9-12-2014

A Geospatial Approach to Wildlife and Wilderness Management

Naresh Nepal

Naresh Nepal

Candidate

Department of Economics

Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Alok K. Bohara , Chairperson

Jennifer A. Thacher

Grant Harris

Wendy L. Hansen

A GEOSPATIAL APPROACH TO WILDLIFE AND WILDERNESS MANAGEMENT

by

NARESH NEPAL

B.A. Economics, Tribhuvan University, 1998
M.A. Economics, Tribhuvan University, 2001

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

**Doctor of Philosophy
Economics**

The University of New Mexico
Albuquerque, New Mexico

July 2014

DEDICATION

To my late father.

ACKNOWLEDGEMENTS

A Geospatial Approach to Wildlife and Wilderness Management

By

Naresh Nepal

B.A. Economics, Tribhuvan University, 1998

M.A. Economics, Tribhuvan University, 2001

Ph.D. Economics. University of New Mexico, 2014

ABSTRACT

The populations in the western United States have increased significantly over the last few decades resulting in increased development pressures on undeveloped and pristine lands. Population growth and increased human activities have also changed the overall ecosystems more rapidly and extensively in recent years than in the past, emphasizing the need for protecting natural land and ecosystems. This dissertation conducts analyses to highlight the importance of protected areas and wildlife habitat.

Chapter 1 explores the use of Poisson and negative binomial regression models to examine winter habitat use by mountain goats in the Kenai Mountains of South-Central Alaska. Using GPS collared locations data, these models produce parameter estimates similar to discrete choice models, popular in resource selection studies, but with less computational complexity. The potential serial (or spatial) correlations present in the data are controlled for by incorporating spatial effects in a Gaussian conditional autoregressive framework. The results support the existing literature on mountain goat habitat use with most of the terrain features statistically significant across individual goat models. The distance to the nearest escape terrain is found to be the most consistent and highly significant determinant of goat habitat selection, where individual goats tend to increase the number of visits by 37.5 to 71.6 percent more to a particular location that is 100m closer to escape terrain. The statistical significance of spatial parameter highlights the importance of neighborhood effects in habitat selection by mountain goats.

Chapter 2 seeks to achieve two objectives. First, using regression based test of equivalence, the predicted habitat use by mountain goats are compared for two different sources of data. For mountain goats, the findings suggest that the predicted habitat uses from GPS collar and aerially surveyed locations are equivalent. Second, using the aerially surveyed goat locations data, the Bayesian inference techniques with respect to negative binomial regressions are employed to explore the effects of winter recreation on mountain goat habitat selection. In addition to the landscape features, the model comparison based on Bayes factor

suggests that human recreation is an important factor affecting goats' habitat use. Goats tend to avoid areas with higher human recreational activities represented by ski-tracks, and hiking trails; increase in the ski-tracks area by one square km within a buffer of 5 km reduces goat count by approximately 2.5 percent at a particular geographic location. This highlights the potential conflict between human recreation and winter habitat and provides an input into policy discussions on conservation of mountain goat habitat.

Chapter 3 of this dissertation analyzes the statistical properties of estimated parameters in the hedonic studies with special reference to the distance variables used to represent the proximity to environmental amenities or dis-amenities. Using Monte Carlo simulations, the findings suggest that when there are single locations available for each of the landmarks in the study area, including two or more distance variables in regression results in inconsistent parameter estimates associated with the distance variables. Nevertheless, when multiple locations are available for each of the landmark types and the distance variables capture the proximity to the nearest landmark location of each type, the parameters are consistently estimated. These findings are important in studies focusing on the estimation of welfare values based on distance variables.

The major objective of Chapter 4 is to examine the effects of proximity to wilderness areas on residential property values in Yavapai County, Arizona by utilizing geographic information system and recently developed spatial models. On average, a residential property located one mile closer to a nearest wilderness

area is estimated to have values $4730 and $5024 higher based on Euclidean and road distance as a measure of proximity respectively.

Table of Contents

List of Figures

List of Tables

List of Acronyms

AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion
CAR	Conditional Autoregressive
DEM	Digital Elevation Model
DGP	Data Generating Process
DIC	Deviance Information Criterion
GIS	Geographic Information Systems
GM	Generalized Moment
GMM	Generalized Method of Moment
GPS	Global Positioning Systems
GSTSLS	Generalized Spatial Two Stage Least Squares
HAC	Heteroskedasticity and Autocorrelation Consistent
HET	Heteroskedastic Error Term
MCMC	Markov chain Monte Carlo
MCP	Minimum Convex Polygons
MH	Metropolis Hastings
ML	Maximum Likelihood
MLE	Maximum Likelihood Estimate
OLS	Ordinary Least Squares
S2SLS	Spatial Two Stage Least Squares
VRM	Vector Ruggedness Measure
WTP	Willingness to Pay

CHAPTER 1

Habitat Selection by Mountain Goats: An application of Bayesian MCMC inference in spatial count model using GPS collared data

1.1 Introduction

The amount and distribution of escape terrain (steep, broken and sloped topography), forage, and snow depth are the primary determinants of mountain goat winter range (Chadwick, 2002; Gross et al., 2002; Schoen and Kirchoff, 1982). Goats either frequent such alpine areas where strong winds expose forage (Fox, 1983), especially south facing wind-swept ridges (Hansen and Archer, 1981), or migrate to lower elevation forests and thick patches of alder to escape deep snow (Fox et al., 1989; Hebert and Turnbull, 1977). These same locations may overlap with human recreation, or potentially be altered by climate change. For outdoor recreation, the rate and intensity of participation among Alaskans are higher than for residents of other states and are projected to continue (Bowker, 2001) suggesting an increasing demand for backcountry recreation (eg., heliskiing, cross-country skiing and snow machining) within mountain goat ranges (steep, high-elevation terrain) of the Kenai Peninsula. The Chugach National Forest and the Alaska Department of Fish and Game, manage these habitats and mountain goat numbers on the Kenai Peninsula (South-central, Alaska). Both agencies are concerned about increasing human activity displacing

goats from important wintering areas, and long-term climatic events potentially reducing goat habitats. Both factors may negatively affect mountain goat populations, whose numbers on the Kenai Peninsula have been declining for decades (McDonough and Selinger, 2008).

For agencies to effectively manage mountain goats, they require information describing the relationship between these species and the human use and climatic factors affecting them. A first step lies in quantifying and evaluating which variables or combination of variables that best predict habitats for mountain goats on the Kenai Peninsula. Such habitat information enables subsequent predictive modeling for how human use or climate change may affect the quality and quantity of goat habitat, while providing a basis to evaluate how changing environments may affect the presence of mountain goat assemblages on the Peninsula.

This chapter addresses these concerns by generating a habitat model for mountain goats inhabiting the Kenai Peninsula. This study uses count regression modeling technique for GPS-collared (global positioning systems) location data and extends the use of spatial analysis to incorporate spatial correlation present in the data. More precisely, Poisson and Negative Binomial models are used in the spatial framework based on Gaussian Conditional Autoregressive (CAR) formulation. Although some recent studies have used count regression in the habitat selection and use studies based on GPS-collared data (Harris et al., 2014; Nielson and Sawyer, 2013), to the best of my knowledge, the application of CAR formulation to account for spatial correlation has not been used in the habitat

selection/use studies and is a novel method used in this study. The results from the model outcomes are then used to produce suitable habitat for mountain goats in the mountains of the Kenai Peninsula. The following section provides a brief review of selected literatures followed by the methods used in this study. Study area and data are described next followed by results. The final section provides discussions and conclusion.

1.2   Review of Selected Literatures

There exists a wide selection of literature analyzing the habitat selection and use for different species. The following sections provide a brief discussion of some selected literatures that focus on the habitat use by different species.

1.2.1 Studies on habitat selection by species other than mountain goat

Wintle et al. (2005) review fauna habitat modeling and model evaluation methods including a case study of habitat modeling for seven species in New South Wales (NSW). The purpose of the authors is to put together recent developments in wildlife habitat modeling into one modeling and evaluation framework and present those methods that seem to be most appropriate for predicting the distribution of species habitat in a conservation planning context in which technical expertise is limited. The most important factor determining

which modeling method to apply in any given situation depends on the nature of the available data. There are five main levels of data availability: 1) little or no data, 2) presence-only (or ad hoc) data, 3) presence-absence (or binary) data, 4) ordinal categorical data, and 5) count data. Out of these five levels of data, the situations with ordinal categorical and/or count data are not dealt by the authors because these situations arise very rarely in conservation planning due to prohibitive costs associated with capture of the data.

Pereira and Itami (1991) use multivariate statistical techniques to develop habitat suitability models for the Mt. Graham red squirrel. The Mr. Graham red squirrel, an endangered species, is very selective in choosing locations, not only for midden but also for general activity areas. The major objectives of the study are to predict the probability of red squirrel presence or absence, based on a series of environmental and location-based descriptor variables. In addition to the multivariate statistical models, the authors use Habitat Evaluation Procedures/Habitat Suitability Indices (HEP/HIS) in their analysis. On the basis of the impact assessment results, the authors argue that the Mt. Graham red squirrel is in danger of extinction, and preservation of its habitat is a necessary condition for long-term survival.

Harris et al. (2008) examine elephant habitat use in two different ecosystems: arid Etosha National park in Namibia and tropical woodlands of South Africa (Tembe Elephant Park) and Mozambique (Maputo Elephant Reserve). The major objective of the study is to quantify the habitat characteristics and express similarities in elephant habitat choices as 'rules'. This

is done by considering the distances that elephants move together with features of landscape-data. GIS data are combined with three most appropriate variables: proportions of different vegetation types, access to water, and distances from human settlements. Other things being equal, the elephants are found to move short distances, keep close to water, select the highest vegetation cover, and avoid people.

1.2.2 Studies on mountain goats and their habitat use

Côté and Festa (2001) assess the effects of maternal characteristics, forage quality and population density on kid birthdate, mass and survival in a population of marked mountain goats in Alberta. They find that the mountain goats are constrained to give birth in a short birth season synchronized with forage productivity. From 1993 to 1997, the median birthdate did not vary statistically. Moreover, the authors do not find significant variation in the kid birthdate even when the effects of maternal age, social rank, previous breeding experience (primiparous vs multiparous) and density are considered. For kid mass, it is found that the mountain goat kids gain mass linearly during summer. Similarly, overall kid survival was 78.5 percent to weaning and 60.3 percent to 1 year.

Singer and Doherty (1985) reports mountain goat responses to highway underpasses, road construction activity, lead-in fencing, and restrictive walls as the mountain goats traveled to and from a natural mineral lick in Glacier

National Park, Montana. Mountain goat selection of crossing routes with security cover suggests that they are most disturbed during the pre-construction period. The use of more open exposed route was found to be increased during general construction. After completion of both bridges, only 5 of 250 groups moved around the ends of the facilities to cross the highway, but they were singles and only 0.4 percent of all crossing goats.

Gross et al (2002) use logistic regression to develop habitat models from observation of mountain goats in alpine habitats near Mr. Evans, Colorado. The goal of the authors is to determine whether habitat use by mountain goats could be predicted from GIS-based data and to evaluate seasonal differences in the intensity and spatial patterns of habitat use. The results of the study show large differences between typical group size and average group size indicating that distributions of sizes are skewed toward large groups. A comparison of attributes of active sites and the entire study area revealed differences in the distribution of predictor variables between areas used and those of the entire study area. Observed mountain goats selected sites closer to escape terrain and at intermediate elevations within the study site. Moreover, predicted habitat suitability increased with proximity to escape terrain and with southern exposure. The coefficients of elevation and slope are positive whereas the coefficients of squares of elevation and slope are negative indicating that habitat use by mountain goats is most strongly associated with mid-elevations and intermediate slopes. The results show little difference between summer and winter habitat used by mountain goats.

Poole et al. (2009) examine winter habitat selection and wintering strategies by mountain goats in two adjacent areas of southeastern British Columbia using 15 GPS collared mountain goats in each areas. Using the multiple logistic regression analysis, the authors find topographic variables among the most important determinants, where mountain goats tend to select upper mid-elevations and warmer aspects. Moreover, goats tend to be closer to the escape terrain possibly to avoid predators.

Multiple logistic regressions and discrete choice modeling (resource selection function) have been the common statistical method employed in the studies of mountain goat habitat use where the goat locations data become available. As mentioned above, some recent studies have employed negative binomial regression to GPS collared data. This paper takes into account the spatial CAR formulation in the count regression models by using goat locations to examine the habitat use by mountain goat.

1.3   Methods

For mountain goats, a good deal of research uses remotely sensed images, geographical information systems (GIS) and GPS data to identify suitable habitat of ungulates. The GPS collar data helps researchers track individual animals over a given period and allows for modeling resource use by capturing serial or spatial correlation, resulting from animal movements within a geographical region (Legendre and Fortin, 1989; Lichstein et al., 2002; Ramsey and Usner, 2003). It

is highly likely that species more often use resources that are close to its current location, and GPS collar data, therefore, permits incorporating these spatial components into the modeling. Discrete choice modeling is a widely used technique that accommodates the probability of individual choices, and the movement of animals can be included in the model to control the resulting serial correlation. (Cooper and Millspaugh, 1999; Harris et al., 2008; Manly et al., 2002; Nielson et al., 2009).

In this study, spatial count regression models (Poisson and negative binomial regression models) are employed using GPS-based goat locations. These count regression models produce coefficient estimates similar to those generated by discrete choice models, but in a computationally efficient way. More precisely, when the probability of individual choices is exclusively the function of choice specific attributes, Poisson regression produces identical parameter estimates as the discrete choice model (Guimaraes et al., 2003). The relatively simple count regression models allow researchers to extend these models to incorporate neighboring cells' spillover effects. In the presence of spatial interdependency, the inclusion of spatial autocorrelation into regression model helps improve the consistency of coefficient estimates (Dormann, 2007).

As mentioned above, this study uses Poisson and Negative Binomial model in the spatial framework based on Gaussian Conditional Autoregressive (CAR) formulation. Moreover, this study extends the use of spatial analysis to Poisson and negative binomial regression models which are useful not only for this study which uses GPS based locations data, but also for the data representing species

distributions exhibiting discrete values at different locations (such as bird counts, aerially surveyed mammals, etc).

1.3.1 Basic Model

Following Cooper and Millspaugh (1999), the basic model starts with the discrete choice model that specifies the probability of choosing a particular alternative derived under the random utility theoretical framework popular in economic theories. Under the random utility framework, the movement of an ungulate from any location to a particular location j can be viewed as the utility of location j being higher than all other alternatives. Therefore, if U_i denotes the utility from location i, the probability that the ungulate chooses a particular location j is given by

$$
\begin{aligned}
P_j &= \Pr\left(U_j > U_i\right), \forall j \neq i \\
&= \Pr\left(U_j - U_i\right), \forall j = i
\end{aligned}
$$

In the absence of spatial effects, it may be assumed that utility from location j depends exclusively on the landscape characteristics of that location so that the utility obtained by an ungulate by choosing location j can be represented as

$$
U_j = \beta x_j + \varepsilon_j \tag{1.1}
$$

where, x_j is the vector of observed variables on attributes of alternative choice j, ε_j is the unobserved error term and β is the parameter to be estimated. If ε_j are

9

independent and have an extreme-type-value 1 Weibull distribution, the probability of choosing site j among all other alternatives is given by

$$P_j = \frac{\exp(\beta x_j)}{\sum_{j=1}^{J} \exp(\beta x_j)} \qquad (1.2)$$

where, J is the total number of available alternative locations.

If the probability of choosing a particular alternative depends exclusively on the vector of that alternative's attributes, the results from standard conditional logit model (or discrete choice model) could be identically estimated by running a simple Poisson regression model with the response variable being the number of times a particular location is chosen (see Appendix A for the equivalency between Poisson and Conditional logit model). The transition from a conditional logit to a Poisson model has at least one notable advantage: it helps to overcome the problem of estimation associated with the large number of choices by reducing the computational complexity. Partly because of this added benefit, it allows researchers to include spatial covariates to explain the effects of neighboring locations' attributes on an ungulate's choices.

Suppose Y_i is a response variable representing the number of times an ungulate appears in a particular location j (j = 1, 2, ..., J). Suppose explanatory variables are denoted by x_j, the landscape characteristics of location j. Then the standard Poisson distribution for the response variable is given by

$$P(Y = y) = \frac{\exp(-\lambda)\lambda^y}{y!} \qquad (1.3)$$

where, $\lambda = E(Y) = V(Y)$. One limitation of the Poisson model is that the mean and variance are equal. The negative binomial model is an alternative to the Poisson model that allows the variance to deviate from mean and captures cross-section heterogeneity. The negative binomial distribution for the response variable Y is given by

$$\Pr(Y = y) = \frac{\Gamma(y+\theta)}{y!\Gamma(\theta)}\left(\frac{\theta}{\theta+\lambda}\right)^{\theta}\left(\frac{\lambda}{\theta+\lambda}\right)^{y} \qquad (1.4)$$

where, $\lambda = E(Y)$ and $V(Y) = \lambda + \lambda^2\theta$, and θ captures the cross-section heterogeneity.

1.3.2 Model with Spatial Component

The probability distributions given by the conditional logit model (equation (1.2)), Poisson model (equation (1.3)) and negative binomial model (equation (1.4)) assume that the choice decision depends only on the characteristics of location j and therefore ignores the attributes of neighboring locations. However, it is reasonable to assume that the characteristics of neighboring locations have significant influence on animals' choices. This is especially true in case of mountain goats, which prefer landscapes with higher forage availability and at the same time try to remain within the close proximity of the escape terrain to avoid predators (Gross et al., 2002; Poole et al., 2009).

Theoretically, including the attributes of neighboring locations, the utility

function (equation (1.1)) takes the form

$$U_j = \beta x_j + \sum_{i=1}^{J} w_{ji} \beta x_i + \varepsilon_j \qquad (1.5)$$

where, w_{ji} is the weighing sequence defined in terms of the distance between

locations j and i. For estimation purposes, the spatial effects are included into the

count regression models. Therefore, if γ is the vector of spatial random effect for

each location, the parameter estimates follows

$$E(Y \mid \beta) = \exp(x\beta + \gamma) \qquad (1.6)$$

Here, the parameter γ is the assumed to be normally distributed with

$$\gamma \sim N\left(0, \sigma^2 Q^{-1}\right)$$

the precision matrix $Q = Q_{ij}$ given by

$$Q_{ij} = \begin{cases} 1 + |\psi| \cdot N_i & i = j \\ -\psi & i \sim j \\ 0 & otherwise \end{cases}$$

where, the notation $i \sim j$ implies that locations i and j are neighboring locations

and N_i is the number of neighbors of location i. The parameter ψ determines the

overall degree of spatial dependence, so that $\psi = 0$ implies locations are spatially

independent and as ψ increases the degree of spatial dependence increases.

1.3.3 Estimation Method

Using the standard conditional logit model, the estimation method follows maximization of log likelihood given by

$$\ln L = \sum_{j=1}^{J} n_j \ln P_j \qquad (1.7)$$

where, $n_j = 1$ if the ungulate picks choice j and $n_j = 0$ otherwise. If n_j is independently Poisson distributed, following Guimarães et al. (2003), it can be shown that the log likelihood is given by

$$\ln L = \sum_{j=1}^{J} n_j \ln P_j - N + N \ln N - \sum_{j=1}^{J} \ln n_j! \qquad (1.8)$$

The first term on the right hand side of equation (1.8) is exactly identical to the log likelihood in equation (1.7) and the remaining terms are independent of the parameters.[1] The maximization of likelihood therefore provides identical parameter estimates.

In order to incorporate spatial dependency, this study considers models in a Bayesian context. Since the resulting parameters are not analytically tractable because of the high dimensional posterior distributions, parameters are estimated using a Markov chain Monte Carlo (MCMC) method. MCMC generates approximate samples for the posterior distribution from which the mean and

---

[1] The details of equivalence between conditional logit and Poisson models are provided in the Appendix A.

standard error of the parameter estimates are obtained. There are various samplers available, such as Gibbs sampler and Metropolis Hastings (MH) sampler. This study uses R Statistical Software (2013) to estimate parameters by using built-in R package 'spatcounts' developed by Schabenberger (2009). This package uses independence MH sampler, relying on a Student's t-distribution with $v = 20$ degrees of freedom. Were $p(\theta|y)$ the posterior distribution, then the independence MH sampler algorithm used in our models can be summarized in the following steps.

Step 1: Choose a starting value $\theta(0)$ and the number of iterations T. Set $t = 1$.

Setp 2:

a) Calculate the mode θ_{mode} and the inverse curvature at the mode $-H(\theta_{mode})^{-1}$ of the target distribution.

b) Propose a candidate value

$$\hat{\theta} \sim q(\hat{\theta}) = \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{v\pi}} \frac{1}{\sigma} \left(1 + \left(\frac{\hat{\theta}-\theta}{\sigma} \right)^{-2} v^{-1} \right)^{-(v+1)/2}$$

where $\sigma = \left(-\frac{v+1}{v} H(\theta_{mode})^{-1} \right)^{1/2}$.

c) Accept $\hat{\theta}$ with probability $\min\left\{1,\dfrac{p(\hat{\theta}|y)}{p(\theta^{t-1}|y)}\dfrac{q(\theta^{t-1})}{q(\hat{\theta})}\right\}$ and set $\theta^t = \hat{\theta}$,

otherwise set $\theta^t = \theta^{t-1}$.

Step 3: If $t = T$ stop the algorithm. Otherwise set $t = t + 1$ and return to Step 2.

The maximum iteration is set at $T = 20{,}000$ for each of the Poisson and negative binomial models. For the models without spatial components, first models are estimated and corresponding Akaike's Information Criteria (*AIC*) are used to identify the model with the best fit. However, in the context of spatial regression models, Deviance Information Criterion (*DIC*) is used. *DIC* is particularly useful in Bayesian model selection problems and is a popular criterion designed to compare hierarchical models. It is computed as

$$DIC = 2\overline{D(\theta|y)} - D(\bar{\theta})$$

where $\overline{D(\theta|y)}$ is the estimated posterior mean of the deviance and $D(\bar{\theta})$ is the deviance of the estimated posterior means. The model with the lower values of *AIC* and *DIC* are preferred.

1.4   Study Area and Data

The study area includes interior mountains within the Kenai Peninsula of South Central Alaska (Figure 1. 1). This study uses the data obtained from Alaska

15

Department of Fish and Game which deployed Tellus 2D GPS collars (manufactured by Televilt) on 12 mountain goats in two different areas (seven goats in the Gilpatric Mountains and five near Spencer Glacier) in October 2007. The data on ten mountain goats were successfully retrieved in May 2008. Because of GPS errors on some of the goats, this study uses only the goat locations of five and three goats during winters of 2007 and 2008, respectively. Winter period are identified as the period from January 1 to March 31 following Smith (1977) and Poole et al. (2009). These data sets for eight goats during winter consist of mountain goat GPS locations recorded at an interval of one hour with only 7% missing observations. Nielson et al. (2009) show that if 10% or more of fixes are unsuccessful it can bias results but do not investigate the lower levels of missing data in the current sample.

With GPS locations from all of these goats, a minimum convex polygon is created. This area is then buffered by 5 km, the maximum distance moved by a goat in one hour within this data set. The entire area is used to classify the suitable habitat for mountain goats. Nevertheless, the analyses are performed at finer scales, and hence, the regression parameters are estimated using a smaller area for each goat. For the smaller area, minimum convex polygon (MCP) is created from the GPS locations of each goat during the given winter period and a systematic rectangular grid (of 60×60m resolution) is created within each MCP. The number of goat locations recorded within each 60×60m grid cell in the three months period of winter is considered the response (dependent) variable for the count regression models.

Explanatory variables include elevation, distance from escape terrain (DistEsc), aspect, vector ruggedness measure (VRM), shrub and rock. This study uses digital elevation model (DEM) raster images of 60m × 60m resolution to extract slope, aspect and VRM.[2] Aspect is converted to a continuous variable by taking the cosine of the angle so that its value ranges from -1 to +1, representing south to north. VRM is calculated using the ArcGIS Script (Sappington et al., 2007) over a 3 × 3 neighborhood cell, and is converted to a 0-100 scale. Higher values of VRM represents more rugged terrain.

There is no consensus among the existing studies defining escape terrain. For example, Gross et al. (2002) consider slope greater than 33 degrees whereas Poole and Heard (2003) categorize areas with slope greater than 45 degrees as escape terrain. In this study, the landscape grid with slope greater than 40 degrees is arbitrarily considered as the escape terrain and the distances from each cell grid to a nearest escape terrain is calculated (scaled to 100m). Following the standard procedure to include dummy variables in a regression model, only two categorical variables (shrub and rock) are used; the third category (forest) is considered a base category.

---

[2] The variables slope and distance to nearest escape terrain are found to be highly correlated. Distance to a nearest escape terrain is considered more important characteristic based on the existing literature (Poole et al, 2008, Gross et al, 2002) and as such, slope is dropped from the models.

1.5 Results

1.5.1 Spatial count regression

Table 1. 1 and Table 1. 2 report parameter estimates generated by Poisson and Negative Binomial regressions without spatial components, while Table 1. 3 and Table 1. 4 present regression results with spatial components.[3] The coefficients for each goat are estimated separately. Each column contains the parameters and their standard errors for an individual goat.[4] As this study focuses on introducing the spatial count regression model, models are kept simple by including only the most important variables. Based on previous literature (Gross et al., 2002; Poole et al., 2009), positive signs are expected for parameters describing elevation, VRM, rock and shrub, and negative signs for Aspect (cosine of aspect) and DistEsc (distance to nearest escape terrain). With few exceptions, most coefficients are statistically significant and have the expected signs.

Results from simple count regression models are not comparable directly with the spatial count regressions because the two methods are different and additional parameters are included in the spatial count regression models. The

---

[3] Results from discrete choice models are presented in Appendix B. When distances from previous locations are not included, the coefficients from discrete choice models are exactly identical to the Poisson model without intercept term.

[4] Attempts are made to consider pooling the data and estimating the joint coefficients but rejected based on the likelihood ratio test.

parameter ψ is statistically significant emphasizing the spatial dependency

(Table 1. 3 and Table 1. 4). Moreover, if the insignificant coefficients are ignored,

and models are compared based on the estimated coefficients about whether they

meet the expected sign, the spatial count regression could be considered to better

predict goat choices. For example, all of the statistically significant coefficients

for elevation are positive in the spatial count regression models but some of the

coefficients have negative signs in simple regressions.

Coefficients for continuous variables in a count regression model can be

loosely interpreted as the rate ratio corresponding to a one-unit difference in the

predictor. In all the models of this study, the dependent variable is the number of

times a goat visited a particular location (grid-cell). For estimation, the

explanatory variables are deviated from their mean values. Therefore, the

statistically significant coefficient for elevation in the Poisson model (Table 1. 3)

ranges from 0.5 to 0.94 across different models which implies that goats

increases the intensity of use by 1.65 to 2.6 times in areas that are one unit

higher than the mean elevation values. Similarly, the numbers of visits to a

particular location by goats are 1.12 to 1.54 times higher for 1 percentage point

greater values in VRM index from their mean values.

The coefficients on DistEsc are most consistent and statistically significant

throughout the regression models emphasizing the importance of escape terrain

on goat preferences. The coefficients on DistEsc from spatial Poisson regression

across models ranges from -1.26 to -0.47 indicating the number of visits to

increase by 37.5 to 71.6 percent when a location is closer to the escape terrain

19

measured by one unit greater than their mean values. The negative and significant parameters corresponding to Aspect implies that mountain goats prefer cliffs with southerly aspect. Similarly, mountain goats prefer barren rock and shrub as implied by positive signs on their corresponding coefficients across individual models.

1.5.2 Habitat Classification Map

This study uses the regression coefficients from spatial Poisson and negative binomial models to classify suitable habitat for mountain goats inhabiting the Kenai Peninsula (Figure 1. 2). First, separate maps describing habitat suitability for each mountain goat are obtained by using the coefficients from each regression model. Raster grids from each model are then combined to gain a single raster image representing the average values. The five different ranges of suitability are classified with quantile classification in ArcGIS. Figure 1. 2 also shows actual goat locations overlaid on the classified map. It is found that 84 and 79 percent of goat locations are contained within the predicted top 20 percent (most preferred) areas from spatial Poisson and negative binomial models respectively.

1.6 Discussions and Conclusion

Habitat models producing accurate maps help wildlife managers define important habitats thereby helping conserve wildlife populations. The aim of the present study is to classify suitable mountain goat habitat in Kenai Peninsula

and, while doing so, provide an improved estimation procedure with the GPS based location data, namely spatial count regression. These count regression models also produce regression coefficients similar to the discrete choice models.

One issue in using GPS collar for generating species and habitat relationships is serial correlation in animal movements (Ramsey and Usner, 2003). It is highly likely that two consecutive locations will be close to one another as there is always a maximum feasible distance moved by an animal. One approach to control for such correlation would be to include the distance moved by an ungulate as an explanatory variable in the regression model. Discrete choice models have this noticeable advantage over simple Poisson models, since the distance moved by an animal could be incorporated into the model. However, this requires maximization of the likelihood function, which increases the computation burden and becomes infeasible while dealing with large data sets. The alternative to a maximum likelihood method, such as a conditional logit model that can be fitted with some standard statistical packages, requires much effort in data manipulation and organization before running the estimation procedure, especially when the distance moved by an animal is to be included. Incorporating distance as a variable in count regression models is theoretically very complicated, if not impossible. However, count regression models allow researchers to address the issue of spatial or serial correlations by including the effects of neighboring locations into the model.

As this study introduces spatial effects into the model, models are kept simple by including mainly terrain features. Instead, emphasis is given more on

the implementation of methods to incorporate spatial effects for the data derived from GPS collared animal locations rather than to explore any additional determinants of goat habitat use. However, depending on the availability of data, these models can be extended to include variables reflecting human activities and climate change. Knowledge of the ecological foundations forming mountain goat habitat fosters management of these areas, and enables ecologists to assess how such recreation overlaps with them. Accurate maps of goat habitat describing locations animals frequent during winter would help attenuate such effects. Moreover, such areas can serve as scientific controls, or benchmarks, to evaluate and compare seasonal movements and distributions of animals in the absence of human disturbance, since recreational activities hold potential to displace goats from preferred wintering areas. If so, reducing the quality or quantity of winter ranges can detrimentally affect these ungulate populations (McDonough and Selinger, 2008).

Table 1.1: Coefficient Estimates from Poisson Regression for Individual Goats (without spatial component)

	G2	G5	G6	G840	G870	G5-8	G6-8	G870-8
Constant	1.141***	-0.503*	-1.457***	0.917***	-6.597***	2.066***	0.339*	-2.534***
	(0.149)	(0.252)	(0.184)	(0.099)	(0.289)	(0.175)	(0.160)	(0.201)
Elevation	-0.033*	-0.118***	0.051***	-0.050***	0.802***	-0.089***	0.036*	0.355***
	(0.017)	(0.010)	(0.014)	(0.010)	(0.031)	(0.021)	(0.015)	(0.024)
Aspect	0.663***	-2.977***	-0.950***	-0.374***	-4.289***	0.078	-1.060***	-2.812***
	(0.053)	(0.244)	(0.038)	(0.081)	(0.238)	(0.092)	(0.038)	(0.161)
VRM	-0.095***	-0.502***	0.125***	0.198***	0.806***	0.002	0.206***	0.800***
	(0.018)	(0.039)	(0.010)	(0.015)	(0.023)	(0.039)	(0.009)	(0.023)
DistEsc	-1.303***	-0.956***	-0.543***	-0.962***	-1.577***	-0.855***	-0.614***	-1.193***
	(0.049)	(0.037)	(0.015)	(0.038)	(0.044)	(0.019)	(0.021)	(0.030)
Rock	0.856***	0.529**	0.754***	-2.817***		1.829***	-0.919***	
	(0.073)	(0.170)	(0.110)	(0.320)		(0.089)	(0.079)	
Shrub	0.392***	0.502***	0.762***	0.558***	-2.108***	2.321***	-0.387***	-0.807***
	(0.075)	(0.053)	(0.108)	(0.049)	(0.098)	(0.075)	(0.071)	(0.082)
ln$L$	-3137	-4271	-6042	-4857	-2874	-2339	-6027	-3227
AIC	6288	8557	12097	9728	5761	4693	12068	6467
N	803	1009	5541	1631	345	834	2306	628

Notes: Standard errors in parentheses. $\cdot\, p < 0.10$, $^* p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$. Elevation and Distance from escape terrain (DistEsc) were measured in 100m. Aspect was Cosine of Aspect from DEM. VRM was an index from 0-100, higher value denoting more rugged surface. Rock and Shrub were dummy variable, 1 if the center of the cell of 60m represented Rock and Shrub, 0 otherwise. The column headings (G2, G5, ...) represent individual goats.

Table 1. 2: Coefficient Estimates from Negative Binomial Regression for Individual Goats (without spatial component)

	G2	G5	G6	G840	G870[a]	G5-8	G6-8	G870-8[a]
Constant	-0.825	0.715	-2.877***	-1.153**	2.531**	0.267	-1.200	-1.605*
	(0.488)	(0.572)	(0.535)	(0.404)	(0.936)	(0.647)	(0.656)	(0.640)
Elevation	0.157**	-0.202***	0.047	0.066	-0.651***	0.054	0.072	-0.052
	(0.056)	(0.040)	(0.042)	(0.043)	(0.115)	(0.069)	(0.062)	(0.082)
Aspect	0.799***	-1.796***	-1.108***	-1.305***	-3.812***	-0.180	-0.739***	-2.952***
	(0.137)	(0.469)	(0.099)	(0.272)	(0.564)	(0.267)	(0.121)	(0.368)
VRM	-0.056	-0.368***	0.256***	0.380***	0.940***	0.177	0.201***	0.938***
	(0.051)	(0.111)	(0.044)	(0.084)	(0.146)	(0.140)	(0.057)	(0.134)
DistEsc	-1.119***	-1.004***	-0.576***	-0.857***		-0.546***	-0.502***	
	(0.105)	(0.101)	(0.036)	(0.090)		(0.067)	(0.055)	
Rock	0.706**	0.128	2.402***	-2.162***		1.932***	0.277	
	(0.252)	(0.627)	(0.315)	(0.460)		(0.359)	(0.326)	
Shrub	0.997***	0.951***	1.905***	1.419***	0.400	2.140***	0.907**	-0.177
	(0.220)	(0.225)	(0.309)	(0.236)	(0.446)	(0.261)	(0.307)	(0.372)
θ	4.42***	6.67***	14.28***	12.5***	3.571***	6.25***	13.7***	6.859***
	(0.37)	(0.55)	(0.89)	(1.043)	(0.431)	(0.625)	(0.938)	(0.067)
lnL	-1088	-1197	-2390	-1197	-1010	-904	-1520	-1717
AIC	2192	2410	4796	2409	1022	1824	3056	1730
N	803	1009	5541	1631	345	834	2306	628

Notes: Standard errors in parentheses. · $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. See notes of table 1 for variables measurements. [a] Model that include all variables did not converge, perhaps due to smaller sample size.

