E.ON Energy Research Center
FCN | Institute for Future Energy Consumer Needs and Behavior

FCN Working Paper No. 9/2009

A Comparison of Three Methods of Estimation in the Context of Spatial Modeling

Gaurav Ghosh and Fernando Carriazo

November 2009

Institute for Future Energy Consumer Needs and Behavior (FCN)

Faculty of Business and Economics / E.ON ERC

RWTH AACHEN UNIVERSITY

fcn Institute for Future Energy Consumer Needs and Behavior

FCN Working Paper No. 9/2009

A Comparison of Three Methods of Estimation in the Context of Spatial Modeling

November 2009

Authors' addresses:

Gaurav Ghosh
Institute for Future Energy Consumer Needs and Behavior (FCN)
Faculty of Business and Economics / E.ON Energy Research Center
RWTH Aachen University
Mathieustrasse 6
52074 Aachen, Germany
E-mail: gghosh@eonerc.rwth-aachen.de

Fernando Carriazo
Economic Research Service
U.S. Department of Agriculture
1800 M Street NW,
Washington D.C. 20036, U.S.A.
E-mail: fcarriazo@ers.usda.gov

Publisher: Prof. Dr. Reinhard Madlener
Chair of Energy Economics and Management
Director, Institute for Future Energy Consumer Needs and Behavior (FCN)
E.ON Energy Research Center (E.ON ERC)
RWTH Aachen University
Mathieustrasse 6, 52074 Aachen, Germany
Phone: +49 (0) 241-80 49820
Fax: +49 (0) 241-80 49829
Web: www.eonerc.rwth-aachen.de/fcn
E-mail: post_fcn@eonerc.rwth-aachen.de

A Comparison of Three Methods of Estimation in the Context of Spatial Modeling*

Gaurav S. Ghosh† Fernando Carriazo‡

November 2009

Abstract

We empirically compare the accuracy and precision of representative Least Squares, Maximum Likelihood and Bayesian methods of estimation. Using an approach similar to the jackknife, each method is repeatedly applied to subsamples of a data set on the property market in Bogotá, Colombia to generate multiple estimates of the underlying explanatory spatial hedonic model. The estimates are then used to predict prices at a fixed set of locations. A nonparametric comparison of the estimates and predictions suggests that the Bayesian method performs best overall, but that the Likelihood method is most suited to estimation of the independent variable coefficients. Significant heterogeneity exists in the specific test results.

Keywords: Spatial Econometrics, Bayesian Statistics, Hedonic Valuation

*The authors appreciate the helpful comments of Murali Haran, James Shortle and participants at the AAEA and NAREA Meetings in July 2007 and 2008 respectively. Neither the FCN nor the USDA are responsible for opinions expressed herein.

†Institute for Future Energy Consumer Needs and Behavior (FCN), Faculty of Business and Economics, E.ON Energy Research Center, RWTH Aachen University, Mathieustrasse 6, 52074 Aachen, Germany. Corresponding Author, email: gghosh@eonerc.rwth-aachen.de

‡Economic Research Service, US Department of Agriculture, 1800 M Street NW, Washington DC 20036, USA

1 Introduction

Methods for estimation and prediction under conditions of spatial correlation can be placed into two broad categories. The first set of methods, which include Least Squares and Maximum Likelihood, derive from the Classical approach to statistical inference, and the second set, which include Hierarchical models, derive from the Bayesian approach. Building on seminal texts like Fisher (1925, 1935) and Neyman and Pearson (1933), Classical methods have historically predominated. Today, however, with concomitant developments in simulation methods and computational power, Hierarchical Bayesian methods are extensively applied in fields as disparate as geography (e.g., Ben-Jamma, Marino, and Loaiciga, 1995; Diggle and Ribeiro, 2002), bio-statistics (e.g. Waller and Gotway, 2004) and demography (e.g., Assunção, Potter, and Cavenaghi, 2002; Banerjee, Wall, and Carlin, 2003). Methodological choice has become largely a matter of familiarity and problem context, and Bayesian and Classical methods are often seen as close substitutes.

The use of Hierarchical Bayesian models is comparably rare when analyzing spatial problems in economics, except for some studies in areas like valuation and property market analysis (e.g. Clapp, Kim, and Gelfand, 2002; Gelfand et al., 2004). Given the broad trends in other disciplines, it is reasonable to expect greater future use of Hierarchical Bayesian models by economists. With this expectation in mind, a comparison of Least Squares, Maximum Likelihood and Hierarchical Bayesian methods as applied to spatial economics problems will be instructive. Such a comparison will shed light on methodological choice given the exigencies of the problem. After all, as Cressie (1993) warns, the foundational differences between the Classical and Bayesian Schools impact the interpretation and quality of estimates, predictions and inferences, as measured by criteria like unbiasedness, efficiency and robustness.

