separate markets here.

*Does the Green Space Matter?*

*Land Ownership Variables*: Model II shows that all the land ownership types, i.e. state trust land, NPS land, USFS land and local parks have a positive, significant coefficient estimates. Here, coefficients of %_State Ttrust, %_NPS, %_BLM and % USFS land are significant at 99% C.L. whereas the coefficient of %_local parks is significant at 90% C.L. All the land ownership variables reappear in the Model V, where %_Forest Service land is not included for multicollinearity issues. In Model V again all land ownership variables, except for %_Local Parks are statistically significant and bear a positive premium on home values. %_Local Parks is found to be insignificant in Model IV too. Interestingly, although state trust land is different in structure and is bound to be developed at some point in future, does have an economic premium. Further, it can be concluded here that among the land ownership variables of green space %_State Trust, %_National Park, %_BLM and %_USFS have a definite premium on housing prices.

*Land Cover Variables*: The land cover variables, all or a subset of them, are included in Models III, IV and V. The variables in each case are included keeping in mind- the purpose of the model specification, multicollinearity conditional number and the goodness-of-fit. In Model III among nine (9) of these variables only three (3) have a positive and significant coefficient, i.e. %_Shrub/Scrub, %_Grasslands and %_Pasture. In Model IV only %_Developed land does have a negative, significant (@90% C.I.) impact on home values while in Model V %_Grasslands has a positive and significant premium on home values @95% C.L. where %_Pasture has a apositve and significant coefficient estimate @~95% C.L. (%_Shrub/Scrub was not included in Model V). Therefore, shrubs, scrubs, grasslands and pasture-lands are considered as natural amenities among all other types in a semi-arid region.

*Biodiversity Potential*: The biodiversity index, in Model IV, has a positive and significant (@99% C.I.) coefficient estimate, pointing out towards the fact that biodiversity-rich areas have higher home values. This carries forward the evidence found in literature for higher home prices for quality vegetation rather than just quantity in terms of proximity or percent land cover.

*A note on Land Ownership vs. Land Cover Variables*: The significant coefficient estimates of green space in Model V reveal that if the ownership of land and land cover type were to compete in the real estate market, the actual physical condition of land will have a higher premium. This argument is supported by the fact that coefficient estimate for %_Grasslands - land cover variable is much higher than any of its significant land ownership counterpart. %_Pasture, too, has a coefficient estimate approximately equal to %_BLM land and higher than other land ownership variables. This argument being inconclusive will again be explored below under the umbrella of marginal effects for this model.

The sections that follow, report marginal effects and an estimation bias of OLS versus Spatial Lag models for continuous variables of interest (Green Space) and the aforementioned dummy variables.

**Marginal Benefits of the Green Space: An Interpretation for model estimates (1)**

It is important to note that spatial lag models provide a powerful tool to demarcate localized and neighborhood effects termed as "direct" and "indirect" marginal effects. The idea behind the neighborhood effects also termed as 'spatial spillovers' is that cross-partial derivatives, of the dependent variable with respect to the independent variables of interest, are no longer zero (LeSage and Pace, 2009).

Mathematically,

From (5) → $\mathbf{y} = (\mathbf{I} - \rho\mathbf{W})^{-1}\,\mathbf{X}\boldsymbol{\beta} + \mathbf{v}$ **(10)**

Where, $\mathbf{v} = (\mathbf{I} - \rho\mathbf{W})^{-1}\,\boldsymbol{\varepsilon}$

Mathematically, the marginal effects as pointed out above can be demarcated as under:

**Direct Effect at location i:** $dy_i/dX_{ik}$ **where i = 1 to 613** **(11)**

**Indirect Effect at location i:** $\sum_{j=2,\ldots,n} dy_j/dX_{jk}$ **where j = 1 to 613 and i ≠ j** **(12)**

**Total Effects at location i = [Direct Effect] + [Indirect Effect]**

$$dy_i/dX_k = [dy_i/dx_{ik}] + [\textstyle\sum_{j=2,\ldots,n} dy_j/dx_{jk}] \quad \textbf{(13)}$$

Where, $\mathbf{x}_{ik}$ is a (n x 1) column vector of one of the explanatory variables; $\mathbf{y}_{ik}$ column vector of one of the dependent variable, both at location i.

From (10) $dy/dX_k = (\mathbf{I} - \rho\mathbf{W})^{-1} * \beta_k$; **(14)**

Where, $\mathbf{(I - \rho W)^{-1}}$ is a (n x n) matrix and $\beta_k$ is the estimated coefficient, a constant and a scalar.

Now, the expression $[\mathbf{(I - \rho W)^{-1}} * \beta_k]$ in (11) can be re-written as:

$$\mathbf{(I - \rho W)^{-1} * \beta_k = (I + \rho W + (\rho W)^2 + (\rho W)^3 + \ldots) * \beta_k} \quad \mathbf{(15)}$$

$$\mathbf{= \{[DIRECT\ EFFECT] + [INDIRECT\ EFFECTS]\}} \quad \mathbf{(16)}$$

The direct effect or the statistically significant own-partial derivative is the impact on the dependent variable by changes in the characteristics of the own region. On the other hand, the indirect effect or the non-zero cross-partial derivative can be interpreted as the impacts from changes in an own region characteristic on other regions ***OR*** changes in other regions' characteristics on the own region. Technically, from equation (15) for any one location - sum of the diagonal elements of the L.H.S. multiplied by the estimated coefficient ($\beta_k$) is termed as direct marginal effects and the sum of non-diagonal elements multiplied by $\beta_k$ is termed as the indirect effects. To evaluate these for a location "i" means summing these diagonal and non-diagonal elements across "$i^{th}$" row of the (n xn) matrix (613 x 613) in this case). It must be noted here that since the diagonal elements of the weights matrix **W** are all zeros there is no spatial spillovers from first-order neighbors. But, since the diagonal elements of $\mathbf{W^2}$ are not equal to zero, there are spatial spillovers from second-order neighbors and similarly from third-order neighbors (from $\mathbf{W^3}$) and so on. Although, since **W** is row-standardized, it is fairly easy to visualize that its square, cube and so on, will yield to a very small number OR a very small "feedback effect" as literature terms it (LeSage and Pace 2009). The spatial spillovers die very fast and the simulation studies (LeSage and Pace, 2009) report that the these die out to almost zero till third-order or fourth-order neighbors. The direct and indirect effects in this thesis are calculated using the spatial econometrics package in R. The associated standard errors and p-values reported are evaluated using simulation techniques by the canned package in R itself.

For the sake of simplicity, in the following part of this report, direct effects will be referred to as $\beta_{DIRECT}$; indirect effects as $\beta_{INDIRECT}$ and total effects as $\beta_{TOTAL}$.

Now, in case of a semi-log model at hand,

$$\beta_{DIRECT} = d\,(\log_N Price_i) / dX_{ik} \quad \text{where } i = 1 \text{ to } 613 \tag{17}$$

$$\beta_{INDIRECT} = \sum_{j=2,\ldots,n} d\,(\log_N Price_i) / dX_{jk} \quad \text{where } j = 1 \text{ to } 613 \text{ and } i \neq j. \tag{18}$$

$$\beta_{TOTAL} = \beta_{DIRECT} + \beta_{INDIRECT} \tag{19}$$

*Interpretation:* Given the semi-logarithmic functional form of the models, $\beta_{DIRECT}$ here represents the continuous rate of change in home values at location "i" given an infinitesimal change in the characteristic at "i". Similarly, $\beta_{INDIRECT}$ is the continuous rate of change in home values at "i", given an infinitesimal change in the characteristics of all locations, other than "i". Further, a bias is to be calculated from OLS vs. Spatial Lag using $\beta_{DIRECT}$ and not $\beta_{TOTAL.}$

**% BIAS (Spatial Lag vs. OLS) =**

$$BIAS = \frac{\beta_{Direct} - \beta_{OLS}}{\beta_{OLS}} \text{ X } 100 \tag{20}$$

The direct, indirect and total marginal effects, along with the potential bias for OLS vs. Spatial Lag is calculated and discussed as under, where the model results are listed in Appendix A:

## MODEL – II:

***Table 8: Marginal Effects - Land Ownership Variables***

| Variables | DIRECT | INDIRECT | TOTAL |
|---|---|---|---|
| **%_Trust** | 0.006 | 0.001 | 0.007 |
| **%_NPS** | 0.009 | 0.002 | 0.011 |
| **%_BLM** | 0.015 | 0.004 | 0.018 |
| **%_USFS** | 0.005 | 0.001 | 0.006 |
| **%_Local Parks** | 0.007 | 0.002 | 0.008 |

***Table 9: Simulated p-values for MEs - Land Ownership Marginal Effects***

| Varaibles | DIRECT | INDIRECT | TOTAL |
|---|---|---|---|
| **%_Trust** | 0.000 | 0.007 | 0.000 |
| **%_NPS** | 0.000 | 0.017 | 0.000 |
| **%_BLM** | 0.001 | 0.036 | 0.001 |
| **%_USFS** | 0.001 | 0.025 | 0.001 |
| **%_Local Parks** | 0.058 | 0.121 | 0.060 |

The above marginal effects suggest that major impact lies in the direct effects and rather a very small, although significant except %_Local Parks, comes from the spatial spillovers. It is also observable here that where the coefficient estimate for %_Local Parks is insignificant there is an improvement on the significance levels of direct and total effects of this variable and can be considered rather significantly different from zero (0). Overall, a direct, indirect and total continuous rate of change in home values is reported to be (0.06, 0.001, 0.007) respectively for a marginal change in respective cover of %_State Trust land. On similar grounds thedirect, indirect and total marginal effects respectively are (0.009, 0.002, 0.011) for %_NPS, (0.015, 0.004, 0.018) for %_BLM, (0.005, 0.001, 0.006) for %_USFS and (0.007, 0, 0.008) for %_Local Parks.

The bias from OLS estimates is reported to vary from -7.9% to +4.4%.