Table 1.3: Coefficient Estimates from Spatial Poisson Regression for Individual Goats (with spatial component)

Variables	G2	G5	G6	G840	G870	G5-8	G6-8	G870-8
Elevation	0.5**	0.22	0.84***	0.32	0.33	0.72*	0.94***	0.62**
	(0.189)	(0.173)	(0.229)	(0.267)	(0.254)	(0.308)	(0.247)	(0.225)
CosAspect	0.98**	-2.01**	-1.48***	-1.04·	-2.63***	0.32	-0.93**	-1.86***
	(0.351)	(0.707)	(0.28)	(0.589)	(0.548)	(0.574)	(0.286)	(0.476)
VRM	0.13*	-0.31*	0.17*	0.36*	0.41**	0.28	0.05	0.43*
	(0.065)	(0.155)	(0.083)	(0.145)	(0.157)	(0.23)	(0.077)	(0.135)
DistEsc	-1.13***	-1.02***	-1.06***	-1.26***	-0.71**	-0.9***	-0.47**	-0.52**
	(0.226)	(0.226)	(0.208)	(0.331)	(0.249)	(0.214)	(0.179)	(0.191)
Rock	0.21	0.98	1.22·	-0.41	--	0.68	1.12*	--
	(0.277)	(0.875)	(0.631)	(0.634)		(0.578)	(0.491)	
Shrub	0.54·	0.2	1.16*	0.78*	-0.23	0.97*	1.13**	-0.31
	(0.293)	(0.33)	(0.579)	(0.346)	(0.463)	(0.413)	(0.437)	(0.373)
γ	-2.26	-2.89	-8.57**	-5.36*	-2.03	-3.38	-5·	-2.2
	(1.968)	(2.248)	(3.082)	(2.711)	(1.518)	(2.286)	(2.551)	(1.517)
ψ	94.41*	72.3**	117.52**	85.55***	41.06·	80.91*	94.55**	56.1·
	(42.85)	(25.4)	(40.59)	(25.61)	(22.68)	(31.41)	(31.21)	(31.44)
lnL	-616	-649	-1087	-586	-308	-483	-753	-469
AIC	1244	1309	2187	1183	626	979	1518	948
DIC	1515	1607	2783	1482	741	1193	1906	1133
N	803	1009	5541	1631	345	834	2306	628

Notes: Standard errors in parentheses. $\cdot p < 0.10$, $^* p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$. See notes of table 1 for variables measurements.

25

Table 1.4: Coefficient Estimates from Spatial Negative Binomial Regression for Individual Goats (with spatial component)

Variables	G2	G5	G6	G840	G870	G5-8	G6-8	G870-8
Elevation	0.36	0.21	0.94***	0.24	0.47	0.72*	0.87***	0.58**
	(0.232)	(0.191)	(0.244)	(0.248)	(0.286)	(0.302)	(0.227)	(0.222)
CosAspect	0.87*	-1.99**	-1.31***	-0.75	-2.18***	0.33	-0.82**	-1.78***
	(0.39)	(0.762)	(0.306)	(0.602)	(0.56)	(0.567)	(0.297)	(0.446)
VRM	0.12·	-0.32*	0.12	0.3*	0.34*	0.26	0.04	0.42**
	(0.072)	(0.154)	(0.077)	(0.14)	(0.166)	(0.21)	(0.081)	(0.129)
DistEsc	-1.05***	-0.91***	-0.87***	-1.08***	-0.55**	-0.57**	-0.37*	-0.41*
	(0.257)	(0.252)	(0.197)	(0.277)	(0.186)	(0.176)	(0.167)	(0.173)
Rock	0.21	0.81	1.02*	-0.34	--	0.69	0.9·	--
	(0.296)	(0.897)	(0.516)	(0.673)		(0.543)	(0.494)	
Shrub	0.54·	0.2	0.91·	0.81*	-0.23	0.91*	0.94*	-0.27
	(0.299)	(0.34)	(0.47)	(0.388)	(0.452)	(0.401)	(0.452)	(0.368)

Table 1.4 (contd.)

Variables	G2	G5	G6	G840	G870	G5-8	G6-8	G870-8
θ	5.36***	6.49***	7.41	4.42**	5.08**	9.09**	12.59***	7.26**
	(2.011)	(2.747)	(4.453)	(1.587)	(2.403)	(2.15)	(3.584)	(2.924)
γ	-2.14	-2.79	-8.29**	-5.18*	-1.78	-3.23	-4.9*	-2.11
	(1.847)	(2.197)	(3.129)	(2.61)	(1.456)	(2.208)	(2.467)	(1.553)
ψ	103.6*	67.46	130.28*	78.61**	32.08	93.53***	72.85***	47.5***
	(44.41)	(51.92)	(51.63)	(25.32)	(26.6)	(26.7)	(15.26)	(13.03)
lnL	-696	-724	-1192	-669	-353	-538	-798	-517
AIC	1407	1463	2398	1353	718	1090	1609	1047
DIC	1685	1778	3060	1676	835	1315	2011	1232
N	803	1009	5541	1631	345	834	2306	628

Notes: Standard errors in parentheses. $\cdot\, p < 0.10$, $^{*}\,p < 0.05$, $^{**}\,p < 0.01$, $^{***}\,p < 0.001$. See notes of table 1 for variables measurements.

27

Figure 1. 1: Study Area. Kenai Peninsula Area of Alaska with GIS location of mountain goats.
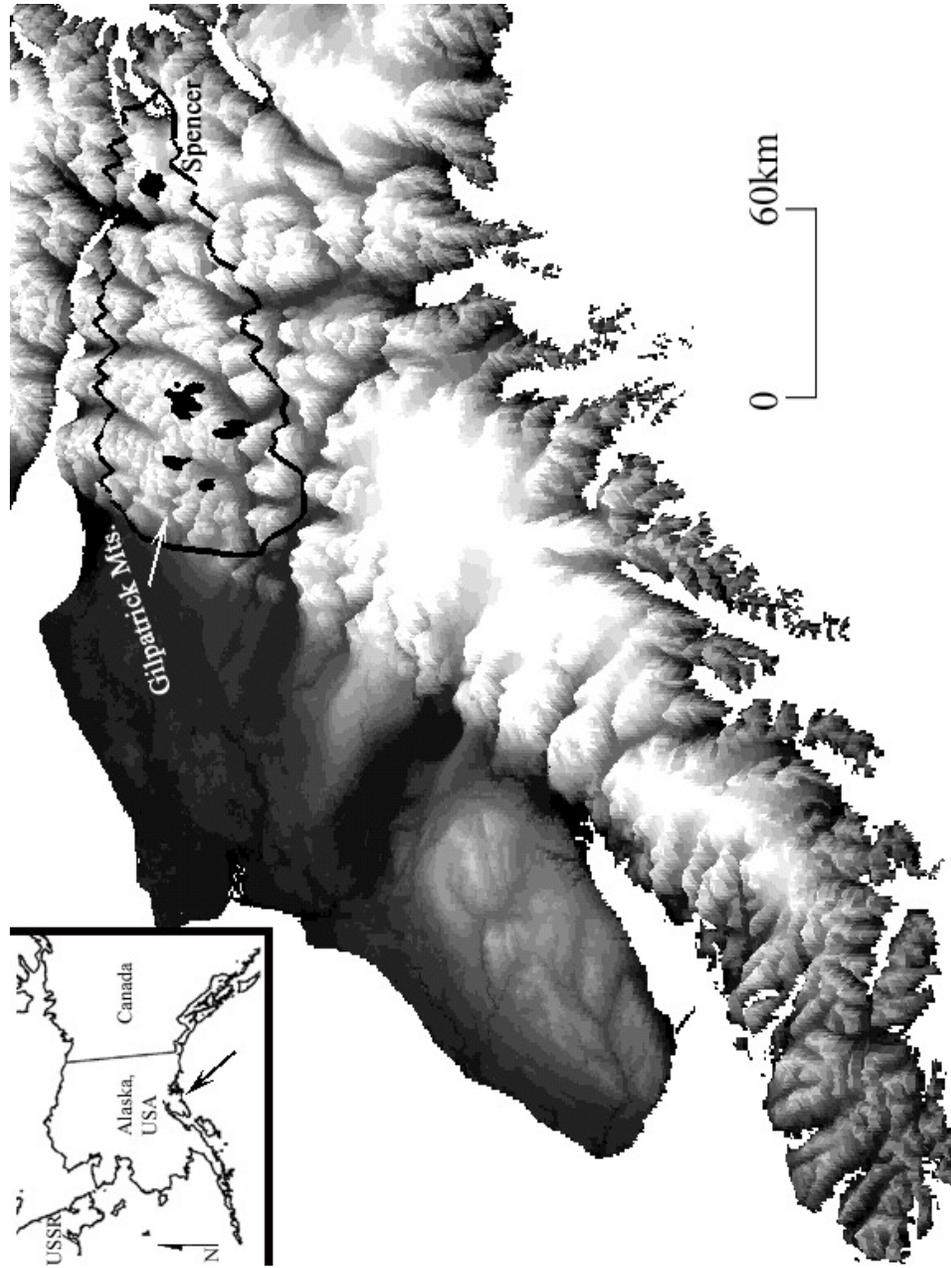
Figure 1. 2: Classification of Suitable Habitat. These classifications are based on the spatial Poisson and Negative Binomial regression models and show the five classes of suitability within the five km buffers of all the goat locations observed in the area.

Suitable Habitat

Poisson Model

Negative Binomial Model

Legend

- Actual Goat Locations
- 1 Least Preferred
- 2
- 3
- 4
- 5 Most Preferred

0   18,500   37,000   74,000
Meters

N
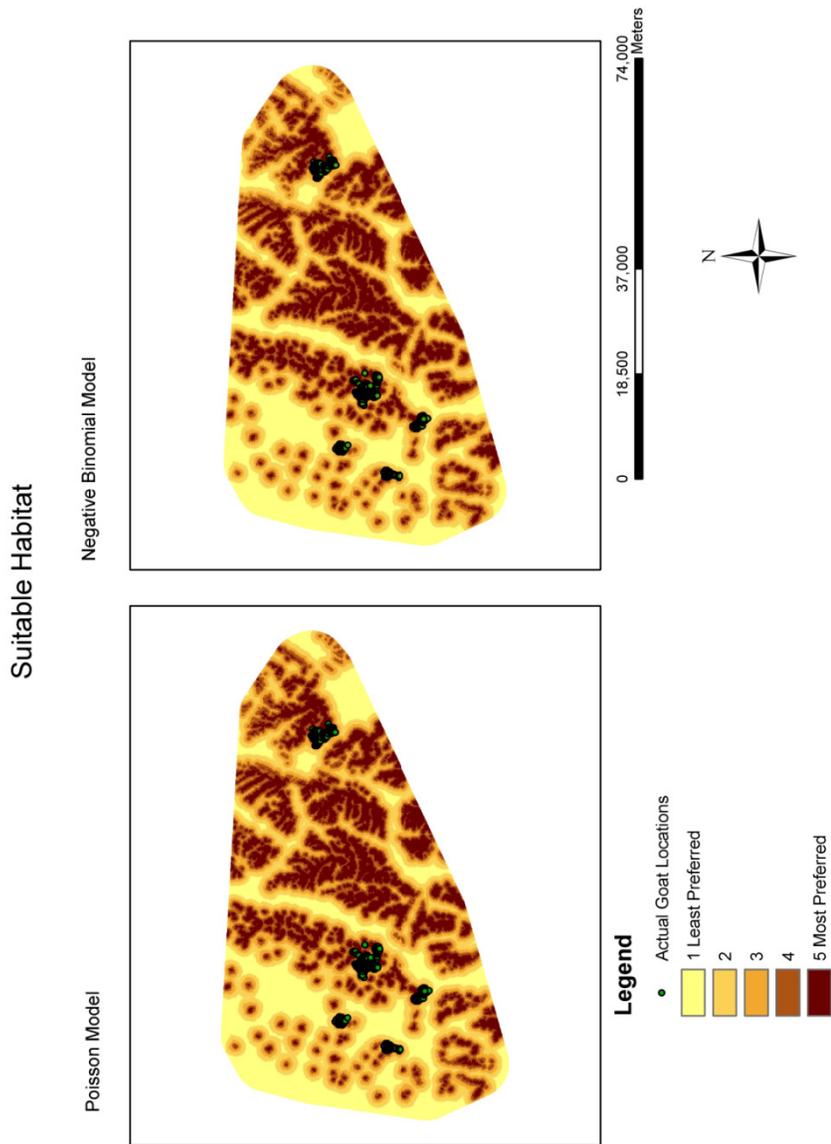
CHAPTER 2

Does winter recreation influence mountain goat habitat selection?

2.1 Introduction

In Alaska, mountain goats hold high value to locals and visitors for both consumptive (hunting) and non-consumptive (watching) purposes. The Chugach National Forest and the Alaska Department of Fish and Game manage mountain goat habitat and goat numbers on the Kenai Peninsula. Both agencies are concerned about the increasing human activity displacing goats from wintering areas. Mountain goat, being one of the big game animals, shares the total revenue generated from outdoor recreation such as hunting and wildlife viewing. According to National Survey of Fishing, Hunting and Wildlife Associated Recreation 2011, one million residents and non-residents spent $3.4 billion on wildlife recreation in Alaska. Given the huge economic benefits of wildlife, and that the wildlife population has direct relationship with wildlife associated recreational activities, maintaining sustainable wildlife population and its habitat is crucial to continue generating economic revenues from this sector.

This study seeks to achieve two main objectives relating to wildlife habitat use with special emphasis on mountain goat habitat in Kenai Mountains of Alaska. First, regression based test of equivalency is used to compare the

predictions of intensity of habitat use by mountain goats from model outcomes using two different sources of data, GPS collared locations and aerially surveyed locations. Although both sources of data contain actual locations of mountain goats, they fall under different study designs popularly employed in resource selection studies as described by Manly et al. (2002); aerial surveys are examples of Design I where measurements are made at the population level whereas GPS collared locations are applications of Design II and Design III with individual animals being identified or collected for the study. Typically Design I are inexpensive compared to Design II and III because individual animals are not captured, collected or relocated (Manly et al., 2002, p. 8). The GPS collared locations data, on the other hand, have advantages particularly in terms of increased precision and accuracy and are increasingly used in habitat selection studies (Whittington et al., 2005; Sawyer et al., 2006; Frair et al., 2010). Nevertheless, the use of GPS technology in habitat selection studies has its own limitation arising from the tradeoff between sample size and the cost of GPS units (Hebblewhite and Haydon, 2010). Comparison of resource selection by animal from different sample designs is expected to be helpful to researchers for considering analysis options when data are constrained by either the costs or the availability of information on available/use resources in the study area.

The second objective central to this study is to examine the effect of human winter recreation on mountain goat habitat selection. While factors affecting wildlife population are numerous, conflict of resource use may be considered as one of the potential components. For instance, geographical

locations suitable for mountain goat winter habitat may be used by human winter recreationist thereby reducing the home range of these animals. Understanding the influence of human disturbance on habitat of mountain goat is fundamental to achieving the objective of conserving these ungulates. To this end, Bayesian model comparison method is employed to highlight the significance of conflict between winter recreation and mountain goat habitat use.

2.2   Data and Study area

The study area includes the mountains of Kenai Peninsula, Alaska. Within Kenai Mountains, goat locations data obtained from two different sources and methods are used to examine habitat use by mountain goat. First, GPS Collar data of five mountain goats in Gilpatrick Mountains and near Spencer Glacier are used to examine habitat use by goats. These data contains actual locations of animals during the winter period (January-March) of 2007 and 2008, where GPS fixes occur at one hour intervals. The second set of data includes locations of mountain goats observed from aerial flight survey in winter period of 2005 in the Kenai Mountains. Since GPS collared goat locations are clustered in small ranges, the study area is defined by creating Minimum Convex Polygon (MCP) using all goat locations observed in aerial survey, which covers wider geographical area than the GPS locations.

For GPS Collared location data, Minimum Convex Polygons (MCPs) are created to represent home range for each goat separately. Then geographical

characteristics of the landscape within each MCP are extracted from DEM raster image of 60m×60m. Number of cells within each MCP varied from 345 to 5541 for different goats. Since this study focuses on employing count regression (Poisson and Negative Binomial) models to estimate the parameters of interest, a count-variable is generated by counting the number of times a goat appears in a particular grid cell during the study period of three months. This count-variable is used as response (dependent) variable in our regression models, which represents the intensity of habitat use by mountain goat at any given location.

For aerially surveyed data, the data set consists of a total of 288 goat locations observed during the winter period in 2005. These data include the locations as well as the number of goats observed at each of the locations. In order to consider available/unused resources, using each raster grid within the entire study area would not be feasible due to relatively larger geographical area resulting in extremely large number of grid cells within the study area. Instead, geographical points equal in number of goat locations are randomly generated using ArcMap, and are considered as locations of unused resources. The goat points and random points are combined and corresponding landscape characteristics are extracted from the raster image of 60m×60m for each point.

This study uses DEM raster image of cell size 60m to extract slope, aspect and Vector Ruggedness Measure (VRM). VRM is calculated using the ArcGIS Script (Sappington et al., 2007) over a 3 × 3 neighborhood cells, finally converting to 0-100 scale, higher values corresponding to more rugged surface (broken terrain). Distance from escape terrain, described as steep slopes of

broken rocky terrain (Gross et al., 2002), has been found to be one of the most important factors affecting goats' decision on habitat use in the existing literature (Gross et al., 2002; Poole and Heard, 2003; Poole et al., 2009). Landscape grids with slope greater than 40 degrees are defined as escape terrain. From each raster cell, distance to a nearest escape terrain is calculated in ArcMAP using 'near' tool. Aspect is converted to a continuous variable by taking Cosine of the angle so that its value ranges from -1 to +1 representing from south to north.

Additional variables corresponding to available landscape cover types are also included as explanatory variables in the regression models. Following the standard statistical procedure to treat binary independent variables in a regression model, only two features, rock and shrub, are included because within each MCP of individual goats, other features are either not available or, when available, are inadequate in number of observations to be included into the models. The same cover-type features are included in the models utilizing the aerial data so as to make the results comparable to those obtained from GPS collar data. A brief description of variables used in this study is provided in Table 2. 1.

In order to examine the effect of winter recreation on mountain goat habitat, the data on ski-tracks observed during the flight survey are used. The area of ski-tracks within 5 km buffers of each of goats' locations and random points are calculated. In addition to ski area, two variables corresponding to hiking trails in the study area are considered: distance to the nearest trail from each location, and the trail-length within 5 km buffer of each location.

2.3 Methods

This study uses Poisson and negative binomial models to estimate the coefficients. For the count regression models, suppose Y_j is a response variable and represents the goat count in a particular location j (j = 1, 2, ..., J). The standard Poisson distribution for the response variable is

$$P(Y = y) = \frac{\exp(-\lambda)\lambda^y}{y!} \tag{2.1}$$

where, $\lambda = E(Y)$ and $V(Y) = \lambda$. Similarly the negative binomial (NB) distribution for the response variable is

$$\Pr(Y = y) = \frac{\Gamma(y+\theta)}{y!\Gamma(\theta)}\left(\frac{\theta}{\theta+\lambda}\right)^\theta\left(\frac{\lambda}{\theta+\lambda}\right)^y \tag{2.2}$$

where θ captures the cross-section heterogeneity. The mean and variance of NB distribution are $\lambda = E(Y)$ and $V(Y) = \lambda + \lambda^2\theta$ respectively. The equation being estimated is then written as

$$f(\lambda) = \ln(\lambda) = \boldsymbol{\beta}\mathbf{x}$$

where, $\boldsymbol{\beta}$ and $\mathbf{x}$ are the coefficient and variable matrices respectively.

2.3.1 Estimation

This study uses 'glm' and 'glm.nb' functions in R (R Core Team, 2013) to obtain parameter estimates of Poisson and Negative Binomial regression models. However, the program codes are written in R to maximize the likelihood function for the discrete choice models (R Codes are provided in Appendix D). For GPS collar data, the regression coefficients are estimated individually for each goat.[5] The regression coefficients from GPS collar data and aerial survey data are not directly comparable as they are performed at different geographical scales. Therefore, first classified raster grids are generated with equal interval of values from 0 to 100 to represent predicted habitat use by mountain goats from both sources of data. Then regression based test of equivalence (described below) is performed. Finally, the negative binomial regressions are estimated using Bayesian method for aerially observed mountain goat locations with and without some additional features representing the proxy related to human recreational activities.

---

[5] An attempt was made to pool the data for all goats, but rejected the model based on likelihood ratio test.

2.3.2 Regression-based test of equivalence for model results from two data sets

Before using the aerially surveyed data, the regression-based equivalence test is performed to confirm that the results from two different data sets produce predictions that are statistically equivalent. In order to achieve this, I adopted the method from Robinson et al (2005), the procedure of which is briefly described below.

1. Poisson and negative binomial models without spatial components are estimated and predicted values are mapped to a raster image. These predictions from GPS data and aerial data are then transformed to 1-100 range from the raster grid with higher values corresponding to more suitable habitat.

2. The results from GPS data are deviated from mean to ensure that the estimates of slope and intercept are independent. This allowed me to perform test of equivalence on both intercept and slope independently.

3. The regions of equivalence are established at 10 percentage for both intercept and slope $\left(RE^-, RE^+ \right)$.

4. Linear regressions are fitted with mean deviated GPS model results as predictor variable and aerial model results as response variable.

5. Two one-sided 95% confidence intervals for both intercept and slope are calculated $\left(CE^-, CE^+ \right)$

The regression based test of equivalence suggests that if the computed confidence intervals fall within the regions of equivalence, the predictions from two approaches are equivalent.

2.3.3 Bayesian Estimation

Bayesian inference approach starts with the probability model for data y, which is specified by a vector of d unknown parameters $\boldsymbol{\theta} = \left(\theta_1, \theta_2, ..., \theta_d\right)$ considered as random variables. Before data are observed, prior beliefs about the parameter vector are represented by prior probability distribution $p(\boldsymbol{\theta})$. The prior distribution reflects the researcher's uncertainty about the parameters and is subjective in that any two persons may have different prior beliefs about the same parameter vector $\boldsymbol{\theta}$. Once the data are observed, the prior beliefs are revised using Bayes' theorem to construct the conditional distribution of parameter $p(\boldsymbol{\theta}|y)$ called posterior distribution. From the Bayes theorem, we have,

$$p(\boldsymbol{\theta}|y) = \frac{p(y|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(y)}$$

where, $p(y) = \int p(y|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}$ is the density of the marginal distribution of y. For estimation purposes, the posterior distribution is obtained up to a constant of proportionality depending on the data and is written as

$$p(\mathbf{\theta}|y) \propto p(y|\mathbf{\theta})p(\mathbf{\theta})$$

Therefore, the posterior distribution is proportional to the likelihood times the prior.

In this paper, Bayesian inference technique is used to estimate the parameter vector for negative binomial regression with the likelihood

$$p(y|\mathbf{\theta}) = \prod_{i=1}^{n} \frac{\Gamma(y_i + \alpha)}{\Gamma(\alpha)y_i!} q_i^{\alpha}(1-q_i)^{y_i}$$

where,

$$q_i = \frac{\alpha}{\alpha + \lambda_i}$$

and

$$\lambda_i = \exp(\mathbf{x}_i' \mathbf{\beta})$$

Here, the model specific parameter vector is $\mathbf{\theta} = (\mathbf{\beta}, \alpha)$, where α is the over-dispersion parameter. The prior for $\mathbf{\beta}$ and α are chosen as $\mathbf{\beta} \sim N(0, \sigma^{-1}\mathbf{I})$ and $\alpha \sim Gamma(a,b)$. Since the posterior distribution is not analytically tractable, Markov chain Monte Carlo (MCMC) method is applied to generate approximate samples from the posterior distribution. Particularly, this study employs Random Walk Metropolis Algorithm within the negative binomial regression model where $(\mathbf{\beta}|\alpha)$ and $(\alpha|\mathbf{\beta})$ are drawn with two different random walks.

2.3.4 Model Comparison using Bayes Factor

The posterior mean and corresponding standard deviation of model specific parameters reflect the importance of individual parameters associated with the variables entering into the model. In addition, it is often of interest to compare the fit of two or more competing models, which allows researchers to select the best model for the given data. The model comparison problem, in this paper, is addressed to emphasize the role of human recreational activities on mountain goat habitat selection. In the Bayesian context, the pair-wise model selection proceeds by comparing models through their posterior odds ratio given by

$$\text{Posterior odds} = \text{Prior odds} \times \text{Bayes Factor}$$

More precisely, the posterior odds ratio of two models M_1 and M_0 is written as

$$\frac{p(M_1 \mid y)}{p(M_0 \mid y)} = \frac{p(M_1)}{p(M_0)} \frac{p(y \mid M_1)}{p(y \mid M_0)} \qquad (2.3)$$

where,

$$p(y \mid M_i) = \int p(\mathbf{\theta}_i \mid M_i) f(y \mid M_i, \mathbf{\theta}_i) d\mathbf{\theta}_i \qquad (2.4)$$

is the marginal likelihood of model M_i, and $\mathbf{\theta}_i$ is the model specific parameter vector (Chib and Jeliazkov, 2001). If one assigns equal priors to the models, the prior odds ratio (first fraction on the right hand side of equation (2.3)) drops out

so that the posterior odds ratio equals Bayes Factor (second fraction on the right hand side of equation (2.3)), which, in turn, is the ratio of marginal likelihoods.

The marginal likelihood of each model is approximated using Laplace method[6]. Representing the marginal likelihood as

$$p(y) = \int p(\boldsymbol{\theta}) \exp\left[\ln f(y \mid \boldsymbol{\theta})\right] d\boldsymbol{\theta}$$

and integrating the quadratic approximation for the log-likelihood centered at its mode, the maximum likelihood estimate (MLE), the approximated marginal-likelihood is given by

$$\tilde{p}(y) = p(\tilde{\boldsymbol{\theta}}) p(y \mid \tilde{\boldsymbol{\theta}}) \frac{(2\pi)^{d/2}}{\sqrt{\det(\mathbf{Q})}}$$

where $\tilde{\boldsymbol{\theta}}$ is the MLE and $\mathbf{Q}$ is the observed information matrix (negative Hessian of the log-likelihood) evaluated at MLE. The approximated marginal likelihood is then used to compute Bayes factor.

Raftery (1996), following Jeffreys (1961), proposes the scale for interpreting Bayes factor. If BF_{10} denotes the Bayes Factor for comparing model M_1 against model M_0 (defined as the ratio of marginal likelihoods of M_1 to M_0),

$$2\ln(BF_{10}) < 0 \qquad \text{Negative evidence for } M_1 \text{ (supports } M_0)$$

---

[6] I modified the R code for the function rnegbinRw from R package bayesm by Rossi (2012) so as to extract marginal likelihood from Laplace method. See Appendix for the detail code.

$$0 \leq 2\ln\left(BF_{10}\right) < 2.2 \qquad \text{Very weak evidence for } M_1$$

$$2.2 \leq 2\ln\left(BF_{10}\right) < 6 \qquad \text{Positive (moderate) evidence for } M_1$$

$$6 \leq 2\ln\left(BF_{10}\right) < 10 \qquad \text{Strong evidence for } M_1$$

$$10 < 2\ln\left(BF_{10}\right) \qquad \text{Very strong evidence for } M_1$$

2.4 Results

Estimated coefficients for Poisson and negative binomial models for each of the individual goats are shown in Table 2. 2 and Table 2. 3 respectively. With few exceptions, most of the estimated coefficients for individual goats meet expected signs. For example, goats prefer higher elevation, with rugged terrain features and areas closer to the escape terrain. These model results are used to create map layers of predicted habitat use by mountain goats. Then, for each of the models (Poisson and negative binomial) the predicted maps are combined to gain a single layer of predicted habitat use for the entire study area. Figure 2. 2 and Figure 2. 4 show the predicted habitat use obtained from GPS collared data using Poisson and negative binomial models respectively.

Similar analysis is performed for aerially surveyed goat locations data. Table 2. 4 shows the estimated coefficients from Poisson and negative binomial regressions. The parameter θ, which captures the cross-section heterogeneity, is

statistically significant and highlights the importance of negative binomial

models for these data. The coefficients of the estimated models are used to create

map layer, which are shown in Figure 2. 3 and Figure 2. 5. The maps layers from

two different data sources are visibly similar for both Poisson and negative

binomial models. In order to establish the equivalency between these predicted

map layers, the regression based test of equivalency is performed separately for

each of the count models as described in Section 2.3.2.

The results of comparison between predicted habitat-use from two sources

of data are shown in Table 2. 5. At the mean values, the confidence intervals are

contained within the regions of equivalence for both intercept and slope for each

of the models. This suggests that the predicted habitat use by mountain goats

from two different sources of data are statistically equivalent suggesting that both

types of data are useful in habitat use studies on mountain goat.

2.4.1 Bayesian model comparison

Table 2. 6 reports the posterior mean and standard deviation for four

models. Model M_0 includes only the variables representing landscape

characteristics, while models M_1, M_2, and M_3 include an additional variable as a

proxy for capturing human related recreational activities. All of the statistically

significant coefficient estimates have expected signs. Moreover, the negative and

significant estimates of SkiArea and TrailLength in models M_1 and M_2, and the

positive and significant estimates of TrailDist in Model M_3 signifies the adverse

effect of human activities on the goat habitat use. Particularly, the posterior mean of -0.0253 for SkiArea in model M_1 implies that as Ski-track area increases by one square km within a buffer of 5 km of a given geographical location, one would expect, on an average, the goat count to decline by 2.5 percentage. Similarly, the model coefficients for TrailLength from M_2 can be interpreted as when trail-length increases by 1 km within a buffer of 5 km of a location, the mean goat count is expected to decline by 3.18 percentage. As expected, the coefficient on TrailDist is positive implying that the goat numbers are expected to be higher at locations farther away from hiking trail. In particular, other things remaining same, the goat numbers are expected to be 5.9 percent higher at a location that is 1 km farther away from a nearest hiking trail.

Table 2. 7 shows the logarithm of marginal likelihoods and the corresponding twice logarithm of Bayes Factor for the three models. For each of the three models (M_1, M_2, and M_3), twice log of Bayes Factor is greater than 10 indicating that the data shows very strong evidence for these models as compared to the base model M_0.

2.5   Discussion and Conclusion

The decline in the number of goats due to increasing human recreation may result in the loss of net economic benefit to society. Only in the presence of unlimited geographically suitable locations, such decline in number of goat could represent a displacement from one location to the other, potentially offsetting

such losses. Swanson et al. (1989) reports net willingness to pay for mountain goat hunting of $73 and $99 per trip (1984-85 unadjusted value) for residents and non-residents of South-East Alaska respectively. If we assume that the increase in winter recreation and the resulting decrease in goat numbers lead to a decline in 100 resident and 50 non-resident hunting trips, there will be a loss of $12,250 in the net willingness to pay value from mountain goat hunting. However, this change in net value does not take into account the loss of value from source other than hunting, such as wildlife viewing. On the other hand, gain in the economic value from winter recreation could potentially reduce the loss of net economic value. The interaction of human recreation with goat habitat calls for an intensive benefit-cost analysis which could provide an insight to land management agencies. While almost all of the outdoor recreations hold economic value (market and/or non-market), the resource allocation decision has to be made so as to maximize the net economic benefit when these recreations overlap in a given geographic region. The interaction of human winter recreation and mountain goat habitat use necessitate a detailed cost benefit analysis comparing the net economic benefits generated from various outdoor recreation activities together with the conservation of wildlife habitat.