In this paper we report findings from a study where representative Least Squares, Maximum Likelihood and Hierarchical Bayesian methods are applied to a spatial hedonic model of a property market in Bogotá, Colombia, and then the resultant estimates and predictions are

compared nonparametrically. The aim is to determine whether the foundational differences between the two Schools, expertly discussed by Bernardo and Smith (1994), Jeffreys (1961) and others, manifest in significantly different outcomes – parameter and prediction estimates – within the context of property market analysis.

A comparison of the estimates yields inferences on the relative quality of the methods. Quality estimates are defined as accurate and precise. Accuracy is inversely proportional to bias and precision is inversely proportional to variability. Knowledge of the quality of a method is important because it affects the reliability of the attendant inferences and prescriptions. Uncertainty in results implies uncertain inference and uncertainty in the consequences of policy prescriptions, which is undesirable from a decision-making perspective (Beron et al., 2004).

We use a large sample approach to compare the three methods of estimation. Each method is applied to multiple subsamples drawn without replacement from the Bogotá property market sample, which in turn is drawn from the property population. This yields many sets of parameter and prediction estimates for each method. A comparison of these sets yields inferences on the relative qualities of the methods. Discussion of the sampling method follows in Section 2.3.

The results show that the chosen Least Squares method of estimation performs worse than the chosen Maximum Likelihood and Hierarchical Bayesian methods on almost all counts. The Likelihood and Bayesian methods perform similarly, but in aggregate the Bayesian method is found to be superior, supplying the highest quality estimates of model parameters and price predictions. However, the Likelihood method is found to be superior when only estimation of the parameters associated with independent variables is important. Discussion of results is found in Section 5.

2 Analytical Framework

In this section we develop the economic and econometric models, explain our the sampling strategy and discuss the methods of estimation chosen for empirical comparison. With respect to notation, we use the convention that symbols refer to vectors or matrices when emboldened and to scalar variables otherwise.

2.1 The Hedonic Property Value Model

The hedonic model is a common economic model widely used in the analysis of property markets. An economic formalization of this model was first derived in Rosen (1974), but earlier applications include Tiebout (1956), who used it to study the implicit market for neighborhoods, and Ridker and Henning (1967), who found a strong statistical relationship between housing values and air quality in metropolitan St. Louis. The latter is the first application of hedonic property value models to environmental valuation. Since then hedonic theory has been used to value environmental factors like air quality (e.g., Chay and Greenstone, 2005; Kim, Phipps, and Anselin, 2003; Smith and Deyak, 1975), water quality (e.g., Leggett and Bockstael, 2000; Poor et al., 2001), agricultural externalities (e.g., Palmquist, Roka, and Vukina, 1997; Ready and Abdalla, 2005), noise (e.g., Baranzini and Ramirez, 2005; Cohen and Coughlin, 2007) and proximity to hazardous facilities (e.g., Brasington and Hite, 2005; Nelson, Genereux, and Genereux, 1992), through their impacts on property values.

Hedonic price theory formalizes the notion that a differentiated commodity is in essence a package of n attributes, $\boldsymbol{x} = (x_1, \cdots, x_n)$ and its price $p = p(\boldsymbol{x})$ a function of the levels of these attributes. When sufficiently large numbers of $\boldsymbol{x}$ are available such that consumers can choose from a wide variety of alternative packages, hedonic price theory stipulates that at the market equilibrium p equals each consumer's bid or value function for the differentiated commodity (Rosen, 1974). Implicit market valuation of attributes becomes possible since

individual valuations are equalized at the margin and set to the marginal price at equilibrium. In other words, at equilibrium, $p_{x_i} = \partial p/\partial x_i$ where p_{x_i} is the marginal implicit price of attribute x_i and $\partial p/\partial x_i$ is the derivative of market price p with respect to x_i. This implicit price is then used to value the attribute. Interested readers may refer to Freeman (2003) and Taylor (2003) for details.

Hedonic price theory is applied to property market analysis because properties are easily modeled as differentiated goods. Attributes consist of structural and neighborhood characteristics such as house size, crime rates and distance to local amenities. The true relationship between attributes and house price is unknown, except that it is monotonically increasing in the desirable attributes (Palmquist, 1999). The relationship is approximated through a reduced form relationship between the attributes and price, which by assumption is the market equilibrium price. The coefficients in the reduced form are estimated and then interpreted as (some function of) the implicit price of the associated attribute. Hedonic price theory allows for the valuation of non-market attributes through the use of market information on related goods.

For this analysis we assume that the relationship between property price and attributes is approximated by a semi-log reduced form equation. If the general form of the regression equation is $y = \boldsymbol{X}\boldsymbol{\beta} + \epsilon$ then under the semi-log specification $y = \ln p$ and $\boldsymbol{X} = \{\boldsymbol{x'}, \ln \boldsymbol{x''}\}$ where $\boldsymbol{x'}$ ($\boldsymbol{x''}$) is the vector of attributes that do not enter (enter) the regression equation in the log scale. The semi-log specification is commonly used in hedonic analysis (e.g., Baranzini and Ramirez, 2005; Pope, 2005) because it permits intuitive interpretation of the coefficients associated with the attributes, which serve as the independent variables in the model. The coefficients are interpreted as implicit price elasticities (semi-elasticities) when the independent variables enter (do not enter) the regression equation in the log scale (Greene, 2003, pg. 123) and as median impacts when the independent variables are dummy variables (Gujarati, 2003, pg. 320). Also, by using log prices, the residuals (and by inference the errors) are normalized and the effects of outliers are reduced.