## MODEL - III

**Table 10: Marginal Effects - Land Cover Variables**

| VARIABLES | DIRECT | INDIRECT | TOTAL |
|---|---|---|---|
| %_Open Water | 0.067 | 0.016 | 0.083 |
| %_Developed Land | | | |
| % Barren Land | -0.001 | 0.000 | -0.001 |
| % Deciduous Forest | -0.010 | -0.002 | -0.012 |
| % Evergreen Forest | 0.011 | 0.002 | 0.013 |
| % Shrub/Scrub | 0.015 | 0.004 | 0.019 |
| % Grassland | 0.048 | 0.011 | 0.059 |
| % Pasture | 0.016 | 0.004 | 0.020 |
| % Cultivated | 0.013 | 0.003 | 0.016 |
| % Wetlands | -0.319 | -0.075 | -0.395 |

**Table 11: Simulated p-values for MEs - Land Cover Variables**

| VARIABLES | DIRECT | INDIRECT | TOTAL |
|---|---|---|---|
| %_Open Water | 0.446 | 0.504 | 0.454 |
| %_Developed Land | | | |
| % Barren Land | 0.818 | 0.881 | 0.829 |
| % Deciduous Forest | 0.683 | 0.709 | 0.686 |
| % Evergreen Forest | 0.304 | 0.344 | 0.306 |
| % Shrub/Scrub | 0.000 | 0.001 | 0.000 |
| % Grassland | 0.022 | 0.050 | 0.022 |
| % Pasture | 0.024 | 0.052 | 0.024 |
| % Cultivated | 0.277 | 0.296 | 0.273 |
| % Wetlands | 0.206 | 0.252 | 0.206 |

Model III reveals an economic significance for %_Shrubs/Scrubs, %_Grasslands and %_Pasture. The direct, indirect and total marginal effects, respectively for %_Shrubs/Scrubs are (0.015, 0.004, 0.019), %_Grasslands are as high as (0.048, 0.011, 0.059) and %_Pasture are (0.016, 0.004, 0.020). The potential bias from OLS models varies from -6% to +9%.

## MODEL IV:

**Table 12: Marginal Effects - Biodiversity Index**

| VARIABLES | DIRECT | INDIRECT | TOTAL |
|---|---|---|---|
| PCT_LOCAL PARKS | 0.003 | 0.001 | 0.004 |
| PCT_BARREN | -0.009 | -0.002 | -0.011 |
| PCT_DEVELOPED | -0.006 | -0.001 | -0.008 |
| BIODIVERSITY INDEX | 0.003 | 0.001 | 0.004 |

**Table 13: Simulated p-values for MEs - Biodiversity Index**

| VARIABLES | DIRECT | INDIRECT | TOTAL |
|---|---|---|---|
| PCT_LOCAL PARKS | 0.286 | 0.380 | 0.294 |
| PCT_BARREN | 0.439 | 0.503 | 0.444 |
| PCT_DEVELOPED | 0.056 | 0.152 | 0.063 |
| BIODIVERSITY INDEX | 0.006 | 0.014 | 0.004 |

The main purpose of specifying MODEL IV is to quantify the premiums paid on the quality of vegetation. This takes the analysis a step further given the fact that it can be concluded from here that people not only pay a premium for the size of the amenity or proximity to it, but also for the factors like biomass, biodiversity etc. The direct, indirect and total effects for the biodiversity index are found to be (0.003, 0.001, 0.004). Another observation here is that %_Developed land has a penalty, implying a premium for open space but this result cannot be stated as robust given the fact that %_Developed land coefficient estimate is highly unstable. The potential bias from OLS models here is -9%.

# MODEL V:

**Table 14: MARGINAL EFFECTS - Land Ownership vs. Land Cover variables**

| | DIRECT | INDIRECT | TOTAL |
|---|---|---|---|
| ***Land OwnerShip Variables*** | | | |
| % State Trust | 0.005 | 0.001 | 0.007 |
| % National Park | 0.009 | 0.002 | 0.012 |
| % BLM | 0.012 | 0.003 | 0.016 |
| % Forest Service | | | |
| % Local Parks | 0.005 | 0.001 | 0.006 |
| ***Land Cover Variables*** | | | |
| %_Open Water | -0.027 | -0.007 | -0.035 |
| %_Developed Land | | | |
| % Barren Land | -0.002 | 0.000 | -0.002 |
| % Deciduous Forest | 0.002 | 0.000 | 0.002 |
| % Evergreen Forest | 0.010 | 0.003 | 0.012 |
| % Shrub/Scrub | | | |
| % Grassland | 0.049 | 0.013 | 0.061 |
| % Pasture | 0.011 | 0.003 | 0.014 |
| % Cultivated | 0.016 | 0.004 | 0.020 |
| % Wetlands | -0.270 | -0.070 | -0.340 |

**Table 15: Simulated p-values for MEs - Land Ownership vs. Land Cover Variables**

| | DIRECT | INDIRECT | TOTAL |
|---|---|---|---|
| ***Land OwnerShip Variables*** | | | |
| **% State Trust** | 0.005 | 0.001 | 0.007 |
| **% National Park** | 0.009 | 0.002 | 0.012 |
| **% BLM** | 0.012 | 0.003 | 0.016 |
| **% Forest Service** | | | |
| **% Local Parks** | 0.005 | 0.001 | 0.006 |
| ***Land Cover Variables*** | | | |
| **%_Open Water** | -0.027 | -0.007 | -0.035 |
| **%_Developed Land** | | | |
| **% Barren Land** | -0.002 | 0.000 | -0.002 |
| **% Deciduous Forest** | 0.002 | 0.000 | 0.002 |
| **% Evergreen Forest** | 0.010 | 0.003 | 0.012 |
| **% Shrub/Scrub** | | | |
| **% Grassland** | 0.049 | 0.013 | 0.061 |
| **% Pasture** | 0.011 | 0.003 | 0.014 |
| **% Cultivated** | 0.016 | 0.004 | 0.020 |
| **% Wetlands** | -0.270 | -0.070 | -0.340 |

The findings of Model V are rather interesting. It seems here that the Land Cover variables or the actual attribute of land has an overall higher premium than the Land Ownership variables in terms of direct effects. Whereas, the spillover effects or the neighborhood effects are clearly dominated by the Land Ownership variables when it comes to premiums paid for home values and there is essentially no premium for green space in the neighborhood coming from the actual physical attribute of the land. The potential bias from the OLS model here ranges from -7.2% to +10%.

**Dummy Variables (for Submarkets): An Interpretation of the model estimates (2)**

It is important to consider the marginal benefits of the dichotomous variables separately, as the above interpretation used for continuous variables are not applicable for discrete one. Here, a *continuous rate of change in prices* for a **discrete** change in these is evaluated across the models and **NOT** a marginal continuous change. Here, the information provided for premium or penalty can be interpreted as a change in home values when switching from Market A (D=0) to Market B (D=1). Interestingly, there are direct, indirect and total effects reported for these dummy variables. The direct effect can be interpreted as continuous rate of change in home values when the location itself switches from Market A to Market B. The indirect effects, on the other hand, imply a continuous rate of change in home values when the neighboring units switch from Market A to market B.

Mathematically,

$\text{Log}_N \text{Price}_i |_{D=1} = \mathbf{(I - \rho W)^{-1}.}\ \beta_D.1 + \mathbf{X\beta}$

$\text{Log}_N \text{Price}_i |_{D=0} = \mathbf{(I - \rho W)^{-1}.}\ \beta_D.0 + \mathbf{X\beta}$

Therefore, for location i:

$\text{Log}_N \text{Price}_i |_{D=1} - \text{Log}_N \text{Price}_i |_{D=0} = \mathbf{(I - \rho W)^{-1}}.\ [(\beta_D.1 + \mathbf{X\beta}) - (\beta_D.0 + \mathbf{X\beta})]$

$\text{Log}_N (\text{Price}_i |_{D=1} / \text{Price}_i |_{D=0}) = \beta_D.\ \mathbf{(I - \rho W)^{-1}\ .}\beta_D$ **(21)**

By the way of model specification, here the entire study area has been broken into different sub-markets or spatial regimes. Models I – V test for existence of submarkets (i) on the U.S.-Mexico

Border; (ii) within Catalina Foothills + Tanque Verde school districts; (iii) near Tucson Intl. Airport and Davis Monthan International Airbase, (iv) between Pima and Pinal County vs. Santa Cruz and Cochise County. Dummy variables are set up to test for the above four (4) possible submarkets. Estimation of the models suggest that there is a significant difference in housing market structure on the border and within the school districts but the counties does not form two separate submarkets. Apart from this, TIA_DM_MAS that controls for the industrial existence near some housing areas also tests for existence of a submarket. Model estimation results reveal that its coefficient is significant across models. Hence, by the way of modeling specification, three separate submarkets are found in the study area: (i) on the border; (ii) within the Catalina Foothills and Tanque Verde school districts; (iii) near the Tucson International Airport and Davis Monthan Military Airbase. The associated marginal effects, their p-values and a discussion follow as under:

**Table 16: MEs and their p-values for Dichotomous Variables - Model I**

| VARIABLE | DIRECT | p-value | INDIRECT | p-value | TOTAL | p-value |
|---|---|---|---|---|---|---|
| D_ON THE BORDER | -0.34 | 0.00 | -0.07 | 0.00 | -0.41 | 0.00 |
| TIA_DM_MAS | -0.19 | 0.00 | -0.04 | 0.00 | -0.23 | 0.00 |
| D_SCHL_DISTRICT | 0.16 | 0.01 | 0.03 | 0.01 | 0.19 | 0.01 |
| D_PIMA_PINAL | -0.27 | 0.00 | -0.06 | 0.00 | -0.33 | 0.00 |

**Table 17: MEs and their p-values for Dichotomous Variables - Model II**

| VARIABLE | DIRECT | p-value | INDIRECT | p-value | TOTAL | p-value |
|---|---|---|---|---|---|---|
| D_ON THE BORDER | -0.19 | 0.06 | -0.05 | 0.07 | -0.24 | 0.05 |
| TIA_DM_MAS | -0.16 | 0.00 | -0.04 | 0.02 | -0.20 | 0.00 |
| D_SCHL_DISTRICT | 0.18 | 0.00 | 0.04 | 0.01 | 0.23 | 0.00 |
| D_PIMA_PINAL | -0.05 | 0.49 | -0.01 | 0.56 | -0.06 | 0.50 |