Table 2. 1: Descriptions of Variables

Variables	Description of Variables
Count	Number of times the goat is present in particular cell (Dependent Variable in all count regression models)
Elevation	Elevation (in 100 m)
ElevationSq	Square of Elevation
Aspect	Cosine of Aspect
VRM	Vector Ruggedness Measure (0-100)
DistEsc	Distance from nearest Escape Terrain (in 100m); Escape Terrain represented by grid cells with Slope 40 degrees or more
Shrub	Shrub
Rock	Rock
SkiArea	Area of ski-tracks formed by winter recreationists within a buffer of 5 km (in square km)
TrailLength	Total trail length within 5 km buffer (in km)
TrailDist	Distance to the nearest trail from each location (in km)

Table 2. 2: Poisson Regression Results (GPS data)

	G2_07	G5_07	G6_07	G840_07	G870_07	G5_08	G6_08	G870_08
Constant	-15.180*** (0.838)	-5.619*** (0.368)	-14.778*** (0.981)	-1.989*** (0.289)	-47.340*** (1.352)	-16.419*** (1.139)	-62.764*** (2.215)	-16.199*** (0.724)
DistEsc	-1.352*** (0.049)	-0.598*** (0.041)	-0.524*** (0.015)	-0.981*** (0.038)	-0.645*** (0.046)	-0.766*** (0.018)	-0.625*** (0.022)	-0.766*** (0.033)
Elevation	3.922*** (0.195)	1.326*** (0.075)	2.529*** (0.176)	0.740*** (0.073)	10.889*** (0.318)	4.626*** (0.277)	11.542*** (0.399)	3.841*** (0.177)
ElevationSq	-0.233*** (0.011)	-0.112*** (0.006)	-0.113*** (0.008)	-0.048*** (0.004)	-0.622*** (0.019)	-0.281*** (0.016)	-0.521*** (0.018)	-0.217*** (0.011)
Aspect	0.701*** (0.052)	-3.638*** (0.255)	-0.955*** (0.037)	-0.223** (0.081)	-3.867*** (0.237)	-0.441*** (0.099)	-0.975*** (0.037)	-2.527*** (0.152)
VRM	0.212*** (0.021)	-0.485*** (0.038)	0.166*** (0.010)	0.286*** (0.017)	1.267*** (0.030)	0.338*** (0.042)	0.280*** (0.008)	0.979*** (0.024)
Rock	0.807*** (0.067)	0.581*** (0.170)	0.653*** (0.111)	-2.675*** (0.319)	--	0.900*** (0.091)	-0.646*** (0.080)	--
Shrub	0.251** (0.076)	0.807*** (0.054)	0.673*** (0.109)	0.408*** (0.050)	-1.556*** (0.080)	1.357*** (0.079)	0.081 (0.072)	-0.875*** (0.077)
Log-likelihood	-2849	-4051	-5918	-4790	-1998	-2137	-5118	-3006
AIC	5713	8118	11853	9596	4011	4291	10252	6025
N	803	1009	5541	1631	345	834	2306	628

Table 2. 3: Negative Binomial Regression Results (GPS Data)

	G2_07	G5_07	G6_07	G840_07	G5_08
Constant	-17.680***	-6.305***	-19.998***	-5.816***	-15.456***
	(2.027)	(1.021)	(2.520)	(1.148)	(2.322)
DistEsc	-1.071***	-0.596***	-0.560***	-0.919***	-0.523***
	(0.104)	(0.122)	(0.036)	(0.091)	(0.064)
Elevation	4.384***	1.658***	3.256***	1.471***	4.141***
	(0.493)	(0.229)	(0.458)	(0.293)	(0.575)
ElevationSq	-0.257***	-0.127***	-0.148***	-0.092***	-0.251***
	(0.029)	(0.017)	(0.021)	(0.018)	(0.034)
Aspect	0.842***	-2.829***	-1.155***	-1.077***	-0.335
	(0.130)	(0.500)	(0.098)	(0.271)	(0.263)
VRM	0.336***	-0.572***	0.310***	0.408***	0.516***
	(0.063)	(0.112)	(0.045)	(0.091)	(0.149)
Rock	0.638**	-0.115	2.512***	-1.856***	1.476***
	(0.238)	(0.610)	(0.330)	(0.467)	(0.337)
Shrub	0.408	1.001***	1.879***	1.210***	1.414***
	(0.222)	(0.218)	(0.325)	(0.243)	(0.243)
θ	0.273***	0.166***	0.074***	0.084***	0.202***
	(0.024)	(0.013)	(0.005)	(0.007)	(0.021)
Log-likelihood	-1052	-1169	-2365	-1189	-880
AIC	2122	2355	4749	2395	1778
N	803	1009	5541	1631	834

Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 2. 4: Poisson and Negative Binomial Regression Results (Aerial Data)

	Poisson	Neg. Bin
Constant	-2.752***	-3.452***
	(0.484)	(0.683)
DistEsc	-0.203***	-0.205***
	(0.026)	(0.035)
Elevation	0.678***	0.814***
	(0.101)	(0.145)
ElevationSq	-0.034***	-0.040***
	(0.005)	(0.008)
Aspect	-0.354***	-0.336***
	(0.063)	(0.096)
VRM	0.040*	0.035
	(0.018)	(0.030)
Rock	0.063	0.084
	(0.103)	(0.164)
Shrub	0.086	0.077
	(0.085)	(0.139)
Theta		1.086***
		(0.138)
Log-likelihood	-925	-801
AIC	1867	1621
N	576	576

Standard errors in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 2. 5: Summary of equivalence-based regression results

Model Coefficients	Region of Equivalence		Confidence Interval	
	RE-	RE+	CI-	CI+
Poisson Model				
Intercept	62.4847	76.3702	75.0585	75.0721
Slope	0.9000	1.1000	0.9781	0.9790
NB Model				
Intercept	65.2849	79.7926	75.7422	75.7584
Slope	0.9000	1.1000	1.0886	1.0897

Table 2. 6: Posterior mean and standard deviation of negative binomial regression coefficients

	M_0	M_1	M_2	M_3
Constant	-0.4869	-0.3033	-0.4705	-0.9036**
	(0.2659)	(0.2784)	(0.2686)	(0.2849)
Elevation	0.1084***	0.1068***	0.1132***	0.103***
	(0.0262)	(0.0269)	(0.0259)	(0.0262)
CosAspect	-0.2662**	-0.2053*	-0.2506*	-0.2448*
	(0.0986)	(0.1012)	(0.0984)	(0.0993)
DistEsc	-0.247***	-0.2486***	-0.2438***	-0.254***
	(0.0336)	(0.0336)	(0.0342)	(0.0346)
VRM	-0.0035	-0.0028	-0.0091	0.008
	(0.0299)	(0.0303)	(0.0306)	(0.0306)
Rock	0.0698	0.0821	0.0921	0.0012
	(0.1733)	(0.1744)	(0.1746)	(0.1734)
Shrub	0.2209	0.3352*	0.35*	0.3154*
	(0.1409)	(0.145)	(0.1475)	(0.1429)
SkiArea		-0.0253***		
		(0.0071)		
TrailLength			-0.0323**	
			(0.0118)	
TrailDist				0.057***
				(0.0124)
α	0.9655***	1.0012***	0.9825***	1.0141***
	(0.1219)	(0.1267)	(0.1251)	(0.1298)

Standard deviations of the posterior mean are in parentheses. The stars, provided for reader's convenience, are defined as $^{*} p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$. The variables SkiArea, TrailLength, and TrailDist are measured in sq.km., km. and km. respectively.

Table 2. 7: Logarithm of marginal likelihood and twice log of Bayes Factors

Models	$\ln(ML)$	$2\ln(BF_{10})$ (against Model M_0)
M_0	-787.96	--
M_1	-775.19	25.55
M_2	-778.73	18.47
M_3	-772.38	31.17

Figure 2. 1: Study area and goat locations observed from flight survey

Study Area

Legend

Goat Locations

Study Area

Kenai Peninsula Area

0   10,000 20,000   40,000   60,000   80,000
                                                          Meters
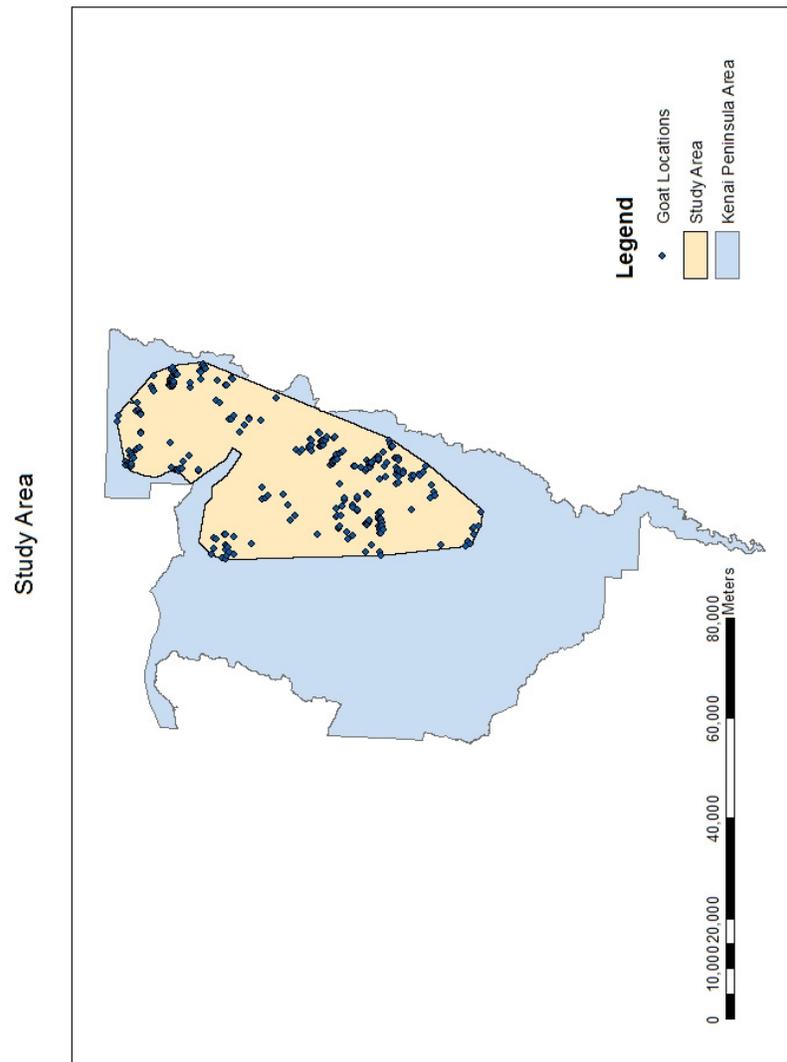
Figure 2. 2: Mountain goat habitat classification (Poisson Model - GPS Collar data)

Habitat Classification (Poisson Model - GPS Collar data)

Poisson (GPS) Classification

Least Preferred

Most Preferred

● Actual Goat Locations

0   11,000   22,000   44,000
Meters

N
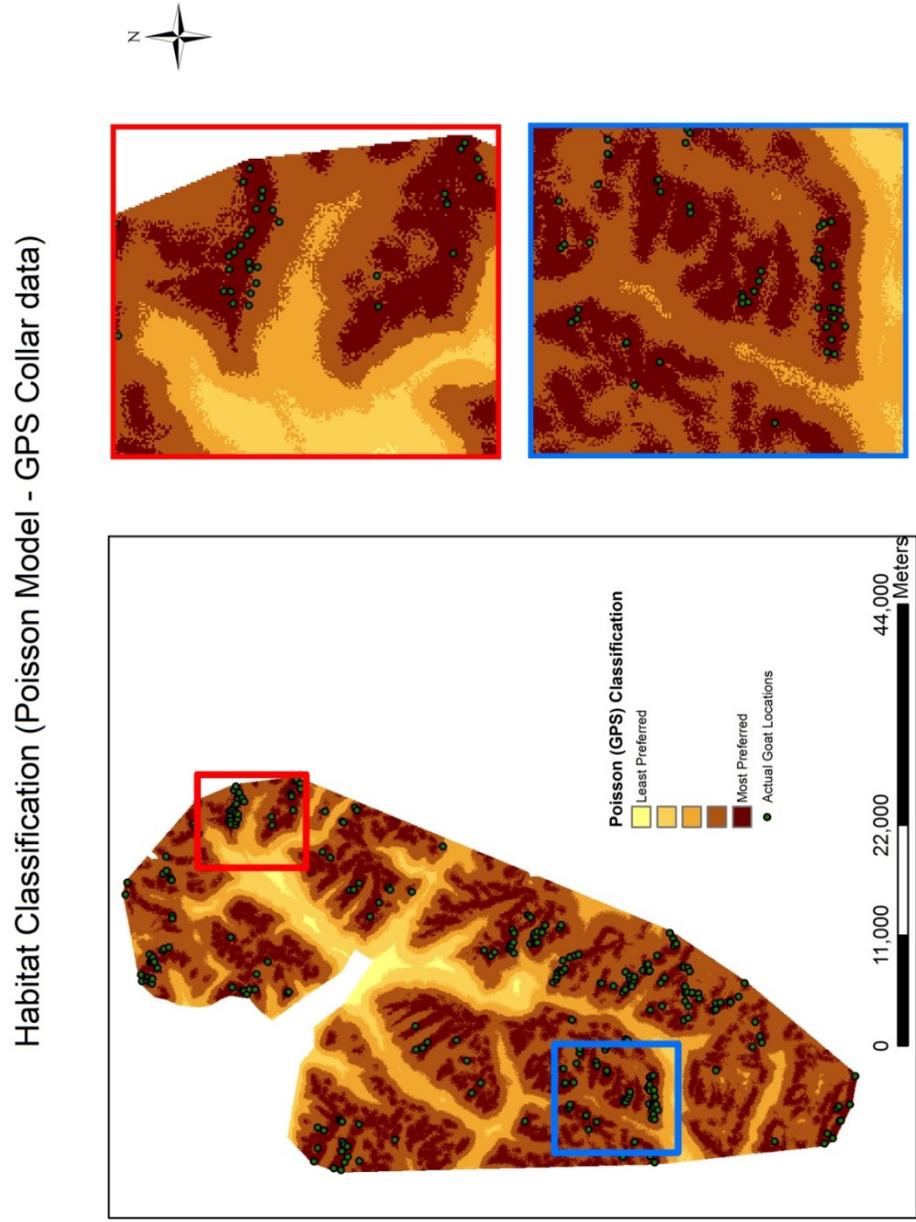
Figure 2. 3: Mountain goat habitat classification (Poisson Model - Aerial data)
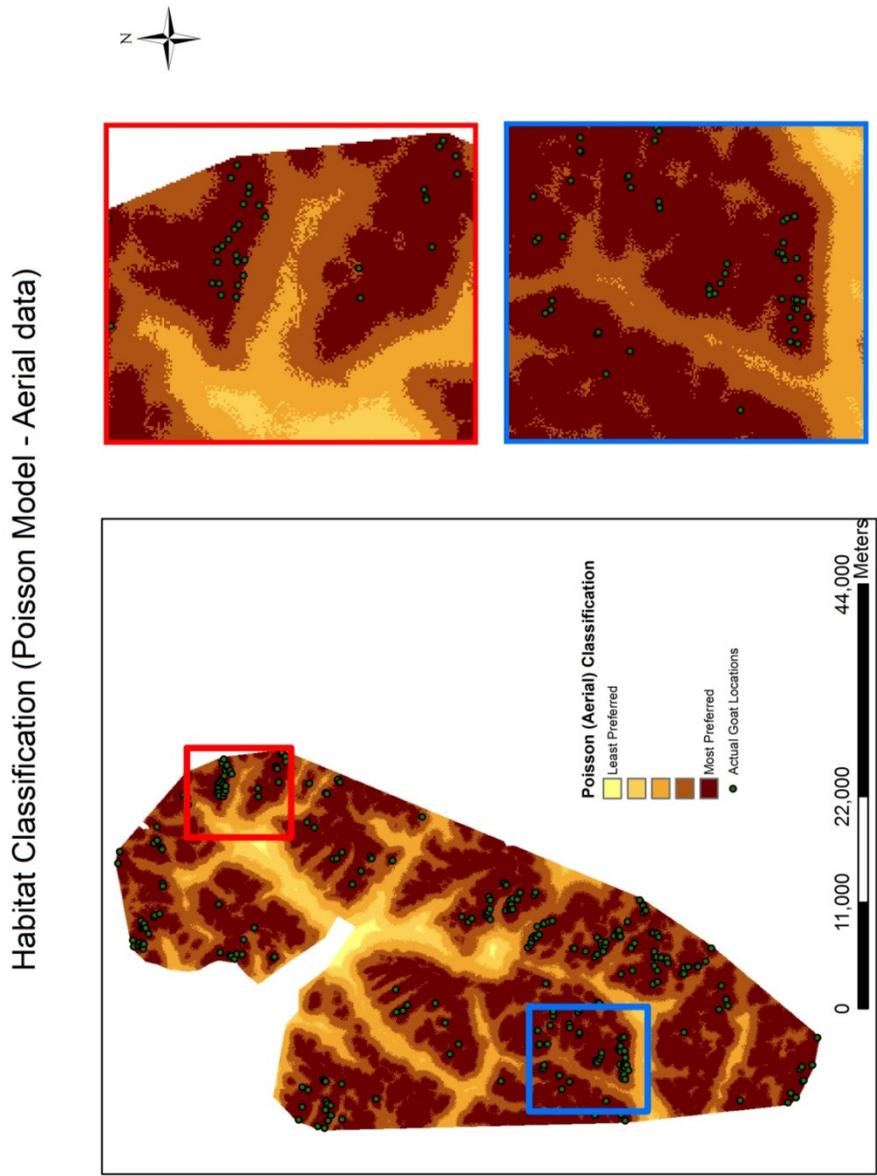
Figure 2. 4: Mountain goat habitat classification (NB Model - GPS Collar data)

Habitat Classification (NB Model - GPS Collar data)

N

NB (GPS) Classification

Least Preferred

Most Preferred

• Actual Goat Locations

0   11,000   22,000   44,000
Meters
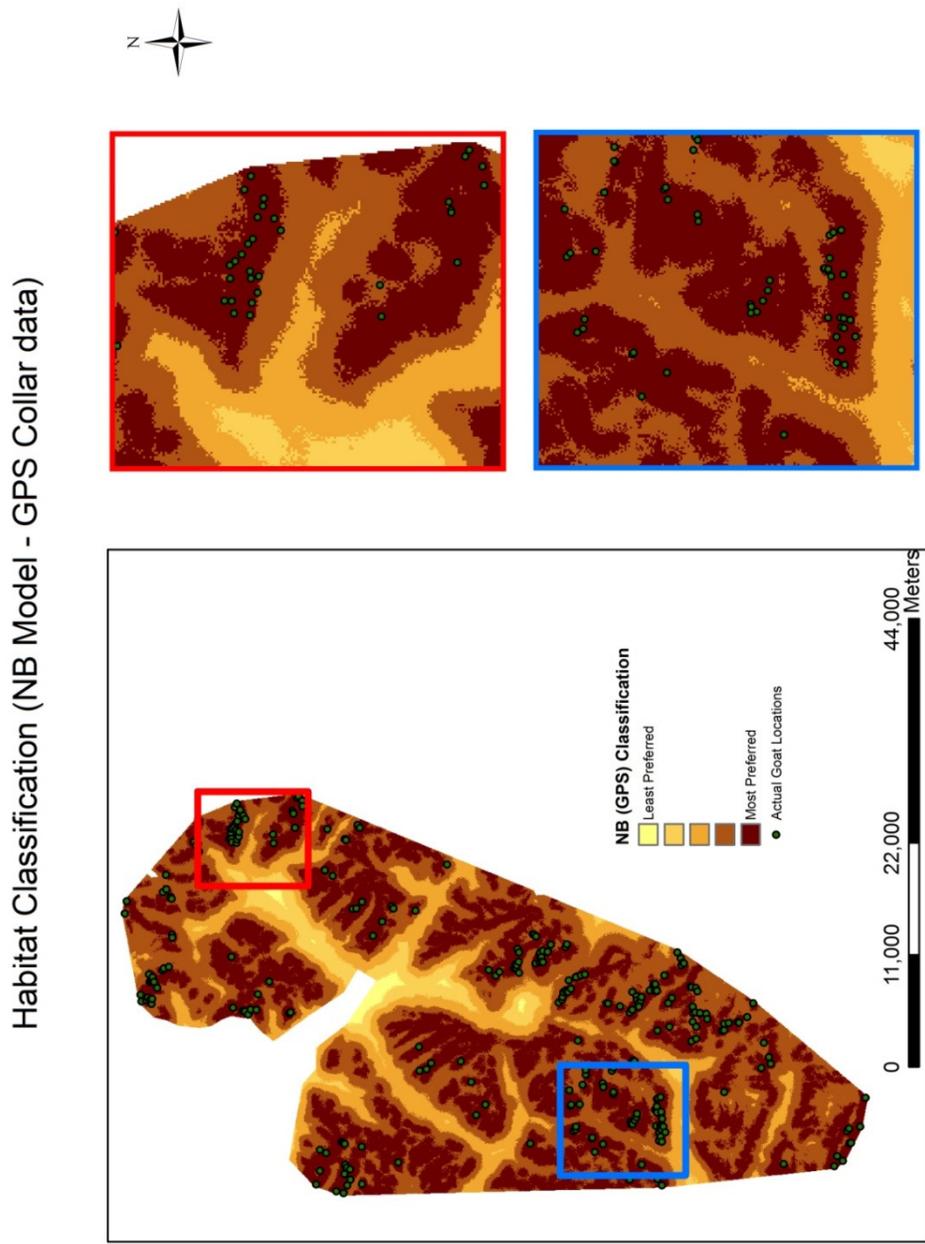
Figure 2. 5: Mountain goat habitat classification (NB Model - Aerial data)

Habitat Classification (NB Model - Aerial data)
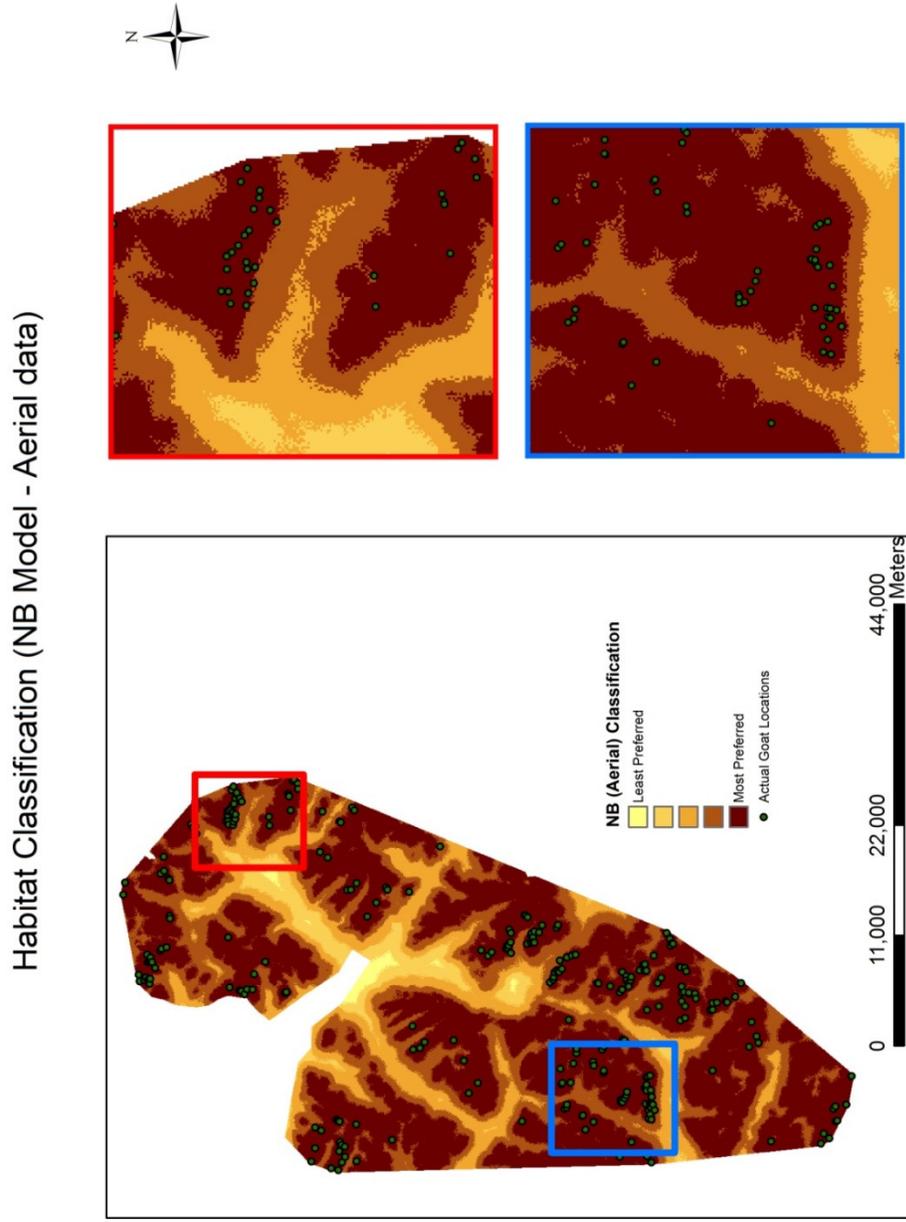
CHAPTER 3

Statistical behavior of distance variables in hedonic property value models

3.1   Introduction

The use of distance variable as a measure of proximity to environmental amenity or dis-amenity is common in hedonic studies (Brasington and Hite, 2005; Noonan et al., 2007; Neumann et al., 2009). However, the effects of distances to amenities (or disamenities) on housing prices are generally not consistent indicators of the true price impact of those amenities (or disamenities), especially when more than two landmark locations posing amenities (or disamenities) to the house values are included in the model specification (Ross et al., 2011). The inconsistency behavior of the distance variables raises concerns about the implication of welfare estimates from studies that primarily employ more than two distance variables.

The main objective of this study is to examine the behavior of statistical properties, particularly in terms of the estimated coefficients and standard errors, of distance variables in hedonic studies under two different scenarios. First case describes an instance where there are unique locations of multiple types of landmarks (such as wilderness area, central business district, airport) and

distance variables associated with each of those unique locations appear in the hedonic regression models. Under such case this study finds that the estimated coefficients behave inconsistently as soon as more than two distance variables are included in the model specification. Second example focuses on the case where there are multiple locations for each type of different landmarks (such as five high schools, four parks in the study area) and the proximity to each of these landmarks is measured by distance to the nearest landmark location. When these proximity variables as measured by distance to the nearest landmark locations appear in the hedonic regression models, the inconsistency behavior no longer remains as long as all of the true distance variables are included. To demonstrate this point, a set of simulations are conducted to assess the statistical properties of the estimated coefficients of proximity or distance variables. Next section provides illustrative example of unique and multiple landmark locations followed by two sets of simulations. The final section concludes.

3.2   Illustrative Example

The distinction between unique landmarks and multiple landmark locations of a given type can be explained with the help of the following illustrative example. Figure 3. 1 shows the spatial locations of housing units in a given geographical space and three types of landmarks: points A, B and C. Panel (a) of Figure 3. 1 has unique locations of the landmarks, where there is exactly one landmark location of a given type (A_1, B_1 and C_1). The irresolvable

identification limitations in the hedonic price analysis exist when the distance to each of these landmark locations appear in the regression model. Panel (b) of the figure shows multiple landmark locations of a given type. For example, A_1, A_2, and A_3 are three locations of the landmark type A, such as wilderness area. When the hedonic price model includes only the distance to a nearest landmark, the coefficients can be estimated consistently, provided that all of such distance variables appear in the regression model that have true effect on the house price. In other words, if all of the landmarks play important role in the determination of house price, the nearest distance to each of the landmarks must appear in the regression model to ensure that the coefficients are consistently estimated.

3.3   Monte Carlo Simulations

The simulations performed in this study follow Ross et al. (2011) in terms of choosing parameter values and simulation procedure. Particularly two sets of simulations are conducted. In the first set of simulations, the unique landmark locations are generated during the data generating process (DGP) and the distance to these landmark locations appear in determining true value of house price. The coefficients are also estimated for distance to these unique landmarks in addition to some hypothetical landmark locations.

The second set of simulations are constructed in the similar way except that the distance to the nearest landmark of given type appear in both true DGP and in the estimation of various regression models.

3.3.1 Monte Carlo Simulation I

In each iteration of Monte Carlo simulations, 400 random locations are generated on the space whose co-ordinates are randomly drawn from uniform distribution between -10 and 10. These random locations are deemed to represent house locations, whose values depend on an intercept term, a random variable x and distances to a pre-specified number of landmarks. The landmark locations are randomly generated over the geographical space whose co-ordinates are uniformly distributed between -20 and 20. Denoting the house price by Y, the true data generating process is based on the following equations:

$$Y = \alpha + \beta x + \gamma_1 D^A + \varepsilon \tag{3.1}$$

$$Y = \alpha + \beta x + \gamma_1 D^A + \gamma_2 D^B + \varepsilon \tag{3.2}$$

$$Y = \alpha + \beta x + \gamma_1 D^A + \gamma_2 D^B + \gamma_3 D^C + \varepsilon \tag{3.3}$$

where, $x \sim N(0,1)$, $\varepsilon \sim N(0,2)$ and D^A, D^B, and D^C are distance to the true landmark locations (A, B, and C). When present in the equation of true DGP, the parameters values are set to $\alpha = 1$, $\beta = 2$, $\gamma_1 = -0.25$, $\gamma_2 = -0.1$, and $\gamma_3 = -0.2$. The distance coefficients are negative to imply that the landmarks are amenities. In each trial, additional sets of hypothetical landmark locations are generated, in addition to the true landmark locations. The distance to these hypothetical landmark locations are denoted by d^a, d^b, d^c and so on. In each of the 10,000

trials, once the dataset is generated, a linear model is estimated for a given set of explanatory variables that include a combination of true and/or hypothetical distance variables and the estimated coefficients are stored. The results of the first sets of simulations are shown in Table 3. 1, Table 3. 2, and Table 3. 3.

Table 3. 1 shows the means and standard deviations of the estimated coefficients for different linear models, where the true DGP includes distance to only one landmark location (D^A) as is shown in equation (3.1). The results suggest that when the regression model includes two or less number of distance variables, the coefficients are consistently estimated. However, when more than two distance variables appear in the regression model, the estimated coefficients are no longer statistically significant, especially in terms of distance to the true landmark location. In addition, the estimated coefficient for the intercept term remains no longer statistically significant when the regression model contains two or more distance variables.

Table 3. 2 reports the means and standard errors of the estimated coefficients for various linear models, where the true DGP includes distances to two landmark locations (D^A and D^B) given by equation (3.2). The estimated coefficients behave in the similar fashion even when the true DGP contains two distance variables. The most important and interesting point to note here is that even if the distance variables D^A and D^B appear in the true DGP, the estimated coefficient for D^B is not statistically significant when the regression model has exactly same variables as in true DGP, although the mean value of the estimated coefficient is very close to the true value used in the true DGP (Model 3 in Table

3. 2). When more than two distance variables appear in the regression model, the standard deviation of true distance variables are high enough to make them statistically insignificant.

The statistical properties of the estimated regression coefficients of distance variables shows further inconsistency as more distance variables are included in the true DGP. This is reported in Table 3. 3, which shows means and standard deviations of the estimated coefficients for various models based on the true DGP given by equation (3.3). When there are three distance variables, none of the distance variables are statistically significant, whether or not those variables appear in the true DGP. All of these simulations imply that when the regression model includes two or more distance variables corresponding to unique locations in the space, the statistical significance of the estimated coefficient does not convey the true importance of the landmark locations.

3.3.2 Monte Carlo Simulation II

In the next set of Monte Carlo simulations, the distance variables entering into the true DGP as well as the regression models are slightly different in nature. Here, the assumption is that there are multiple landmarks of a given type located within the study area and the value of a house depends only on the proximity to the nearest landmark. For example, if the study area of a hedonic model analysis is a city, where there are multiple public parks, it is reasonable to assume that the

house price is affected by the characteristics suggesting how far the nearest public park is located, but is not significantly affected by any other public parks.

In each trial of the simulation, 400 random points are generated to represent house locations on the space whose co-ordinates are randomly drawn from uniform distribution between -10 and 10. Additionally, three landmark locations of each type are randomly generated over the geographical space whose co-ordinates are uniformly distributed between -20 and 20. For example, if A denotes a given landmark, then three points, such as A_1, A_2, and A_3, are generated over the space and the distance to the nearest A is calculated for each house location.

Denoting the house price by Y, the true data generating process is based on the following equations:

$$Y = \alpha + \beta x + \gamma_1 D^{nA} + \varepsilon \qquad (3.4)$$

$$Y = \alpha + \beta x + \gamma_1 D^{nA} + \gamma_2 D^{nB} + \varepsilon \qquad (3.5)$$

$$Y = \alpha + \beta x + \gamma_1 D^{nA} + \gamma_2 D^{nB} + \gamma_3 D^{nC} + \varepsilon \qquad (3.6)$$

where, $x \sim N(0,1)$, $\varepsilon \sim N(0,2)$ and D^{nA}, D^{nB}, and D^{nC} are distances to the nearest true landmark locations A, B, and C. When present in the equation of true DGP, the parameters values are set to $\alpha = 1$, $\beta = 2$, $\gamma_1 = -0.25$, $\gamma_2 = -0.1$, and $\gamma_3 = -0.2$. In each trial, additional sets of hypothetical landmark locations are generated; again three locations for each type of hypothetical landmark. The distance to the nearest of these hypothetical landmark locations are denoted by

d^{na}, d^{nb}, d^{nc} and so on. In each of the 10,000 trials, once the dataset is generated, a linear model is estimated for a given set of explanatory variables that include a combination of true and/or hypothetical distance variables and the estimated coefficients are stored. The results of the simulations are shown in Table 3. 4, Table 3. 5, and Table 3. 6.

Table 3. 4 reports the mean and standard deviations of the estimated coefficients for various models when the true DGP includes only one distance variable given by equation (3.4). As the results clearly indicate, the coefficients are consistently estimated for any true variable appearing in the regression model. One more point to note here is that the inclusion of addition distance variables relating to the nearest hypothetical landmark locations does not influence the level of significance of the true distance and non-distance coefficient estimates.

Table 3. 5 and Table 3. 6 report the results from simulations where the true DGP includes two and three distance variables corresponding to the nearest true landmark locations. The means of coefficient estimates and their standard deviations indicate that the coefficients are consistently estimated as long as true distance variables are included in the regression model. The estimated coefficients for distance variable D^{nB} in Model 3 of Table 3. 6 have very high standard deviation implying that the variable is not statistically significant. However, this may have been caused by the problem of omitted variable bias as true DGP includes three distance variables whereas regression model includes only two. When all of the true distance variables are included in the regression

65

model the problem of inconsistency disappears. Moreover, as expected, the standard deviations of the estimated coefficients of true distance variables increase when more distance variables (relating to hypothetical landmark locations) are added to the regression model. However, the cost of including additional distance variables is not as high as compared to omitting true distance variables. In any case, the coefficients of distance variables corresponding to hypothetical landmark locations turn out to be statistically insignificant as expected.

3.4   Conclusion

Hedonic price models are one of the revealed preference methods used in the non-market valuation studies. These models are useful in estimating the welfare measures as property values can capture the value of all possible changes in environmental quality at a housing site in a single number (Freeman, 2003). Hedonic studies often use distance variables as a measure of proximity to environmental amenities/dis-amenities which proves useful in policy implication. For example, local government may realize increased tax revenue by protecting areas with richer environmental amenities, if such amenities raise values of nearby properties.

Given the importance of distance variables, the reliability of estimated coefficients is very important for gaining optimum policy decisions. This study attempts to answer the questions raised in recent literatures on including

distance variables in hedonic studies. If the objective of any hedonic study is to capture the optimal location rather than estimating the welfare estimates corresponding to the distance variable itself, one suggestion would be to use longitude and latitude in the regression as suggested by Ross et al. (2011). However, if the objective is to estimate welfare measures from distance variables, it is suggested that researchers should be cautious to use them only when multiple types of landmark locations are present in the study area and the distance to nearest landmark location for each type is considered.

Table 3. 1: Means and Standard deviations of Monte Carlo Simulation (Number of landmarks of a given type = 1, True number of Landmark locations in DGP = 1)

	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	-4.812*	1.004**	1.001	1.021	0.972
	(2.053)	(0.332)	(0.746)	(1.550)	(2.645)
x	1.999***	1.999***	2.001***	1.999***	2.000***
	(0.118)	(0.099)	(0.100)	(0.100)	(0.101)
D^A		-0.250***	-0.250*	-0.251	-0.249
		(0.019)	(0.111)	(0.257)	(0.424)
d^a			-0.001	-0.001	-0.001
			(0.112)	(0.411)	(0.377)
d^b				0.001	0.002
				(0.386)	(0.377)
d^c	-0.002				-0.001
	(0.167)				(0.441)
R Squared	0.452	0.593	0.594	0.595	0.596
N	400	400	400	400	400

Notes: True DGP : $Y = \alpha + \beta x + \gamma_1 D^A + \varepsilon$, where $\alpha = 1$, $\beta = 2$, $\gamma_1 = -0.25$ and

$\varepsilon \sim N(0,2)$. Standard deviations in parentheses. $^* p < 0.05$, $^{**} p < 0.01$, $^{***} p <$

0.001.