2.2 The Spatial Error Model

In hedonic property models, it is assumed that spatial dependence is driven by a spatial diffusion process, which can be of two types. First, it can be the spillover effect of neighboring property prices, and second, it can be the spillover effect of unobserved attributes associated with neighboring properties. The first type of spatial process[1] may be thought of as a neighborhood or locality effect, whereby all properties in a neighborhood or locality command positively correlated prices merely by virtue of their existence in the same neighborhood. This type of spatial process is modeled through inclusion in the trend component of the econometric model,[2] with price being a function of a spatially weighted average of neighboring prices (Anselin, 1988). High (low) property prices of neighbors act as positive (negative) externalities in price formation for a given property. Since the dependent variable is lagged, such models are called spatial lag models.

The second type of spatial process may be thought of as an omitted variable effect. It exists because of spatial interdependencies among unobserved or poorly observed property attributes. Given the correlation between omitted variables, the spatial process is modeled through the error component (Gelfand et al., 2004; Kim, Phipps, and Anselin, 2003). The model is commonly referred to as a spatial error model. In both spatial lag and error models, spatial correlation is modeled through a weights matrix, which defines the correlation between all pairs of location in the data set. Following Tobler's First Law of Geography – "everything is related to everything else, but near things are more related than distant things" – correlation is generally modeled as an inverse function of distance (Tobler, 1970).

We build a spatial error model of the property market in Bogotá because our sample is relatively sparse and because we assume that we do not have data on all determinants of house price, which implies the existence of omitted variables. Let $\ln p$ be a spatial process,

[1] As per convention a stochastic process in more than one dimension is called a spatial process or random field (Banerjee, Carlin, and Gelfand, 2004, pg. 21).

[2] Consider the standard econometric model of the form $y = X\beta+\epsilon$. We refer to $X\beta$ as the trend component and ϵ as the error component.

$\{\ln p(\boldsymbol{s}) : \boldsymbol{s} \in \boldsymbol{D} \subseteq \mathbb{R}^2_+\}$, where $\boldsymbol{s}$ is a generic location in $\boldsymbol{D}$, a fixed subset in two dimensional Euclidean space. $\boldsymbol{D}$ corresponds to the geographical window, Bogotá, over which the spatial process is defined. Given the semi-log specification in the economic model, the following is true, i.e. $\ln p(\boldsymbol{s}) = \boldsymbol{X}(\boldsymbol{s})\boldsymbol{\beta} + \epsilon(\boldsymbol{s})$, where $\boldsymbol{X}(\boldsymbol{s}) \in \mathbb{R}^m$ is an m-vector of attributes, $\boldsymbol{\beta} \in \mathbb{R}^m$ is an m-vector of parameters and $\epsilon(\boldsymbol{s})$ is the error. We assume that $\epsilon(\boldsymbol{s})$ is distributed normally, which implies that a set of errors $\boldsymbol{\epsilon}(\boldsymbol{s}) = \{\epsilon(\boldsymbol{s}_1), \cdots, \epsilon(\boldsymbol{s}_n)\}$ corresponding to any set of locations $\{\boldsymbol{s}_1, \cdots, \boldsymbol{s}_n\}$ is distributed multivariate normal, $\boldsymbol{\epsilon}(\boldsymbol{s}) \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$. It follows that $\ln p(\boldsymbol{s}) \sim N(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$. The normality assumption simplifies the distribution theory and is justified by the Central Limit Theorem (Banerjee, Carlin, and Gelfand, 2004, pg. 31).

The error $\epsilon(\boldsymbol{s})$ consists of a measurable spatial effect $w(\boldsymbol{s})$ and an idiosyncratic nugget effect $\eta(\boldsymbol{s})$, i.e. $\epsilon(\boldsymbol{s}) = w(\boldsymbol{s}) + \eta(\boldsymbol{s})$. The nugget effect $\eta(\boldsymbol{s})$ captures measurement error and microscale variation, which is correlation over distances smaller than the spatial scale of the model. It is modeled as a normally distributed uncorrelated pure error term, $\boldsymbol{\eta}(\boldsymbol{s}) \sim N(\boldsymbol{0}, \tau^2\boldsymbol{I})$, where τ^2, also called the *nugget*, is a parameter capturing variability because of nonspatial error. The spatial correlation in the model is captured by the measurable spatial effect $w(\boldsymbol{s})$ and is modeled as a zero-centered, stationary multivariate normal process with $\boldsymbol{w}(\boldsymbol{s}) \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{H}(\boldsymbol{\phi}))$. $\boldsymbol{H}(\cdot)$ is a covariance matrix, containing information on covariance between all pairs of locations in the geographical window. σ^2, also called the *partial sill*, is a parameter affecting the magnitude of the spatial correlation, and $\boldsymbol{\phi}$ is the set of *range* parameters, defining the neighborhood within which a property is correlated with other properties. Since $\boldsymbol{\epsilon}(\boldsymbol{s}) \sim N(\boldsymbol{0}, \boldsymbol{\Sigma})$ and $\epsilon(\boldsymbol{s}) = w(\boldsymbol{s}) + \eta(\boldsymbol{s})$ it follows that