**Table 18: MEs and their p-values for Dichotomous Variables - Model III**

| VARIABLE | DIRECT | p-value | INDIRECT | p-value | TOTAL | p-value |
|---|---|---|---|---|---|---|
| D_ON THE BORDER | -0.37 | 0.00 | -0.09 | 0.00 | -0.46 | 0.00 |
| TIA_DM_MAS | -0.16 | 0.00 | -0.04 | 0.02 | -0.20 | 0.00 |
| D_SCHL_DISTRICT | 0.19 | 0.00 | 0.04 | 0.01 | 0.23 | 0.00 |
| D_PIMA_PINAL | 0.00 | 0.81 | 0.00 | 0.84 | 0.00 | 0.82 |

**Table 19: MEs and their p-values for Dichotomous Variables - Model IV**

| VARIABLE | DIRECT | p-value | INDIRECT | p-value | TOTAL | p-value |
|---|---|---|---|---|---|---|
| D_ON THE BORDER | -0.25 | 0.02 | -0.05 | 0.09 | -0.31 | 0.03 |
| TIA_DM_MAS | -0.16 | 0.00 | -0.03 | 0.03 | -0.19 | 0.00 |
| D_SCHL_DISTRICT | 0.15 | 0.01 | 0.03 | 0.02 | 0.18 | 0.01 |
| D_PIMA_PINAL | 0.13 | 0.08 | 0.03 | 0.14 | 0.16 | 0.08 |

**Table 20: MEs and their p-values for Dichotomous Variables - Model V**

| VARIABLE | DIRECT | p-value | INDIRECT | p-value | TOTAL | p-value |
|---|---|---|---|---|---|---|
| D_ON THE BORDER | -0.37 | 0.01 | -0.10 | 0.05 | -0.47 | 0.01 |
| TIA_DM_MAS | -0.17 | 0.00 | -0.04 | 0.02 | -0.21 | 0.00 |
| D_SCHL_DISTRICT | 0.18 | 0.00 | 0.05 | 0.00 | 0.23 | 0.00 |
| D_PIMA_PINAL | -0.17 | 0.23 | -0.04 | 0.27 | -0.21 | 0.23 |

The very first observation for submarket variables is that the results obtained here, reject the hypothesis of submarkets between PIMA-PINAL vs. SANTA CRUZ – COCHISE County. Further, there is a definite submarket at the US-MEXICO BORDER (<5 miles) where there is a continuous rate of decline in the home values by at least 34% if the CBG itself switches to this submarket (direct effect) and also, there is a fall in prices by at least 7% if neighbors shift into this submarket (indirect effect). For a particular CBG, being in the submarket of Tucson Intl. Airport and Davis Monthan Airbase, a similar interpretation follows that if the CBG itself falls into this submarket there is a decline in home values by at least 16% and if neighbors fall into this submarket there will be a decline of at least 3% in home values. The dummy for Catalina Foothills and Tanque Verde reveals a jump in prices by at least 15% is the CBG itself switches to these submarkets and of at least 3% if neighbors fall into it.

## CHAPTER 5

### *Discussion and Policy Implications:*

This study deals with a semi-arid region in southern Arizona, along the Santa Cruz River Basin. Importance of this study derives from the fact that numerous budget-intensive programs are in place to resolve water and land allocation issues. The purpose of these is to conserve and keep the environment healthy. This project aims at answering the question: do land-use patterns in a river watershed (lying in the semi-arid region of southern Arizona) drive housing prices? By the way of housing prices, consumer preferences for natural amenities are derived for which marginal economic benefits are being calculated. Hedonic pricing method has been employed for the above purpose with an application of spatial econometrics. Census data aggregated as the block group level has been used with 613 observations in total. Tools of spatial econometrics can test if or not the areas under study can be considered as isolated units or they possess certain degree of connectedness. Moran's I test statistic has been calculated for this purpose which confirms such a connection among different units, also termed as spatial autocorrelation. Degree of spatial connectedness is found to be 0.60, not controlling for which can bias the results obtained by traditional hedonic models that employ OLS estimation technique. Existence of localized submarkets has also been detected in the process. Moving towards the border and near Tucson International Airport / Davis Monthan, military airbase significantly drops housing prices whereas Catalina Foothills and Tanque Verde school districts have high premiums to housing values. The models correcting for spatial effects used here are IV-2SLS robust models which are capable of controlling for other statistical issues like heteroskedasticity and variable endogeneity.

The variables included in the study control for the structural, neighborhood, contractual attributed and natural amenity attributes of homes. The natural amenity variables are classified as land-ownership variables, land cover variables and a biodiversity index. Land ownership variables control for the percentage of land that lies under State Trust, National Park Service, US Forest Service, BLM and Local Parks. Land cover variables control for percentage of land in a CBG classified as open water surface, barren land, deciduous forest, evergreen forest, shrubs, scrubs, grasslands, pastures, cultivated lands and wetlands. The estimated models here reveal that designated State Trust, NPS and USFS lands do have a premium on home values. There premiums for this type of land exist for at least less than half the CBGs. The % age cover for each these land types is highly skewed with a median of zero (0). However, the CBGs that have certain amount of these land types to begin with, do have response from home consumers upon a marginal change in the respective attributes.

The Land Ownership variables in Model II are all found to produce a positive externality on home values resulting in economic premiums. Comparing across models %_State Trust Land, %_NPS land, %_BLM Land and %_USFS Lands are found to have definite premiums where %_Local Parks are significant at slightly lower significance levels. Similarly, the land cover variables – shrub, scrubs, grassland, and pasture land cover types have an economic premium for the area. Both direct and indirect marginal effects are evaluated for all these variables demarcating localized effects and neighborhood effects or 'spatial spillovers.

The biodiversity index being statistically significant and positive implies premiums for home values. This variable addresses the quality vs. quantity issue and adheres to the hypothesis recently found in literature that homebuyers not only value size and proximity of green space but also the "degree of greenness" of the vegetation covers. Hence, the people who live in biodiversity rich areas are ready to pay more to retain the amount of the same.

Finally, Model V is designed to compare the preferences for land-ownership types vs. land cover types among home buyers. Results from these models reveal that Land Cover variables dominate when it comes to localized effects while Land Ownership variables or the 'brand' of the land dominates with regards to indirect effects. This implies that public lands do account spatial spillovers while actual physical attribute of the land does not.

The methodology, results and policy implications found in this study are more generalizable than most other hedonic studies. Hedonic studies to date use household data that is acquired from different, non-uniform sources, whereas this project uses census data. Use of census data assures use of methodology in future application since there remains a consistency in the data generating process. Also, this study covers a large area, which accounts for preferences of homebuyers who reside very close to and very far away from greener areas and everyone in between, which usually is not found in household level studies.

Compare and Contrast with Schultz and King (2001): This study is mainly based on the structure laid down by Schultz and King (2001), relying upon the aggregation level found most appropriate for Tucson metropolitan area. This study advances the analysis of Schultz and King (2001) by including the entire Santa Cruz Watershed and thus urban and rural areas. The current study also conducts a much sophisticated analysis based on differentiating the affects land ownership (vs.) land cover variables. Also, here a biodiversity index calculated answers the question of premiums on quality of green space, in contrast to quantity (distance/proximity etc.), not included in the latter. The study Schultz and King (2001) though similar in structure and region (for the most part), does both contrasts and conforms on different grounds with the current study on its findings. While the current study proposes that as quality of greenness increases, the premium on home values increases, latter finds higher grade wildlife has a negative effect and moderate one has a high premium. The findings conform for large and protected natural resource in Schultz and King (2001) with the public lands being a definite amenity in the current study. Here, the public lands are differentiated unlike Schultz and King (2001) who seem to club each into a single category. Also, latter finds homebuyers either uninterested or disinterested by regional and district parks and nearest neighborhood park, whereas here Local Parks are found to be an amenity, though not as strong as other public lands. Other findings like negative influence of industrial areas are conformable to a discount in home values caused by proximity to Tucson International Airport and DM Airbase in the current study. Also, this study controls for school districts and includes explicit variables for demography, income levels and proximity to the US-Mexico International Border to suffice for dummy variables (market segments) in Schultz and King (2001) as an attempt to explain differences due to crime rates, quality of school indices and cultural and race measures.

***Caveats and Future Research:***

An issue that arises while estimating these models is reporting and interpreting the estimated coefficients and respective elasticities for CBGs lying on the boundary of the study area. Given the fact that this study acknowledges and corrects for "neighborhood effects" or "spatial spillovers", care must be taken for boundary areas that might be experiencing spatial spillovers from adjoining CBGs, which are out of focus for this study. One way of getting around this issue is to choose a larger area of study and focus on results of the interested region. However, this is not a robust technique as the spatial spillovers extend to neighbors of higher degrees and it is not possible to know where to stop while defining the "larger area of study".

Some other techniques like GWR (Geographically Weighted Regression) and/or estimation using panel data by spatial panel modeling might provide better and more comprehensive policy implications. The theory of Spatial Autocorrelation Heteroskedasticity Coefficient (SAHC), which is a recent addition in the literature, can be used to achieve efficient estimates.

In context to future research, this effort will be extended to the Mexican side of the Santa Cruz River Watershed and the estimated elasticities will be an input for the Santa Cruz Watershed Portfolio Model at the Western Geographic Science Center, U.S.G.S.

## REFERENCES

**Acharya, G. and Bennett, L.L.,** "Valuing open space and land-use patterns in urban watersheds". *Journal of Real Estate Finance and Economics* 22 (2–3), 221–237. 2001.