Table 3. 2: Means and Standard deviations of Monte Carlo Simulation (Number of landmarks of a given type = 1, True number of Landmark locations in DGP = 2)

	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	-4.726	-0.647	1.009	1.013	1.028
	(3.391)	(1.290)	(0.778)	(1.576)	(2.594)
x	1.999***	2.002***	2.001***	2.000***	2.000***
	(0.116)	(0.104)	(0.101)	(0.101)	(0.101)
D^A		-0.252***	-0.251**	-0.250	-0.246
		(0.069)	(0.093)	(0.173)	(0.505)
D^B			-0.099	-0.101	-0.105
			(0.093)	(0.192)	(0.355)
d^a				0.000	0.004
				(0.187)	(0.483)
d^b	-0.013				-0.004
	(0.179)				(0.459)
R Squared	0.484	0.586	0.604	0.605	0.607
N	400	400	400	400	400

Notes: True DGP: $Y = \alpha + \beta x + \gamma_1 D^A + \gamma_2 D^B + \varepsilon$, where $\alpha = 1$, $\beta = 2$, $\gamma_1 = -0.25$, $\gamma_2 = -0.1$ and $\varepsilon \sim N(0,2)$. Standard deviations in parentheses. $^* p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$.

Table 3. 3: Means and Standard deviations of Monte Carlo Simulation (Number of landmarks of a given type = 1, True number of Landmark locations in DGP = 3)

	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	-7.984	-3.871	-1.942	0.985	1.000
	(4.170)	(2.804)	(4.201)	(1.617)	(2.654)
x	2.001***	2.001***	2.001***	2.000***	2.000***
	(0.123)	(0.110)	(0.105)	(0.102)	(0.100)
D^A		-0.261	-0.265	-0.248	-0.245
		(0.150)	(0.628)	(0.253)	(0.452)
D^B			-0.116	-0.097	-0.102
			(0.628)	(0.292)	(0.538)
D^C				-0.204	-0.199
				(0.321)	(0.524)
d^a	-0.019				-0.005
	(0.225)				(0.466)
R Squared	0.477	0.561	0.615	0.637	0.639
N	400	400	400	400	400

Notes: True DGP: $Y = \alpha + \beta x + \gamma_1 D^A + \gamma_2 D^B + \gamma_3^C + \varepsilon$, where $\alpha = 1$, $\beta = 2$, $\gamma_1 = -0.25$

, $\gamma_2 = -0.1$, $\gamma_3 = -0.2$ and $\varepsilon \sim N(0,2)$. Standard deviations in parentheses. $^*p <$

0.05, $^{**}p < 0.01$, $^{***}p < 0.001$.

Table 3. 4: Means and Standard deviations of Monte Carlo Simulation (Number of landmarks of a given type = 3, True number of Landmark locations in DGP = 1)

	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	-1.630	1.002***	0.999*	0.995	0.994
	(1.533)	(0.290)	(0.473)	(0.690)	(0.960)
x	1.999***	1.999***	2.001***	2.000***	2.001***
	(0.109)	(0.100)	(0.100)	(0.099)	(0.100)
D^{nA}		-0.250***	-0.250***	-0.250***	-0.250***
		(0.027)	(0.032)	(0.038)	(0.046)
d^{na}			0.000	0.000	0.000
			(0.032)	(0.038)	(0.046)
d^{nb}				0.000	-0.000
				(0.038)	(0.046)
d^{nc}	-0.002				0.001
	(0.125)				(0.046)
R Squared	0.471	0.558	0.559	0.561	0.562
N	400	400	400	400	400

Notes: True DGP : $Y = \alpha + \beta x + \gamma_1 D^{nA} + \varepsilon$, where $\alpha = 1$, $\beta = 2$, $\gamma_1 = -0.25$ and

$\varepsilon \sim N(0,2)$. Standard deviations in parentheses. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p <$

0.001.

Table 3. 5: Means and Standard deviations of Monte Carlo Simulation (Number of landmarks of a given type = 3, True number of Landmark locations in DGP = 2)

	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	-2.713	-0.063	0.997*	1.000	1.007
	(1.649)	(0.680)	(0.478)	(0.690)	(0.985)
x	1.999***	2.000***	2.000***	1.998***	2.000***
	(0.111)	(0.102)	(0.101)	(0.101)	(0.101)
D^{nA}		-0.251***	-0.250***	-0.250***	-0.250***
		(0.056)	(0.032)	(0.039)	(0.046)
D^{nB}			-0.100**	-0.100**	-0.100*
			(0.032)	(0.038)	(0.046)
d^{na}				-0.000	-0.000
				(0.038)	(0.046)
d^{nb}	-0.000				0.000
	(0.134)				(0.046)
R Squared	0.466	0.552	0.567	0.567	0.569
N	400	400	400	400	400

Notes: True DGP: $Y = \alpha + \beta x + \gamma_1 D^{nA} + \gamma_2 D^{nB} + \varepsilon$, where $\alpha = 1$, $\beta = 2$, $\gamma_1 = -0.25$, $\gamma_2 = -0.1$ and $\varepsilon \sim N(0,2)$. Standard deviations in parentheses. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Table 3. 6: Means and Standard deviations of Monte Carlo Simulation (Number of landmarks of a given type = 3, True number of Landmark locations in DGP = 3)

	Model 1	Model 2	Model 3	Model 4	Model 5
Constant	-4.812*	-2.190	-1.119	0.994	1.012
	(2.053)	(1.406)	(1.773)	(0.698)	(0.962)
x	1.999***	1.999***	2.001***	1.999***	2.000***
	(0.118)	(0.108)	(0.106)	(0.102)	(0.101)
D^{nA}		-0.251*	-0.253*	-0.250***	-0.251***
		(0.115)	(0.110)	(0.038)	(0.045)
D^{nB}			-0.101	-0.100**	-0.100*
			(0.109)	(0.038)	(0.046)
D^{nC}				-0.200***	-0.201***
				(0.038)	(0.046)
d^{na}	-0.002				0.000
	(0.167)				(0.045)
R Squared	0.452	0.527	0.554	0.594	0.595
N	400	400	400	400	400

Notes: True DGP: $Y = \alpha + \beta x + \gamma_1 D^{nA} + \gamma_2 D^{nB} + \gamma_3^{nC} + \varepsilon$, where $\alpha = 1$, $\beta = 2$,

$\gamma_1 = -0.25$, $\gamma_2 = -0.1$, $\gamma_3 = -0.2$ and $\varepsilon \sim N(0,2)$. Standard deviations in

parentheses. $^* p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$.

Figure 3. 1: Illustrative example of house and landmark locations
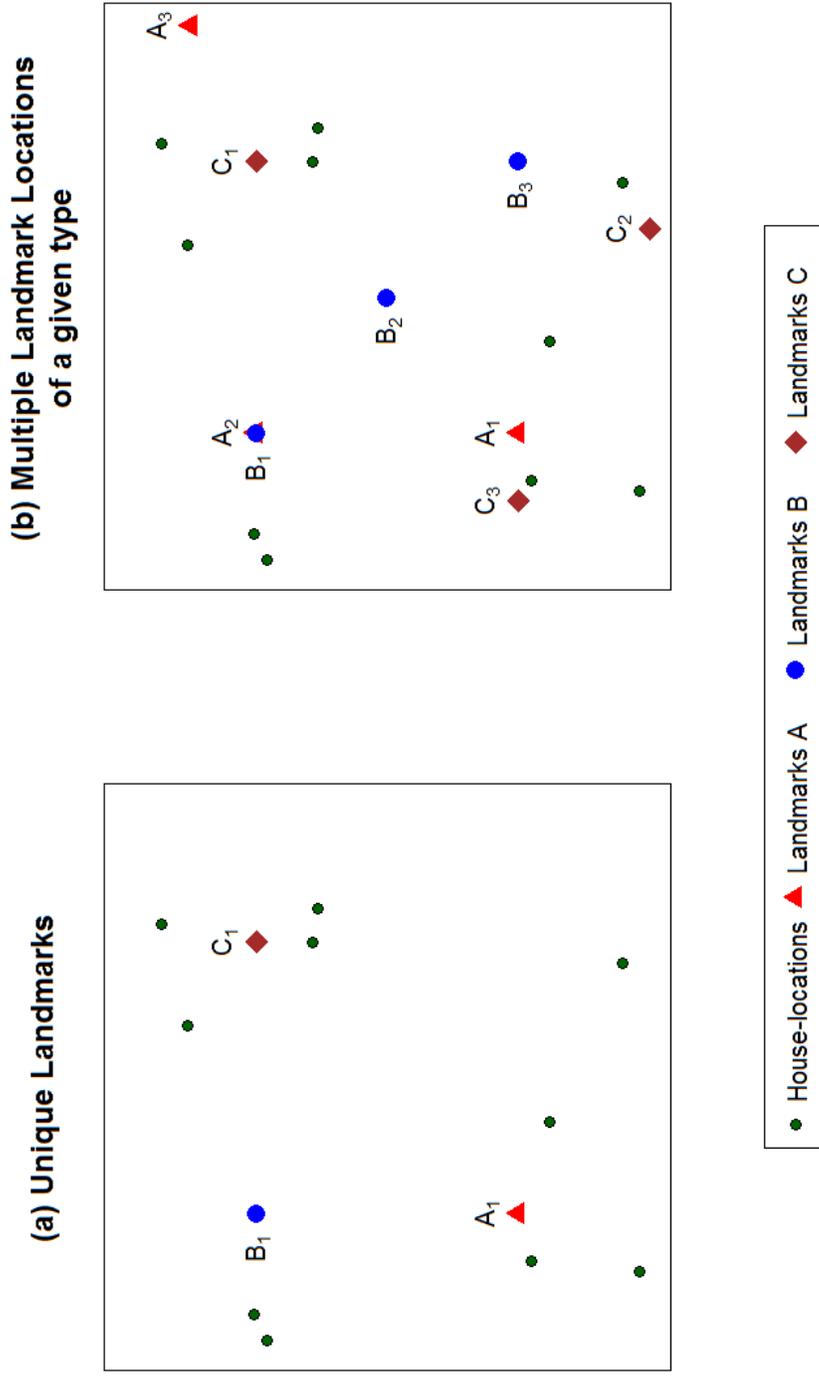
CHAPTER 4

Effects of proximity to wilderness areas on residential property

values: An application of spatial hedonic model

4.1   Introduction

The natural protected areas such as wilderness[7] are valued by society for

various reasons including the provision of recreational opportunities, educational

and philosophical values, protection of biodiversity and the economic benefits to

areas near wilderness (Wright, 2000). While a myriad of benefits associated with

the designated wilderness areas has been identified (Morton, 1998), there often

exists the contested policy debate, which arise from the economic tradeoffs that

---

[7] The Wilderness Act of 1964 defines the wilderness as 'an area where the earth

and its community of life are untrammeled by man'. This further is elaborated to

mean the area which 1) generally appears to have been affected primarily by the

forces of nature, with the imprint of man's work substantially unnoticeable, 2)

has outstanding opportunities for solitude or a primitive and unconfined type of

recreation, 3) has at least five thousand acres of land or is of sufficient size as to

make practicable its preservation and use in an unimpaired condition, and 4)

may also contain ecological, geological, or other features of scientific,

educational, scenic, or historical value.

these protected areas, on the one hand, limit the commercial activities and adversely affect the employment and income generation through extraction industries such as logging and mining (Patric and Harbin, 1998), and on the other hand, preserve natural amenities and provide recreational opportunities that could attract tourists, migrants as well as new businesses (Duffy-Deno, 1998).

The economic values that people place on natural and environmental amenities may also play important role in making decisions regarding choices for residential location. While traditional models of urban economics focus on the tradeoff between housing costs and commuting costs (Wu and Gopinath, 2008), a growing number of recent studies have explored the role of environmental amenities as a possible source of households' choice in making decision regarding residential location (Schmidt and Courant, 2006; Hand et al., 2008a; Izón et al., 2010). As forest resources and protected areas provide many opportunities for recreation as well as hold other non-market benefits, individuals are even willing to accept lower wages to live in close proximity to such areas (Schmidt and Courant, 2006).

The case of protected wilderness areas is particularly important in the western United States because most of the wilderness areas are located in these regions. Population growth in the West and the South regions were almost four

times over the last decade[8] as compared to other regions resulting in increased development pressures on undeveloped lands.

Unlike the market goods, the market for environmental amenities does not yield an observable price, urging researchers to seek alternative methods in order to elicit values of environmental resources. Researchers have used various techniques to measure the benefits of environmental goods such as contingent valuation, choice experiments, travel costs, hedonic pricing methods, etc. Using the hedonic price methods, this study focuses on employing geographic information system (GIS) and recently developed spatial econometric methods in estimating the marginal implicit price of living in proximity to congressionally designated wilderness area. GIS and non-market valuation methods are particularly useful in providing information to public land managers and are considered helpful in making many natural resource management decisions. Recent developments in the spatial econometrics have utilized various techniques to provide spatially-explicit analyses and representations of non-market values of different environmental and natural amenities (Baerenklau et al., 2010). The use of spatial econometrics in the valuation of environmental amenities has been

---

[8] Population growth in South and West regions over the period from 2000 to 2010 were 14.3 and 13.8 percent, whereas in Northeast and Midwest regions population growth were 3.2 and 3.9 percent respectively. (Source: U.S. Census Bureau, Census 2010, Census 2000)

particularly important because of their ability to deal with spatial dependency and spatial heterogeneity present on the data.

This study uses single family residential property value data to examine the effects of proximity to wilderness areas on house prices in Yavapai county of Arizona. Two measures of proximity to wilderness areas are considered: Euclidean distance and driving distance to a nearest wilderness area. Using the generalized methods of moments (GMM), spatial hedonic methods are applied to estimate the parameters for variables of interest and the results are compared with ordinary least squares (OLS) estimates. In addition, spatial bootstrap techniques are used to test for the spatial dependence. Bootstrap methods are based on the empirical distribution of the estimated parameters and is more flexible than the theoretical asymptotic approach in that in a heteroskedastic and non-normal distributional environment, the spatial bootstrap test has a better overall performance compared to the asymptotic counterpart (Lin et al., 2011).

4.2   Hedonic Model

The hedonic price theory provides the basis for deriving welfare measures from observed differences in prices of the houses (Freeman, 2003). The fundamental notion of hedonic price models, more formally introduced by Rosen (Rosen, 1974), lies on the attributes approach to consumer theory (Lancaster, 1966) where utility is the function of attributes of goods. In the context of measuring welfare values pertaining to environmental amenities and dis-

amenities, the central assumption of hedonic price theory is based on the hypothesis that nearby features relating to such environmental characteristics are capitalized into the property values. The ability of the econometric models to decompose the property values to reflect those of individual attributes has made hedonic model a very popular method of non-market valuation. A good deal of empirical research on residential property values have applied hedonic price theory to derive welfare measures of numerous public goods such as open space, water bodies (lakes and rivers), wetlands, urban parks, forest, air quality, etc (Doss and Taff, 1996; Bastian et al., 2002; Chay and Greenstone, 2005; Hand et al., 2008b; Cho et al., 2011; Larson and Perrings, 2013).

4.2.1 Basic Theory

The basic theory behind the hedonic pricing model can be explained by using the utility maximization problem of an individual consumer. If an individual owns the property j, her utility function can be written as (Freeman, 2003):

$$U = U\left(S_j, N_j, Q_j, X\right)$$

where S_j, N_j and Q_j are the vector of structural, neighborhood and environmental characteristics respectively and X is a Hicksian composite good (representing goods other than the property j). If P_j denotes the price of the

property *j*, the individual will maximize her utility $U(.)$ subject to the budget constraint:

$$M - P_j - X = 0$$

Assuming that the property or housing market is in equilibrium with all the individuals maximizing their utility, and that the market clears at the existing prices, the hedonic price function can be written as

$$P_j = P_j\left(S_j, N_j, Q_j\right) \tag{4.1}$$

The utility maximization of the individual can be used to derive marginal implicit price of the particular characteristics of the property. For example, the first order conditions of the utility maximization problem lead to

$$\frac{\partial U / \partial q}{\partial U / \partial X} = \frac{\partial P_j}{\partial q}$$

where, $\partial P_j / \partial q$ is the marginal implicit price of characteristic q_j. Importantly, this marginal implicit price can be interpreted as the willingness to pay (WTP) value for one unit increase in the characteristic q_j.

Although researchers have tried various functional forms for the hedonic price function, the semi-log model (transforming only dependent variable) has been most frequently used in the literature. Alternatively, one could use Box-Cox transformation to decide between linear or log-linear form of the model. The econometric model for the hedonic price function (4.1) can be written more generally as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \qquad\qquad (4.2)$$

where, $\mathbf{Y} = \ln \mathbf{P}$ (semi-log model) or $\mathbf{Y} = \mathbf{P}$ (simple linear model) denotes the vector of dependent variable, $\mathbf{X}$ is the matrix of explanatory variables that includes all of structural, neighborhood and environmental variables, $\boldsymbol{\beta}$ is the vector of parameters and $\mathbf{u}$ is the vector of stochastic errors.

4.2.2 Spatial Regression

With the advance of spatial econometric techniques (Anselin, 1988), recent literatures on hedonic pricing models have been able to address the econometric issues involving spatial correlation present in the data. Particularly, two questions are addressed. First question deals with identifying whether or not there is spatial correlation present in the data. Assuming an affirmative answer to this question as suggested by some statistical tests, second question is to potentially recognize the suitable spatial model for the given data. In a typical spatial regression model, there are two basic ways to incorporate spatial effects suggesting two basic reasons for explaining the presence of spatial correlation: the spatial-lag model and the spatial-error model. The spatial-lag model are designed to incorporate spatial autoregressive effects present in the dependent variable and assumes that the spatially weighted average of property prices in a neighborhood affects the price of each property. The spatial-error model, which takes into account the spatial autoregressive effects through the stochastic errors,

assumes that the errors are correlated over geographical space due to omitted variables.

When the true data generating process (DGM) follows one of the spatially dependent models, inappropriate model specification results in either the biased coefficient estimates or the loss of efficiency of the estimates. The cost of ignoring spatial dependence in the dependent variable is relatively high as they produce biased estimates if this type of dependence is ignored (LeSage and Pace, 2009, p. 156).

4.2.3 Spatial Lag Model

Formally, a spatial lag model is expressed as (Anselin, 1988):

$$\mathbf{Y} = \lambda\mathbf{WY} + \mathbf{X\beta} + \mathbf{u} \qquad (4.3)$$

The motivation for a spatial-lag specification is based on the assumption that the house price is a function of its own characteristics as well as the characteristics of neighboring properties (Anselin and Lozano-Gracia, 2009). This is also evident from the reduced form, which can be written as:

$$\mathbf{Y} = \mathbf{AX\beta} + \mathbf{\varepsilon} \qquad (4.4)$$

where $\mathbf{Y}$ is $n\times1$ column vector of log of prices, $\mathbf{A} = \left[\mathbf{I} - \lambda\mathbf{W}\right]^{-1}$ is $n\times n$ inverse matrix with λ being a spatial autocorrelation parameter and $\mathbf{W}$ an $n\times n$ spatial weight matrix, $\mathbf{X}$ is $n\times k$ matrix of explanatory variables including all of

structural, neighborhood and environmental variables, $\boldsymbol{\beta}$ is $k \times 1$ vector of parameters to be estimated and $\boldsymbol{\varepsilon} = \left[\mathbf{I} - \lambda \mathbf{W} \right]^{-1} \mathbf{u}$ is $n \times 1$ column vector of stochastic errors. In this form of spatial-lag specification the stochastic error terms $(\boldsymbol{\varepsilon})$ are assumed independently and identically distributed with zero mean and constant variance. Alternatively, equation (4.3) can be written in the compact form as

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u} \tag{4.5}$$

with $\mathbf{Z} = \left[\mathbf{X}, \mathbf{WY} \right]$ and $\boldsymbol{\gamma} = \left[\boldsymbol{\beta}, \lambda \right]$, which suggests the presence of a potential correlation between $\mathbf{WY}$ and $\mathbf{u}$ in a typical spatial-lag specification and motivating the use of instrumental variable approach in the estimation procedure.

The spatial model differs from a non-spatial model mainly due to the assumption that the house price at a given location depends not only on the characteristics of the house at the same location but also on the characteristics of houses in neighboring areas. For a spatial lag specification, this implies that the total effect on the house price of a marginal change in an independent variable comprises of own effect and spillover effects. Expanding equation (4.4),

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1k} \\ X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Now, if $\mathbf{X}_k$ denotes $n \times 1$ vector of kth explanatory variable, the Jacobian matrix of $\mathbf{Y}$ with respective of $\mathbf{X}'_k$ is given by

$$\frac{\partial \mathbf{Y}}{\partial \mathbf{X}'_k} = \begin{bmatrix} \partial Y_1 / \partial X_{1k} & \partial Y_1 / \partial X_{2k} & \cdots & \partial Y_1 / \partial X_{nk} \\ \partial Y_2 / \partial X_{1k} & \partial Y_2 / \partial X_{2k} & \cdots & \partial Y_2 / \partial X_{nk} \\ \vdots & \vdots & \ddots & \vdots \\ \partial Y_n / \partial X_{1k} & \partial Y_n / \partial X_{2k} & \cdots & \partial Y_n / \partial X_{nk} \end{bmatrix}$$

Thus, each row of the Jacobian matrix suggests that the house price of location j depends not only on the characteristics of location j, but also other locations (Kim et al., 2003). Thus the total impact of a marginal change in one of the explanatory variable on the housing price can be broken up into direct and indirect impacts. The direct impact refers to the changes in the housing price of location j due to a marginal change in one housing characteristic (say driving distance to a nearest wilderness area) of location j. The indirect impact, on the other hand, is the sum of the induced impacts, i.e., the change in the price of a property at location j resulting from a marginal change the housing characteristic of other locations.

In terms of the estimated parameters, the Jacobian matrix can be written as

$$\frac{\partial \mathbf{Y}}{\partial \mathbf{X}'_k} = \begin{bmatrix} \beta_k a_{11} & \beta_k a_{12} & \cdots & \beta_k a_{1n} \\ \beta_k a_{21} & \beta_k a_{22} & \cdots & \beta_k a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_k a_{n1} & \beta_k a_{n2} & \cdots & \beta_k a_{nn} \end{bmatrix} = \beta_k \mathbf{A}$$

Therefore, the marginal effects from a log-linear hedonic pricing model is

$$\frac{\partial \mathbf{Y}}{\partial \mathbf{X}'_k} = \beta_k \left[\mathbf{I} - \rho \mathbf{W} \right]^{-1} \tag{4.6}$$

4.2.4 Spatial Error Model

The spatial-error model incorporates the spatial error autocorrelation which results when omitted variables follow a spatial structure such that the error variance-covariance matrix is no longer diagonal (Anselin and Lozano-Gracia, 2009). The spatial-error model can be formally written as

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ \mathbf{u} &= \rho \mathbf{W}\mathbf{u} + \boldsymbol{\varepsilon} \end{aligned} \tag{4.7}$$

where $\mathbf{u}$ is the vector of error terms, ρ is the spatial correlation parameter, and $\boldsymbol{\varepsilon} = \{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$ is the vector of independently distributed error terms with mean zero. Depending on the specification, the variances of error terms $\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$ may be assumed non-constant so as to allow for heteroskedasticity. The reduced form of the spatial-error model can be written as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\zeta} \tag{4.8}$$

where, $\boldsymbol{\zeta} = \mathbf{M}\boldsymbol{\varepsilon} = \left[\mathbf{I} - \rho \mathbf{W} \right]^{-1} \boldsymbol{\varepsilon}$. In contrast to the spatial-lag model, the change in the dependent variable with respect to a marginal change in one of the housing characteristics for a spatial-error model is given by the corresponding coefficient estimate, as is evident from equation (4.8).

85

4.2.5 Spatial lag model with auto-correlated disturbance term

Spatial-lag model can be combined with spatial error model to account for the presence of autocorrelation in both dependent variable and disturbance term, which can be formally written as

$$\mathbf{Y} = \lambda \mathbf{WY} + \mathbf{X\beta} + \mathbf{u}$$
$$\mathbf{u} = \rho \mathbf{Wu} + \mathbf{\varepsilon}$$

(4.9)

which can be compactly written as

$$\mathbf{Y} = \mathbf{Z\gamma} + \mathbf{\zeta}$$

(4.10)

4.3 Estimation Method

This study focuses on estimating and comparing various spatial hedonic models with the non-spatial simple regression model. First, non-spatial simple hedonic model is estimated by ordinary least squares (OLS) method and the results are used to test for spatial correlation and the presence of heteroskedasticity. Using the OLS models, bootstrap method is employed to test for the spatial correlation, which suggests the use of spatial lag model with autocorrelation error distribution. In order to account for the spatial dependency, various spatial models are estimated. Particularly, four spatial models are estimated: 1) spatial two stage least squares (S2SLS), 2) spatial two stage least squares with a heteroskedasticity and autocorrelation consistent (HAC) estimator of the variance covariance matrix (S2SLS-HAC), 3) generalized spatial two stage

least squares (GS2SLS), and 4) generalized spatial two stage least squares with heteroskedastic error term (GS2SLS-HET).

Estimations of all of the spatial models are based on the generalized method of moments (GMM). The alternative to GMM estimator is to employ the (quasi) maximum likelihood (ML) procedure. Although the asymptotic properties of the ML estimator have been established (Lee, 2004), the cost of computing the estimator is significantly high when the sample size is large and some of the assumptions are stronger than those required by GMM (Piras, 2010). This is because the ML estimation procedure requires the maximization of likelihood function that involves matrix operations the dimension of which depends on the sample size. In response to this, Kelejian and Prucha (1999) introduced GMM estimator for the autoregressive parameter with the disturbance generated by autoregressive process and prove the asymptotic properties under homoscedasticity assumptions. Kelejian and Prucha (2010) developed a methodology for GMM estimator with heteroskedastic innovation that is both consistent and asymptotically normal.

4.3.1 Spatial two stage least squares (S2SLS)

This method is an extension of the conventional two-stage least square procedure, where the instruments are selected based on the spatial weight matrix. In a typical setting, the matrix of instruments can be defined as

$$\mathbf{H} = \left(\mathbf{X}, \mathbf{WX}, \mathbf{W}^2\mathbf{X}, ..., \mathbf{W}^\tau\mathbf{X} \right)$$

where, $\tau \leq 2$ and the S2SLS parameter estimate is given by

$$\hat{\mathbf{\gamma}}_{S2SLS} = \hat{\mathbf{\sigma}}^2 \left(\hat{\mathbf{Z}}^T\mathbf{Z} \right)^{-1} \hat{\mathbf{Z}}^T\mathbf{Y}$$

where $\hat{\mathbf{Z}} = \mathbf{PZ} = \left(\mathbf{X}, \widehat{\mathbf{WY}} \right)$, $\widehat{\mathbf{WY}} = \mathbf{PWY}$ and $\mathbf{P} = \mathbf{H} \left(\mathbf{H}^T\mathbf{H} \right)^{-1} \mathbf{H}^T$. Statistical inference is generally based on the asymptotic variance covariance matrix

$$Var\left(\hat{\mathbf{\gamma}}_{S2SLS} \right) = \hat{\mathbf{\sigma}}^2 \left(\hat{\mathbf{Z}}^T\mathbf{Z} \right)^{-1}$$

with $\hat{\mathbf{\sigma}}^2 = \dfrac{\mathbf{e}^T\mathbf{e}}{n}$ and $\mathbf{e} = \mathbf{Y} - \mathbf{Z}\hat{\mathbf{\gamma}}_{S2SLS}$ (Kelejian and Prucha, 1999; Piras, 2010).

4.3.2 Spatial two stage least squares with heteroskedastic and autocorrelation consistent estimator (S2SLS-HAC)

The spatial HAC estimator, proposed by Kelejian and Prucha (2007), is based on the estimated disturbances and allows for heteroskedasticity of unknown forms across spatial units. The disturbance vector is assumed to be generated by the process

$$\mathbf{\epsilon} = \mathbf{R}\mathbf{\xi}$$

where $\boldsymbol{\xi}$ is a vector of innovations and $\mathbf{R}$ is an $n \times n$ non-stochastic matrix with unknown elements. The asymptotic distribution of corresponding IV estimators will have the variance-covariance (VC) matrix given by

$$\boldsymbol{\Psi} = n^{-1}\mathbf{H}^{T}\boldsymbol{\Sigma}\mathbf{H}$$

where $\boldsymbol{\Sigma} = \mathbf{R}\mathbf{R}^{T}$ is the variance covariance of $\boldsymbol{\xi}$. Kelejian and Prucha (2007) propose to estimate the (r, s) element of $\boldsymbol{\Psi}$ by

$$\hat{\psi}_{rs} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}h_{ir}h_{js}\hat{\varepsilon}_{i}\hat{\varepsilon}_{j}K\left(\frac{d_{ij}^{*}}{d}\right)$$

where the subscripts refer to the elements of the matrix of instruments $\mathbf{H}$ and residuals $\hat{\boldsymbol{\varepsilon}}$, and $K(\ )$ is a Kernel function used to form the weights for different covariance elements which depends on the distance measures. Finally, the asymptotic variance covariance matrix of the parameter vectors is given by

$$\hat{\boldsymbol{\Phi}} = n\left(\hat{\mathbf{Z}}^{T}\hat{\mathbf{Z}}\right)^{-1}\mathbf{Z}^{T}\mathbf{H}\left(\mathbf{H}^{T}\mathbf{H}\right)^{-1}\hat{\boldsymbol{\Psi}}\left(\mathbf{H}^{T}\mathbf{H}\right)^{-1}\mathbf{H}^{T}\mathbf{Z}\left(\hat{\mathbf{Z}}^{T}\hat{\mathbf{Z}}\right)^{-1}$$

4.3.3 Generalized spatial two stage least square estimators

The S2SLS-HAC estimator discussed in the previous section assumes that the disturbance terms follow the spatial autoregressive structure generated by the some unknown process. An alternative to this is to allow the disturbance terms to take a specific form as

$$\boldsymbol{\varepsilon} = \rho\mathbf{W}\boldsymbol{\varepsilon} + \boldsymbol{\xi} \qquad (4.11)$$

89

If $\boldsymbol{\xi}$ are assumed to be homoscedastic, the GSTSLS method is used (Kelejian and Prucha, 1999). On the other hand, if $\boldsymbol{\xi}$ are assumed to be independent (with zero mean and non-constant variance σ_i^2), GSTSLS-HET method is used to capture the heteroskedastic error structure(Kelejian and Prucha, 2010).

4.3.4 Spatial Weights

The spatial weights used in this study are based on the nearest six neighbors. The number of neighbors is based on the average number of neighbors obtained after creating Thiessen Polygons by using all the house locations in the study area. The models and marginal effects using Thiessen Polygons are also estimated and the results are provided in Appendix C.

4.3.5 The Marginal Effects

One of the objectives of this paper is to estimate the marginal effects of a change in an environmental variable on the house price. Under the assumption that every household is in equilibrium with respect to a given set of house prices so that the prices clear for a given supply, the marginal implicit price may be interpreted as a measure of a household's marginal utility (Koschinsky et al., 2012). In a log-linear specification of a non-spatial model such as equation (4.2), the marginal effect of a given characteristic, say x_i, is given by

$$MWTP_{x_i} = \frac{\partial Y}{\partial x_i} = \frac{\partial lnP}{\partial x_i} = \hat{\beta}_i P$$

This welfare measure assumes that there are no spillover effects. However, when the data possesses spatial dependency, the marginal effects must be calculated to incorporate any spillover effects. Whenever the estimation model includes spatial lag parameter λ, the marginal willingness to pay is given by equation (4.6) which can be re-written for a given variable x_i as

$$MWTP_{x_i} = \frac{\partial Y}{\partial x_i} = \frac{\partial lnP}{\partial x_i} = \hat{\beta}_i \left(\frac{1}{1-\lambda} \right) P$$

4.3.6 Bootstrap test for Spatial Dependency

Bootstrapping is a non-parametric approach to statistical inference that substitutes computation for more traditional distributional assumptions and asymptotic results. One of the important advantages of bootstrap techniques is that it does not require distributional assumptions, the bootstrap can provide more accurate inferences when the data are not well behaved or when the sample size is small.

The bootstrap procedure applied in this paper is based on Lin et al (2011), which can be summarized in the following steps:

1. Estimate the OLS model $y = X\beta + \varepsilon$, and compute relevant test statistics.

2. Transform the regression residuals $\hat{\varepsilon} = y - X\hat{\beta}$ into re-scaled, re-centered, leverage-adjusted series:

$$e_i^* = \sqrt{\frac{N}{N-1}} \left(\frac{\hat{\varepsilon}}{\sqrt{1-h_i^2}} - \frac{1}{N} \sum_{j=1}^{N} \frac{\hat{\varepsilon}}{\sqrt{1-h_j^2}} \right), i = 1,2,...,N$$

where, h is the diagonal vector of hat matrix $X(X'X)^{-1}X'$

3. Resample e^* with replacement to obtain $\hat{e}^*$. To account for the general heteroskedasticity, let $\tilde{e}_i^* = \hat{e}_i^* v_i$ with v_i a random variable satisfying:

$$v_i = \begin{cases} -1 & \text{with probability 1/2} \\ 1 & \text{with probability 1/2} \end{cases}$$

4. Using the fixed spatial weight matrix W and exogenous explanatory variables X, the bootstrapped sample is $(\tilde{y}, X)$, where $\tilde{y} = X\hat{\beta} + \tilde{e}^*$. Using this bootstrap sample, estimate model $\hat{\tilde{y}} = X\hat{\tilde{\beta}} + \hat{\tilde{e}}$ and compute relevant test statistic.

5. Repeat steps 3 and 4 to generate large number of bootstrap observations of the test-statistics.

Once the bootstrapped test statistics are generated, the empirical p-values are computed, which are given by $P(I) = \#\{I_j > I_o\}$, where I_j is the test-statistic in jth bootstrap procedure. Noting that the Moran's index I can take positive or negative value, two-tail test is performed and as such the p-value for Moran's I is defined by $2\min(P(I_o), 1 - P(I_o))$.