$$\boldsymbol{\Sigma} = \sigma^2\boldsymbol{H}(\boldsymbol{\phi}) + \tau^2\boldsymbol{I} \quad \text{where } [\boldsymbol{H}(\boldsymbol{\phi})]_{ij} = \rho(\boldsymbol{\phi}; \boldsymbol{s}_i - \boldsymbol{s}_j) \tag{1}$$

$[\boldsymbol{H}(\boldsymbol{\phi})]_{ij}$ is the covariance between locations i and j, as defined by the covariance function $\rho(\boldsymbol{\phi}; \boldsymbol{h})$, which defines the correlation structure and how it decays over the Euclidean distance $\boldsymbol{h} = \boldsymbol{s}_i - \boldsymbol{s}_j$ between any two locations i and j. We use variogram analysis to determine

that $\rho(\boldsymbol{\phi}, \boldsymbol{h})$ is of the exponential form.[3] The exponential variogram has a single range parameter, $\boldsymbol{\phi} = \{\phi\}$. Interested readers may refer to Banerjee, Carlin, and Gelfand (2004) for an in-depth discussion of the variogram.

In summary, $\ln p$ is assumed to be a spatial process with trend $\boldsymbol{X}(\boldsymbol{s})\boldsymbol{\beta}$ and error $\epsilon(\boldsymbol{s})$. The error $\epsilon(\boldsymbol{s})$ is zero-centered and distributed multivariate normal with $\epsilon(\boldsymbol{s}) \sim N(0, \sigma^2\boldsymbol{H}(\phi) + \tau^2\boldsymbol{I})$ where $[\boldsymbol{H}(\phi)]_{ij} = \rho(\phi; \boldsymbol{s}_i - \boldsymbol{s}_j)$ is ij^{th} element of the covariance matrix. Variogram analysis indicates that spatial covariance decays exponentially over distance. For the rest of this discussion we refer to the set of error parameters as $\boldsymbol{\alpha} = \{\tau^2, \sigma^2, \phi\}$ and the set of all model parameters as $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$.

2.3 Sampling Strategy

Characterizing the spatial error as a stochastic process or random field implies that a sample of n locations is a *single* realization from an n-dimensional distribution (Banerjee, Carlin, and Gelfand, 2004, pg. 31). One cannot know if this single realization is a representative draw from the underlying population. This causes two problems. First, inference about the underlying population is not robust except under the assumption of stationarity (Schabenberger and Gotway, 2005). Although almost universally used, this assumption is often untenable. Second, even when stationarity does exist, reliable inference is only feasible for the trend parameters and not for the error parameters. The distributional properties of $\boldsymbol{\alpha}$ estimators are not well understood under the stochastic process characterization, which implies that their significance cannot be determined (Cressie, 1993, pg. 99). Asymptotic variances of the error-related parameters, as calculated in Upton and Fingleton (1985) are of no use (Griffith, 1988, pg. 25). These variances are for the limiting case $k \to \infty$ and should not be used if $k < 30$, where k is the number of realizations of the n-dimensional

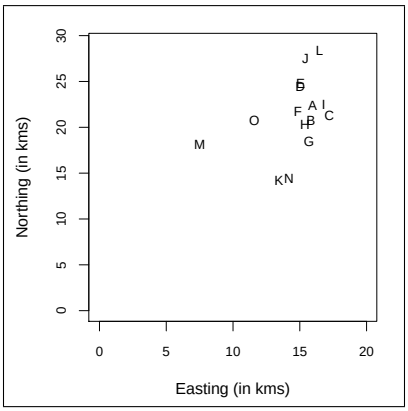
[3] The variogram measures the variance in the error as a function of distance or spatial lag. Under an exponential distribution, $\rho(\phi; \boldsymbol{h}) = \begin{cases} \tau^2 + \sigma^2(1 - \exp(-\phi\boldsymbol{h})) & \text{if } \boldsymbol{h} > \boldsymbol{0} \\ 0 & \text{otherwise} \end{cases}$

distribution.

Both problems vanish when a large sample of n-dimensional realizations is drawn from the underlying population. As a consequence of the Law of Large Numbers, a large sample allows the use of asymptotic sample theory for inference about model parameters. Hence we conduct a large sample analysis of the property market in Bogotá. The sample consists of $k = 38$ independent and identically distributed subsamples drawn randomly and without replacement from the original data set. Each subsample is of size $n = 100$, which implies that each subsample is a 100-dimension realization of the underlying stochastic process. Under the thumb rule that the sampling distribution of the sample mean of a non-normal parent distribution is approximately normal when the sample size is greater than 30 – in which case normality-based large sample estimation theory is applicable – the sample may be considered "large."