**Anderson, S. T. and West, S. E.,** "Open Space, Residential Property Values, and Spatial Context" *Regional Science and Urban Economics* 36(2006):773

**Anselin L.,** "Spatial Econometrics: Methods and Models", *Kluwer Academic,* Dordrecht, 1988(a).

**Anselin, L., Bera, A., Florax, R. and M. Yoon** "Simple Diagnostic Tests For Spatial Dependence". *Regional Science and Urban Economics* 26, 77–104. 1996.

**Anselin, L.** "Spatial Econometrics" *Center For Spatially Integrated Social Science. 1999.* http://www.csiss.org/learning_resources/content/papers/baltchap.pdf

**Aten, B** "A Mapping of International Prices" *Review of Income and Wealth*, 1996.

**Bark-Hodgins, R. H. and Colby, B. G.,** "An Economic Assessment of the Sonoran Desert Conservation Plan", *Natural Resources Journal*, 46:709–725. 2006.

**Bark-Hodgins, R. H., Osgood, D. E. and Colby, B. G.** "Remotely Sensed Proxies for Environmental Amenities in Hedonic Analysis: What does Green Mean?" In J. I. Caruthers and B. Mundy, eds., *Environmental Valuation: Intraregional and Interregional Perspectives,* London: Ash gate, 2006, 191–210.

**Bark, R. H., Osgood, D. E., Colby, B. G., Katz, G and Stromberg, J.** "Habitat Preservation and Restoration: Do Homebuyers Have Preferences for Quality Habitat?" *Ecological Economics* 68(2009):1465–1475. Available Online: 17 November 2008.

**Bark, Rosalind H., Osgood, Daniel E., Colby, Bonnie G. and Halper, Eve B**. "How Do Homebuyers Value Different Types Of Green Space?" *Journal of Agricultural and Resource Economics,* 36(2):395–415 Western Agricultural Economics Association. 2011.

**Bateman, I.J., Jones, A.P., Lovett, A.A., Lake, I.R. and Day, B.H., 2002** "Applying Geographical Information Systems (GIS) to environmental and resource economics" *Environmental and Resource Economics,* 22, 219–269.

**Benirschka, M. and Binkley, J.K.** "Land Price Volatility In A Geographically Dispersed Market", *American Journal of Agricultural Economics,* 76, 185–95. 1994.

**Benson, E.D., Hansen, J.L., Arthur, J., Schwartz, L., Smersh, G.T.,** "Pricing residential amenities: the value of a view". *Journal of Real Estate Finance and Economics* 16, 55–73. 1998.

**Bin, O. and Polasky, S.,** "Evidence on the amenity value of wetlands in a rural setting". *Journal of Agricultural and Applied Economics* 37(3): 589-602. 2005.

**Bin, O., Crawford, T. W., Kruse, J. B. and Landry, C. E.** "Viewscapes and Flood Hazard: Coastal Housing Market Response to Amenities and Risk." *Land Economics* 84:434–448. 2008

**Bitter, Christopher, Mulligan, Gordon F., Dall'erba, Sandy**, "Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method", *Journal of Geographical Systems* (2007) 9:7–27. 2006.

**Black, S. E.,** "Do Better Schools Matter? Parental Valuation of Elementary Education", *Quarterly Journal of Economics* 114(1999):577–599.

**Blomquist, G.C.,** "The Effect of Electric Utility Power Plant Location on Area Property Value". *Land Economics, 50:1, 97-100.* 1974.

**Blomquist, G.C., Berger, M.C., and Hoehn, J.P.,** "New Estimates Of Quality Of Life In Urban Areas". *American Economic Review*, 78:1, 89-107. 1988.

**Boyle, B.A. and Kiel, K.A.,** "A Survey of House Price Hedonic Studies of the Impact of Environmental Externalities". *Journal of Real Estate Literature*, 9(2), 117-144. 2001.

**Brasington, D.M. and Hite D**. "Demand for Environmental Quality: A Spatial Hedonic Analysis" *Regional Science and Urban Economics* 35(1), 2005, p. 57-82.

**Case, A. C.** 'Spatial Patterns in Household Demand', *Econometrica* **59**, 953–965. 1991.

**Case, A., Rosen, H.S. and Hines, J.R**. "Budget spillovers and fiscal policy interdependence: evidence from the States" *Journal of Public Economics* 52, 285–307. 1993.

**Cheshire, P., Sheppard, S.**, "Capitalizing The Value Of Free Schools: The Impact Of Supply Characteristics And Uncertainty" *Economic Journal*, 114 (499), F397–F424 (1). 2004.

**Chiodo, A.J., Hernández-Murillo, R., and Owyang, M.T.,**"Nonlinear Hedonics And The Search For School District Quality". Working Paper 2003-039D, *Federal Reserve Bank of St. Louis*: pp 1–31. 2005. (http://research.stlouisfed.org/wp/2003/2003-039.pdf.)

**Cliff, A. and Ord, J.K.,** "Spatial Process: Models and Applications. London: Pion. 1981.

**Colwell, Peter F. and Dilmore, G**. "Who was First? An Examination of Early Hedonic Study" *Land Economics*, 75(4), November 1999, pp. 620-6.

**Court A.T.** "Hedonic Price Indexes With Automotive Examples". In *The Dynamics of Automobile Demand*, New York, NY: General Motors, 1939.

**David, E.L.,** "Lakeshore Property Values: A Guide to Public Investment in Recreation", *Water Resources Research*, 4:4, 697-707. 1968.

**Doss, C. R. and Taff, S. J.** 'The Influence of Wetland Type and Wetland Proximity on Residential Property Values', *Journal of Agricultural and Resource Economics* **21**, 120–129. 1996.

**Downes, T. A. and Zabel, J. E.** "The Impact of School Characteristics on House Prices: Chicago 1987-1991." *Journal of Urban Economics* 52:1–25. 2002.

**Dubin, R. and Sung, C.**, "Spatial variation in the price of housing: rent gradients in non-mono-centric cities". *Urban Studies,* 24:193–204. 1987.

**Follain, J.R. and Malpezzi, S.** "Dissecting Housing Value and Rent", Washington D.C.: *The Urban institute, 1980.*

**Follain, James R. and Jimenez, E.** "The Demand For Housing Characteristics in Developing Countries" *Urban Studies*, 22(5), 1985, pp 421-32. (1985b).

**Freeman A.** "The Measurement of Environmental and Resource Values". *Resources for the Future*, Washington. 2003.

**Geoghegan, J., Wainger, L.A., Bockstael, N.E.,** "Analysis spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS". *Journal of Ecological Economics* 23, 251–264. 1997.

**Geoghegan, J.,** 'The Value of Open Spaces in Residential Land Use', *Land Use Policy*. 2002.

**Getis, Arthur,** "Problems Associated with Spatial Pattern Analysis [Flash Player File]". *Introduction to Spatial Pattern Analysis in a GIS Environment*, Summer Workshop at the Center For Spatially Integrated Social Science (CSISS), University of California, Santa Barbara. 2002. Retrieved from: http://www.csiss.org/streaming_video/2002/spa.html

**Goodman, A.,** "Housing submarkets within urban areas—definitions and evidence", *Journal of Regional Science,* 21:175–185. 1981.

**Goodman, A.,** "Housing Market Segmentation", *Journal of Housing Economics,* 7:121–143. 1998.

**Goodman, Allen C**. "Andrew Court and the Invention of hedonic Price Analysis". *Journal of Urban Economics,* 44(2), September 1998, pp. 291-8. 1998.

**Haas, G.C.** "Sales Prices as a Basis for Farm Land Appraisal" Technical Bulletin 9, St. Paul: The University of Minnesota Agricultural Experiment Station, 1922.

**International Boundary and Water Commission (IBWC)** "Enlargement of the International Facilities for the Treatment of Nogales, Arizona and Nogales, Sonora Sewage, El Paso" *Minute 227 International Boundary and Water Commission* 1967.

**Irwin, E. G.,** "Using Spatial Data and Methods to Study Rural-Urban Change". Paper presented at *Rural Policy: Issues, Data Needs and Data Access Conference*. Washington D.C. 2000.

**Irwin, E.G. and Bockstael, N.E.,** "The problem of identifying landuse spillovers: measuring the effects of open space on residential property values". *American Journal of Agricultural Economics* 83 (3), 698–704. 2001.

**Irwin, E. G. and Geoghegan, J**., 'Theory, Data, Methods: Developing Spatially Explicit Economic Models of Land-Use Change', *Agriculture, Ecosystems and Environment* **85**(1), 7–23. 2001.

**Kim, Chong Won, Phipps, Tim T. and Anselin L.,** "Measuring Benefits of Air Quality Improvement: A Spatial Hedonic Approach", *Journal of Environmental Economics and Management, 45, pp. 24-39,* 2003.

**Lancaster, Kelvin J.** "A New Approach To Consumer Theory" *Journal of Political Economy,* 1966. 74, 132-57

**Leech, D. and Campos, E.,** "Is comprehensive education really free?: a case-study of the effects of secondary school admissions policies on house prices in one local area". Journal *of the Royal Statistical Society*. Series A (Statistics in Society) 166 (1), 135–154. 2003.

**Leggett, C.G. and Bockstael, N.E.,** "Evidence of the effects of water quality on residential land prices". *Journal of Environmental Economics and Management* 39 (2), 121–144. 2000.

**LeSage, James P. and Pace, R. Kelly,** "Introduction to Spatial Econometrics". *CRC Press, Taylor & Francis Group,: Boca Raton, FL*. 2009

**Mahan, B. L., Polasky, S. and Adams, R. M.**, "Valuing Urban Wetlands: A Property Price Approach." *Land Economics* 76:100–113. 2000.

**Malpezzi S.**"Hedonic Pricing Models: A Selective and Applied Review", prepared for *Housing Economics*: Essays in Honor of Duncan Maclennan, 2002. http://www.bus.wisc.edu/realestate.

**Manski, C.F.,** "Approximations of Endogenous Social Effects: The Reflection Problem", *Review of Economic Studies* 60, 531-42. 1993.

**Metz, W. C. and Clark, D. E**. "The Effect of Decisions about Spent Nuclear Fuel Storage on Residential Property Prices", *Risk Analysis* **17**, 571–582. 1997.