4.4   Data for Empirical Estimation

The area of study includes the Yavapai county of Arizona (Figure 4. 1). The dataset used in this study is assembled from various sources with series of structural, neighborhood and environmental variables for each single family housing units. The data on housing characteristics including sale prices originate from the Yavapai County Parcel Dataset, which are obtained from Yavapai County office, Arizona. The dataset includes 3169 single-family housing units sold within a year round between October 2008 and October 2009. Since some of the house prices are reported as low as $1 indicating a possibility of gifts/donations, which would not represent the true value, potential outliers are removed by using a cut off value of $40,000. Thus the sale prices in the final sample of 2738 houses ranges from $40,000 to $2,880,000, with a mean value of $233,600. Table 4. 1 reports the names and description of variables used in this study. The structural variables of the housing units included in this analysis are land parcel size, total floor area of the housing structure, story height and house age of the house. The parcel database did not contain other structural variables such as number of rooms and number of bathrooms, which prevented me from including such variables in the analysis.

In addition to the parcel data, some other GIS data used in this study are also obtained from Yavapai County office, such as hiking trail, water body and land use. The GIS data on wilderness areas are downloaded from The University

93

of Montana's Wilderness.net (http://www.wilderness.net). Similarly, the GIS data on US forest areas are obtained from land use data published by Arizona Land Resource Information System in cooperation with Bureau of Land Management, Arizona State Office. Currently, there are 757 wilderness areas in the National Wilderness Preservation System that are located in 44 states including Puerto Rico with the total area of 109,511,966 acres. There are 19 designated wilderness areas in the Yavapai county, five of which are managed by Bureau of Land Management (BLM) and remaining 14 are managed by US forest service. The total designated area of the wilderness in Yavapai County is 551.5 sq. miles, which is approximately 7 percentage of the total area in the County.

In order to capture the neighborhood characteristics, four variables are used from two additional sources of data. First three variables Urban, PopDensity and White are generated from Census Data 2010 to capture the Census Block Group characteristics. Additionally, school districts may potentially have influence on individual's choice of location and thereby affecting home prices. A comprehensive school district test score index is created from 2012 AIMS (Arizona's Instrument to Measure Standard) and is included in the regression models to capture the potential neighborhood effect.

All of the environmental variables included in the models are the distance variables measuring the proximity to natural amenities. Each of the distance variables representing the proximity to a nearest amenity is calculated by utilizing the tools available in ESRI's ArcGIS software. The "near" tool in ArcGIS is used to calculate the Euclidean distance to the nearest amenities. Similarly, the

driving distances are calculated by utilizing the OD (Origin Destination) cost
matrix function available in the Network Analysis Extension in ArcGIS. In
addition to the variables presented in Table 4. 1, the house prices may also be
affected by characteristics such as crime rates, hazardous waste disposal sites,
etc. However, these information are either available in the aggregate form for the
entire county or are not available in the region.

4.5   Empirical Results

4.5.1 Model Results

As mentioned above, the major objective of this paper is to examine the
effect of proximity to a wilderness area on residential house prices. Attempts are
made to achieve this objective by estimating various regression models with and
without spatial component. Although the proximities to nearest water bodies,
hiking trails and forest areas are also included in the regression models and the
results are briefly discussed below, this study focuses more on the two measures
of proximity to wilderness area: Euclidean distance and driving distance to the
nearest wilderness area from each house location.

Before running the spatial regression, bootstrap test of spatial dependency
is performed. Table 4. 2 report the test statistics for Moran I and Lagrange
Multiplier for lag and error models with bootstrap confidence intervals. Both lag
and error model specifications are statistically significant except for the robust
lag model. Although robust standard deviation turns out to be statistically

insignificant, the generalized models show statistical significant lag parameter across all spatial models and therefore lag model specifications are retained.

Table 4. 3 reports the estimated coefficients and standard errors from OLS and four variations of spatial models. The estimated coefficients for structural and neighborhood variables are robust across models. Generally the parameter values are slightly larger for OLS method as compared to S2SLS models. Although the standard errors are greater for spatial models, they are only marginally greater for most of the parameters and do not affect the statistical significance of the estimated coefficients except for some environmental (or proximity/distance) variables. For example, the distance to the nearest water body is statistically significant at 1 percent for OLS and S2SLS, but is significant at 5 and 10 percent for GSTSLS and GSTSLS-HET methods. All of the statistically significant distance variables have negative sign showing that the property values are usually higher if they are located closer to these amenities.

The model results in Table 4. 4 differ from Table 4. 3 in terms of the three distance variables; the distances to nearest major road, nearest forest area and nearest wilderness area, which are the driving distances instead of Euclidean distances used in Table 4. 3. The difference between Euclidian distance and driving distance are not only in terms of the length, but they may vary dis-proportionately for different property locations especially when the distances are of reasonable length. In other words, if the Euclidian distance to the nearest wilderness area from property A is longer than from property B, it does not necessarily imply that the driving distance from property A is longer than from

property B. Moreover, the driving distance to the nearest wilderness area is measured from house location to the access point of the wilderness area whereas the Euclidean distance is measured from the property location to the nearest point of the wilderness area. This distinction is shown in Figure 4. 2.

The estimated coefficients in Table 4. 4 are similar in sign and significance for structural and neighborhood characteristics. Again, the distances to the nearest water body and trail are statistically significant and have expected sign. The driving distance to the nearest forest area is no longer statistically significant for all of the spatial models. One possible explanation for such a difference in Euclidian versus driving distance with respect to forest is that the forest area provides fewer recreational opportunities therefore generating insignificant effect corresponding to driving distance, but the physical house location within the vicinity of forest area could potentially raise household utility level through the sense of having an open space surrounding the property.

4.5.2 Marginal effects

The marginal effects are calculated for all of the estimated coefficients, and are broken up into direct and indirect effects for all the spatial models. The direct effects indicate the actual change in the property value due to an one unit change in explanatory variables whereas indirect effects give the spillover effects on the property values through the spatial dependence. Table 4. 5 and Table 4. 6 report the direct and indirect effects based on the model estimates of Table 4. 3. Since

97

OLS does not take into account the spatial dependency, the marginal effects reported are the total effects and are same in Table 4. 5 and Table 4. 6. For distance to the nearest wilderness area, these results show that the direct increase in property values ranges from $3539 to $3978 across models if the property is located one mile close to the wilderness area. Similarly, the increase in property values through spillover effects ranges from $451 to $752 across four spatial models.

The marginal effects for the proximity to wilderness area as is measured by the driving distances are reported in Table 4. 7 and Table 4. 8. These effects are slightly higher than that of the Euclidian distances. The direct marginal effects of the proximity of wilderness area range from $3824 to $4201 and the indirect marginal effects range from $486 to $823 across models. The 95% confidence intervals for all of the marginal effects are computed by using Krinsky Robb method (Krinsky and Robb, 1986). The confidence intervals for marginal effects corresponding to driving distance to nearest wilderness area indicate the statistical significance of the estimated values.

4.6   Conclusions

This paper uses spatial hedonic models to analyze the effect of proximity to designated wilderness area on the residential property values in the Yavapai County, Arizona. The data are assembled from various sources with a series of structural, neighborhood and environmental variables. Although the most

important variable of interest is the designated wilderness area, proximity to other environmental variables are also included (water body, hiking trail, parks, and forests). Proximity to the major roads, forest and wilderness area are measured in terms of Euclidian distance as well as driving distance.

Bootstrap method of statistical test for spatial dependency is performed, which suggests the presence of spatial autocorrelation in both dependent variable and disturbance term. In order to account for the spatial dependency, four variations of spatial models are employed (spatial two stage least squares with and without heteroskedasticity and generalized spatial two stage least squares with and without heteroskedasticity) and the results are compared with the basic linear regression. All of the spatial models are estimated using generalized method of moments (GMM) in R statistical software. The results from spatial models are used to calculate the marginal willingness to pay for the variables of interest which are broken into the direct effects and indirect effects. Krinsky and Robb method is used to calculate the confidence interval for both direct and indirect effects.

The findings of this paper confirm the importance of wilderness area to home-owners, which can be valuable to policy makers. However, the estimated coefficients on distances to wilderness area provide only a partial measure of the total economic value of such lands. The actual magnitudes of the value might be much larger and demands a detailed cost benefits analysis of wilderness management.

Table 4. 1: Variables names and descriptions

Variables	Descriptions
Dependent variable	
SalesPrice	Natural logarithm of sales price ($) of the single family homes
Structural Characteristics	
ParcelSize	Total land parcel size in acres
Height	Height of the roof line in feet
FlArea	Natural logarithm of total floor area in square feet of the housing units
HAge	Age of the house on 2009 in years (2009 – construction year)
Elevation	Elevation of house location (in km)
Neighborhood Variables	
Urban	Indicator variable for house located in urban areas
PopDensity	Block group population density in Census 2010
White	Proportion of White population in Census 2010 (in 100)
TestScoreIndex	Comprehensive school district test score index constructed by using 2012 AIMS (Arizona's Instrument to Measure Standard) result (0-100 scale)
Environmental Variables and Variables of interest	
WaterDist	Euclidian distance from house location to a nearest water body (in miles)
TrailDist	Euclidian distance from house location a nearest hiking/biking trail (in miles)
ParkDist	Euclidian distance from house location to a nearest parks (in miles)
WilderDist	Euclidian distance from house location to a nearest designated wilderness area (in miles)

Table 4. 1 (contd.)

Variables	Descriptions
RoadDist	Euclidian distance from house location to a nearest Major Road (in miles)
ForestDist	Euclidian distance from house location to a nearest US Forest Area (in miles)
RoadDDist	Driving distance from house location to a nearest Major Road (in miles)
ForestDDist	Driving distance from house location to a nearest forest area (in miles)
WilderDDist	Driving distance from the house location to the nearest wilderness area (in miles)

Table 4. 2: Computed spatial test statistics and bootstrapped confidence intervals

Test Statistic	Estimate (p-value)	Bootstrapped 95% CI [0.025% 0.975%] (p-value)
Moran-I	0.2694 (0.000)	[-0.02403, 0.01578] (0.000)
LM-error	681.0467 (0.000)	[0.00130, 5.65685] (0.000)
LM-lag	375.6682 (0.000)	[0.00095, 5.44244] (0.000)
LM-error	305.6711 (0.000)	[0.00105, 5.35781] (0.000)
LM-lag	0.2927 (0.588)	[0.00089, 5.17939] (0.590)

Table 4. 3: Estimation results from OLS, S2SLS, S2SLS-HAC, GSTSLS, GSTSLS-HET

Variables	OLS	S2SLS	S2SLS-HAC	GSTSLS	GSTSLS-HET
Constant	4.775***	3.972***	3.972***	4.069***	4.089***
	(0.201)	(0.409)	(0.566)	(0.456)	(0.510)
Structural Characteristics					
ParcelSize	0.102***	0.101***	0.101***	0.133***	0.137***
	(0.014)	(0.019)	(0.024)	(0.016)	(0.021)
Height	0.126***	0.114***	0.114***	0.097***	0.096***
	(0.010)	(0.015)	(0.018)	(0.010)	(0.013)
FlArea	0.527***	0.505***	0.505***	0.506***	0.505***
	(0.020)	(0.038)	(0.051)	(0.021)	(0.034)
HAge	-0.004***	-0.004***	-0.004***	-0.006***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Elevation	-0.011	-0.017	-0.017	0.018	0.023
	(0.042)	(0.046)	(0.059)	(0.062)	(0.076)
Neighborhood Variables					
Urban	-0.036*	-0.023	-0.023	-0.027	-0.028
	(0.019)	(0.018)	(0.020)	(0.024)	(0.024)
PopDensity	0.007	0.005	0.005	0.008	0.008
	(0.006)	(0.005)	(0.006)	(0.007)	(0.008)
White	0.024***	0.021***	0.021***	0.015***	0.015***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
TestScoreIndex	0.008***	0.007***	0.007***	0.007***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table 4. 3 (contd.)

Variables	OLS	S2SLS	S2SLS-HAC	GSTSLS	GSTSLS-HET
Environmental Variables and Variables of Interest					
WaterDist	-0.021***	-0.019***	-0.019***	-0.015**	-0.015*
	(0.005)	(0.005)	(0.006)	(0.008)	(0.008)
TrailDist	-0.023***	-0.020***	-0.020***	-0.020***	-0.020***
	(0.004)	(0.005)	(0.006)	(0.006)	(0.007)
ParkDist	0.004	0.003	0.003	0.003	0.003
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
RoadDist	0.010	0.008	0.008	0.005	0.005
	(0.008)	(0.009)	(0.011)	(0.011)	(0.013)
ForestDist	-0.015***	-0.013***	-0.013***	-0.011**	-0.011**
	(0.003)	(0.003)	(0.004)	(0.005)	(0.005)
WilderDist	-0.018***	-0.015***	-0.015***	-0.017***	-0.017***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)
λ		0.115**	0.115**	0.158***	0.162***
		(0.047)	(0.057)	(0.044)	(0.049)
ρ				0.424	0.453***
					(0.040)

Standard errors in parentheses. $^{*} p < 0.10$, $^{**} p < 0.05$, $^{***} p < 0.01$.

Table 4. 4: Estimation results from OLS, S2SLS, S2SLS-HAC, GSTSLS, GSTSLS-HET

Variables	OLS	S2SLS	S2SLS-HAC	GSTSLS	GSTSLS-HET
Constant	4.921***	4.127***	4.127***	4.159***	4.168***
	(0.201)	(0.403)	(0.557)	(0.441)	(0.503)
Structural Variables					
ParcelSize	0.130***	0.125***	0.125***	0.147***	0.150***
	(0.014)	(0.018)	(0.023)	(0.016)	(0.021)
Height	0.125***	0.113***	0.113***	0.097***	0.096***
	(0.010)	(0.014)	(0.018)	(0.010)	(0.013)
FlArea	0.521***	0.500***	0.500***	0.502***	0.502***
	(0.020)	(0.037)	(0.052)	(0.021)	(0.034)
HAge	-0.004***	-0.004***	-0.004***	-0.006***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Elevation	-0.004	-0.005	-0.005	0.028	0.032
	(0.039)	(0.042)	(0.050)	(0.057)	(0.070)
Neighborhood Variables					
Urban	-0.037*	-0.024	-0.024	-0.029	-0.030
	(0.019)	(0.018)	(0.020)	(0.024)	(0.024)
PopDensity	0.008	0.006	0.006	0.008	0.008
	(0.006)	(0.005)	(0.006)	(0.007)	(0.007)
White	0.024***	0.021***	0.021***	0.015***	0.015***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
TestScoreIndex	0.007***	0.007***	0.007***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table 4. 4 (contd.)

Variables	OLS	S2SLS	S2SLS-HAC	GSTSLS	GSTSLS-HET
Environmental Variables and Variables of Interest					
WaterDist	-0.028***	-0.025***	-0.025***	-0.021***	-0.021**
	(0.005)	(0.005)	(0.006)	(0.007)	(0.008)
TrailDist	-0.019***	-0.016***	-0.016**	-0.015**	-0.015**
	(0.004)	(0.005)	(0.006)	(0.006)	(0.008)
ParkDist	0.005	0.004	0.004	0.003	0.003
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
RoadDDist	0.009**	0.008	0.008	0.005	0.005
	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)
ForestDDist	-0.005**	-0.004*	-0.004	-0.003	-0.003
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)
WilderDDist	-0.018***	-0.016***	-0.016***	-0.018***	-0.018***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)
λ		0.114**	0.114**	0.162***	0.167***
		(0.045)	(0.055)	(0.042)	(0.048)
ρ				0.420	0.439***
					(0.041)

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 4.5: Marginal Effects for different estimators (Direct Effects)

Variables	OLS	S2SLS	S2SLSHAC	GSTSLS	GSTSLSHET
ParcelSize	23938	23642	23642	31264	32128
	[17737, 30185]	[15216, 31943]	[12548, 34873]	[23907, 38320]	[22435, 41716]
Height	29455	26784	26784	22789	22533
	[24907, 33885]	[20037, 33438]	[18467, 35060]	[18005, 27359]	[16575, 28402]
FlArea	123016	118245	118245	118669	118491
	[114095, 132137]	[100469, 135489]	[94038, 142381]	[108714, 128130]	[102587, 133582]
HAge	-944	-925	-925	-1290	-1325
	[-1198, -697]	[-1230, -619]	[-1319, -517]	[-1553, -1018]	[-1648, -980]
Urban	-2551	-4023	-4023	4214	5298
	[-21970, 16805]	[-25296, 17108]	[-30459, 22810]	[-23874, 31683]	[-29335, 40159]
Elevation	-8436	-5386	-5386	-6291	-6541
	[-17033, 122]	[-13532, 2738]	[-14197, 3845]	[-17110, 4667]	[-17485, 4313]
PopDensity	1626	1246	1246	1885	1980
	[-978, 4232]	[-1215, 3727]	[-1566, 4114]	[-1549, 5233]	[-1440, 5436]
White	5705	4929	4929	3614	3445
	[4922, 6458]	[3847, 5990]	[3580, 6263]	[2458, 4774]	[2114, 4771]
TestScoreIndex	1940	1720	1720	1665	1664
	[1689, 2193]	[1359, 2066]	[1253, 2172]	[1246, 2078]	[1110, 2214]

Table 4. 5 (contd.)

Variables	OLS	S2SLS	S2SLSHAC	GSTSLS	GSTSLSHET
WaterDist	-4890	-4531	-4531	-3570	-3456
	[-7227, -2577]	[-6847, -2114]	[-7434, -1649]	[-7072, -77]	[-7322, 409]
TrailDist	-5311	-4762	-4762	-4770	-4763
	[-7101, -3493]	[-6891, -2603]	[-7651, -1832]	[-7508, -1967]	[-8104, -1342]
ParkDist	961	665	665	669	688
	[-920, 2865]	[-924, 2295]	[-1282, 2636]	[-1703, 3067]	[-1619, 3047]
RoadDist	2264	1777	1777	1111	1099
	[-1207, 5647]	[-2209, 5719]	[-3223, 6703]	[-3678, 5865]	[-4958, 7270]
ForestDist	-3530	-2941	-2941	-2640	-2603
	[-5015, -2057]	[-4402, -1476]	[-4746, -1163]	[-4930, -324]	[-4979, -188]
WilderDist	-4138	-3531	-3531	-3937	-3978
	[-5453, -2814]	[-5028, -2048]	[-5358, -1659]	[-5938, -1918]	[-6458, -1452]

Table 4. 6: Marginal Effects for different estimators (Indirect Effects)

Variables	OLS	S2SLS	S2SLSHAC	GSTSLS	GSTSLSHET
ParcelSize	23938	3021	3021	5721	6075
	[17593, 30364]	[1968, 4136]	[1618, 4463]	[4488, 7193]	[4342, 8073]
Height	29455	3423	3423	4170	4261
	[25014, 33848]	[2603, 4349]	[2403, 4550]	[3354, 5126]	[3200, 5492]
FlArea	123016	15110	15110	21715	22405
	[113884, 131935]	[13089, 17555]	[12235, 18410]	[20338, 23926]	[19878, 25889]
HAge	-944	-118	-118	-236	-250
	[-1198, -696]	[-159, -80]	[-172, -67]	[-290, -192]	[-318, -190]
Urban	-2551	-514	-514	771	1002
	[-21977, 16845]	[-3205, 2242]	[-4010, 2997]	[-4565, 6053]	[-5406, 7800]
Elevation	-8436	-688	-688	-1151	-1237
	[-17226, 196]	[-1785, 324]	[-1888, 457]	[-3247, 856]	[-3352, 840]
PopDensity	1626	159	159	345	374
	[-1010, 4278]	[-157, 482]	[-206, 540]	[-261, 982]	[-285, 1047]
White	5705	630	630	661	651
	[4929, 6489]	[501, 780]	[466, 814]	[452, 893]	[409, 919]
TestScoreIndex	1940	220	220	305	315
	[1685, 2193]	[176, 268]	[163, 282]	[232, 388]	[214, 425]

Table 4. 6 (contd.)

Variables	OLS	S2SLS	S2SLSHAC	GSTSLS	GSTSLSHET
WaterDist	-4890	-579	-579	-653	-654
	[-7194, -2586]	[-891, -278]	[-970, -211]	[-1340, -4]	[-1405, 88]
TrailDist	-5311	-609	-609	-873	-901
	[-7097, -3447]	[-894, -342]	[-998, -232]	[-1410, -371]	[-1575, -266]
ParkDist	961	85	85	122	130
	[-862, 2851]	[-125, 296]	[-174, 342]	[-325, 580]	[-326, 597]
RoadDist	2264	227	227	203	208
	[-1224, 5753]	[-284, 748]	[-410, 866]	[-718, 1105]	[-953, 1361]
ForestDist	-3530	-376	-376	-483	-492
	[-5028, -2061]	[-575, -187]	[-613, -143]	[-914, -59]	[-956, -44]
WilderDist	-4138	-451	-451	-720	-752
	[-5441, -2836]	[-657, -265]	[-709, -217]	[-1124, -347]	[-1250, -287]

Table 4.7: Marginal Effects for different estimators (Direct Effects)

Variables	OLS	S2SLS	S2SLSHAC	GSTSLS	GSTSLSHET
ParcelSize	30297	29167	29167	34531	35148
	[24071, 36434]	[20978, 37349]	[18544, 39289]	[27349, 41567]	[25756, 44647]
Height	29138	26468	26468	22639	22404
	[24689, 33559]	[19947, 33097]	[17952, 34617]	[17805, 27333]	[16459, 28299]
FlArea	121719	116950	116950	117746	117646
	[112533, 130966]	[99501, 133854]	[92894, 140614]	[107771, 126772]	[101657, 133171]
HAge	-1025	-1006	-1006	-1332	-1363
	[-1267, -784]	[-1316, -695]	[-1412, -597]	[-1593, -1064]	[-1685, -1029]
Urban	-1032	-1258	-1258	6448	7458
	[-18773, 16582]	[-20595, 17883]	[-23823, 21489]	[-19738, 33132]	[-25574, 39770]
Elevation	-8558	-5699	-5699	-6841	-7091
	[-16990, 153]	[-13953, 2404]	[-14922, 3296]	[-17359, 4203]	[-18048, 3705]
PopDensity	1837	1464	1464	1925	1986
	[-745, 4420]	[-1010, 3911]	[-1401, 4332]	[-1406, 5236]	[-1518, 5436]
White	5716	4915	4915	3633	3471
	[4924, 6484]	[3857, 5958]	[3564, 6241]	[2464, 4777]	[2170, 4776]
TestScoreIndex	1748	1525	1525	1437	1431
	[1497, 1999]	[1181, 1863]	[1092, 1954]	[1011, 1853]	[895, 1940]

Table 4. 7 (contd.)

Variables	OLS	S2SLS	S2SLSHAC	GSTSLS	GSTSLSHET
WaterDist	-6446	-5855	-5855	-4979	-4872
	[-8693, -4191]	[-8275, -3493]	[-8813, -2932]	[-8350, -1628]	[-8706, -1124]
TrailDist	-4393	-3839	-3839	-3513	-3472
	[-6196, -2601]	[-5975, -1657]	[-6701, -967]	[-6339, -785]	[-6870, 19]
ParkDist	1277	935	935	650	632
	[-554, 3136]	[-783, 2630]	[-1064, 2970]	[-1719, 2916]	[-1911, 3121]
WilderDDist	2212	1765	1765	1227	1163
	[4, 4431]	[-589, 4151]	[-1206, 4746]	[-1769, 4208]	[-2257, 4388]
ForestDDist	-1128	-909	-909	-767	-752
	[-2115, -128]	[-1897, 102]	[-2188, 311]	[-2245, 733]	[-2439, 863]
RoadDDist	-4262	-3824	-3824	-4171	-4201
	[-5185, -3315]	[-4888, -2789]	[-5082, -2465]	[-5521, -2722]	[-5885, -2528]

Table 4. 8: Marginal Effects for different estimators (Indirect Effects)

Variables	OLS	S2SLS	S2SLSHAC	GSTSLS	GSTSLSHET
ParcelSize	30297	3709	3709	6521	6889
	[23952, 36383]	[2709, 4810]	[2390, 5130]	[5261, 8033]	[5174, 8919]
Height	29138	3366	3366	4275	4391
	[24804, 33537]	[2589, 4277]	[2356, 4468]	[3449, 5263]	[3297, 5657]
FlArea	121719	14872	14872	22237	23058
	[112694, 130855]	[12818, 17238]	[12025, 18151]	[20813, 24549]	[20357, 26571]
HAge	-1025	-128	-128	-252	-267
	[-1272, -770]	[-170, -89]	[-181, -76]	[-308, -205]	[-339, -204]
Urban	-1032	-160	-160	1218	1462
	[-18748, 16923]	[-2673, 2319]	[-3103, 2822]	[-3770, 6238]	[-4919, 7993]
Elevation	-8558	-725	-725	-1292	-1390
	[-17113, -48]	[-1772, 308]	[-1888, 431]	[-3394, 797]	[-3603, 723]
PopDensity	1837	186	186	363	389
	[-684, 4429]	[-127, 505]	[-181, 551]	[-261, 1006]	[-295, 1093]
White	5716	625	625	686	680
	[4931, 6488]	[498, 769]	[460, 802]	[476, 922]	[433, 948]
TestScoreIndex	1748	194	194	271	281
	[1493, 2002]	[153, 241]	[141, 253]	[194, 357]	[181, 390]

Table 4. 8 (contd.)

Variables	OLS	S2SLS	S2SLSHAC	GSTSLS	GSTSLSHET
WaterDist	-6446	-745	-745	-940	-955
	[-8679, -4202]	[-1057, -448]	[-1140, -370]	[-1612, -297]	[-1711, -209]
TrailDist	-4393	-488	-488	-664	-680
	[-6231, -2549]	[-770, -217]	[-861, -134]	[-1223, -135]	[-1400, -16]
ParkDist	1277	119	119	123	124
	[-481, 3128]	[-104, 343]	[-143, 385]	[-325, 566]	[-367, 622]
WilderDDist	2212	224	224	232	228
	[-8, 4471]	[-78, 536]	[-154, 613]	[-321, 799]	[-427, 877]
ForestDDist	-1128	-116	-116	-145	-147
	[-2123, -131]	[-248, 13]	[-281, 43]	[-439, 145]	[-481, 188]
RoadDDist	-4262	-486	-486	-788	-823
	[-5216, -3327]	[-632, -356]	[-662, -323]	[-1080, -526]	[-1172, -497]

Figure 4. 1: Study Area (Yavapai County, Arizona)

Study Area (Yavapai County, Arizona)

Legend

YavaRes0809
MajorRoads
WildernessArea
USFS
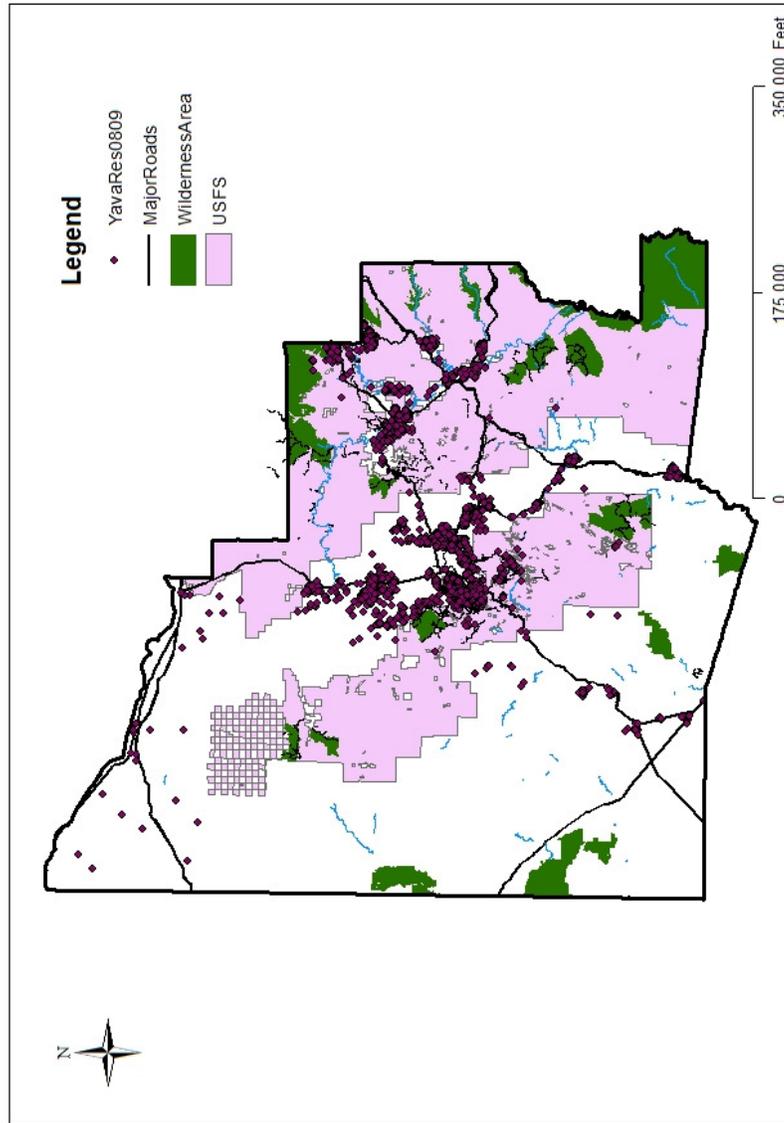
Figure 4. 2: Euclidian distance *vs.* Driving distance. The Euclidian distance is shorter not only because it measures along the straight line, but also because it measures the distance between two points, which are different from the driving distance points.

116
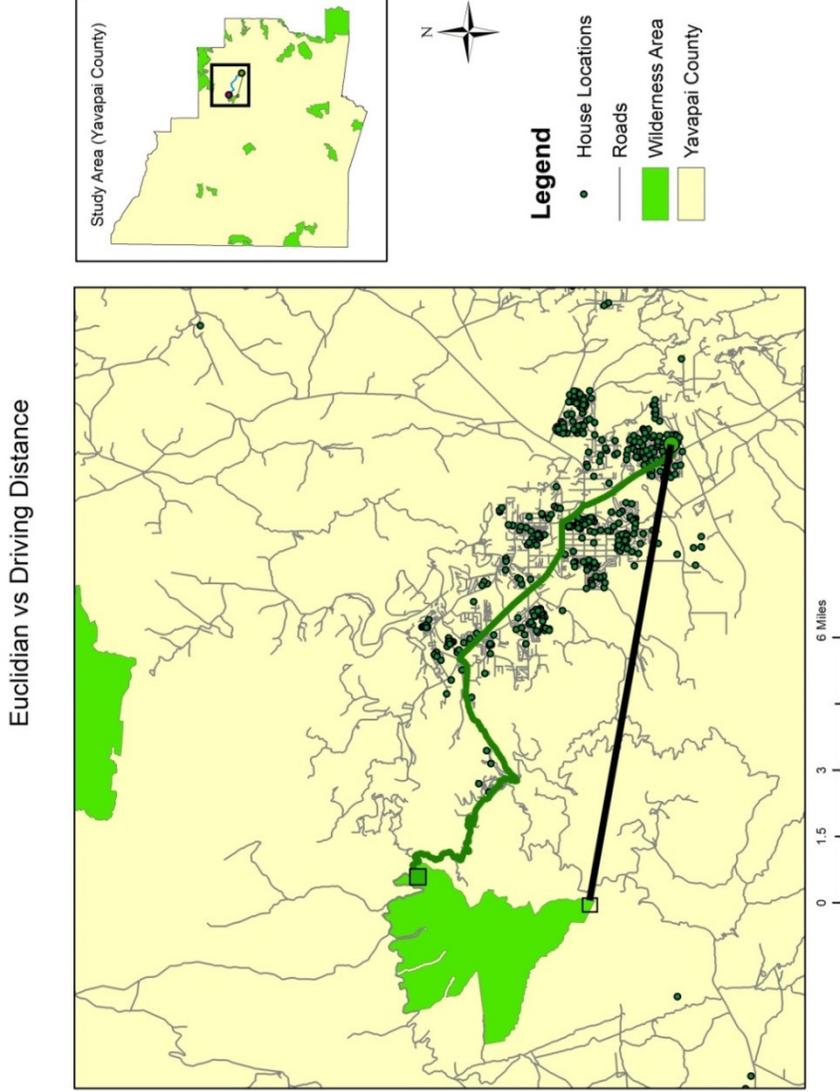
CHAPTER 5

Conclusion and Future Research

5.1   Dissertation Summary

The objective of this dissertation is to highlight the importance of wildlife habitat and protected areas. It focuses on providing geo-spatial methods and applications in relation to wildlife habitat selection, human-wildlife interactions and economic benefits of protected wilderness areas. Chapter 2 and 3 focus on the intensity of habitat use by mountain goats in Kenai Peninsula, Alaska, and Chapter 3 and 4 focus on hedonic methods in analyzing the effects of proximity to natural amenities on house prices.

The aim of Chapter 1 is to provide an improved method in habitat selection/use study with special focus on mountain goat habitat use. Using the GPS collared goats' locations, the intensity of habitat use by mountain goats are predicted. Poisson and negative binomial regression models are extended to incorporate spatial effects under Gaussian CAR approach. These models highlight the importance of nearby areas (spatial dependence) in the habitat selection by animals. With the advance in GPS technology, there is an increasing trend in using satellite telemetry data in wildlife habitat selection/use studies. The locations from GPS collar are accurate and are very useful in studying the habitat

selection and movement behavior of animals. This, however, poses challenges in developing methodologies that accommodate the advancement in available data. Researchers have used various methods to understand the habitat selection/use by animals using GPS collar data. The methods applied in this chapter are expected to be useful for other similar studies involving GPS locations of animals.

Chapter 2 focuses on comparing habitat use prediction from two sources of data and exploring the effects of winter recreational activities on mountain goat habitat selection. Regression based test of equivalency is performed to compare the predicted habitat use based on the model outcomes from GPS collar and Aerial survey of goat locations. While this may or may not be valid for all animals, the map layers obtained from two different survey methods for mountain goats are found to be equivalent, which is important for researchers focusing on one particular survey method to create suitable habitat map for mountain goats. This chapter also explores the effects of winter recreational activities on goat habitat use. As goats tend to avoid areas utilized by human winter recreationists, wildlife managers may take advantage of the outcomes of this research in identifying more suitable goat habitat and preserving those areas from recreational and other forms of human disturbances. The Bayesian method of model comparison is another contribution this chapter, which is expected to be useful in selecting appropriate model in studies using negative binomial regressions.

Chapter 3 explores the inconsistency issues associated with the distance variables used in hedonic price models. In particular, simulations are performed

to examine the statistical behaviors of estimated coefficients when explanatory variables include one or more distance variables. The coefficients of distance variables turn out to be inconsistent when there are multiple points representing landmarks (referring to environmental amenities or disamenities) and distances to two or more of such landmarks appear in the model specifications. However, when there are multiple locations for each type of unique landmark locations, the coefficients are correctly estimated with expected standard deviations as long as the distance variables appearing into the regression models are distances to the nearest landmark locations. The results are particularly important because many studies have major objective of estimating welfare measures based on the distance variables and as such the use of distance variables may not be completely avoided. Therefore, the implications of this chapter will be useful in future hedonic studies using distance variables.