The three estimation methods are applied to each of the 38 subsamples and the parameter estimates, $\hat{\boldsymbol{\theta}}$, are recorded. $\hat{\boldsymbol{\theta}}$ is then used to predict prices at 15 locations, chosen randomly and without replacement from the data set. Properties at these locations were not included in the subsamples used to estimate $\hat{\boldsymbol{\theta}}$. The spatial distribution of the predictee locations, indexed $\{A, B, \cdots, O\}$, is shown in Figure 1. The 38 sets of estimations and predictions

Figure 1: Prediction Locations



are then compared nonparametrically to enable inference about the relative quality of each estimation method in the context of spatial hedonic modeling.

The sampling scheme is similar to the bootstrap and the jackknife procedures, which though more comprehensive are infeasible in the current context. The bootstrap was not used because it requires sampling with replacement, which means that an observation can figure more than once in a subsample. As a result, the spatial configuration of the sample is distorted, which affects inference (Griffith, 1988, pg. 28). The jackknife is not used because of the size of the data set. The jackknife requires all subsamples of a given size to be analyzed. Given that the sample size is $N = 1853$, if subsample size were $n = 1852$, then ${}^{N}C_{n} = 1852$ subsamples exist and hence 1852 regressions are necessary for a jackknife estimation. Similarly, if the subsample size were $n = 1851$, then ${}^{N}C_{n} = 1,715,878$ subsamples exist and as many regressions are required. Since Markov Chain Monte Carlo (MCMC) simulations using the entire sample – and needed for the Hierarchical method – take more than three days to run (see Ghosh and Carriazo-Osorio, 2007), the jackknife is too time-consuming. The sampling scheme we use instead may be likened to a *partial jackknife*. Like in the jackknife, equal-sized subsets of the data are sampled, but unlike the jackknife all available subsets are not sampled. The estimates are calculated in a manner analogous to that for the bootstrap or jackknife. They will have the same asymptotic properties as jackknife or bootstrap estimates, but will be less efficient due to the smaller number of sample points. The proposed approach is not as rigorous as the jackknife, but does allow recourse to large sample theory. Similar to the jackknife and bootstrap approaches, the econometric model remains unchanged across the k regressions.

One may note that the problems with estimating error-related parameters under the stochastic process assumption only affects the Classical estimation methods. The MCMC algorithm used in the Hierarchical method easily simulates a sampling distribution of the autoregressive structure (Haining, 1978). The sampling distribution tends towards the true distribution as the sample size increases implying that under the Hierarchical method a good approximation

of the standard error is obtainable even when asymptotic theory is unusable.

2.4 The Methods of Estimation

A number of estimators from the Least Squares, Maximum Likelihood and Bayesian paradigms have been used in the literature. We choose a single estimator from each paradigm for comparison. Details on each estimator and explanations of why they were chosen follow.

2.4.1 Iterated Weighted Least Squares with the Cressie (1985) Loss Function

From among the various Least Squares estimation methods the Iterated Weighted Least Squares (IWLS) method is chosen because it is a "pragmatic compromise between efficiency (GLS) and simplicity (OLS)" (Cressie, 1993, pg. 97). Generalized Least Squares (GLS) cannot be used because the correlation structure is unknown *a priori* and Ordinary Least Squares (OLS) is infeasible because it does not permit spatial correlation. As with all Least Squares based methods, the IWLS method requires minimization of a squared-errors based loss function. A number of loss functions have been proposed in the literature, but we use the Cressie (1985) loss function [see (2)], chosen because its fitting criterion gives more weight to observation pairs with smaller lags, which implies a good fit near the origin. As Stein (1988) notes, this is a good property. It also conforms to Tobler's First Law of Geography, which favors small distance relationships.

The IWLS method with the Cressie (1985) loss function works as follows: First the OLS estimate of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}^{OLS}$, is estimated and the residuals calculated. The residuals are used to estimate the Hawkins-Cressie empirical variogram.[4] The first iteration of error parameter estimates, $\hat{\boldsymbol{\alpha}}^{C(1)}$, are then obtained by substituting the empirical variogram into the Cressie

[4]For details on this variogram see Cressie (1993).