**Michaels, R. and Smith, V.,** "Market segmentation and valuing amenities with hedonic models: the case of hazardous waste sites". *Journal of Urban Economics*, 28:223–242. 1990.

**Mooney, S. and Eisgruber, L. M.,** "The Influence of Riparian Protection Measures on Residential Property Values: The Case of the Oregon Plan for Salmon and Watersheds" *Journal of Real Estate Finance and Economics* 22 (2/3): 273-86. 2001.

**Murdoch, J.C., Rahmatian, M. and Thayer, M. A.** "A spatially autoregressive median voter model of recreation expenditures". *Public Finance Quarterly* 21, 334–350. 1993.

**Netusil, Noelwah R.,** "Economic Valuation of Riparian Corridors and Upland Wildlife Habitat in Urban Watershed", *Journal of Contemporary Water Research and Education*, 134: pp: 39-45. 2006.

**Norman, L. M., Villarreal, M. L., Lara-Valencia, F., Yuan, Y., Nie, W., Wilson, S., Amaya, G., Sleeter, R.**, "Mapping socio-environmentally vulnerable populations access and exposure to ecosystem services at the U.S.–Mexico borderlands". *Applied Geography*. 34(0), 413–424. doi:10.1016/j.apgeog.2012.01.006. (2012)

**Orford, S.** "Valuing The Built Environment*"* Aldershot: Ashgate. 1999.

**Paterson, R.W. and Boyle, K.J.,** "Out of sight, out of mind? UsingGIS to incorporate visibility in hedonic property value models". *Land Economics* 78 (3), 417–425. 2002.

**Powe, N. A., Garrod, G. D., Brunsdon, C. F. and K. G.Willis**, "Using a Geographic Information System to Estimate an Hedonic Price Model of the Benefits of Woodland Access", *Forestry* **70**(2), 139–149. 1997.

**Ridker, R.G. and Henning, J.A.,** "The Determinants of Residential Property Values With Special Reference To Air Pollution". *Review Of Economics and Statistics*, 49:2, 246-57. 1967.

**Rosen, S.** "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* 82(1974):34–55.

**Schnare, A. and Struyk, R.,** "Segmentation in urban housing markets", *Journal of Urban Economics,* 4:146–166. 1976.

**Schultz, S. D. and King, D.A.** "The Use of Census Data for Hedonic Price Estimates of Open-Space Amenities And Land Use." *Journal of Real Estate Finance & Economics* 22:239–252. 2001.

**Sirmans, Stacy, Macpherson, David A. and Zeitz, Emily N** "The Composition of Hedonic Pricing Models," *Journal of Real Estate Literature*, 13, 3–43. 2005

**Smith, V. K.** "The Rise of Regional Relevance", Paper presented at *Resources for the Future Seminar*. Washington D.C. 1996.

**Smith, V.K., Poulos, C., Kim, H.,** "Treating open space as an urban amenity". *Resource and Energy Economics* 24 (1–2), 107–129. 2002.

**Sonoran Institute,** "A Living River- Charting the Health of the Upper Santa Cruz River 2010 Water Year"- The 3rd annual living river report published by *The Sonoran Institute*. Retrieved from http://sonoraninstitute.org/library/reports.html

**Sprouse T.W**. "Water Issues On the Arizona – Mexico Border: The Santa Cruz, San Pedro and Colorado Rivers" *An Issue Paper. Water Resource Research Center* 2005

**Topa, Giorgio** "Social interactions, local spillovers and unemployment", Ph.D. Dissertation, Department of Economics, University of Chicago. 1996.

**Turner, M. G., Romme, W. H., Gardner, R. H., O'Neill, R. V. and Kratz, T. K.** "A Revised Concept of Landscape: Equilibrium, Disturbance and Stability on Sealed Landscapes", L*andscape Ecology* **8**, 213–217. 1993.

**Villarreal, M.L., Norman, L.M., Wallace, C.S.A., and van Riper, Charles, III,** "A multi-temporal (1979-2009) land-use/land-cover dataset of the bi-national Santa Cruz Watershed" *U.S. Geological Survey Open-File Report* 2011-1131, 26 p. and database. 2011.

**Villarreal, M.L., Norman, L.M., Boykin, K.G., and Wallace, C.S.A.,** "Biodiversity losses and conservation tradeoffs: assessing future growth scenarios for a North American transnational trade corridor". *International Journal Of Biodiversity Science, Ecosystem Services, and Management,* 2012.

**Waddell, P. and Berry, B. J. L.** "House Price Gradients – The Intersection of Space and Built Form", *Geographical Analysis* **25**, 117–141. 1993.

**Wallace, C.S.A., Villarreal, M.L., and Norman, L.M.,** "Development of a high-resolution binational vegetation map of the Santa Cruz River riparian corridor and surrounding watershed, southern Arizona and northern Sonora, Mexico: *U.S. Geological Survey Open-File Report* 2011-1143, 22 p. and data files. 2011.

**Wallace, H.A.** "Comparative farmland values in Iowa" *Journal of Land and Public Utility Economics*, 2, October 1926. pp. 385-92.

**Witte, Ann D., Sumka, Howard J., Erekson, H.** "An Estimate of a Structural Hedonic Price Model of the Housing Market: An Application of Rosen's Theory of Implicit Markets" *Econometrica*, Vol. 47, No. 5 (Sep., 1979), pp. 1151-1173.

# APPENDIX A

## *I. Econometric Model Estimates – IV- 2SLS Robust Spatial Lag Models*

**Table 21: Model I: IV 2SLS Robust Spatial Lag Estimation**

| Variables | | | Simulation results (robust IV variance matrix): | | | | | |
|---|---|---|---|---|---|---|---|---|
| ***Structural, Neighborhood and Contract Variables*** | | | Impact measures | | | Simulated p-values | | |
| | Estimate | Pr(>\|z\|) | | | | | | |
| **Rho** | 0.181 | 0.001 | | | | | | |
| **Constant** | 9.788 | < 2.2e- | **Direct** | **Indirect** | **Total** | **Direct** | **Indirect** | **Total** |
| **% Owner** | -0.475 | 0.000 | -0.477 | -0.102 | -0.580 | 0.000 | 0.011 | 0.000 |
| **% Vacant** | -0.141 | 0.714 | -0.142 | -0.030 | -0.172 | 0.747 | 0.785 | 0.751 |
| **% Mobile Homes** | -0.844 | < 2.2e- | -0.849 | -0.182 | -1.031 | < 2.22e-16 | 0.006 | 0.000 |
| **Median Rooms (#)** | 0.050 | 0.147 | 0.050 | 0.011 | 0.061 | 0.076 | 0.145 | 0.078 |
| **PCNEW** | 0.666 | 0.053 | 0.669 | 0.143 | 0.813 | 0.063 | 0.114 | 0.063 |
| **PCOLD39** | 0.534 | 0.000 | 0.537 | 0.115 | 0.652 | 0.000 | 0.017 | 0.000 |
| **NOPHONE** | -0.496 | 0.406 | -0.499 | -0.107 | -0.606 | 0.471 | 0.510 | 0.473 |
| **INCPLUMB** | -0.387 | 0.708 | -0.390 | -0.083 | -0.473 | 0.733 | 0.736 | 0.732 |
| **% Flood** | 0.000 | 0.784 | 0.000 | 0.000 | 0.000 | 0.790 | 0.746 | 0.781 |
| **% Minorities** | -0.573 | 0.000 | -0.576 | -0.123 | -0.699 | 0.000 | 0.005 | 0.000 |
| **%_earning_lt$25k** | -0.416 | 0.010 | -0.418 | -0.090 | -0.508 | 0.006 | 0.045 | 0.006 |
| **%_earning_gt$60k** | 1.131 | 0.000 | 1.138 | 0.244 | 1.381 | 0.000 | 0.006 | 0.000 |
| **D_ON THE BORDER** | -0.338 | 0.005 | -0.340 | -0.073 | -0.412 | 0.004 | 0.036 | 0.004 |
| **TIA_DM_MAS** | -0.187 | 0.000 | -0.188 | -0.040 | -0.228 | 0.000 | 0.011 | 0.000 |
| **D_SCHL_DISTRICT** | 0.159 | 0.003 | 0.160 | 0.034 | 0.194 | 0.005 | 0.032 | 0.005 |
| **D_PIMA_PINAL** | -0.272 | 0.001 | -0.274 | -0.059 | -0.332 | 0.001 | 0.019 | 0.001 |