Chapter 4 examines the effects of proximity to wilderness areas on residential property values. Using the GMM approach, spatial hedonic methods are applied to estimate the marginal effects of living in proximity to protect wilderness areas. On average, a residential property located one mile closer to a nearest wilderness area is estimated to have values $4730 and $5024 higher based on Euclidean and road distance as a measure of proximity respectively.

5.2   Avenues for Future Research

This dissertation manuscript can be considered complete in its form, but it presents further avenues for future research. Specific to habitat models (Chapter 1), exploring alternative methods accounting for spatial dependency would allow researchers to choose better method that accurately predict habitat. For example, the alternative spatial error specification such as simultaneous autoregressive formulation can be employed in the count regression models. Similarly, the discrete choice models may be improved by incorporating features accounting for spatial dependency.

Throughout this dissertation, habitat use and selection (Chapter 1 and 2) are modeled for a single species using mountain goats. These models can be improved by considering multiple species utilizing similar resources. Moreover, additional information on climate change can be incorporated to improve prediction of habitat use by animals.

Regarding to wilderness area (Chapter 4), two potential areas of extensions remain for future research. First, the application of hedonic method itself can be extended by combining the information on changes in land use with data on house prices over a longer time period to examine the temporal trends of in welfare values of protected wilderness areas. Second, hedonic method may be combined to other non-market valuation methods to estimate the overall values of wilderness areas.

Appendices

Appendix A: The equivalence between Conditional Logit Model and Poisson Regression

The following derivation is based on Guimarães et al (2003). Suppose the individual choice decisions are based on the vector of choice-specific attribute variables alone so that the probability of choosing a particular location j is

$$p_j = \frac{\exp(\beta x_j)}{\sum_{j=1}^{J} \exp(\beta x_j)}$$

Then log likelihood of the conditional logit model is

$$\ln L_{cl} = \sum_{j=1}^{J} n_j \ln p_j \tag{A.1}$$

where n_j is the number of times location j is chosen.

Now suppose n_j is independently Poisson distributed with $E(n_j) = \lambda_j = \exp(\alpha + \beta x_j)$, the log likelihood function for the Poisson process

$$P(Y = y) = \frac{\exp(-\lambda) \lambda^y}{y!}$$

can be written as

$$
\begin{aligned}
\ln L_p &= \sum_{j=1}^{J} \left(-\lambda_j + n_j \ln \lambda_j - \ln n_j! \right) \\
&= \sum_{j=1}^{J} \left[-\exp(\alpha + \beta x_j) + n_j (\alpha + \beta x_j) - \ln n_j! \right]
\end{aligned}
\tag{A.2}
$$

The first order condition to maximize log likelihood with respect to α is given as

$$\frac{\partial \ln L_p}{\partial \alpha} = \sum_{j=1}^{J} \left[n_j - \exp\left(\alpha + \beta x_j\right) \right] = 0$$

which implies

$$\alpha = \ln N - \ln\left(\sum_{j=1}^{J} \exp\left(\beta x_j\right) \right) \tag{A.3}$$

Now substituting α back to the log likelihood, we get

$$
\begin{aligned}
\ln L_p &= -N + N \ln N - \sum_{j=1}^{J} n_j \ln\left(\sum_{j=1}^{J} \exp\left(\beta x_j\right) \right) + \sum_{j=1}^{J} n_j \beta x_j - \sum_{j=1}^{J} \ln n_j ! \\
&= \sum_{j=1}^{J} n_j \ln p_j - N + N \ln N - \sum_{j=1}^{J} \ln n_j !
\end{aligned}
\tag{A.4}
$$

The first term of equation (A.4) is identical to log likelihood in equation (A.1). The remaining terms are independent of parameters, and hence both methods produce identical estimates. The estimation process for standard conditional logit model is performed by maximizing the log likelihood given by equation (A.1). Writing program codes and maximizing the log likelihood creates computational burden and is time consuming. However, in some statistical softwares (R, Stata, etc,), specific packages have been developed to obtain parameter estimates of conditional logit models. With some efforts in data organization, these canned procedures can help researchers estimate unknown parameters quickly and efficiently.

Appendix B: Discrete Choice Model Coefficients

Table B 1

	G2_07	G5_07	G6_07	G840_07	G870_07	G5_08	G6_08	G870_08
Distance (in 100m) from Previous Location	-0.0257*** (0.0004)	-0.0283*** (0.0004)	-0.0242*** (0.0004)	-0.0293*** (0.0005)	-0.0226*** (0.0004)	-0.0214*** (0.0004)	-0.0222*** (0.0003)	-0.0201*** (0.0003)
DistEsc	-0.7121*** (0.0973)	0.1319 (0.0980)	-0.2951*** (0.0460)	-0.8794*** (0.1000)	-0.6472*** (0.0890)	-0.3413*** (0.0500)	-0.0346 (0.0545)	-0.2010** (0.0660)
Elevation	1.7506*** (0.3792)	1.1960*** (0.1816)	1.5238*** (0.4225)	-0.3839 (0.2841)	5.4049*** (0.4555)	2.3540*** (0.5013)	4.7729*** (0.6578)	2.3756*** (0.2827)
ElevationSq	-0.1015*** (0.0218)	-0.0873*** (0.0150)	-0.0545** (0.0193)	0.0281 (0.0188)	-0.2877*** (0.0294)	-0.1415*** (0.0298)	-0.2002*** (0.0303)	-0.1353*** (0.0187)
Aspect	0.2517 (0.1562)	-2.1341*** (0.3352)	-0.4863*** (0.0679)	0.2381 (0.1553)	-1.6506*** (0.2875)	0.2971 (0.1814)	-0.9324*** (0.0696)	-1.0648*** (0.1919)

Table B 1 (contd.)

	G2_07	G5_07	G6_07	G840_07	G870_07	G5_08	G6_08	G870_08
VRM	0.1296***	-0.0966.	0.042*	0.1558***	0.7212***	0.1544**	0.1404***	0.5446***
	(0.0314)	(0.0535)	(0.0166)	(0.0388)	(0.0381)	(0.0529)	(0.0154)	(0.0321)
Rock	0.3313***	1.1876***	0.3021.	-0.637.		0.3776**	1.174***	
	(0.0954)	(0.3044)	(0.168)	(0.3676)		(0.1342)	(0.138)	
Shrub	0.2919*	0.6533***	0.1774	0.7719***	-0.4941***	0.7499***	0.7202***	-0.368***
	(0.1211)	(0.1074)	(0.1463)	(0.0738)	(0.0815)	(0.1162)	(0.082)	(0.0826)
Log-Likelihood	-4369	-4960	-5866	-3821	-4659	-5161	-5777	-5884
AIC	8754	9937	11748	7658	9332	10339	11570	11781
BIC	8798	9982	11793	7702	9372	10383	11615	11821
N	1849	2103	2128	1918	2046	1908	2125	2057

Appendix C: Spatial weights based on Thiessen Polygon

Table C1: Estimation results from OLS, S2SLS, S2SLS-HAC, GSTSLS, GSTSLS-HET (Spatial Weights based on Thiessen Polygon)

Variables	OLS	S2SLS	S2SLS-HAC	GSTSLS	GSTSLS-HET
Constant	4.775***	3.605***	3.605***	3.835***	3.863***
	(0.201)	(0.434)	(0.581)	(0.408)	(0.480)
Structural Characteristics					
ParcelSize	0.102***	0.101***	0.101***	0.127***	0.131***
	(0.014)	(0.018)	(0.023)	(0.015)	(0.021)
Height	0.126***	0.113***	0.113***	0.108***	0.108***
	(0.010)	(0.014)	(0.017)	(0.010)	(0.013)
FlArea	0.527***	0.497***	0.497***	0.495***	0.494***
	(0.020)	(0.037)	(0.048)	(0.021)	(0.035)
HAge	-0.004***	-0.004***	-0.004***	-0.005***	-0.005***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Elevation	-0.011	-0.016	-0.016	0.038	0.048
	(0.042)	(0.046)	(0.055)	(0.054)	(0.064)
Neighborhood Variables					
Urban	-0.036	-0.019	-0.019	-0.024	-0.025
	(0.019)	(0.018)	(0.020)	(0.022)	(0.022)
PopDensity	0.007	0.006	0.006	0.009	0.009
	(0.006)	(0.005)	(0.006)	(0.007)	(0.007)
White	0.024***	0.020***	0.020***	0.018***	0.017***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
TestScoreIndex	0.008***	0.007***	0.007***	0.007***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table C1 (contd.)

Variables	OLS	S2SLS	S2SLS-HAC	GSTSLS	GSTSLS-HET
Environmental Variables					
WaterDist	-0.021***	-0.019***	-0.019**	-0.017*	-0.017*
	(0.005)	(0.005)	(0.006)	(0.007)	(0.007)
TrailDist	-0.023***	-0.019***	-0.019**	-0.017***	-0.017**
	(0.004)	(0.005)	(0.006)	(0.005)	(0.006)
ParkDist	0.004	0.002	0.002	0.003	0.003
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
RoadDist	0.010	0.007	0.007	0.002	0.002
	(0.008)	(0.009)	(0.010)	(0.009)	(0.012)
ForestDist	-0.015***	-0.012***	-0.012**	-0.012**	-0.012**
	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)
WilderDist	-0.018***	-0.014***	-0.014***	-0.017***	-0.017***
	(0.003)	(0.003)	(0.004)	(0.004)	(0.005)
λ		0.159***	0.159**	0.158***	0.159***
		(0.047)	(0.056)	(0.039)	(0.046)
ρ				0.325	0.401***
					(0.043)

Standard errors in parentheses. $^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$.

Table C 2: Estimation results from OLS, S2SLS,S2SLS-HAC, GSTSLS, GSTSLS-HET (Spatial weights based on Thiessen Polygon)

Variables	OLS	S2SLS	S2SLS-HAC	GSTSLS	GSTSLS-HET
Constant	4.921***	3.883***	3.883***	4.003***	4.012***
	(0.201)	(0.437)	(0.586)	(0.404)	(0.484)
Structural Characteristics					
ParcelSize	0.130***	0.124***	0.124***	0.145***	0.148***
	(0.014)	(0.018)	(0.022)	(0.015)	(0.020)
Height	0.125***	0.113***	0.113***	0.108***	0.107***
	(0.010)	(0.014)	(0.017)	(0.010)	(0.013)
FlArea	0.521***	0.495***	0.495***	0.492***	0.491***
	(0.020)	(0.037)	(0.048)	(0.020)	(0.035)
HAge	-0.004***	-0.004***	-0.004***	-0.005***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Elevation	-0.004	-0.004	-0.004	0.038	0.045
	(0.039)	(0.042)	(0.048)	(0.050)	(0.060)
Neighborhood Variables					
Urban	-0.037	-0.023	-0.023	-0.027	-0.028
	(0.019)	(0.018)	(0.020)	(0.022)	(0.022)
PopDensity	0.008	0.007	0.007	0.009	0.010
	(0.006)	(0.005)	(0.006)	(0.007)	(0.007)
White	0.024***	0.020***	0.020***	0.018***	0.018***
	(0.002)	(0.002)	(0.003)	(0.002)	(0.003)
TestScoreIndex	0.007***	0.006***	0.006***	0.006***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)

Table C2 (contd.)

Variables	OLS	S2SLS	S2SLS-HAC	GSTSLS	GSTSLS-HET
Environmental Variables					
WaterDist	-0.028***	-0.025***	-0.025***	-0.023***	-0.023**
	(0.005)	(0.005)	(0.006)	(0.007)	(0.007)
TrailDist	-0.019***	-0.015**	-0.015**	-0.013*	-0.013*
	(0.004)	(0.005)	(0.006)	(0.005)	(0.006)
ParkDist	0.005	0.004	0.004	0.004	0.004
	(0.004)	(0.004)	(0.004)	(0.005)	(0.005)
RoadDDist	0.009*	0.007	0.007	0.005	0.005
	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)
ForestDDist	-0.005*	-0.004	-0.004	-0.004	-0.004
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
WilderDDist	-0.018***	-0.016***	-0.016***	-0.017***	-0.017***
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
λ		0.141**	0.141*	0.153***	0.155***
		(0.046)	(0.056)	(0.038)	(0.046)
ρ				0.333	0.389***
					(0.044)

Standard errors in parentheses. $^{*} p < 0.10$, $^{**} p < 0.05$, $^{***} p < 0.01$.

Table C3: Marginal Effects for different estimators (Direct Effects)

Variables	OLS	S2SLS	S2SLSHAC	GSTSLS	GSTSLSHET
ParcelSize	23938	23615	23615	29860	30847
	[17591, 30394]	[14928, 32057]	[12688, 34071]	[22850, 36670]	[21226, 40339]
Height	29455	26577	26577	25405	25266
	[25032, 33868]	[20078, 32857]	[18632, 34400]	[20572, 29940]	[18999, 31220]
FlArea	123016	116702	116702	116189	116028
	[113769, 132161]	[98823, 133144]	[94419, 138578]	[106383, 125162]	[99728, 131091]
HAge	-944	-933	-933	-1218	-1257
	[-1188, -692]	[-1239, -623]	[-1314, -539]	[-1472, -952]	[-1583, -934]
Urban	-2551	-3770	-3770	8927	11293
	[-21563, 16588]	[-25043, 17653]	[-29669, 21243]	[-15638, 33528]	[-17881, 40435]
Elevation	-8436	-4529	-4529	-5713	-5979
	[-16855, 102]	[-12658, 3551]	[-13412, 4331]	[-15618, 4328]	[-16194, 3805]
PopDensity	1626	1398	1398	2019	2132
	[-896, 4260]	[-1000, 3816]	[-1366, 4127]	[-1083, 5110]	[-979, 5350]
White	5705	4679	4679	4151	4048
	[4937, 6478]	[3617, 5710]	[3383, 5943]	[3089, 5200]	[-2815, 5259]
TestScoreIndex	1940	1627	1627	1536	1516
	[1688, 2196]	[1250, 1986]	[1178, 2086]	[1157, 1882]	[1050, 1976]

Table C3 (contd.)

Variables	OLS	S2SLS	S2SLSHAC	GSTSLS	GSTSLSHET
WaterDist	-4890	-4411	-4411	-4014	-3934
	[-7135, -2655]	[-6727, -2085]	[-7217, -1503]	[-7084, -890]	[-7376, -509]
TrailDist	-5311	-4410	-4410	-4024	-3921
	[-7084, -3518]	[-6569, -2235]	[-7167, -1653]	[-6417, -1621]	[-6756, -1086]
ParkDist	961	521	521	724	752
	[-909, 2821]	[-1126, 2129]	[-1428, 2410]	[-1434, 2990]	[-1467, 2964]
RoadDist	2264	1614	1614	536	399
	[-1127, 5706]	[-2352, 5550]	[-3162, 6334]	[-3726, 4781]	[-5147, 5840]
ForestDist	-3530	-2737	-2737	-2856	-2873
	[-5033, -2038]	[-4192, -1240]	[-4480, -996]	[-4818, -827]	[-4981, -776]
WilderDist	-4138	-3341	-3341	-3976	-4084
	[-5402, -2805]	[-4809, -1862]	[-5030, -1590]	[-5716, -2247]	[-6127, -1988]

Table C4: Marginal Effects for different estimators (Indirect Effects)

Variables	OLS	S2SLS	S2SLSHAC	GSTSLS	GSTSLSHET
ParcelSize	23938	4343	4343	5454	5676
	[17653, 30201]	[2838, 6026]	[2365, 6506]	[4258, 6899]	[4007, 7610]
Height	29455	4887	4887	4640	4649
	[25054, 33835]	[3787, 6212]	[3523, 6495]	[3903, 5615]	[3596, 5937]
FlArea	123016	21461	21461	21220	21348
	[113608, 132212]	[18799, 25205]	[17717, 26127]	[19961, 23476]	[18930, 24883]
HAge	-944	-171	-171	-222	-231
	[-1197, -694]	[-234, -118]	[-248, -103]	[-277, -179]	[-299, -173]
Urban	-2551	-693	-693	1630	2078
	[-21578, 16355]	[-4587, 3292]	[-5598, 4083]	[-2963, 6302]	[-3376, 7684]
Elevation	-8436	-833	-833	-1043	-1100
	[-17179, 282]	[-2361, 711]	[-2555, 811]	[-2970, 809]	[-3062, 751]
PopDensity	1626	257	257	369	392
	[-1005, 4258]	[-190, 716]	[-253, 779]	[-206, 967]	[-177, 991]
White	5705	860	860	758	745
	[4927, 6486]	[684, 1082]	[640, 1126]	[580, 971]	[534, 989]
TestScoreIndex	1940	299	299	281	279
	[1690, 2195]	[236, 376]	[220, 392]	[220, 355]	[200, 375]

Table C4 (contd.)

Variables	OLS	S2SLS	S2SLSHAC	GSTSLS	GSTSLSHET
WaterDist	-4890	-811	-811	-733	-724
	[-7163, -2620]	[-1276, -381]	[-1375, -299]	[-1314, -183]	[-1397, -98]
TrailDist	-5311	-811	-811	-735	-721
	[-7117, -3486]	[-1245, -420]	[-1357, -307]	[-1210, -300]	[-1288, -190]
ParkDist	961	96	96	132	138
	[-871, 2814]	[-210, 404]	[-272, 460]	[-287, 559]	[-267, 552]
RoadDist	2264	297	297	98	73
	[-1183, 5666]	[-427, 1051]	[-597, 1221]	[-702, 885]	[-968, 1097]
ForestDist	-3530	-503	-503	-522	-529
	[-5012, -2076]	[-800, -232]	[-851, -183]	[-914, -164]	[-937, -151]
WilderDist	-4138	-614	-614	-726	-751
	[-5462, -2807]	[-902, -362]	[-954, -302]	[-1068, -410]	[-1176, -383]

Table C5: Marginal Effects for different estimators (Direct Effects)

Variables	OLS	S2SLS	S2SLSHAC	GSTSLS	GSTSLSHET
ParcelSize	30297	28990	28990	33946	34689
	[23939, 36509]	[20891, 36942]	[18392, 39215]	[26792, 40832]	[25387, 43820]
Height	29138	26566	26566	25311	25182
	[24800, 33593]	[20169, 32933]	[18666, 34486]	[20635, 29710]	[18970, 31104]
FlArea	121719	116079	116079	115316	115182
	[112601, 130766]	[98886, 132762]	[93420, 137521]	[105467, 124397]	[99127, 130514]
HAge	-1025	-1012	-1012	-1270	-1303
	[-1271, -780]	[-1319, -701]	[-1390, -631]	[-1535, -1000]	[-1624, -979]
Urban	-1032	-899	-899	8831	10487
	[-19178, 16500]	[-19655, 18167]	[-23584, 21071]	[-14221, 32129]	[-16540, 37959]
Elevation	-8558	-5281	-5281	-6300	-6520
	[-17172, 1]	[-13397, 2864]	[-14275, 3727]	[-16409, 3973]	[-16454, 3492]
PopDensity	1837	1632	1632	2167	2262
	[-819, 4437]	[-840, 4050]	[-1074, 4369]	[-1043, 5312]	[-902, 5352]
White	5716	4775	4775	4219	4118
	[4943, 6508]	[3719, 5798]	[3460, 6027]	[3162, 5262]	[2898, 5307]
TestScoreIndex	1748	1470	1470	1377	1361
	[1495, 2001]	[1124, 1814]	[1034, 1909]	[1013, 1735]	[899, 1795]

134

Table C5 (contd.)

Variables	OLS	S2SLS	S2SLSHAC	GSTSLS	GSTSLSHET
WaterDist	-6446	-5751	-5751	-5442	-5373
	[-8667, -4201]	[-8105, -3411]	[-8606, -2817]	[-8450, -2414]	[-8762, -1973]
TrailDist	-4393	-3613	-3613	-3095	-2988
	[-6242, -2592]	[-5820, -1447]	[-6284, -878]	[-5504, -617]	[-5993, -33]
ParkDist	1277	835	835	872	869
	[-531, 3086]	[-899, 2533]	[-1160, 2862]	[-1291, 3092]	[-1540, 3246]
WilderDDist	2212	1669	1669	1236	1186
	[69, 4377]	[-684, 4009]	[-1246, 4552]	[-1439, 3882]	[-1836, 4239]
ForestDDist	-1128	-853	-853	-865	-872
	[-2111, -131]	[-1849, 179]	[-2075, 376]	[-2195, 491]	[-2385, 608]
RoadDDist	-4262	-3716	-3716	-4016	-4048
	[-5226, -3306]	[-4762, -2640]	[-4961, -2426]	[-5262, -2730]	[-5479, -2560]

Table C6: Marginal Effects for different estimators (Indirect Effects)

Variables	OLS	S2SLS	S2SLSHAC	GSTSLS	GSTSLSHET
ParcelSize	30297	4655	4655	5948	6202
	[24168, 36568]	[3415, 6079]	[3078, 6449]	[4832, 7342]	[4669, 8008]
Height	29138	4266	4266	4435	4502
	[24836, 33617]	[3285, 5418]	[3083, 5648]	[3712, 5377]	[3500, 5751]
FlArea	121719	18638	18638	20207	20593
	[112619, 130829]	[16271, 21858]	[15355, 22678]	[18992, 22367]	[18163, 23988]
HAge	-1025	-162	-162	-223	-233
	[-1276, -775]	[-217, -115]	[-229, -103]	[-275, -180]	[-298, -178]
Urban	-1032	-144	-144	1548	1875
	[-18534, 16202]	[-3252, 3066]	[-3819, 3392]	[-2537, 5823]	[-3125, 6977]
Elevation	-8558	-848	-848	-1104	-1166
	[-17235, 99]	[-2220, 488]	[-2335, 608]	[-2934, 705]	[-3047, 654]
PopDensity	1837	262	262	380	404
	[-734, 4405]	[-130, 683]	[-181, 725]	[-174, 946]	[-167, 987]
White	5716	767	767	739	736
	[4951, 6497]	[609, 958]	[570, 1000]	[569, 940]	[531, 971]
TestScoreIndex	1748	236	236	241	243
	[1500, 1999]	[185, 298]	[173, 311]	[182, 313]	[166, 331]

Table C6 (contd.)

Variables	OLS	S2SLS	S2SLSHAC	GSTSLS	GSTSLSHET
WaterDist	-6446	-923	-923	-954	-961
	[-8673, -4210]	[-1331, -549]	[-1426, -475]	[-1512, -431]	[-1600, -356]
TrailDist	-4393	-580	-580	-542	-534
	[-6202, -2558]	[-948, -237]	[-1047, -161]	[-1002, -120]	[-1096, -4]
ParkDist	1277	134	134	153	155
	[-584, 3121]	[-151, 427]	[-191, 471]	[-249, 563]	[-281, 604]
WilderDDist	2212	268	268	217	212
	[-18, 4412]	[-108, 656]	[-185, 752]	[-250, 703]	[-327, 767]
ForestDDist	-1128	-137	-137	-152	-156
	[-2093, -128]	[-308, 24]	[-341, 67]	[-399, 85]	[-430, 114]
RoadDDist	-4262	-597	-597	-704	-724
	[-5182, -3309]	[-783, -436]	[-816, -401]	[-950, -493]	[-1013, -470]

Appendix D: R Codes

```
#######################
#######################
R Codes used in Chapter 1
#######################
#######################

# Set working directory
setwd("D:/naresh/...")


# Load required packages
library(MASS)
library(sp)
library(mgcv)
library(spdep)
library(BayesX)
library(spatcounts)


# Load data (repeat this process for each individual goats)
features <- read.table(file="2007g2f.txt", header=T, sep="\t")
locations <- read.table(file="2007g2.txt", header=T, sep="\t")


# Scaling the variables
features$elev <- features$elevation/100
features$elevsq <- features$elev^2
features$caspect <- cos(features$aspect*pi/180)
features$vrm100 <- features$vrm*100
features$distesc10 <- features$distesc/100
features$rock <- as.numeric(features$lc_code==24)
features$shrub <- as.numeric(features$lc_code >= 14 & features$lc_code <=15)


# Creating X matrix
Xdf <- subset(features, select = c("elev","elevsq","caspect",
              "vrm100","distesc10","rock","shrub"))
X <- data.frame(Xdf)


# Deviate from mean
```

```
X.mdv <- X - matrix(mean(X), nrow=nrow(X), ncol=ncol(X), byrow=T)

# Creating response variable
features$ct <- rep(NA, nrow(features))
for(i in 1:nrow(features)){
        features$ct[i] <- length(na.omit(locations$cellid[locations$cellid==
                features$cellid[i]]))
}


# Creating Y as Yin in spatcounts
Y <- data.frame(features$ct)


# Spatial regions
region <- 1:nrow(features)
region <- data.frame(region)
coordinates(features) <- c("ptx","pty")
coordinates(locations) <- c("ptx","pty")


# Neighbor list
f_nb <- dnearneigh(features, 0, 85)


# Weight Matrix
weightmat <- nb2listw(f_nb, style="B")


# Creating gmat as sim.gmat in spatcounts
gmat <- listw2mat(weightmat)
gmat <- data.frame(gmat)


# --------
# The following series of commands are used to create
# nmat as sim.nmat in spatcounts
nb.map <- nb2gra(f_nb)
nlist <- get.neighbor(nb.map, 1:nrow(features))
nlist <- lapply(nlist,function(v){if(length(v)<2)v[2]<-"0"
 v})
nlist <- lapply(nlist,function(v){if(length(v)<3)v[3]<-"0"
 v})
nlist <- lapply(nlist,function(v){if(length(v)<4)v[4]<-"0"
 v})
nlist <- lapply(nlist,function(v){if(length(v)<5)v[5]<-"0"
 v})
```

```r
nlist <- lapply(nlist,function(v){if(length(v)<6)v[6]<-"0"
  v})
nlist <- lapply(nlist,function(v){if(length(v)<7)v[7]<-"0"
  v})
nlist <- lapply(nlist,function(v){if(length(v)<8)v[8]<-"0"
  v})
nlist.mat <- do.call(rbind,nlist)

n.neigh <- card(f_nb)
nmat <- cbind(1:nrow(nlist.mat), nlist.mat[,], n.neigh)
nmat <- apply(nmat, 2, as.numeric)
rownames(nmat) <- 1:nrow(nmat)
colnames(nmat) <- c("V1","V2","V3","V4","V5",
          "V6","V7","V8","V9","V10")
nmat <- data.frame(nmat) # This is the final nmat as sim.nmat in spatcounts
# --------

# Poisson model estimation (for negative binomial, replace "Poi" by "NB")
poi.sp1 <- est.sc(Y, ~ X.mdv[,1] + X.mdv[,3] + X.mdv[,4] + X.mdv[,5] +
    X.mdv[,6] + X.mdv[,7] - 1,
       region, model="Poi", gmat, nmat, totalit=20000)


# DIC Computation
DIC.poi <- DIC(Y, ~ X.mdv[,1] + X.mdv[,3] + X.mdv[,4] + X.mdv[,5] +
    X.mdv[,6] + X.mdv[,7] - 1,
       region, poi.sp1, burnin = 1000)


# Burn-in range
range <- 1001:20000


# Log Likelihood matrix (Individual Log Likelihood)
ll.poi <- LogLike(Y, ~ X.mdv[,1] + X.mdv[,3] + X.mdv[,4] + X.mdv[,5] +
    X.mdv[,6] + X.mdv[,7] - 1,
       region, poi.sp1, burnin = 1000)


# Mean Log Likelihood for all iterations
m.ll.poi <- mean(apply(ll.poi$ll,2,sum))


# Parameter Estimates
beta <- round(apply(poi.sp1$beta[,range],1,mean),2)
```

```
# Mean gamma and psi
gama1 <- round(apply(poi.sp1$gamma[,range],1,mean),2)
psi1 <- mean(poi.sp1$psi[range])


# Standard deviation of psi
sdpsi <- round(sd(poi.sp1$psi[range]),2)


# Number of parameters
k <- length(beta)


# Standard Errors, Z-scores and Confidence Interval
st.d <- round(apply(poi.sp1$beta[,range],1,sd),3)
zscore <- round(beta/st.d, 2)
ci.poi <- round(apply(poi.sp1$beta[,range],1,quantile, c(0.025, 0.975)), 3)


# Results for parameter estimates
result.poi1 <- data.frame(beta, st.d, zscore, t(ci.poi))


# AIC Calculation
aic <- - 2 * m.ll.poi + 2 * k


# Results for gamma and psi
gmpsi <- round(c(mean(gama1), psi1, aic, DIC.poi$DIC),2)
sdgmpsi <- round(c(sd(gama1), sdpsi, NA, NA),3)
zgmpsi <- round(gmpsi/sdgmpsi, 2)
cigmpsi <- cbind(round(quantile(gama1,probs=c(0.025,0.975)),3),
        round(quantile(poi.sp1$psi[range],probs=c(0.025,0.975)),3), NA, NA)


result.poi2 <- data.frame(gmpsi, sdgmpsi, zgmpsi,t(cigmpsi))
colnames(result.poi2) <- c("beta", "st.d", "zscore", "X2.5.", "X97.5.")
result.poi <- rbind(result.poi1,result.poi2)
colnames(result.poi) <- c("Coefficients", "St. Error",
                "Z-Score", "Conf.int 0.025", "0.975")
vname <- c("Elevation", "CosAspect",
      "VRM", "DistEsc", "Rock", "Shrub", "Gamma", "Psi", "AIC", "DIC")
rownames(result.poi) <- vname

# save results and run model for another goat
# This creates table for a single goat model (Table 1.3)
write.table(result.poi,
```

```
file = "D:/naresh/.../g2.txt",
sep = "\t")
```

```
########################
########################
# R Codes used in Chapter 2
########################
########################

# Set Working Directory
setwd("D:/naresh/...")

# Load Required Packages
library(MASS)


#=======================================
# ---------------------- Functions --------------------------
# The following functions are heavily borrowed from
# package "bayesm" and modified to obtain marginal
# likelihood
#=======================================

llnegbin <- function(par, X, y, nvar) {
 beta <- par[1:nvar]
        alpha <- exp(par[nvar + 1]) + 1e-50
        mean <- exp(X %*% beta)
        prob <- alpha/(alpha + mean)
        prob <- ifelse(prob < 1e-100, 1e-100, prob)
        out <- (dnbinom(y, alpha, prob, log = TRUE))
        return(sum(out))
}

lpostbetai <- function(beta, alpha, X, y, betabar, A) {
        lambda <- exp(X %*% beta)
        p <- alpha/(alpha + lambda)
        residual <- as.vector(beta - betabar)
        sum(alpha * log(p) + y * log(1 - p)) - 0.5 * (t(residual) %*%
        A %*% residual)
}

lpostalpha <- function(alpha, beta, X, y, a, b) {
        sum(log(dnbinom(y, size = alpha, mu = exp(X %*% beta)))) +
        (a - 1) * log(alpha) - b * alpha
}

negbinmcmcRW <- function (y,X,NITER,
     marginal.likelihood = c("none","Laplace")) {
        nvar <- ncol(X)
        nobs <- length(y)
```

```
betabar <- rep(0, nvar)
par <- rep(0, (nvar + 1))
beta0 <- rep(0, nvar)

mle <- optim(par, llnegbin, X = X, y = y, n
 var = nvar, method = "L-BFGS-B",
         upper = c(Inf, Inf, Inf, log(1e+08)), hessian = TRUE,
         control = list(fnscale = -1))

beta_mle <- mle$par[1:nvar]
alpha_mle <- exp(mle$par[nvar + 1])
varcovinv <- -mle$hessian

betacvar <- s_beta * solve(varcovinv[1:nvar, 1:nvar])
betaroot <- t(chol(betacvar))
alpha <- alpha_mle
alphacvar <- s_alpha/varcovinv[nvar + 1, nvar + 1]
alphacroot <- sqrt(alphacvar)

beta <- beta0
oldlpostbeta <- 0
nacceptbeta <- 0
nacceptalpha <- 0
clpostbeta <- 0
alphadraw <- rep(0, NITER)
betadraw <- matrix(double(NITER*nvar),ncol=nvar)
llike <- rep(0, NITER)
logmarglike <- NULL

for (r in 1:NITER) {
        betac <- beta + betaroot %*% rnorm(nvar)
        oldlpostbeta <- lpostbetai(beta, alpha, X, y, betabar, A)
        clpostbeta <- lpostbetai(betac, alpha, X, y, betabar, A)
        ldiff <- clpostbeta - oldlpostbeta
        acc <- min(1, exp(ldiff))
        if (acc < 1) {
                unif <- runif(1)
        }
        else {
                unif <- 0
        }
        if (unif <= acc) {
                beta <- betac
                nacceptbeta <- nacceptbeta + 1
        }
        logalphac <- rnorm(1, mean = log(alpha), sd = alphacroot)
        oldlpostalpha <- lpostalpha(alpha,beta, X, y, a, b)
```

```r
            clpostalpha <- lpostalpha(exp(logalphac), beta, X, y, a, b)
            ldiff <- clpostalpha - oldlpostalpha
            acc <- min(1, exp(ldiff))
            if (acc < 1) {
                    unif <- runif(1)
            }
            else {
                    unif <- 0
            }
            if (unif <= acc) {
                    alpha <- exp(logalphac)
                    nacceptalpha <- nacceptalpha + 1
            }

            betadraw[r, ] <- beta
            alphadraw[r] <- alpha
            llike[r] <- llnegbin(c(beta, alpha), X, y, nvar)
        }

        if (marginal.likelihood == "Laplace") {
            theta.tilde <- mle$par
            Sigma.tilde <- varcovinv
            logmarglike <- (length(theta.tilde)/2) * log(2 * pi) +
                            log(sqrt(det(Sigma.tilde))) +
        llnegbin(theta.tilde,X,y,nvar)
        }

        return(list(llike = llike, betadraw = betadraw,
         alphadraw = alphadraw,
            logmarglike = logmarglike,
            acceptrbeta = nacceptbeta/NITER * 100,
            acceptralpha = nacceptalpha/NITER * 100))
}




# -----------------------------------------
# Read data
data50 <- read.table("D:/naresh/.../data50.txt",
        header = T, sep = "\t")


# Scaling the variables (data50)
data50$elevation1 <- data50$elevation/100
data50$elevationsq <- data50$elevation1^2
data50$cosaspect <- cos(data50$aspect*pi/180)
```

```
data50$vrm1 <- data50$vrm*100
data50$distesc1 <- data50$distesc/100
data50$traildist1 <- data50$traildist/1000


# Create Dummy variables (data50)
data50$shrub <- as.numeric(
        data50$lc_code >= 14 & data50$lc_code <= 15)
data50$rock <- as.numeric(data50$lc_code == 24)

# Scale ski-area and trail length from sq.m and m
#  to sq.km and km respectively
data50$tlength5km <- data50$tlength5km/1000
data50$skiarea5km <- data50$skiarea5km/1000000


#Data
# Construct matrix for X variables
X1 <- cbind('constant' = rep(1,nrow(data50)),
        as.matrix(subset(data50,select=c(
        'elevation1', 'cosaspect','distesc1','vrm1',
        'rock','shrub')))
        )
X2 <- cbind('constant' = rep(1,nrow(data50)),
        as.matrix(subset(data50,select=c(
        'elevation1', 'cosaspect','distesc1','vrm1','rock',
        'shrub','skiarea5km')))
        )
X3 <- cbind('constant' = rep(1,nrow(data50)),
        as.matrix(subset(data50,select=c(
        'elevation1', 'cosaspect','distesc1','vrm1','rock',
        'shrub','tlength5km')))
        )
X4 <- cbind('constant' = rep(1,nrow(data50)),
        as.matrix(subset(data50,select=c(
        'elevation1', 'cosaspect','distesc1','vrm1','rock',
        'shrub','traildist1')))
        )


#=======
# Codes for X = X1; for other models, replace X1 by X2, X3 and X4

X <- X1
nobs <- nrow(X)
nvar <- ncol(X)
```

```
NITER <- 50000

# Prior
A = 0.01 * diag(nvar)
a = 0.5
b = 0.1
s_alpha <- 2.93
s_beta <- 2.93/sqrt(nvar)

# Store m1 and run model to with m2, m3 and m4 for X2, X3 and X4
m1 <- negbinmcmcRW(data50$count,X,NITER,marginal.likelihood="Laplace")

range = 5000:NITER # burn-in iteration

# ==========
# Printing table of results
# Repeat the following codes by replacing m1 for m2, m3 and m4
n <- nobs
theta <- round(apply(m1$betadraw[range,],2,mean),4)
SEs <- round(apply(m1$betadraw[range,],2,sd),4)
k <- length(theta)
tval <- theta/SEs
pval <- 2*(1-pt(abs(tval), n-k))
star <- rep(" ,k)
for(j in 1:k){
        star[j] <- if(pval[j]<0.001){"***"}
        else if(pval[j] < 0.01){"**"}
        else if(pval[j] < 0.05){"*"}
        else if(pval[j] < 0.1){"."}
        else {""}
}

# Create table of result for a given model (Table 2.6)
cat(paste("\n", colnames(X),"\t",
        theta,star, "\n", "\t","(",SEs,")","\n",
        sep=""), file = "resultm1.txt")


# posterior mean and standard deviation of alpha
mean(m1$alphadraw[range])
sd(m1$alphadraw[range])
mean(m2$alphadraw[range])
sd(m2$alphadraw[range])
mean(m3$alphadraw[range])
sd(m3$alphadraw[range])
mean(m4$alphadraw[range])
sd(m4$alphadraw[range])
```

```
# marginal likelihood
m1$logmarglike
m2$logmarglike
m3$logmarglike
m4$logmarglike

logBayesFactor.m2 <- m2$logmarglike-m1$logmarglike
logBayesFactor.m3 <- m3$logmarglike-m1$logmarglike
logBayesFactor.m4 <- m4$logmarglike-m1$logmarglike

# 2 ln BF
2 * logBayesFactor.m2
2 * logBayesFactor.m3
2 * logBayesFactor.m4
```

```
########################
########################

R Codes used in Chapter 3
########################
########################


#####################################
-------------------Functions----------------------------
#####################################

# Extract coefficients from regression
lm.coef <- function(model) {
  beta <- coef(model)
  smry <- summary(model)
  r2 <- smry$r.squared
  k <- smry$df[1]
  n <- sum(smry$df[1:2])
  return(c(beta, r2 = r2, n = n, k = k))
}

# Generate Data and Run Regression returning coefficients of the regression
# This function does not handle cases when more than three distance
# variables are included. For such cases, alternate codes are available upon
# request.

reg.lm.dist <- function(Hloc, n = n, n.lm = n.lm,
    beta1 = beta1, beta2 = beta2, gamma1 = gamma1, gamma2 = gamma2,
    gamma3 = gamma3, nD = nD, reg.rhs = reg.rhs) {

  nD <- nD
  reg.rhs <- reg.rhs

  n.lm <- n.lm
  A <- data.frame(
    x = runif(n.lm, -20, 20),
    y = runif(n.lm, -20, 20))

  B <- data.frame(
    x = runif(n.lm, -20, 20),
    y = runif(n.lm, -20, 20))

  C <- data.frame(
    x = runif(n.lm, -20, 20),
    y = runif(n.lm, -20, 20))
```

```r
DA <- crossdist(Hloc[, 1], Hloc[, 2], A[, 1], A[, 2])
min.DA <- apply(DA, 1, min) # nearest A

DB <- crossdist(Hloc[, 1], Hloc[, 2], B[, 1], B[, 2])
min.DB <- apply(DB, 1, min) # nearest B

DC <- crossdist(Hloc[, 1], Hloc[, 2], C[, 1], C[, 2])
min.DC <- apply(DC, 1, min) # nearest C

x <- rnorm(n)
err <- 2

if(!is.na(nD)){
  if(nD > 3)
    stop("can not handle more than 3 types of landmarks")

  if(nD == 1){
    Y <- beta1 + beta2 * x + gamma1 * min.DA +
      rnorm(n, mean = 0, sd = err)
  }
  if(nD == 2){
    Y <- beta1 + beta2 * x + gamma1 * min.DA + gamma2 * min.DB +
      rnorm(n, mean = 0, sd = err)
  }
  if(nD == 3){
    Y <- beta1 + beta2 * x + gamma1 * min.DA +
      gamma2 * min.DB + gamma3 * min.DC +
      rnorm(n, mean = 0, sd = err)
  }
}
if(is.na(nD)){
  Y <- beta1 + beta2 * x + gamma1 * min.DA +
    gamma2 * min.DB + gamma3 * min.DC +
    rnorm(n, mean = 0, sd = err)
}


a <- data.frame(
  x = runif(n.lm, -20, 20),
  y = runif(n.lm, -20, 20))

b <- data.frame(
  x = runif(n.lm, -20, 20),
  y = runif(n.lm, -20, 20))

c <- data.frame(
  x = runif(n.lm, -20, 20),
```

```
  y = runif(n.lm, -20, 20))

 da <- crossdist(Hloc[, 1], Hloc[, 2], a[, 1], a[, 2])
 min.da <- apply(da, 1, min) # nearest A

 db <- crossdist(Hloc[, 1], Hloc[, 2], b[, 1], b[, 2])
 min.db <- apply(db, 1, min) # nearest B

 dc <- crossdist(Hloc[, 1], Hloc[, 2], c[, 1], c[, 2])
 min.dc <- apply(dc, 1, min) # nearest C

 td <- c("min.DA", "min.DB", "min.DC")
 fd <- c("min.da", "min.db", "min.dc")

 fm1 <- "Y ~ x"
 form <- as.formula(paste(fm1, paste(
    if(is.na(reg.rhs[1])) {paste("")} else {paste(td[1:reg.rhs[1]], collapse = "+")},
    if(is.na(reg.rhs[2])) {paste("")} else {paste(fd[1:reg.rhs[2]], collapse = "+")},
    sep = if(is.na(reg.rhs[1]) | is.na(reg.rhs[2])) {""} else {" + "}),
    sep = if(is.na(reg.rhs[1]) & is.na(reg.rhs[2])) {""} else {" + "}))

 olsmain <- lm(form)
 lm.coef(olsmain)
}

# Simulate regression r times and stores the results as a list
mcmc.lm.dist <- function(r, n = 400, n.lm = 1,
   beta1 = 1, beta2 = 2, gamma1 = -.25, gamma2 = -.1,
   gamma3 = -.2, nD = NA, reg.rhs = c(NA, NA)) {
 Hloc <- lapply(1:r, function(i) data.frame(x = runif(n, -10, 10), y = runif(n, -10,
10)))
 betas <- sapply(Hloc, reg.lm.dist, n = n, n.lm = n.lm, nD = nD,
        beta1 = beta1, beta2 = beta2, gamma1 = gamma1, gamma2 = gamma2,
        gamma3 = gamma3, reg.rhs = reg.rhs)
coef.list <- lapply(seq_len(nrow(betas)), function(i) betas[i, ])
names(coef.list) <- rownames(betas)
return(coef.list)
}

# Set the working directory
setwd("D:/naresh/...")

# Load Required R-packages and source file
library(spatstat)
library(memisc)

#=====================================
```

```
#--------------------------------------
# Part A: Create Illustrative Graph
# --------------------------------
```

```
# Set seed to make it reproducible
set.seed(5125)
```

```
# The following values are used for illustrative graph
```

```
# Number of house locations
n <- 10
```

```
# House Locations randomly created over the space [(-2,-2), (2,2)]
Hloc <- data.frame(
 x = runif(n, -2, 2),
 y = runif(n, -2, 2))
```

```
# Landmark A locations
A <- data.frame(
 x = c(-1, -1, 2),
 y = c(-1, 1, 1.5))
```

```
# Landmark B locations
B <- data.frame(
 x = c(-1, 0, 1),
 y = c(1, 0, -1))
```

```
# Landmark C locations
C <- data.frame(
 x = c(1, 0.5, -1.5),
 y = c(1, -2, -1))
```

```
#----------------------------------------------------------------------------
# Print all the graphs to a single file
png(file = paste0("../results/", "SimLoc11.png"),
 width = 1200, height = 800, res = 120)
```

```
# Set graphical parameter for plot region
op <- par(
 oma = c(3, 0, 3, 0),
 mfrow = c(1, 2)
)
```

```
# Plot house locations and unique landmark locations
plot(Hloc, xlab = "", ylab = "", pch = 21, bg = "darkgreen",
 axes = FALSE, xlim = c(-2,2), ylim = c(-2,2))
points(A[1, ], col = "red", cex = 1.5, pch = 17)
```