(1985) Loss Function (2), which is minimized.

$$LOSS(\boldsymbol{\theta}) = \sum_j |N(\boldsymbol{h}(j))| \left\{ \frac{\bar{\gamma}(\boldsymbol{h}(j))}{\gamma(\boldsymbol{h}(j);\boldsymbol{\theta})} - 1 \right\}^2 \tag{2}$$

The j's in (2) are the bins that the spatial lags, $\boldsymbol{h}$, have been divided into. Let the zeroth iteration of $\boldsymbol{\theta}$ estimates be $\hat{\boldsymbol{\theta}}^{(0)} = \{\hat{\boldsymbol{\beta}}^{OLS}, \hat{\boldsymbol{\alpha}}^{C(1)}\}$. $\hat{\boldsymbol{\theta}}^{(0)}$ is substituted into $(2\bar{\gamma}(\boldsymbol{h}) - 2\gamma(\boldsymbol{h};\boldsymbol{\theta}))'\boldsymbol{\Sigma}(\boldsymbol{\alpha})^{-1}(2\bar{\gamma}(\boldsymbol{h}) - 2\gamma(\boldsymbol{h};\boldsymbol{\theta}))$, which when minimized supplied the first GLS estimate of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}^{C(1)}$. Using the same iterative process, $\hat{\boldsymbol{\theta}}^{(1)} = \{\hat{\boldsymbol{\beta}}^{C(1)}, \hat{\boldsymbol{\alpha}}^{C(1)}\}$ is used to estimate $\hat{\boldsymbol{\theta}}^{(2)}$ and so on. The iterations continue until convergence and the determination of the Cressie (1985) IWLS (C) estimates, $\hat{\boldsymbol{\theta}}^{C} = \{\hat{\boldsymbol{\beta}}^{C}, \hat{\boldsymbol{\alpha}}^{C}\}$.

2.4.2 Restricted Maximum Likelihood

The Classical Maximum Likelihood estimator of the error parameters, $\boldsymbol{\alpha}$, is severely biased because of the presence of $\boldsymbol{\beta}$ in the likelihood function (Mardia and Marshall, 1984). To correct this bias, Patterson and Thompson (1971) devised the Restricted Maximum Likelihood (R) estimator, which works by applying the Maximum Likelihood method to error contrasts rather than the data itself and hence is independent of $\boldsymbol{\beta}$. An error contrast is a linear combination $\boldsymbol{a}'\boldsymbol{Q}$ with two properties: first $E(\boldsymbol{a}'\boldsymbol{Q}) = 0$ and second $\boldsymbol{a}'\boldsymbol{X} = \boldsymbol{0}'$. Assuming that log of property price $\ln \boldsymbol{p} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}(\boldsymbol{\alpha}))$, R estimates of $\boldsymbol{\theta}$ are obtained by minimizing the negative log-likelihood function in (3) where $\boldsymbol{\Pi}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} - \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}$.

$$\mathfrak{L}(\boldsymbol{\theta}) = \frac{n-(p+1)}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{X}'\boldsymbol{X}| + \frac{1}{2}\log|\boldsymbol{\Sigma}(\boldsymbol{\theta})| + \frac{1}{2}\log|\boldsymbol{X}'\boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\boldsymbol{X}| + \frac{1}{2}\boldsymbol{Q}'\boldsymbol{\Pi}(\boldsymbol{\theta})\boldsymbol{Q} \tag{3}$$

Prediction variances and the bias in $\boldsymbol{\alpha}$ estimates are lower under the R method than under the Classical Maximum Likelihood method.

2.4.3 Hierarchical Bayes

The Hierarchical Bayesian method is chosen from the Bayesian paradigm. It is the standard method used in the area of spatial statistics. The Hierarchical Bayesian (H) method differs from Least Squares and Maximum Likelihood methods in that two levels of stochasticity are assumed. On the first level, $\ln p(\boldsymbol{s})$ is assumed to be normally distributed, conditional on $\boldsymbol{\theta}$, the independent variables, $\boldsymbol{X}(\boldsymbol{s})$, and the spatial effects vector $\boldsymbol{W}(\boldsymbol{s})$. On the second level, $\boldsymbol{W}(\boldsymbol{s})$ is assumed to be normally distributed, conditional on $\boldsymbol{\theta}$, which is also stochastic. In Bayesian methods, model parameters are treated as random effects, whereas in the Least Squares and Maximum Likelihood methods model parameters are fixed.

Estimation under the H method is as follows: first, $\boldsymbol{\theta}$ is assigned a prior distribution. Next data is used to confront the prior and update beliefs about the behavior of $\boldsymbol{\theta}$. The updated beliefs take the form of a posterior distribution for $\boldsymbol{\theta}$. The process of moving from prior to posterior beliefs is underpinned by Bayes' Theorem, which states that the posterior distribution of $\boldsymbol{\theta}$ conditional on $\ln \boldsymbol{p}$ and $\boldsymbol{X}$ is

$$\pi(\boldsymbol{\theta}|\ln \boldsymbol{p}, \boldsymbol{X}) = \frac{f(\ln \boldsymbol{p}|\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{X}) f(\boldsymbol{W}|\boldsymbol{\theta}) f(\boldsymbol{\theta})}{\int f(\ln \boldsymbol{p}|\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{X}) f(\boldsymbol{W}|\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta}} \tag{4}$$

where $f(\ln \boldsymbol{p}|\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{X})$ is the prior conditional distribution assumed for $\ln \boldsymbol{p}$, $f(\boldsymbol{W}|\boldsymbol{\theta})$ is the conditional prior for $\boldsymbol{W}$ and $f(\boldsymbol{\theta})$ is the prior on $\boldsymbol{\theta}$. *A priori* it is assumed that the parameters are independent: $f(\boldsymbol{\theta}) = f(\boldsymbol{\beta}) f(\sigma^2) f(\tau^2) f(\phi)$.