**Table 22: MODEL II - IV 2SLS Robust Spatial Lag Estimation**

| Variables | | | Simulation results (robust IV variance matrix): | | | | | |
|---|---|---|---|---|---|---|---|---|
| *Structural, Neighborhood and Contract Variables* | | | | | | | | |
| | Estimate | Pr(>\|z\|) | Impact measures (lag, exact): | | | Simulated p-values: | | |
| | | | Direct | Indirect | Total | Direct | Indirect | Total |
| Rho | 0.202 | 0.000 | | | | | | |
| Constant | 9.223 | < 2.2e-16 | | | | | | |
| % Owner | -0.471 | 0.000 | -0.474 | -0.116 | -0.590 | 0.000 | 0.014 | 0.000 |
| % Vacant | -0.305 | 0.406 | -0.307 | -0.075 | -0.383 | 0.210 | 0.253 | 0.209 |
| % Mobile Homes | -0.954 | < 2.2e-16 | -0.961 | -0.234 | -1.195 | 0.00 | 0.005 | 0.000 |
| Median Rooms (#) | 0.053 | 0.110 | 0.053 | 0.013 | 0.066 | 0.180 | 0.203 | 0.176 |
| PCNEW | 0.332 | 0.276 | 0.335 | 0.082 | 0.417 | 0.173 | 0.247 | 0.180 |
| PCOLD39 | 0.519 | 0.000 | 0.523 | 0.128 | 0.650 | 0.000 | 0.015 | 0.000 |
| NOPHONE | -0.753 | 0.173 | -0.758 | -0.185 | -0.943 | 0.167 | 0.233 | 0.169 |
| INCPLUMB | -1.344 | 0.170 | -1.354 | -0.330 | -1.684 | 0.198 | 0.259 | 0.202 |
| % Flood | 0.001 | 0.222 | 0.001 | 0.000 | 0.001 | 0.142 | 0.226 | 0.149 |
| % Minorities | -0.533 | 0.000 | -0.537 | -0.131 | -0.668 | 0.000 | 0.001 | 0.000 |
| %_earning_lt$25k | -0.207 | 0.160 | -0.208 | -0.051 | -0.259 | 0.252 | 0.307 | 0.255 |
| %_earning_gt$60k | 1.174 | 0.000 | 1.183 | 0.289 | 1.471 | 0.00 | 0.004 | 0.000 |
| D_ON THE BORDER | -0.190 | 0.059 | -0.191 | -0.047 | -0.238 | 0.056 | 0.069 | 0.049 |
| TIA_DM_MAS | -0.160 | 0.000 | -0.161 | -0.039 | -0.200 | 0.000 | 0.020 | 0.000 |
| D_SCHL_DISTRICT | 0.180 | 0.001 | 0.181 | 0.044 | 0.225 | 0.000 | 0.011 | 0.000 |
| D_PIMA_PINAL | -0.048 | 0.552 | -0.048 | -0.012 | -0.060 | 0.489 | 0.560 | 0.496 |
| *Land OwnerShip Variables* | | | | | | | | |
| % State Trust | 0.006 | 0.000 | 0.006 | 0.001 | 0.007 | 0.000 | 0.007 | 0.000 |
| % National Park | 0.009 | 0.000 | 0.009 | 0.002 | 0.011 | 0.000 | 0.017 | 0.000 |
| % BLM | 0.015 | 0.000 | 0.015 | 0.004 | 0.018 | 0.001 | 0.036 | 0.001 |
| % Forest Service | 0.005 | 0.001 | 0.005 | 0.001 | 0.006 | 0.001 | 0.025 | 0.001 |
| % Local Parks | 0.007 | 0.080 | 0.007 | 0.002 | 0.008 | 0.058 | 0.121 | 0.060 |

**Table 23: MODEL III: IV 2SLS Robust Spatial Lag Estimation**

| Variables | | | Simulation results (robust IV variance matrix): | | | | | |
|---|---|---|---|---|---|---|---|---|
| ***Structural, Neighborhood and Contract Variables*** | | | | | | | | |
| | Estimate | Pr(>\|z\|) | Impact measures (lag, exact): | | | Simulated p-values: | | |
| **Rho** | 0.20 | 0.00 | **Direct** | **Indirect** | **Total** | **Direct** | **Indirect** | **Total** |
| **Constant** | 9.34 | < 2.2e-16 | | | | | | |
| **% Owner** | -0.48 | 0.00 | -0.49 | -0.11 | -0.60 | 0.00 | 0.00 | 0.00 |
| **% Vacant** | -0.49 | 0.18 | -0.49 | -0.12 | -0.61 | 0.15 | 0.22 | 0.16 |
| **% Mobile Homes** | -0.98 | < 2.2e-16 | -0.98 | -0.23 | -1.21 | < 2.22e-16 | 0.00 | < 2.22e-16 |
| **Median Rooms (#)** | 0.04 | 0.20 | 0.04 | 0.01 | 0.05 | 0.16 | 0.18 | 0.16 |
| **PCNEW** | 0.02 | 0.94 | 0.02 | 0.01 | 0.03 | 0.91 | 0.94 | 0.91 |
| **PCOLD39** | 0.55 | 0.00 | 0.55 | 0.13 | 0.68 | 0.00 | 0.00 | 0.00 |
| **NOPHONE** | -0.66 | 0.25 | -0.66 | -0.16 | -0.82 | 0.20 | 0.26 | 0.21 |
| **INCPLUMB** | -0.93 | 0.33 | -0.94 | -0.22 | -1.16 | 0.32 | 0.34 | 0.32 |
| **% Flood** | 0.00 | 0.94 | 0.00 | 0.00 | 0.00 | 0.87 | 0.88 | 0.87 |
| **% Minorities** | -0.58 | 0.00 | -0.59 | -0.14 | -0.73 | 0.00 | 0.00 | 0.00 |
| **%_earning_lt$25k** | -0.26 | 0.09 | -0.26 | -0.06 | -0.32 | 0.12 | 0.17 | 0.13 |
| **%_earning_gt$60k** | 1.05 | 0.00 | 1.05 | 0.25 | 1.30 | 0.00 | 0.00 | 0.00 |
| **D_ON THE BORDER** | -0.37 | 0.00 | -0.37 | -0.09 | -0.46 | 0.00 | 0.00 | 0.00 |
| **TIA_DM_MAS** | -0.16 | 0.00 | -0.16 | -0.04 | -0.20 | 0.00 | 0.02 | 0.00 |
| **D_SCHL_DISTRICT** | 0.18 | 0.00 | 0.19 | 0.04 | 0.23 | 0.00 | 0.01 | 0.00 |
| **D_PIMA_PINAL** | 0.00 | 1.00 | 0.00 | 0.00 | 0.00 | 0.81 | 0.84 | 0.82 |
| ***Land Cover Variables*** | | | | | | | | |
| **%_Open Water** | 0.07 | 0.55 | 0.07 | 0.02 | 0.08 | 0.45 | 0.50 | 0.45 |
| **%_Developed Land** | | | | | | | | |
| **% Barren Land** | 0.00 | 0.96 | 0.00 | 0.00 | 0.00 | 0.82 | 0.88 | 0.83 |
| **% Deciduous Forest** | -0.01 | 0.78 | -0.01 | 0.00 | -0.01 | 0.68 | 0.71 | 0.69 |
| **% Evergreen Forest** | 0.01 | 0.28 | 0.01 | 0.00 | 0.01 | 0.30 | 0.34 | 0.31 |
| **% Shrub/Scrub** | 0.02 | 0.00 | 0.02 | 0.00 | 0.02 | 0.00 | 0.00 | 0.00 |
| **% Grassland** | 0.05 | 0.02 | 0.05 | 0.01 | 0.06 | 0.02 | 0.05 | 0.02 |
| **% Pasture** | 0.02 | 0.01 | 0.02 | 0.00 | 0.02 | 0.02 | 0.05 | 0.02 |
| **% Cultivated** | 0.01 | 0.41 | 0.01 | 0.00 | 0.02 | 0.28 | 0.30 | 0.27 |
| **% Wetlands** | -0.32 | 0.19 | -0.32 | -0.08 | -0.39 | 0.21 | 0.25 | 0.21 |

**Table 24: MODEL IV: IV 2SLS Robust Spatial Lag Estimation**

| Variables | | | | Simulation results (robust IV variance matrix): | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ***Structural, Neighborhood and Contractual Variables*** | | | | | | | | | | |
| | Estimate | Pr(>\|z\|) | | Impact measures (lag, exact): | | | | Simulated p-values: | | |
| | | | | | | | | | | |
| **Rho** | 0.18 | 0.00 | | | | | | | | |
| **Constant** | 9.38 | < 2.2e-16 | | **Direct** | **Indirect** | **Total** | | **Direct** | **Indirect** | **Total** |
| **% Owner** | -0.48 | 0.00 | | -0.48 | -0.10 | -0.58 | | 0.00 | 0.02 | 0.00 |
| **% Vacant** | -0.46 | 0.19 | | -0.46 | -0.10 | -0.56 | | 0.14 | 0.20 | 0.14 |
| **% Mobile Homes** | -1.00 | < 2.2e-16 | | -1.00 | -0.21 | -1.21 | | < 2.22e-16 | 0.00 | < 2.22e-16 |
| **Median Rooms (#)** | 0.04 | 0.22 | | 0.04 | 0.01 | 0.05 | | 0.32 | 0.35 | 0.32 |
| **PCNEW** | 0.04 | 0.90 | | 0.04 | 0.01 | 0.05 | | 0.95 | 0.99 | 0.95 |
| **PCOLD39** | 0.56 | 0.00 | | 0.57 | 0.12 | 0.69 | | 0.00 | 0.01 | 0.00 |
| **NOPHONE** | -0.58 | 0.31 | | -0.58 | -0.12 | -0.71 | | 0.32 | 0.38 | 0.32 |
| **INCPLUMB** | -0.89 | 0.33 | | -0.90 | -0.19 | -1.09 | | 0.31 | 0.34 | 0.31 |
| **% Flood** | 0.00 | 0.64 | | 0.00 | 0.00 | 0.00 | | 0.54 | 0.58 | 0.54 |
| **% Minorities** | -0.63 | 0.00 | | -0.64 | -0.14 | -0.77 | | 0.00 | 0.00 | 0.00 |
| **%_earning_lt$25k** | -0.23 | 0.11 | | -0.23 | -0.05 | -0.28 | | 0.14 | 0.18 | 0.14 |
| **%_earning_gt$60k** | 1.01 | 0.00 | | 1.01 | 0.22 | 1.23 | | 0.00 | 0.01 | 0.00 |
| **D_ON THE BORDER** | -0.25 | 0.02 | | -0.25 | -0.05 | -0.31 | | 0.02 | 0.09 | 0.03 |
| **TIA_DM_MAS** | -0.16 | 0.00 | | -0.16 | -0.03 | -0.19 | | 0.00 | 0.03 | 0.00 |
| **D_SCHL_DISTRICT** | 0.15 | 0.00 | | 0.15 | 0.03 | 0.18 | | 0.01 | 0.02 | 0.01 |
| **D_PIMA_PINAL** | 0.13 | 0.15 | | 0.13 | 0.03 | 0.16 | | 0.08 | 0.14 | 0.08 |
| ***Land Own. + Land Cover*** | | | | | | | | | | |
| **PCT_LOCAL PARKS** | 0.00 | 0.23 | | 0.00 | 0.00 | 0.00 | | 0.29 | 0.38 | 0.29 |
| **PCT_BARREN** | -0.01 | 0.43 | | -0.01 | 0.00 | -0.01 | | 0.44 | 0.50 | 0.44 |
| **PCT_DEVELOPED** | -0.01 | 0.07 | | -0.01 | 0.00 | -0.01 | | 0.06 | 0.15 | 0.06 |
| **BIODIVERSITY INDEX** | 0.00 | 0.01 | | 0.00 | 0.00 | 0.00 | | 0.01 | 0.01 | 0.00 |