152

```
points(B[1, ], col = "blue", cex = 1.5, pch = 16)
points(C[1, ], col = "brown", cex = 2, pch = 18)
title("(a) Unique Landmarks")
text(A[1,1],A[1,2]+0.3,expression(A[1]))
text(B[1,1]-0.3,B[1,2],expression(B[1]))
text(C[1,1],C[1,2]+0.3,expression(C[1]))
box()

# Plot house locations and multiple landmark locations
plot(Hloc, xlab = "", ylab = "", pch = 21, bg = "darkgreen",
  axes = FALSE, xlim = c(-2,2), ylim = c(-2,2), bty = "l")
points(A, col = "red", cex = 1.5, pch = 17)
points(B, col = "blue", cex = 1.5, pch = 16)
points(C, col = "brown", cex = 2, pch = 18)
title(paste0("(b) Multiple Landmark Locations", "\n",
  "of a given type"))
text(A[1,1],A[1,2]+0.3,expression(A[1]))
text(B[1,1]-0.3,B[1,2],expression(B[1]))
text(C[1,1],C[1,2]+0.3,expression(C[1]))
text(A[2,1],A[2,2]+0.3,expression(A[2]))
text(B[2,1]-0.3,B[2,2],expression(B[2]))
text(C[2,1],C[2,2]+0.3,expression(C[2]))
text(A[3,1],A[3,2]+0.3,expression(A[3]))
text(B[3,1]-0.3,B[3,2],expression(B[3]))
text(C[3,1],C[3,2]+0.3,expression(C[3]))
box()

# Reset graphical parameters to allow writing title
par(op)

# Print title of the graph in the graphical device above the plot region
mtext("Illustration of House and Landmark locations", cex = 2,
  line = 1.5, font = 2)

# Reset graphical parameters to print legend
op <- par(usr = c(0,1,0,1), xpd = NA)

# Plot legend of the graph
legend(x = 0.5, y = 0,
  c("House-locations", "Landmarks A", "Landmarks B", "Landmarks C"),
  pch = c(21, 17, 16, 18), pt.bg = "darkgreen",
  pt.cex = c(1, 1.5, 1.5, 2),
  col = c("black", "red", "blue", "brown"),
  yjust = 1, xjust = 0.5, xpd = NA, horiz = TRUE
)

# Close the device, so the graph is saved as a png file. (Figure 4.1)
```

```
dev.off()

# Reset seed
rm(.Random.seed)

# End Part A

# =========================
# Part B: Monte Carlo Simulation
# ------------------------------

# Choose parameters to generate data for the hedonic specification
#
# P = b1 + b2*x + g1*dA + g2*dB + g3*dC + e
#
# where P = House Price
#      x = random variable x ~ N(0,1)
#     dA = distance to the nearest A
#     dB = distance to the nearest B
#     dC = distance to the nearest C
#      e = random error term, e ~ N(0,2)

mcmc.out.sum <- function(mcmc.out){

  out <- sapply(mcmc.out, mean)
  out.coef <- out[! names(out) %in% c("r2", "n", "k")]
  n <- out[["n"]]
  k <- out[["k"]]
  out.se <- sapply(mcmc.out, sd)
  outcoef.se <- out.se[! names(out) %in% c("r2", "n", "k")]
  out.t <- out.coef/outcoef.se
  out.p <- 2 * pt(abs(out.t), df = n-k, lower.tail = F)
  r2 <- out[["r2"]]
  df = data.matrix(data.frame(out.coef, outcoef.se,
    out.t, out.p))
  colnames(df) <- c("Coef", "SE", "z", "Pr(>|z|)")
  reslist <- list(df = df, r2 = r2, n = n)
  class(reslist) <- "mcmc"
  return(reslist)
}

getSummary.mcmc <-
 function(obj, alpha = 0.05, ...)
{
  smry <- obj
  coef <- smry$df
   lower <- qnorm(p = alpha/2, mean = coef[, 1], sd = coef[, 2])
```

```
  upper <- qnorm(p = 1 - alpha/2, mean = coef[, 1], sd = coef[, 2])
  coef <- cbind(coef, lower, upper)
  colnames(coef) <- c("est", "se", "stat", "p", "lwr", "upr")
  n <- smry$n
  r2 <- smry$r2
  sumstat <- c("N" = n, "R-Squared" = r2)
  list(coef = coef, sumstat = sumstat)
 }
setSummaryTemplate(mcmc = c("R Squared" = "($R-Squared:f)", "N" =
"($N:d)"))

b1 <- 1
b2 <- 2
g1 <- -0.25
g2 <- -0.10
g3 <- -0.2

# ==========
mcmc.out1 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 1, nD = 1,
  reg.rhs = c(NA, 1),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out1.11 <- mcmc.out.sum(mcmc.out1)

mcmc.out2 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 1, nD = 1,
  reg.rhs = c(1, NA),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out2.11 <- mcmc.outsum(mcmc.out2)

mcmc.out3 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 1, nD = 1,
  reg.rhs = c(2, NA),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out3 <- mcmc.out.sum(mcmc.out3)
out3.11 <- mcmc.out.sum(mcmc.out3)

mcmc.out4 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 1, nD = 1,
  reg.rhs = c(3, NA),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out4.11 <- mcmc.out.sum(mcmc.out4)

mcmc.out5 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 1, nD = 1,
  reg.rhs = c(3, 1),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out5.11 <- mcmc.out.sum(mcmc.out5)

tab1nlm1 <- mtable("del" = out5.11, "TD1-1" = out1.11, "TD1-2" = out2.11, "TD1-3"
= out3.11,
  "TD1-4" = out4.11, "TD1-5" = out5.11); tab1nlm1
```

```
#----------------------------

mcmc.out1 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 1, nD = 2,
  reg.rhs = c(NA, 1),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out1.21 <- mcmc.out.sum(mcmc.out1)

mcmc.out2 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 1, nD = 2,
  reg.rhs = c(1, NA),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out2.21 <- mcmc.out.sum(mcmc.out2)

mcmc.out3 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 1, nD = 2,
  reg.rhs = c(2, NA),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out3.21 <- mcmc.out.sum(mcmc.out3)

mcmc.out4 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 1, nD = 2,
  reg.rhs = c(3, NA),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out4.21 <- mcmc.out.sum(mcmc.out4)

mcmc.out5 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 1, nD = 2,
  reg.rhs = c(3, 1),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out5.21 <- mcmc.out.sum(mcmc.out5)

tab2nlm1 <- mtable("del" = out5.21,"TD2-1" = out1.21, "TD2-2" = out2.21, "TD2-
3" = out3.21,
  "TD2-4" = out4.21, "TD2-5" = out5.21); tab2nlm1

#-----------------------------------------

mcmc.out1 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 1, nD = 3,
  reg.rhs = c(NA, 1),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out1.31 <- mcmc.out.sum(mcmc.out1)

mcmc.out2 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 1, nD = 3,
  reg.rhs = c(1, NA),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out2.31 <- mcmc.out.sum(mcmc.out2)

mcmc.out3 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 1, nD = 3,
  reg.rhs = c(2, NA),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
```

```
out3.31 <- mcmc.out.sum(mcmc.out3)

mcmc.out4 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 1, nD = 3,
  reg.rhs = c(3, NA),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out4.31 <- mcmc.out.sum(mcmc.out4)

mcmc.out5 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 1, nD = 3,
  reg.rhs = c(3, 1),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out5.31 <- mcmc.out.sum(mcmc.out5)

tab3nlm1 <- mtable("del" = out5.31,"TD3-1" = out1.31, "TD3-2" = out2.31, "TD3-
3" = out3.31,
  "TD3-4" = out4.31, "TD3-5" = out5.31); tab3nlm1


#======================================

mcmc.out1 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 3, nD = 1,
  reg.rhs = c(NA, 1),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out1.15 <- mcmc.out.sum(mcmc.out1)

mcmc.out2 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 3, nD = 1,
  reg.rhs = c(1, NA),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out2.15 <- mcmc.out.sum(mcmc.out2)

mcmc.out3 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 3, nD = 1,
  reg.rhs = c(2, NA),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out3.15 <- mcmc.out.sum(mcmc.out3)

mcmc.out4 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 3, nD = 1,
  reg.rhs = c(3, NA),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out4.15 <- mcmc.out.sum(mcmc.out4)

mcmc.out5 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 3, nD = 1,
  reg.rhs = c(3, 1),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out5.15 <- mcmc.out.sum(mcmc.out5)

tab1nlm5 <- mtable("del" = out5.15, "TD1-1" = out1.15, "TD1-2" = out2.15, "TD1-
3" = out3.15,
  "TD1-4" = out4.15, "TD1-5" = out5.15); tab1nlm5
```

```
#-----------------------------------------
mcmc.out1 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 3, nD = 2,
  reg.rhs = c(NA, 1),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out1.25 <- mcmc.out.sum(mcmc.out1)

mcmc.out2 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 3, nD = 2,
  reg.rhs = c(1, NA),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out2.25 <- mcmc.out.sum(mcmc.out2)

mcmc.out3 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 3, nD = 2,
  reg.rhs = c(2, NA),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out3.25 <- mcmc.out.sum(mcmc.out3)

mcmc.out4 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 3, nD = 2,
  reg.rhs = c(3, NA),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out4.25 <- mcmc.out.sum(mcmc.out4)

mcmc.out5 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 3, nD = 2,
  reg.rhs = c(3, 1),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out5.25 <- mcmc.out.sum(mcmc.out5)

tab2nlm5 <- mtable("del" = out5.25, "TD2-1" = out1.25, "TD2-2" = out2.25,
"TD2-3" = out3.25,
  "TD2-4" = out4.25, "TD2-5" = out5.25); tab2nlm5

#-----------------------------------------

mcmc.out1 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 3, nD = 3,
  reg.rhs = c(NA, 1),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out1.35 <- mcmc.out.sum(mcmc.out1)

mcmc.out2 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 3, nD = 3,
  reg.rhs = c(1, NA),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out2.35 <- mcmc.out.sum(mcmc.out2)

mcmc.out3 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 3, nD = 3,
  reg.rhs = c(2, NA),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
```

```
out3.35 <- mcmc.out.sum(mcmc.out3)

mcmc.out4 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 3, nD = 3,
  reg.rhs = c(3, NA),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out4.35 <- mcmc.out.sum(mcmc.out4)

mcmc.out5 <- mcmc.lm.dist(r = 10000, n = 400, n.lm = 3, nD = 3,
  reg.rhs = c(3, 1),
  beta1 = b1, beta2 = b2, gamma1 = g1, gamma2 = g2, gamma3 = g3)
out5.35 <- mcmc.out.sum(mcmc.out5)

tab3nlm5 <- mtable("del" = out5.35, "TD3-1" = out1.35, "TD3-2" = out2.35,
  "TD3-3" = out3.35, "TD3-4" = out4.35, "TD3-5" = out5.35); tab3nlm5


# =====================

var.label11 <- c("(Intercept)" = "Constant",
  "min.DA" = "$D^A$",
  "min.DB" = "$d^a$",
  "min.DC" = "$d^b$",
  "min.da" = "$d^c$",
  "min.db" = "$d^d$",
  "min.dc" = "$d^e$")

var.label12 <- c("(Intercept)" = "Constant",
  "min.DA" = "$D^A$",
  "min.DB" = "$D^B$",
  "min.DC" = "$d^a$",
  "min.da" = "$d^b$",
  "min.db" = "$d^c$",
  "min.dc" = "$d^d$")

var.label13 <- c("(Intercept)" = "Constant",
  "min.DA" = "$D^A$",
  "min.DB" = "$D^B$",
  "min.DC" = "$D^C$",
  "min.da" = "$d^a$",
  "min.db" = "$d^b$",
  "min.dc" = "$d^c$")

var.label51 <- c("(Intercept)" = "Constant",
  "min.DA" = "$D^{nA}$",
  "min.DB" = "$d^{na}$",
  "min.DC" = "$d^{nb}$",
  "min.da" = "$d^{nc}$",
  "min.db" = "$d^{nd}$",
```

```
  "min.dc" = "$d^{ne}$")

var.label52 <- c("(Intercept)" = "Constant",
  "min.DA" = "$D^{nA}$",
  "min.DB" = "$D^{nB}$",
  "min.DC" = "$d^{na}$",
  "min.da" = "$d^{nb}$",
  "min.db" = "$d^{nc}$",
  "min.dc" = "$d^{nd}$")

var.label53 <- c("(Intercept)" = "Constant",
  "min.DA" = "$D^{nA}$",
  "min.DB" = "$D^{nB}$",
  "min.DC" = "$D^{nC}$",
  "min.da" = "$d^{na}$",
  "min.db" = "$d^{nb}$",
  "min.dc" = "$d^{nc}$")

(tab1nlm1 <- relabel(tab1nlm1, var.label11))
(tab1nlm5 <- relabel(tab1nlm5, var.label51))
(tab2nlm1 <- relabel(tab2nlm1, var.label12))
(tab2nlm5 <- relabel(tab2nlm5, var.label52))
(tab3nlm1 <- relabel(tab3nlm1, var.label13))
(tab3nlm5 <- relabel(tab3nlm5, var.label53))

write.mtable(tab1nlm1, file = "./sim-text-files/tab1nlm1.txt")
write.mtable(tab2nlm1, file = "./sim-text-files/tab2nlm1.txt")
write.mtable(tab3nlm1, file = "./sim-text-files/tab3nlm1.txt")
write.mtable(tab1nlm5, file = "./sim-text-files/tab1nlm5.txt")
write.mtable(tab2nlm5, file = "./sim-text-files/tab2nlm5.txt")
write.mtable(tab3nlm5, file = "./sim-text-files/tab3nlm5.txt")

doc.file.fun <- function(txt, path){
  tab <- read.table(paste0(path, txt, ".txt"), header = T, sep = "\t")
  doc <- RTF(paste0(path, txt, ".doc"))
  addTable(doc, tab)
  done(doc)
}

# Create Tables 3.1 − 3.6
doc.file.fun(txt = "tab1nlm1", path = "./sim-text-files/")
doc.file.fun(txt = "tab1nlm5", path = "./sim-text-files/")
doc.file.fun(txt = "tab2nlm1", path = "./sim-text-files/")
doc.file.fun(txt = "tab2nlm5", path = "./sim-text-files/")
doc.file.fun(txt = "tab3nlm1", path = "./sim-text-files/")
doc.file.fun(txt = "tab3nlm5", path = "./sim-text-files/")
```

```
########################
########################
R Codes used in Chapter 4
########################
########################

# Load required packages
library(sp, quietly = TRUE)
library(rgeos)
library(maptools); gpclibPermit()
library(spgwr, quietly = TRUE)
library(memisc)
library(rtf)
library(car)
library(doBy)
library(sphet)
library(rgdal)
library(spdep)
library(RANN)
library(lmtest)
library(sandwich)
library(MSBVAR)
library(ggplot2)
library(gridExtra)


# Set working directory
setwd("D:/naresh/...")
getwd() # check working directory

# Source created R functions

##############################
# ------------------Functions-----------------
##############################

# calc moran stat for lm model
moran.lm.stat <- function(model) {
        beta <- model[["coefficients"]]
        N <- length(model$model[[1]])
        eps <- resid(model)
         sigma.sq <- (t(eps) %*% eps) / N

        yhat <- X.mat %*% beta
        Ystar.mat <- model[["y"]]
        J <- (1/(N * sigma.sq)) *
        (t(W.mat %*% yhat) %*% M.mat %*% (W.mat %*% yhat) +
```