Since $\ln p(\boldsymbol{s}) = \boldsymbol{X}(\boldsymbol{s})\boldsymbol{\beta} + \epsilon(\boldsymbol{s})$ and $\epsilon(\boldsymbol{s}) = w(\boldsymbol{s}) + \eta(\boldsymbol{s})$, (1) implies that $\ln \boldsymbol{p}|\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{X} \sim N(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}, \tau^2 \boldsymbol{I})$ and $\boldsymbol{W}|\boldsymbol{\theta} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{H}(\phi))$. We assign σ^2 and τ^2 inverse gamma priors,[5] $\boldsymbol{\beta}$ a flat prior[6] and ϕ a uniform prior.[7] The priors are chosen because they are flexible enough

[5] The inverse gamma distribution is defined over the support $x \in (0, \infty)$ and has the *pdf* $f(x; a, b) = \frac{b^a}{\Gamma(a)} x^{-a-1} \exp\left(\frac{-b}{x}\right)$.

[6] $\boldsymbol{\beta} \sim N(\bar{\boldsymbol{\beta}}, diag(\infty))$. The infinite variance of each element of $\boldsymbol{\beta}$ implies that $\boldsymbol{\beta}$ has zero mass along its entire range. $\boldsymbol{\beta}$ is essentially left undefined.

[7] $\phi \sim U(a, b)$, where a and b define the range.

to approximate a variety of spatial processes and are commonly used in the geostatistical literature (Banerjee, Carlin, and Gelfand, 2004, pg. 131).

From the posterior distribution of $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta}|\ln \boldsymbol{p}, \boldsymbol{X})$ in (4), the posterior distributions of the individual parameters are calculated by marginalization. For example the posterior distribution of ϕ is $\pi(\phi|\ln \boldsymbol{p}, \boldsymbol{X}) = \iiint \pi(\boldsymbol{\theta}|\ln \boldsymbol{p}, \boldsymbol{X})\ d\boldsymbol{\beta}\ d\tau^2\ d\sigma^2$. Once the marginal posterior distributions are known, parameter statistics are easily calculated. For example the posterior mean of ϕ is $\bar{\phi} = \int \phi\ d\pi(\phi|\ln \boldsymbol{p}, \boldsymbol{X})$ and the posterior median is $\tilde{\phi}$ such that $\int_{-\infty}^{\tilde{\phi}} d\pi(\phi|\ln \boldsymbol{p}, \boldsymbol{X}) = 0.5$.

Often $\pi(\boldsymbol{\theta}|\ln \boldsymbol{p}, \boldsymbol{X})$ and the marginal posteriors are not available in closed form, or the closed form solutions are so messy that calculating statistics becomes intractable. Then MCMC methods are used to sample from the posteriors. A correlated sample is collected by iteratively drawing from $f(\ln \boldsymbol{p}|\boldsymbol{\theta}, \boldsymbol{W}, \boldsymbol{X}) f(\boldsymbol{W}|\boldsymbol{\theta}) f(\boldsymbol{\theta})$, which is proportional to $\pi(\boldsymbol{\theta}|\ln \boldsymbol{p}, \boldsymbol{X})$ since its denominator is constant. Draws from $\pi(\boldsymbol{\theta}|\ln \boldsymbol{p}, \boldsymbol{X})$ are not made because calculating the denominator is computationally expensive. The correlated sample is then used to estimate desired statistics.

A number of algorithms can be used to operationalize MCMC sampling. In this study the Gibbs sampler is used to draw the $\boldsymbol{\beta}$ sample and the Metropolis-Hastings algorithm to draw the $\boldsymbol{\alpha}$ sample. The Gibbs sampler is easier to implement, but can only be used when the full conditional distribution $\pi(\beta_i|\boldsymbol{\beta}_{-i}, \boldsymbol{\alpha}, \ln \boldsymbol{p}, \boldsymbol{X})$ is available in closed form; which occurs only when the prior $f(\boldsymbol{\beta})$ and the likelihood $f(\boldsymbol{p}|\boldsymbol{\theta})$ are a conjugate pair (Banerjee, Carlin, and Gelfand, 2004, pg. 113). Since closed forms of the full conditionals of $\boldsymbol{\alpha}$ are not available, the Metropolis-Hastings algorithm is used to obtain samples of $\boldsymbol{\alpha}$. A correlated sample is drawn to ensure rapid convergence of the Metropolis-Hastings algorithm. The quality of the chains is tested through diagnostics. The acceptance rate of the chain determines if the prior is an appropriate approximation of the true parameter distribution. For all simulations the acceptance rates were in the 25-40% range, which is optimal (Gelman, Roberts, and

Gilks, 1996). All chains converged, as indicated by tests devised by Geweke (1992) and Heidelberger and Welch (1983).