**Table 25: MODEL V: IV 2SLS Robust Spatial Lag Estimation**

| Variables | | | Simulation results (robust IV variance matrix): | | | | | |
|---|---|---|---|---|---|---|---|---|
| ***Structural, Neighborhood and Contractual*** | Estimate | Pr(>\|z\|) | | | | | | |
| | | | **Impact measures (lag, exact):** | | | **Simulated p-values:** | | |
| | | | **Direct** | **Indirect** | **Total** | **Direct** | **Indirect** | **Total** |
| **Rho** | 0.213 | 0.000 | | | | | | |
| **Constant** | 9.221 | < 2.2e-16 | | | | | | |
| **% Owner** | -0.49 | 0.00 | -0.49 | -0.13 | -0.62 | 0.00 | 0.01 | 0.00 |
| **% Vacant** | -0.30 | 0.43 | -0.30 | -0.08 | -0.38 | 0.41 | 0.46 | 0.41 |
| **% Mobile Homes** | -0.93 | < 2.2e-16 | -0.94 | -0.25 | -1.19 | < 2.22e-16 | 0.00 | 0.00 |
| **Median Rooms (#)** | 0.05 | 0.11 | 0.05 | 0.01 | 0.07 | 0.10 | 0.16 | 0.11 |
| **PCNEW** | 0.39 | 0.24 | 0.39 | 0.10 | 0.49 | 0.29 | 0.32 | 0.29 |
| **PCOLD39** | 0.52 | 0.00 | 0.52 | 0.14 | 0.66 | 0.00 | 0.01 | 0.00 |
| **NOPHONE** | -0.72 | 0.20 | -0.73 | -0.19 | -0.91 | 0.19 | 0.24 | 0.20 |
| **INCPLUMB** | -0.87 | 0.37 | -0.88 | -0.23 | -1.11 | 0.30 | 0.33 | 0.30 |
| **% Flood** | 0.00 | 0.49 | 0.00 | 0.00 | 0.00 | 0.45 | 0.48 | 0.45 |
| **% Minorities** | -0.50 | 0.00 | -0.51 | -0.13 | -0.64 | 0.00 | 0.00 | 0.00 |
| **%_earning_lt$25k** | -0.25 | 0.10 | -0.25 | -0.07 | -0.32 | 0.12 | 0.14 | 0.11 |
| **%_earning_gt$60k** | 1.16 | 0.00 | 1.17 | 0.31 | 1.48 | 0.00 | 0.00 | 0.00 |
| **D_ON THE BORDER** | -0.37 | 0.01 | -0.37 | -0.10 | -0.47 | 0.01 | 0.05 | 0.01 |
| **TIA_DM_MAS** | -0.17 | 0.00 | -0.17 | -0.04 | -0.21 | 0.00 | 0.02 | 0.00 |
| **D_SCHL_DISTRICT** | 0.18 | 0.00 | 0.18 | 0.05 | 0.23 | 0.00 | 0.00 | 0.00 |
| **D_PIMA_PINAL** | -0.17 | 0.22 | -0.17 | -0.04 | -0.21 | 0.23 | 0.27 | 0.23 |
| ***Land OwnerShip Variables*** | | | | | | | | |
| **% State Trust** | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 |
| **% National Park** | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 | 0.01 | 0.00 |
| **% BLM** | 0.01 | 0.02 | 0.01 | 0.00 | 0.02 | 0.02 | 0.08 | 0.02 |
| **% Forest Service** | | | | | | | | |
| **% Local Parks** | 0.01 | 0.19 | 0.01 | 0.00 | 0.01 | 0.14 | 0.21 | 0.15 |
| ***Land Cover Variables*** | | | | | | | | |
| **%_Open Water** | -0.03 | 0.86 | -0.03 | -0.01 | -0.03 | 0.82 | 0.80 | 0.82 |
| **%_Developed Land** | | | | | | | | |
| **% Barren Land** | 0.00 | 0.88 | 0.00 | 0.00 | 0.00 | 0.85 | 0.90 | 0.86 |
| **% Deciduous Forest** | 0.00 | 0.97 | 0.00 | 0.00 | 0.00 | 0.99 | 0.95 | 0.98 |
| **% Evergreen Forest** | 0.01 | 0.43 | 0.01 | 0.00 | 0.01 | 0.42 | 0.44 | 0.42 |
| **% Shrub/Scrub** | | | | | | | | |
| **% Grassland** | 0.05 | 0.03 | 0.05 | 0.01 | 0.06 | 0.05 | 0.10 | 0.05 |
| **% Pasture** | 0.01 | 0.06 | 0.01 | 0.00 | 0.01 | 0.07 | 0.11 | 0.07 |
| **% Cultivated** | 0.02 | 0.25 | 0.02 | 0.00 | 0.02 | 0.23 | 0.27 | 0.23 |
| **% Wetlands** | -0.27 | 0.34 | -0.27 | -0.07 | -0.34 | 0.44 | 0.46 | 0.44 |

## *II. Econometric Model Estimates –OLS Models & Associated Bias*

**Table 26: MODEL I: OLS Estimation**

| **Variables** | | |
|---|---|---|
| ***Structural, Neighborhood and Contract Variables*** | Estimate | Pr(>\|t\|) |
| **Constant** | 11.940 | < 2e-16 |
| **% Owner** | -0.531 | 0.000 |
| **% Vacant** | -0.113 | 0.634 |
| **% Mobile Homes** | -0.875 | < 2e-16 |
| **Median Rooms (#)** | 0.055 | 0.058 |
| **PCNEW** | 0.676 | 0.006 |
| **PCOLD39** | 0.545 | 0.000 |
| **NOPHONE** | -0.548 | 0.111 |
| **INCPLUMB** | -0.632 | 0.455 |
| **% Flood** | 0.000 | 0.922 |
| **% Minorities** | -0.685 | 0.000 |
| **%_earning_lt$25k** | -0.456 | 0.001 |
| **%_earning_gt$60k** | 1.210 | 0.000 |
| **D_ON THE BORDER** | -0.392 | 0.006 |
| **TIA_DM_MAS** | -0.189 | 0.000 |
| **D_SCHL_DISTRICT** | 0.210 | 0.000 |
| **D_PIMA_PINAL** | -0.312 | 0.012 |
| | | |
| **Multiple R-squared:** | **0.7543** | |
| **Adjusted R-squared:** | **0.7477** | |
| **BIAS (b~direct - b~ols)/ b~ols * 100** | | |
| D_ON THE BORDER | -13.3 | |
| TIA_DM_MAS | -0.6 | |
| D_SCHL_DISTRICT | -23.9 | |
| D_PIMA_PINAL | -12.2 | |

**Table 27: MODEL II: OLS ESTIMATION**

| **Variables** | | |
|---|---|---|
| ***Structural, Neighborhood and Contract*** | Estimate | Pr(>\|t\|) |
| **Constant** | 11.62 | < 2e-16 |
| **% Owner** | -0.54 | 0.00 |
| **% Vacant** | -0.26 | 0.26 |
| **% Mobile Homes** | -1.00 | < 2e-16 |
| **Median Rooms (#)** | 0.06 | 0.03 |
| **PCNEW** | 0.33 | 0.17 |
| **PCOLD39** | 0.53 | 0.00 |
| **NOPHONE** | -0.83 | 0.01 |
| **INCPLUMB** | -1.63 | 0.05 |
| **% Flood** | 0.00 | 0.30 |
| **% Minorities** | -0.66 | 0.00 |
| **%_earning_lt$25k** | -0.24 | 0.08 |
| **%_earning_gt$60k** | 1.27 | < 2e-16 |
| **D_ON THE BORDER** | -0.25 | 0.07 |
| **TIA_DM_MAS** | -0.16 | 0.00 |
| **D_SCHL_DISTRICT** | 0.24 | 0.00 |
| **D_PIMA_PINAL** | -0.10 | 0.46 |
| ***Land OwnerShip Variables*** | | |
| **% State Trust** | 0.01 | 0.00 |
| **% National Park** | 0.01 | 0.04 |
| **% BLM** | 0.01 | 0.06 |
| **% Forest Service** | 0.00 | 0.00 |
| **% Local Parks** | 0.01 | 0.14 |
| **Multiple R-squared:** | 0.7768 | |
| **Adjusted R-squared:** | 0.7688 | |

| **BIAS (b~direct - b~ols)/ b~ols * 100** | |
|---|---|
| **D_ON THE BORDER** | -24.2 |
| **TIA_DM_MAS** | -0.1 |
| **D_SCHL_DISTRICT** | -23.5 |
| **D_PIMA_PINAL** | -50.9 |
| ***Land OwnerShip Variables*** | |
| **% State Trust** | -6.3 |
| **% National Park** | -7.9 |
| **% BLM** | 2.9 |
| **% Forest Service** | 4.4 |
| **% Local Parks** | 4.3 |