161

```
  Tr * sigma.sq)
      eWe <- t(eps) %*% W.mat %*% eps


      moran.I <- (eWe)/(t(eps) %*% eps)
  LM.error <- (eWe/sigma.sq)^2 / Tr
  LM.lag <- (1/(N*J)) *
   ((t(eps) %*% W.mat %*% (Ystar.mat)) / sigma.sq)^2

  LM.error.star <- ((eWe / sigma.sq) -
      (Tr / (N * J)) *
    (t(eps) %*% W.mat %*% Ystar.mat) / sigma.sq)^2 /
      (Tr * (1 - Tr / (N * J)))
  LM.lag.star <-
    (((t(eps) %*% W.mat %*% Ystar.mat) / sigma.sq) -
      (eWe / sigma.sq))^2 /
            (N * J - Tr)
  return(list(moran.I = moran.I[1,1], LM.error = LM.error[1,1],
    LM.lag = LM.lag[1,1],
        LM.error.star = LM.error.star[1,1],
    LM.lag.star = LM.lag.star[1,1]))
}

# re-sample error and apply wild bootstrap for lm model
b.moran.lm <- function(estar, r) {
 e.hat.star <- lapply(1:r,
   function(i) sample(estar, N, replace = T))
 r.moran <- sapply(e.hat.star, b.moran.lm.stat)
 list(moran.I = as.numeric(r.moran[1,]),
   LM.error = as.numeric(r.moran[2,]),
   LM.lag = as.numeric(r.moran[3,]),
   LM.error.star = as.numeric(r.moran[4,]),
   LM.lag.star = as.numeric(r.moran[5, ]))
}

# wild bootstrap for lm model
b.moran.lm.stat <- function(eps) {
 upsilon <- sample(c(-1, 1), N, replace = TRUE)
 e.tilde <- eps * upsilon
 z.star <- X.mat %*% beta.lm + e.tilde
 y.tilde <- z.star
 lm.model <- lm(y.tilde[, 1] ~ X.mat[, -1], y=TRUE)
 moran.lm.stat(lm.model)
}

# Load data (data.sp2 is the SpatialPointsDataFrame with
# features representing each house location, created in
# ESRI's ArcGIS)
```

```
load("DataSP.RData")

# Formulas for different models
fm1 <- (lnsalesPrice ~ parcelSize + storyHeight +
  lnTotalFlArea + houseAge + elevationkm)
fm2 <- update(fm1, ~ . + urban + lnPop10SqMi +
  whiteP + meanTestScoreIndex)
fm3 <- update(fm2, ~ . + waterEucMile + trailEucMile +
  parkEucMile)
fm4 <- update(fm3, ~ . + roadEucMile + forestEucMile +
  wilderEucMile)
fm6 <- update(fm3, ~ . + majorRoadNetworkMile +
  forestNetworkMile + wilderNetworkMile)



# Create neighbor object from SpatialPointDataFrame
# nearest 6 neighbors
yava.nb5 <- knn2nb(knearneigh(data.sp2, k = 6))

# Coordinates of data, IDs
coords <- coordinates(data.sp2)


#-------------------------------------------------
# Weight lists (row standardised and binary)
# weight for nearest 5 neighbors
yava.lw.w <- nb2listw(yava.nb5, style="W")

# OLS Regressions
ols1 <- lm(fm4, data=data.sp2)
ols2 <- lm(fm6, data=data.sp2)

# Spatial 2SLS Models
sp2sls.m1 <- stsls(fm4,
              data = data.sp2, listw = yava.lw.w, robust = T)
sp2sls.m2 <- stsls(fm6,
              data = data.sp2, listw = yava.lw.w, robust = T)



#===========================================================
====
# GM method (sphet) to estimate spatial models
# with heteroskedastic innovation
#---

# Create distance
id <- seq(1: nrow(coords))
tmp <- distance(coords, region.id=id, output = TRUE,
```

```
    type = "NN", nn = 6, shape.name = "shapefile",
    region.id.name = "id",
    firstline = TRUE, file.name = "yava_nn_5.GWT")
coldist <- read.gwt2dist(file = "yava_nn_5.GWT",
    region.id = id, skip = 1)

# STLS estimation with HAC
sphet.m1 <- stslshac(fm4,
      data = data.sp2, listw = yava.lw.w,
      distance = coldist, HAC = TRUE,
        type = "Triangular")
sphet.m2 <- stslshac(fm6,
      data = data.sp2, listw = yava.lw.w,
      distance = coldist, HAC = TRUE,
        type = "Triangular")

# General Spatial 2SLS (spdep)
gstls.m1 <- gstsls(fm4,
      data = data.sp2, listw = yava.lw.w)
gstls.m2 <- gstsls(fm6,
      data = data.sp2, listw = yava.lw.w)

# General Spatial 2SLS-HET (sphet)
gstlshet.m1 <- gstslshet(fm4,
      data = data.sp2, listw = yava.lw.w, initial.value = 0.2)
gstlshet.m2 <- gstslshet(fm6,
      data = data.sp2, listw = yava.lw.w, initial.value = 0.2)

# Write the regression results into a file:
options(signif.symbols = c("***" = 0.01, "**" = 0.05, "*" = 0.1))
M1Table <- mtable("OLS" = ols1, "S2SLS" = sp2sls.m1,
  "S2SLS-HAC" = sphet.m1, "GSTSLS" = gstls.m1,
  "GSTSLS-HET" = gstlshet.m1, summary.stats=FALSE)

label.var <- c("(Intercept)" = "Constant","parcelSize" = "ParcelSize",
"storyHeight" = "Height",
"lnTotalFlArea" = "FlArea",
"houseAge" = "HAge",
"urban" = "Urban",
"elevationkm" = "Elevation",
"lnPop10SqMi" = "PopDensity",
"whiteP" = "White",
"meanTestScoreIndex" = "TestScoreIndex",
"waterEucMile" = "WaterDist",
"trailEucMile" = "TrailDist",
"parkEucMile" = "ParkDist",
"roadEucMile" = "RoadDist",
```

```
"forestEucMile" = "ForestDist",
"wilderEucMile" = "WilderDist",
"wilderNetworkMile" = "WilderDDist",
"forestNetworkMile" = "ForestDDist",
"majorRoadNetworkMile" = "RoadDDist")
M1Table <- relabel(M1Table,label.var)

M2Table <- mtable("OLS" = ols2, "S2SLS" = sp2sls.m2,
  "S2SLS-HAC" = sphet.m2, "GSTSLS" = gstls.m2,
  "GSTSLS-HET" = gstlshet.m2, summary.stats=FALSE)
M2Table <- relabel(M2Table, label.var)

#------
#===========================================

# Mean Sales Price
Pbar <- mean(data.sp2$salesPrice)

# Direct and indirect effects m1 (EucMile)
Im.stslsm1 <- impacts(sp2sls.m1, listw = yava.lw.w)
Im.stslsm2 <- impacts(sp2sls.m2, listw = yava.lw.w)
Im.sphetm1 <- impacts(sphet.m1, listw = yava.lw.w)
Im.sphetm2 <- impacts(sphet.m2, listw = yava.lw.w)
Im.gstlsm1 <- impacts(gstls.m1, listw = yava.lw.w)
Im.gstlsm2 <- impacts(gstls.m2, listw = yava.lw.w)
Im.gstlshetm1 <- impacts(gstlshet.m1, listw = yava.lw.w)
Im.gstlshetm2 <- impacts(gstlshet.m2, listw = yava.lw.w)

# Data Frame
Imdf.stslsm1 <- data.frame(attr(Im.stslsm1,"bnames"),Im.stslsm1[1:3])
Imdf.stslsm2 <- data.frame(attr(Im.stslsm2,"bnames"),Im.stslsm2[1:3])
Imdf.sphetm1 <- data.frame(attr(Im.sphetm1,"bnames"),Im.sphetm1[1:3])
Imdf.sphetm2 <- data.frame(attr(Im.sphetm2,"bnames"),Im.sphetm2[1:3])
Imdf.gstlsm1 <- data.frame(attr(Im.gstlsm1,"bnames"),Im.gstlsm1[1:3])
Imdf.gstlsm2 <- data.frame(attr(Im.gstlsm2,"bnames"),Im.gstlsm2[1:3])
Imdf.gstlshetm1 <- data.frame(attr(Im.gstlshetm1,"bnames"),Im.gstlshetm1[1:3])
Imdf.gstlshetm2 <-
data.frame(attr(Im.gstlshetm2,"bnames"),Im.gstlshetm2[1:3])

# Direct Effects data frame
DirectEffects <- data.frame(row.names = Imdf.stslsm1[, 1])
DirectEffects$S2SLS <- round(Imdf.stslsm1[, 2] * Pbar, 0)
DirectEffects$S2SLSHAC <- round(Imdf.sphetm1[, 2] * Pbar, 0)
DirectEffects$GSTSLS <- round(Imdf.gstlsm1[, 2] * Pbar, 0)
DirectEffects$GSTSLSHET <- round(Imdf.gstlshetm1[, 2] * Pbar, 0)

# InDirect Effects data frame
```

```
InDirectEffects <- data.frame(row.names = Imdf.stslsm1[, 1])
InDirectEffects$S2SLS <- round(Imdf.stslsm1[, 3] * Pbar, 0)
InDirectEffects$S2SLSHAC <- round(Imdf.sphetm1[, 3] * Pbar, 0)
InDirectEffects$GSTSLS <- round(Imdf.gstlsm1[, 3] * Pbar, 0)
InDirectEffects$GSTSLSHET <- round(Imdf.gstlshetm1[, 3] * Pbar, 0)

# Total Effects data frame
TotalEffects <- data.frame(row.names = Imdf.stslsm1[, 1])
TotalEffects$OLS <- round(ols1$coef[-1]*Pbar)
TotalEffects$S2SLS <- round(Imdf.stslsm1[, 4] * Pbar, 0)
TotalEffects$S2SLSHAC <- round(Imdf.sphetm1[, 4] * Pbar, 0)
TotalEffects$GSTSLS <- round(Imdf.gstlsm1[, 4] * Pbar, 0)
TotalEffects$GSTSLSHET <- round(Imdf.gstlshetm1[, 4] * Pbar, 0)

rownames(TotalEffects) <- label.var[which(names(label.var) %in%
rownames(TotalEffects))]
rownames(DirectEffects) <- label.var[which(names(label.var) %in%
rownames(DirectEffects))]
rownames(InDirectEffects) <- label.var[which(names(label.var) %in%
rownames(InDirectEffects))]


##############################################

# Krinsky and Robb method for se of Direct Marginal Effects
# OLS model
ols1.beta <- ols1$coefficients
ols1.vcov <- vcov(ols1)
ols1beta.sim <- rmultnorm(10000, mu = ols1.beta, vmat = ols1.vcov)
ols1.mwtp <- ols1beta.sim * Pbar
ols1.ci <- apply(ols1.mwtp, 2, quantile, c(0.025, 0.975))
ols1.cif <- round(ols1.ci[,-1],0)

# S2SLS
sp2slsm1.beta <- coef(sp2sls.m1)
sp2slsm1.vcov <- summary(sp2sls.m1)$var
sp2slsm1beta.sim <- rmultnorm(10000, mu = sp2slsm1.beta, vmat =
sp2slsm1.vcov)
sp2slsm1.mwtp <- sp2slsm1beta.sim * Pbar
sp2slsm1.ci <- apply(sp2slsm1.mwtp, 2, quantile, c(0.025, 0.975))
sp2slsm1.cif <- round(sp2slsm1.ci[,c(-1,-2)])

# S2SLSHAC
sp2slsm1hac.beta <- coef(sphet.m1)
sp2slsm1hac.vcov <- summary(sphet.m1)$var
sp2slsm1hacbeta.sim <- rmultnorm(10000, mu = sp2slsm1hac.beta, vmat =
sp2slsm1hac.vcov)
```

```r
sp2slsm1hac.mwtp <- sp2slsm1hacbeta.sim * Pbar
sp2slsm1hac.ci <- apply(sp2slsm1hac.mwtp, 2, quantile, c(0.025, 0.975))
sp2slsm1hac.cif <- round(sp2slsm1hac.ci[,c(-1,-2)],0)

# GSTSLS
gstlsm1.beta <- coef(gstls.m1)[-length(coef(gstls.m1))]
gstlsm1.vcov <- gstls.m1$secstep_var
gstlsm1beta.sim <- rmultnorm(10000, mu = gstlsm1.beta, vmat = gstlsm1.vcov)
gstlsm1.mwtp <- gstlsm1beta.sim * Pbar
gstlsm1.ci <- apply(gstlsm1.mwtp, 2, quantile, c(0.025, 0.975))
gstlsm1.cif <- round(gstlsm1.ci[,c(-1,-2)],0)

# GSTSLSHET
gstlsm1het.beta <- coef(gstlshet.m1)
gstlsm1het.vcov <- summary(gstlshet.m1)$var
gstlsm1het.sim <- rmultnorm(10000, mu = gstlsm1het.beta, vmat =
gstlsm1het.vcov)
gstlsm1het.mwtp <- gstlsm1het.sim * Pbar
gstlsm1het.ci <- apply(gstlsm1het.mwtp, 2, quantile, c(0.025, 0.975))
gstlsm1het.cif <- round(gstlsm1het.ci[,c(-1,-17,-18)],0)

# Print single model
cat("\n",
  paste(DirectEffects$GSTSLSHET,"\n","[",round(gstlsm1het.cif[1,],0),", ",
    round(gstlsm1het.cif[2,],0),"]","\n", sep = ""), "\n")

# Print Direct Margina Effects and confidence interval for all models
cat(
  paste("Variables", "\t", "OLS", "\t", "S2SLS", "\t", "S2SLSHAC",
    "\t", "GSTSLS", "\t", "GSTSLSHET", sep = ""), "\n",
  paste(rownames(TotalEffects), "\t", TotalEffects$OLS, "\t",
    DirectEffects$S2SLS, "\t", DirectEffects$S2SLSHAC, "\t",
    DirectEffects$GSTSLS, "\t", DirectEffects$GSTSLSHET, "\n", "\t",
    "[", ols1.cif[1,], ", ", ols1.cif[2,], "]", "\t",
    "[", sp2slsm1.cif[1,],", ", sp2slsm1.cif[2,],"]", "\t",
    "[", sp2slsm1hac.cif[1,],", ", sp2slsm1hac.cif[2,],"]", "\t",
    "[", gstlsm1.cif[1,], ", ", gstlsm1.cif[2,], "]", "\t",
    "[", gstlsm1het.cif[1,], ", ", gstlsm1het.cif[2,], "]",
    "\n", sep = ""),
  file="DirectEffects.txt")

DirectDF <- read.table("DirectEffects.txt", header = TRUE, sep="\t")


### ------------------------------------------------------------
###
###
# Krinsky and Robb method for se of Indirect Marginal Effects
```

```r
# OLS model nothing difference
ols1.beta <- ols1$coefficients
ols1.vcov <- vcov(ols1)
ols1beta.sim <- rmultnorm(10000, mu = ols1.beta, vmat = ols1.vcov)
ols1.mwtp <- ols1beta.sim * Pbar
ols1.ci <- apply(ols1.mwtp, 2, quantile, c(0.025, 0.975))
ols1.cif <- round(ols1.ci[,-1],0)

# S2SLS
sp2slsm1.beta <- coef(sp2sls.m1)
sp2slsm1.vcov <- summary(sp2sls.m1)$var
sp2slsm1beta.sim <- rmultnorm(10000, mu = sp2slsm1.beta, vmat =
sp2slsm1.vcov)
sp2slsm1.mwtpin <- sp2slsm1beta.sim * Pbar * (1/(1-sp2slsm1.beta[1])-1)
sp2slsm1.ciin <- apply(sp2slsm1.mwtpin, 2, quantile, c(0.025, 0.975))
sp2slsm1.cifin <- round(sp2slsm1.ciin[,c(-1,-2)])

# S2SLSHAC
sp2slsm1hac.beta <- coef(sphet.m1)
sp2slsm1hac.vcov <- summary(sphet.m1)$var
sp2slsm1hacbeta.sim <- rmultnorm(10000, mu = sp2slsm1hac.beta, vmat =
sp2slsm1hac.vcov)
sp2slsm1hac.mwtpin <- sp2slsm1hacbeta.sim * Pbar * (1/(1-
sp2slsm1hac.beta[1])-1)
sp2slsm1hac.ciin <- apply(sp2slsm1hac.mwtpin, 2, quantile, c(0.025, 0.975))
sp2slsm1hac.cifin <- round(sp2slsm1hac.ciin[,c(-1,-2)],0)

# GSTSLS
gstlsm1.beta <- coef(gstls.m1)[-length(coef(gstls.m1))]
gstlsm1.vcov <- gstls.m1$secstep_var
gstlsm1beta.sim <- rmultnorm(10000, mu = gstlsm1.beta, vmat = gstlsm1.vcov)
gstlsm1.mwtpin <- gstlsm1beta.sim * Pbar * (1/(1-gstlsm1.beta[1])-1)
gstlsm1.ciin <- apply(gstlsm1.mwtpin, 2, quantile, c(0.025, 0.975))
gstlsm1.cifin <- round(gstlsm1.ciin[,c(-1,-2)],0)

# GSTSLSHET
gstlsm1het.beta <- coef(gstlshet.m1)
gstlsm1het.vcov <- summary(gstlshet.m1)$var
gstlsm1het.sim <- rmultnorm(10000, mu = gstlsm1het.beta, vmat =
gstlsm1het.vcov)
gstlsm1het.mwtpin <- gstlsm1het.sim * Pbar * (1/(1-
gstlsm1het.beta[length(gstlsm1het.beta)-1])-1)
gstlsm1het.ciin <- apply(gstlsm1het.mwtpin, 2, quantile, c(0.025, 0.975))
gstlsm1het.cifin <- round(gstlsm1het.ciin[,c(-1,-17, -18)],0)

# Print single model
cat("\n",
```

```
  paste(InDirectEffects$GSTSLSHET,"\n","[",round(gstlsm1het.cifin[1,],0),", ",
    round(gstlsm1het.cifin[2,],0),"]","\n", sep = ""), "\n")

# Print Direct Margina Effects and confidence interval for all models
cat(
  paste("Variables", "\t", "OLS", "\t", "S2SLS", "\t", "S2SLSHAC",
    "\t", "GSTSLS", "\t", "GSTSLSHET", sep = ""), "\n",
  paste(rownames(TotalEffects), "\t", TotalEffects$OLS, "\t",
    InDirectEffects$S2SLS, "\t", InDirectEffects$S2SLSHAC, "\t",
     InDirectEffects$GSTSLS, "\t", InDirectEffects$GSTSLSHET, "\n", "\t",
    "[", ols1.cif[1,], ", ", ols1.cif[2,], "]", "\t",
    "[", sp2slsm1.cifin[1,],", ", sp2slsm1.cifin[2,],"]", "\t",
    "[", sp2slsm1hac.cifin[1,],", ", sp2slsm1hac.cifin[2,],"]", "\t",
    "[", gstlsm1.cifin[1,], ", ", gstlsm1.cifin[2,], "]", "\t",
    "[", gstlsm1het.cifin[1,], ", ", gstlsm1het.cifin[2,], "]",
    "\n", sep = ""),
  file="InDirectEffects.txt")

InDirectDF <- read.table("InDirectEffects.txt", header = TRUE, sep="\t")
InDirectDF


rtf <- RTF("MWTP.doc", width = 8.5, height = 11, font.size = 12, omi = c(1,1,1,1))
addHeader(rtf, "Table : Marginal Effects for different estimators (Direct Effects)",
  font.size = 12)
addTable(rtf, DirectDF, font.size=12)
addPageBreak(rtf)

addHeader(rtf, "Table : Marginal Effects for different estimators (Indirect
Effects)",
  font.size = 12)
addTable(rtf, InDirectDF, font.size = 12)
done(rtf)

# ==============================
# ==============================

# Marginal Effects with Driving Distance

# Direct Effects data frame
DirectEffects2 <- data.frame(row.names = Imdf.stslsm2[, 1])
DirectEffects2$S2SLS <- round(Imdf.stslsm2[, 2] * Pbar, 0)
DirectEffects2$S2SLSHAC <- round(Imdf.sphetm2[, 2] * Pbar, 0)
DirectEffects2$GSTSLS <- round(Imdf.gstlsm2[, 2] * Pbar, 0)
DirectEffects2$GSTSLSHET <- round(Imdf.gstlshetm2[, 2] * Pbar, 0)

# InDirect Effects data frame
```

```
InDirectEffects2 <- data.frame(row.names = Imdf.stslsm2[, 1])
InDirectEffects2$S2SLS <- round(Imdf.stslsm2[, 3] * Pbar, 0)
InDirectEffects2$S2SLSHAC <- round(Imdf.sphetm2[, 3] * Pbar, 0)
InDirectEffects2$GSTSLS <- round(Imdf.gstlsm2[, 3] * Pbar, 0)
InDirectEffects2$GSTSLSHET <- round(Imdf.gstlshetm2[, 3] * Pbar, 0)

# Total Effects data frame
TotalEffects2 <- data.frame(row.names = Imdf.stslsm2[, 1])
TotalEffects2$OLS <- round(ols2$coef[-1]*Pbar)
TotalEffects2$S2SLS <- round(Imdf.stslsm2[, 4] * Pbar, 0)
TotalEffects2$S2SLSHAC <- round(Imdf.sphetm2[, 4] * Pbar, 0)
TotalEffects2$GSTSLS <- round(Imdf.gstlsm2[, 4] * Pbar, 0)
TotalEffects2$GSTSLSHET <- round(Imdf.gstlshetm2[, 4] * Pbar, 0)

rownames(TotalEffects2) <- label.var[which(names(label.var) %in%
rownames(TotalEffects2))]
rownames(DirectEffects2) <- label.var[which(names(label.var) %in%
rownames(DirectEffects2))]
rownames(InDirectEffects2) <- label.var[which(names(label.var) %in%
rownames(InDirectEffects2))]


#############################################
# Krinsky and Robb method for se of Direct Marginal Effects
# OLS model
ols2.beta <- ols2$coefficients
ols2.vcov <- vcov(ols2)
ols2beta.sim <- rmultnorm(10000, mu = ols2.beta, vmat = ols2.vcov)
ols2.mwtp <- ols2beta.sim * Pbar
ols2.ci <- apply(ols2.mwtp, 2, quantile, c(0.025, 0.975))
ols2.cif <- round(ols2.ci[,-1],0)

# S2SLS
sp2slsm2.beta <- coef(sp2sls.m2)
sp2slsm2.vcov <- summary(sp2sls.m2)$var
sp2slsm2beta.sim <- rmultnorm(10000, mu = sp2slsm2.beta, vmat =
sp2slsm2.vcov)
sp2slsm2.mwtp <- sp2slsm2beta.sim * Pbar
sp2slsm2.ci <- apply(sp2slsm2.mwtp, 2, quantile, c(0.025, 0.975))
sp2slsm2.cif <- round(sp2slsm2.ci[,c(-1,-2)])

# S2SLSHAC
sp2slsm2hac.beta <- coef(sphet.m2)
sp2slsm2hac.vcov <- summary(sphet.m2)$var
sp2slsm2hacbeta.sim <- rmultnorm(10000, mu = sp2slsm2hac.beta, vmat =
sp2slsm2hac.vcov)
sp2slsm2hac.mwtp <- sp2slsm2hacbeta.sim * Pbar
```

```
sp2slsm2hac.ci <- apply(sp2slsm2hac.mwtp, 2, quantile, c(0.025, 0.975))
sp2slsm2hac.cif <- round(sp2slsm2hac.ci[,c(-1,-2)],0)

# GSTSLS
gstlsm2.beta <- coef(gstls.m2)[-length(coef(gstls.m2))]
gstlsm2.vcov <- gstls.m2$secstep_var
gstlsm2beta.sim <- rmultnorm(10000, mu = gstlsm2.beta, vmat = gstlsm2.vcov)
gstlsm2.mwtp <- gstlsm2beta.sim * Pbar
gstlsm2.ci <- apply(gstlsm2.mwtp, 2, quantile, c(0.025, 0.975))
gstlsm2.cif <- round(gstlsm2.ci[,c(-1,-2)],0)

# GSTSLSHET
gstlsm2het.beta <- coef(gstlshet.m2)
gstlsm2het.vcov <- summary(gstlshet.m2)$var
gstlsm2het.sim <- rmultnorm(10000, mu = gstlsm2het.beta, vmat =
gstlsm2het.vcov)
gstlsm2het.mwtp <- gstlsm2het.sim * Pbar
gstlsm2het.ci <- apply(gstlsm2het.mwtp, 2, quantile, c(0.025, 0.975))
gstlsm2het.cif <- round(gstlsm2het.ci[,c(-1,-17,-18)],0)

# Print single model
cat("\n",
  paste(DirectEffects2$GSTSLSHET,"\n","[",round(gstlsm2het.cif[1,],0),", ",
    round(gstlsm2het.cif[2,],0),"]","\n", sep = ""), "\n")

# Print Direct Margina Effects and confidence interval for all models
cat(
  paste("Variables", "\t", "OLS", "\t", "S2SLS", "\t", "S2SLSHAC",
    "\t", "GSTSLS", "\t", "GSTSLSHET", sep = ""), "\n",
  paste(rownames(TotalEffects2), "\t", TotalEffects2$OLS, "\t",
    DirectEffects2$S2SLS, "\t", DirectEffects2$S2SLSHAC, "\t",
    DirectEffects2$GSTSLS, "\t", DirectEffects2$GSTSLSHET, "\n", "\t",
    "[", ols2.cif[1,], ", ", ols2.cif[2,], "]", "\t",
    "[", sp2slsm2.cif[1,],", ", sp2slsm2.cif[2,],"]", "\t",
    "[", sp2slsm2hac.cif[1,],", ", sp2slsm2hac.cif[2,],"]", "\t",
    "[", gstlsm2.cif[1,], ", ", gstlsm2.cif[2,], "]", "\t",
    "[", gstlsm2het.cif[1,], ", ", gstlsm2het.cif[2,], "]",
    "\n", sep = ""),
  file="DirectEffects2.txt")

DirectDF2 <- read.table("DirectEffects2.txt", header = TRUE, sep="\t")


### -------------------------------------------------------------
###
###
# Krinsky and Robb method for se of Indirect Marginal Effects
# OLS model nothing difference
```

```
ols2.beta <- ols2$coefficients
ols2.vcov <- vcov(ols2)
ols2beta.sim <- rmultnorm(10000, mu = ols2.beta, vmat = ols2.vcov)
ols2.mwtp <- ols2beta.sim * Pbar
ols2.ci <- apply(ols2.mwtp, 2, quantile, c(0.025, 0.975))
ols2.cif <- round(ols2.ci[,-1],0)

# S2SLS
sp2slsm2.beta <- coef(sp2sls.m2)
sp2slsm2.vcov <- summary(sp2sls.m2)$var
sp2slsm2beta.sim <- rmultnorm(10000, mu = sp2slsm2.beta, vmat =
sp2slsm2.vcov)
sp2slsm2.mwtpin <- sp2slsm2beta.sim * Pbar * (1/(1-sp2slsm2.beta[1])-1)
sp2slsm2.ciin <- apply(sp2slsm2.mwtpin, 2, quantile, c(0.025, 0.975))
sp2slsm2.cifin <- round(sp2slsm2.ciin[,c(-1,-2)])

# S2SLSHAC
sp2slsm2hac.beta <- coef(sphet.m2)
sp2slsm2hac.vcov <- summary(sphet.m2)$var
sp2slsm2hacbeta.sim <- rmultnorm(10000, mu = sp2slsm2hac.beta, vmat =
sp2slsm2hac.vcov)
sp2slsm2hac.mwtpin <- sp2slsm2hacbeta.sim * Pbar * (1/(1-
sp2slsm2hac.beta[1])-1)
sp2slsm2hac.ciin <- apply(sp2slsm2hac.mwtpin, 2, quantile, c(0.025, 0.975))
sp2slsm2hac.cifin <- round(sp2slsm2hac.ciin[,c(-1,-2)],0)

# GSTSLS
gstlsm2.beta <- coef(gstls.m2)[-length(coef(gstls.m2))]
gstlsm2.vcov <- gstls.m2$secstep_var
gstlsm2beta.sim <- rmultnorm(10000, mu = gstlsm2.beta, vmat = gstlsm2.vcov)
gstlsm2.mwtpin <- gstlsm2beta.sim * Pbar * (1/(1-gstlsm2.beta[1])-1)
gstlsm2.ciin <- apply(gstlsm2.mwtpin, 2, quantile, c(0.025, 0.975))
gstlsm2.cifin <- round(gstlsm2.ciin[,c(-1,-2)],0)

# GSTSLSHET
gstlsm2het.beta <- coef(gstlshet.m2)
gstlsm2het.vcov <- summary(gstlshet.m2)$var
gstlsm2het.sim <- rmultnorm(10000, mu = gstlsm2het.beta, vmat =
gstlsm2het.vcov)
gstlsm2het.mwtpin <- gstlsm2het.sim * Pbar * (1/(1-
gstlsm2het.beta[length(gstlsm2het.beta)-1])-1)
gstlsm2het.ciin <- apply(gstlsm2het.mwtpin, 2, quantile, c(0.025, 0.975))
gstlsm2het.cifin <- round(gstlsm2het.ciin[,c(-1,-17, -18)],0)

# Print single model
cat("\n",
 paste(InDirectEffects2$GSTSLSHET,"\n","[",round(gstlsm2het.cifin[1,],0),", ",
```

```
    round(gstlsm2het.cifin[2,],0),"]","]","\n", sep = ""), "\n")

# Print Direct Margina Effects and confidence interval for all models
cat(
  paste("Variables", "\t", "OLS", "\t", "S2SLS", "\t", "S2SLSHAC",
    "\t", "GSTSLS", "\t", "GSTSLSHET", sep = ""), "\n",
  paste(rownames(TotalEffects2), "\t", TotalEffects2$OLS, "\t",
    InDirectEffects2$S2SLS, "\t", InDirectEffects2$S2SLSHAC, "\t",
    InDirectEffects2$GSTSLS, "\t", InDirectEffects2$GSTSLSHET, "\n", "\t",
    "[", ols2.cif[1,], ", ", ols2.cif[2,], "]", "\t",
    "[", sp2slsm2.cifin[1,],", ", sp2slsm2.cifin[2,],"]", "\t",
    "[", sp2slsm2hac.cifin[1,],", ", sp2slsm2hac.cifin[2,],"]", "\t",
    "[", gstlsm2.cifin[1,], ", ", gstlsm2.cifin[2,], "]", "\t",
    "[", gstlsm2het.cifin[1,], ", ", gstlsm2het.cifin[2,], "]",
    "\n", sep = ""),
  file="InDirectEffects2.txt")

InDirectDF2 <- read.table("InDirectEffects2.txt", header = TRUE, sep="\t")

# End
```

References

Anselin, L., 1988. Spatial econometrics: methods and models. Kluwer Academic Publishers, Dordrecht; Boston.

Anselin, L., Lozano-Gracia, N., 2009. Errors in variables and spatial effects in hedonic house price models of ambient air quality, in: Arbia, G., Baltagi, B.H. (Eds.), Spatial Econometrics, Studies in Empirical Economics. Physica-Verlag HD, pp. 5–34.

Baerenklau, K.A., González-Cabán, A., Paez, C., Chavez, E., 2010. Spatial allocation of forest recreation value. Journal of Forest Economics 16, 113–126. doi:10.1016/j.jfe.2009.09.002

Bastian, C.T., McLeod, D.M., Germino, M.J., Reiners, W.A., Blasko, B.J., 2002. Environmental amenities and agricultural land values: a hedonic model using geographic information systems data. Ecological Economics 40, 337–349. doi:10.1016/S0921-8009(01)00278-6

Bowker, J.M., 2001. Outdoor recreation by Alaskans: projections for 2000 through 2020 (No. PNW-GTR-527), General Technical Report. US Department of Agriculture, Forest Service, Pacific Northwest Research Station, Portland, OR.

Brasington, D.M., Hite, D., 2005. Demand for environmental quality: a spatial hedonic analysis. Regional Science and Urban Economics 35, 57–82. doi:10.1016/j.regsciurbeco.2003.09.001

Chadwick, D.H., 2002. A beast the color of winter: the mountain goat observed. U of Nebraska Press.

Chay, K.Y., Greenstone, M., 2005. Does Air Quality Matter? Evidence from the Housing Market. Journal of Political Economy 113, 376–424. doi:10.1086/427462

Chib, S., Jeliazkov, I., 2001. Marginal Likelihood From the Metropolis–Hastings Output. Journal of the American Statistical Association 96, 270–281. doi:10.1198/016214501750332848

Cho, S.-H., Lambert, D.M., Kim, S.G., Roberts, R.K., Park, W.M., 2011. Relationship between value of open space and distance from housing locations within a community. Journal of Geographical Systems 13, 393–414. doi:10.1007/s10109-010-0126-4

Cooper, A.B., Millspaugh, J.J., 1999. The application of discrete choice models to wildlife resource selection studies. Ecology 80, 566–575.

Côté, S.D., Festa-Bianchet, M., 2001. Birthdate, mass and survival in mountain goat kids: effects of maternal characteristics and forage quality. Oecologia 127, 230–238. doi:10.1007/s004420000584

Dormann, C.F., 2007. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. Global Ecology and Biogeography 16, 129–138.

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

Duffy-Deno, K.T., 1998. The Effect of Federal Wilderness on County Growth in
the Intermountain Western United States. Journal of Regional Science 38,
109–136. doi:10.1111/0022-4146.00084

Fox, J.L., 1983. Constraints on winter habitat selection by the mountain goat
(Oreamnos americanus) in Alaska. University of Washington.

Fox, J.L., Smith, C.A., Schoen, J.W., 1989. Relation between mountain goats and
their habitat in southeastern Alaska.

Frair, J.L., Fieberg, J., Hebblewhite, M., Cagnacci, F., DeCesare, N.J., Pedrotti,
L., 2010. Resolving issues of imprecise and habitat-biased locations in
ecological analyses using GPS telemetry data. Phil. Trans. R. Soc. B 365,
2187–2200. doi:10.1098/rstb.2010.0084

Freeman, A.M., 2003. The Measurements of Environmental and Resource
Values: Theory and Methods. Resources for the Future.

Gross, J.E., Kneeland, M.C., Reed, D.F., Reich, R.M., 2002. GIS-based habitat
models for mountain goats. Journal of Mammalogy 83, 218–228.

Guimaraes, P., Figueirdo, O., Woodward, D., 2003. A tractable approach to the
firm location decision problem. Review of Economics and Statistics 85,
201–204.

Hand, M.S., Thacher, J.A., McCollum, D.W., Berrens, R.P., 2008a. Forest Amenities and Location Choice in the Southwest. Journal of Agricultural and Resource Economics 33, 232–253.

Hand, M.S., Thacher, J.A., McCollum, D.W., Berrens, R.P., 2008b. Intra-Regional Amenities, Wages, and Home Prices: The Role of Forests in the Southwest. Land Economics 84, 635–651. doi:10.3368/le.84.4.635

Hansen, R.M., Archer, S.R., 1981. Range survey of mountain goat wintering areas (No. 40-0120-0-290), Final Report. USDA Forest Service, Chugach National Forest, Seward, Alaska.

Harris, G., Nielson, R.M., Rinaldi, T., Lohuis, T., 2014. Effects of winter recreation on northern ungulates with focus on moose (Alces alces) and snowmobiles. Eur J Wildl Res 60, 45–58. doi:10.1007/s10344-013-0749-0

Harris, G.M., Russell, G.J., van Aarde, R.I., Pimm, S.L., 2008. Rules of habitat use by elephants Loxodonta africana in southern Africa: insights for regional management. Oryx 42, 66–75.

Hebblewhite, M., Haydon, D.T., 2010. Distinguishing technology from biology: a critical review of the use of GPS telemetry data in ecology. Phil. Trans. R. Soc. B 365, 2303–2312. doi:10.1098/rstb.2010.0087

Hebert, D.M., Turnbull, W.G., 1977. A description of southern interior and coastal mountain goat ecotypes in British Columbia, in: Proceedings of the First International Mountain Goat Symposium. pp. 126–146.

Izón, G.M., Hand, M.S., Fontenla, M., Berrens, R.P., 2010. The Economic Value of Protecting Inventoried Roadless Areas: A Spatial Hedonic Price Study in New Mexico. Contemporary Economic Policy 28, 537–553. doi:10.1111/j.1465-7287.2009.00190.x

Kelejian, H.H., Prucha, I.R., 1999. A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. International Economic Review 40, 509–533. doi:10.1111/1468-2354.00027

Kelejian, H.H., Prucha, I.R., 2007. HAC estimation in a spatial framework. Journal of Econometrics 140, 131–154. doi:10.1016/j.jeconom.2006.09.005

Kelejian, H.H., Prucha, I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics 157, 53–67. doi:10.1016/j.jeconom.2009.10.025

Kim, C.W., Phipps, T.T., Anselin, L., 2003. Measuring the benefits of air quality improvement: a spatial hedonic approach. Journal of Environmental Economics and Management 45, 24–39. doi:10.1016/S0095-0696(02)00013-X

Koschinsky, J., Lozano-Gracia, N., Piras, G., 2012. The welfare benefit of a home's location: an empirical comparison of spatial and non-spatial model estimates. Journal of Geographical Systems 14, 319–356. doi:10.1007/s10109-011-0148-6

Krinsky, I., Robb, A.L., 1986. On Approximating the Statistical Properties of Elasticities. The Review of Economics and Statistics 68, 715–19.

Lancaster, K.J., 1966. A New Approach to Consumer Theory. Journal of Political Economy 74, 132–157. doi:10.2307/1828835

Larson, E.K., Perrings, C., 2013. The value of water-related amenities in an arid city: The case of the Phoenix metropolitan area. Landscape and Urban Planning 109, 45–55. doi:10.1016/j.landurbplan.2012.10.008

Lee, L.-F., 2004. Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models. Econometrica 72, 1899–1925. doi:10.1111/j.1468-0262.2004.00558.x

Legendre, P., Fortin, M.J., 1989. Spatial pattern and ecological analysis. Vegetatio 80, 107–138.

LeSage, J., Pace, R.K., 2009. Introduction to Spatial Econometrics, 1st ed. Chapman and Hall/CRC.

Lichstein, J.W., Simons, T.R., Shriner, S.A., Franzreb, K.E., 2002. Spatial autocorrelation and autoregressive models in ecology. Ecological monographs 72, 445–463.

Lin, K.-P., Long, Z.-H., Ou, B., 2011. The Size and Power of Bootstrap Tests for Spatial Dependence in a Linear Regression Model. Computational Economics 38, 153–171. doi:10.1007/s10614-010-9224-0

Manly, B.F.J., McDonald, L.L., Thomas, D.L., McDonald, T.L., Erickson, W.P., 2002. Resource selection by animals: statistical analysis and design for field studies. Nordrecht, The Netherlands: Kluwer.

McDonough, T.J., Selinger, J.S., 2008. Mountain goat management on the Kenai Peninsula, Alaska: a new direction, in: Biennial Symposium of the Northern Wild Sheep and Goat Council. pp. 50–67.

Morton, P., 1998. Economic Benefits of Wilderness: Theory and Practice, The. Denver University Law Review 76, 465.

Neumann, B.C., Boyle, K.J., Bell, K.P., 2009. Property price effects of a national wildlife refuge: Great Meadows National Wildlife Refuge in Massachusetts. Land Use Policy 26, 1011–1019. doi:10.1016/j.landusepol.2008.12.008

Nielson, R.M., Manly, B.F., McDonald, L.L., Sawyer, H., McDonald, T.L., 2009. Estimating habitat selection when GPS fix success is less than 100%. Ecology 90, 2956–2962.

Nielson, R.M., Sawyer, H., 2013. Estimating resource selection with count data. Ecol Evol 3, 2233–2240. doi:10.1002/ece3.617

Noonan, D.S., Krupka, D.J., Baden, B.M., 2007. Neighborhood Dynamics and Price Effects of Superfund Site Clean-Up*. Journal of Regional Science 47, 665–692. doi:10.1111/j.1467-9787.2007.00525.x

Patric, J.H., Harbin, R.L., 1998. Whither wilderness? How much is enough? Heartland Policy Study 89.

Pereira, J.M.C., Itami, R.M., 1991. GIS-based habitat modeling using logistic multiple regression : a study of the Mt. Graham red squirrel. Photogrammetric engineering and remote sensing 57, 1475–1486.

Piras, G., 2010. sphet: Spatial Models with Heteroskedastic Innovations in R. Journal of Statistical Software 35, 1–21.

Poole, K.G., Heard, D.C., 2003. Seasonal habitat use and movements of mountain goats, Oreamnos americanus, in east-central British Columbia. The Canadian Field-Naturalist 117, 565–576.

Poole, K.G., Stuart-Smith, K., Teske, I.E., 2009. Wintering strategies by mountain goats in interior mountains. Canadian Journal of Zoology 87, 273–283.

R Core Team, 2013. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramsey, F.L., Usner, D., 2003. Persistence and heterogeneity in habitat selection studies using radio telemetry. Biometrics 59, 332–340.

Rosen, S., 1974. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. Journal of Political Economy 82, 34–55. doi:10.2307/1830899

Ross, J.M., Farmer, M.C., Lipscomb, C.A., 2011. Inconsistency in Welfare Inferences from Distance Variables in Hedonic Regressions. The Journal of Real Estate Finance and Economics 43, 385–400. doi:10.1007/s11146-009-9221-z

Rossi, P., 2012. bayesm: Bayesian Inference for Marketing/Micro-econometrics.

Sappington, J., Longshore, K.M., Thompson, D.B., 2007. Quantifying landscape
ruggedness for animal habitat analysis: a case study using bighorn sheep
in the Mojave Desert. The Journal of wildlife management 71, 1419–1426.

Sawyer, H., Nielson, R.M., Lindzey, F., McDonald, L.L., 2006. Winter Habitat
Selection of Mule Deer Before and During Development of a Natural Gas
Field. Journal of Wildlife Management 70, 396–403. doi:10.2193/0022-
541X(2006)70[396:WHSOMD]2.0.CO;2

Schabenberger, H., 2009. spatcounts: Spatial count regression.

Schmidt, L., Courant, P.N., 2006. Sometimes Close Is Good Enough: The Value of
Nearby Environmental Amenities*. Journal of Regional Science 46, 931–
951. doi:10.1111/j.1467-9787.2006.00491.x

Schoen, J.W., Kirchoff, M.D., 1982. Habitat use by mountain goats in southeast
Alaska. Final Report, Federal Aid in Wildlife Restoration Projects W-17-10.
W-17-11, W-21-1, and W-21-2, Job 12.4 R. Alaska Department of Fish and
Game, Juneau, Alaska.

Singer, F.J., Doherty, J.L., 1985. Managing Mountain Goats at a Highway
Crossing. Wildlife Society Bulletin 13, 469–477.

Smith, B.L., 1977. Influence of snow conditions on winter distribution, habitat
use, and group size of mountain goats. Report by Montana Cooperative
Research Unit, University of Montana, Missoula, Montana.

Whittington, J., St. Clair, C.C., Mercer, G., 2005. Spatial responses of wolves to roads and trails in mountain valleys. Ecological Applications 15, 543–553. doi:10.1890/03-5317

Wintle, B.A., Elith, J., Potts, J.M., 2005. Fauna habitat modelling and mapping: A review and case study in the Lower Hunter Central Coast region of NSW. Austral Ecology 30, 719–738. doi:10.1111/j.1442-9993.2005.01514.x

Wright, V., 2000. The Aldo Leopold Wilderness Research Institute: a national wilderness research program in support of wilderness management, in: RMRS-P-15-3, Conference Volume 3. Presented at the Wilderness as a place for scientific inquiry, U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Missoula, MT, pp. 260–268.

Wu, J., Gopinath, M., 2008. What Causes Spatial Variations in Economic Development in the United States? American Journal of Agricultural Economics 90, 392–408. doi:10.1111/j.1467-8276.2007.01126.x