**Table 28: MODEL III: OLS Estimation**

| Variables | Estimate | Pr(>\|t\|) |
|---|---|---|
| ***Structural, Neighborhood and Contractual*** | | |
| **Constant** | 11.653 | < 2e-16 |
| **% Owner** | -0.543 | 0.000 |
| **% Vacant** | -0.460 | 0.051 |
| **% Mobile Homes** | -1.020 | < 2e-16 |
| **Median Rooms (#)** | 0.049 | 0.077 |
| **PCNEW** | -0.003 | 0.991 |
| **PCOLD39** | 0.558 | 0.000 |
| **NOPHONE** | -0.730 | 0.026 |
| **INCPLUMB** | -1.240 | 0.126 |
| **% Flood** | 0.000 | 0.833 |
| **% Minorities** | -0.709 | 0.000 |
| **%_earning_lt$25k** | -0.289 | 0.036 |
| **%_earning_gt$60k** | 1.129 | 0.000 |
| **D_ON THE BORDER** | -0.417 | 0.005 |
| **TIA_DM_MAS** | -0.162 | 0.000 |
| **D_SCHL_DISTRICT** | 0.239 | 0.000 |
| **D_PIMA_PINAL** | -0.029 | 0.848 |
| ***Land Cover Variables*** | | |
| **%_Open Water** | 0.071 | 0.738 |
| **%_Developed Land** | | |
| **% Barren Land** | -0.002 | 0.782 |
| **% Deciduous Forest** | -0.008 | 0.850 |
| **% Evergreen Forest** | 0.011 | 0.106 |
| **% Shrub/Scrub** | 0.016 | 0.000 |
| **% Grassland** | 0.044 | 0.019 |
| **% Pasture** | 0.015 | 0.043 |
| **% Cultivated** | 0.011 | 0.792 |
| **% Wetlands** | -0.325 | 0.419 |
| **Multiple R-squared:** | 0.784 | |
| **Adjusted R-squared:** | 0.775 | |

| **BIAS (b~direct - b~ols)/ b~ols * 100** | |
|---|---|
| **D_ON THE BORDER** | -11.66 |
| **TIA_DM_MAS** | 0.71 |
| **D_SCHL_DISTRICT** | -21.99 |
| **D_PIMA_PINAL** | -98.48 |
| ***Land Cover Variables*** | |
| **%_Open Water** | -5.48 |
| **%_Developed Land** | |
| **% Barren Land** | -74.59 |
| **% Deciduous Forest** | 18.85 |
| **% Evergreen Forest** | -3.25 |
| **% Shrub/Scrub** | -6.13 |
| **% Grassland** | 8.83 |
| **% Pasture** | 5.05 |
| **% Cultivated** | 17.46 |
| **% Wetlands** | -1.89 |

**Table 29: MODEL IV: OLS Estimation**

| Variables | | |
|---|---|---|
| ***Structural, Neighborhood and Contract Variables*** | **Estimate** | **Pr(>\|t\|)** |
| Constant | 11.455 | < 2e-16 |
| % Owner | -0.534 | 0.000 |
| % Vacant | -0.440 | 0.053 |
| % Mobile Homes | -1.033 | < 2e-16 |
| Median Rooms (#) | 0.046 | 0.092 |
| PCNEW | 0.029 | 0.908 |
| PCOLD39 | 0.575 | 0.000 |
| NOPHONE | -0.640 | 0.048 |
| INCPLUMB | -1.177 | 0.140 |
| % Flood | -0.001 | 0.482 |
| % Minorities | -0.746 | 0.000 |
| %_earning_lt$25k | -0.264 | 0.051 |
| %_earning_gt$60k | 1.080 | 0.000 |
| D_ON THE BORDER | -0.303 | 0.024 |
| TIA_DM_MAS | -0.157 | 0.000 |
| D_SCHL_DISTRICT | 0.197 | 0.000 |
| D_PIMA_PINAL | 0.116 | 0.354 |
| **LO + LC (SUBSET)** | | |
| PCT_LOCAL PARKS | 0.003 | 0.480 |
| PCT_BARREN | -0.011 | 0.163 |
| PCT_DEVELOPED | -0.006 | 0.084 |
| **BIODIVERSITY INDEX** | 0.003 | 0.002 |
| Multiple R-squared: | 0.7861 | |
| Adjusted R-squared: | 0.7789 | |
| **BIAS (b~direct - b~ols)/ b~ols * 100** | | |
| D_ON THE BORDER | -15.8 | |
| TIA_DM_MAS | 0.2 | |
| D_SCHL_DISTRICT | -23.7 | |
| D_PIMA_PINAL | 14.2 | |
| **LO + LC (SUBSET)** | | |
| PCT_LOCAL PARKS | 12.1 | |
| PCT_BARREN | -17.6 | |
| PCT_DEVELOPED | 3.2 | |
| **BIODIVERSITY INDEX** | -8.9 | |

**Table 30: MODEL V: OLS Estimation**

| Variables | Estimate | Pr(>\|t\|) |
|---|---|---|
| ***Structural, Neighborhood and Contract Variables*** | | |
| **Constant** | 11.741 | < 2e-16 |
| **% Owner** | -0.553 | 0.000 |
| **% Vacant** | -0.252 | 0.296 |
| **% Mobile Homes** | -0.978 | < 2e-16 |
| **Median Rooms (#)** | 0.060 | 0.035 |
| **PCNEW** | 0.388 | 0.114 |
| **PCOLD39** | 0.529 | 0.000 |
| **NOPHONE** | -0.803 | 0.017 |
| **INCPLUMB** | -1.213 | 0.144 |
| **% Flood** | 0.001 | 0.535 |
| **% Minorities** | -0.633 | 0.000 |
| **% earning lt$25k** | -0.285 | 0.044 |
| **% earning gt$60k** | 1.264 | < 2e-16 |
| **D ON THE BORDER** | -0.424 | 0.005 |
| **TIA DM MAS** | -0.168 | 0.000 |
| **D SCHL DISTRICT** | 0.238 | 0.000 |
| **D PIMA PINAL** | -0.211 | 0.165 |
| ***Land OwnerShip Variables*** | | |
| **% State Trust** | 0.006 | 0.000 |
| **% National Park** | 0.010 | 0.034 |
| **% BLM** | 0.012 | 0.119 |
| **% Forest Service** | | |
| **% Local Parks** | 0.005 | 0.302 |
| ***Land Cover Variables*** | | |
| **% Open Water** | -0.023 | 0.919 |
| **% Developed Land** | | |
| **% Barren Land** | -0.004 | 0.643 |
| **% Deciduous Forest** | 0.004 | 0.929 |
| **% Evergreen Forest** | 0.010 | 0.143 |
| **% Shrub/Scrub** | | |
| **% Grassland** | 0.045 | 0.020 |
| **% Pasture** | 0.010 | 0.195 |
| **% Cultivated** | 0.014 | 0.747 |
| **% Wetlands** | -0.274 | 0.509 |
| **Multiple R-square** | 0.776 | |
| **Adjusted R-squared:** | 0.765 | |
| **BIAS (b~direct - b~ols)/ b~ols * 100** | | |
| *D ON THE BORDER* | -12.786 | |
| *TIA DM MAS* | 0.429 | |
| *D SCHL DISTRICT* | -23.762 | |
| *D PIMA PINAL* | -20.295 | |
| ***Land OwnerShip Variables*** | | |
| *% State Trust* | -7.193 | |
| *% National Park* | -6.683 | |
| *% BLM* | 2.693 | |
| *% Forest Service* | | |
| *% Local Parks* | 2.222 | |
| *Land Cover Variables* | | |
| *% Open Water* | 18.008 | |
| *% Developed Land* | | |
| *% Barren Land* | -50.048 | |
| *% Deciduous Forest* | -59.603 | |
| *% Evergreen Forest* | -3.384 | |
| *% Shrub/Scrub* | | |
| *% Grassland* | 8.602 | |
| *% Pasture* | 10.410 | |
| *% Cultivated* | 16.271 | |
| *% Wetlands* | -1.378 | |

# APPENDIX B

## A Note On The "Insignificant" Land Cover Variables:

It is an observation in the models estimated in the study that most of the Land Cover / Land Use variables are insignificant. It is, in general, unlikely that in semi-arid region areas amenities like evergreen forests, deciduous forests, etc. do not have any impact on the values of homes. An explanation towards the insignificance of these variables lies in the fact that there is not much variability in the variables in these regions. Most of the data points for each of these variables are essentially zeros (0), and even those which are not zero (0) are more often than not very close to zero. This argument traces back to one of the most peculiarities of the region, i.e. scarcity of natural amenities.

In order to provide a robust reasoning for the importance of these Land Cover variables, there is a relationship attempted of each of these with the aforementioned Biodiversity Index (Table 31). It is found that the areas with high %_age of land under forests are in fact captured in by the index created for the purpose of this study. The direct relationship of all these variables with biodiversity index suggests that the index is merely, a summary statistic for these variables. Given the fact that (Table 24: MODEL IV: IV 2SLS Robust Spatial Lag Estimation) biodiversity index highly significant and positive in the models estimated before is an evidence of the premium for the Land Cover Variables which are in fact captured by this index alone.

Apart from Land Cover variables, the regression below shows has a dummy variable "RIVER" for all CBGs that lie on the channel of flow of the Santa Cruz River and its major tributaries. RIVER has a positive and significant coefficient pointing out to the fact that there is indeed a premium for being close to the river channel too.

Another variable %_Flood which was again insignificant in the main regression models above, is also slightly significant @90% C.L. which is quite reasonable since this land falls under Federal Emergence Management Agency and is expected to bear some economic value to the homes located here. In fact the slight significance of this variable explains that the premiums for this kind of land can be attributed to its “greenery” potential too.

**Table 31: Biodiversity Potential as a Dependent Variable**

| Variables | Estimate | Pr(>\|t\|) |
|---|---|---|
| **(Intercept)** | 80.85 | 0.00 |
| **PercentFlo** | 0.05 | 0.10 |
| **River** | 3.82 | 0.01 |
| **PCT_OPEN_WATER** | -0.27 | 0.98 |
| **PCT_BARREN LAND** | 0.59 | 0.05 |
| **PCT_DECIDUOS** | 6.92 | 0.00 |
| **PCT_EVERGREEN** | 2.86 | 0.00 |
| **PCT_GRASSLAND** | 4.33 | 0.00 |
| **PCT_PASTURE** | 2.97 | 0.00 |
| **PCT_SHRUB** | 3.09 | 0.00 |
| **PCT_CULTIVATED** | 0.66 | 0.69 |
| **PCT_WETLANDS** | -5.64 | 0.70 |
| **PCT_LOCAL PARKS** | 0.46 | 0.01 |
| **PCT_STATE TRUST** | -0.14 | 0.01 |
| | | |
| **Multiple R-squared** | 0.860 | |
| **Adjusted R-squared** | 0.857 | |