2.4.4 Prediction or Kriging

Since prediction at specified locations is important to many research and policy questions we compare the predictive capabilities of each method. Let $\boldsymbol{s}_0$ be a vector of m locations where prediction is desired and $\boldsymbol{s}_1$ be the vector of the n locations in the sample. Suppressing the location argument $\boldsymbol{s}$ and since $\ln p$ is normally distributed

$$\begin{pmatrix} \ln \boldsymbol{p}_0 \\ \ln \boldsymbol{p}_1 \end{pmatrix} \sim N \left(\begin{pmatrix} \boldsymbol{X}_0\boldsymbol{\beta} \\ \boldsymbol{X}_1\boldsymbol{\beta} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{00} & \boldsymbol{\Sigma}_0 \\ \boldsymbol{\Sigma}_0^T & \boldsymbol{\Sigma} \end{pmatrix} \right)$$

where $\boldsymbol{\Sigma}$ is obtained from (1), $\boldsymbol{\Sigma}_{00}$ is an $m \times m$ matrix of covariances across predictee locations and $\boldsymbol{\Sigma}_0$ an $m \times n$ matrix of covariances between predictee and observed locations. All matrices are defined by the spatial covariance function $\rho(\cdot)$. The normality assumption supplies the closed form for the conditional distribution $\ln \boldsymbol{p}_0 | \ln \boldsymbol{p}_1$, the mean and variance of which are

$$E[\ln \boldsymbol{p}_0 | \ln \boldsymbol{p}_1] = \boldsymbol{X}_0\boldsymbol{\beta} + \boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}(\ln \boldsymbol{p}_1 - \boldsymbol{X}\boldsymbol{\beta}) \tag{5}$$

$$\begin{aligned} V[\ln \boldsymbol{p}_0 | \ln \boldsymbol{p}_1] = {} & \boldsymbol{\Sigma}_{00} - \boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0' \\ & + \left(\boldsymbol{X}_0 - \boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right) \left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1} \left(\boldsymbol{X}_0 - \boldsymbol{\Sigma}_0\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)' \end{aligned} \tag{6}$$

The mean (5) is also called the Best Linear Unbiased Predictor (BLUP) under squared-error loss (Gotway and Cressie, 1993). To get Classical estimates of the kriged means and variances at $\boldsymbol{s}_0$ conditional on data at $\boldsymbol{s}_1$ the Classical parameter estimates ($\hat{\boldsymbol{\theta}}^C$ or $\hat{\boldsymbol{\theta}}^R$ in this paper) are substituted into (5) and (6). When kriging in the Hierarchical framework, the

predictive distribution of the spatial process $\ln p(\boldsymbol{s})$ at the predictee locations $\boldsymbol{s}_0$ is

$$f(\ln \boldsymbol{p}_0 | \ln \boldsymbol{p}_1, \boldsymbol{X}_0, \boldsymbol{X}_1) = \int f(\ln \boldsymbol{p}_0 | \ln \boldsymbol{p}_1, \boldsymbol{X}_0, \boldsymbol{\theta}) \pi(\boldsymbol{\theta} | \ln \boldsymbol{p}_1, \boldsymbol{X}_1) \, d\boldsymbol{\theta}$$

$f(\ln \boldsymbol{p}_0 | \ln \boldsymbol{p}_1, \boldsymbol{X}_0, \boldsymbol{\theta})$ is conditionally normal and generates (5) and (6) through integration. $\pi(\boldsymbol{\theta} | \ln \boldsymbol{p}_1, \boldsymbol{X}_1)$ is known from (4). Draws proportional to $\pi(\boldsymbol{\theta} | \ln \boldsymbol{p}_1, \boldsymbol{X}_1)$ already exist because of the MCMC methods used to estimate $\hat{\boldsymbol{\theta}}^H$. These draws are used to generate draws from the distribution of $\ln \boldsymbol{p}_0 | \ln \boldsymbol{p}_1, \boldsymbol{X}_0, \boldsymbol{\theta}$, which is known to be normally distributed. The latter draws are then used to generate density plots or to estimate statistics related to $\ln \boldsymbol{p}_0$.

3 The Data

The data set consists of 1853 detached or semi-detached residential properties in Bogotá, Colombia. The variables included in the hedonic price function reflect locational, structural and neighborhood characteristics. Neighborhood characteristics include pollution, crime and amenities. Descriptive statistics on all variables is provided in Table 1. All data was provided by `metrocuadrado.com`, a Colombian property price listing service, unless stated otherwise.

Each housing unit is geo-referenced, originally in latitude and longitude, but converted into Bogotá Zone Universal Transverse Mercator (UTM) coordinates. Distortion due to the Earth's curvature is lower under the UTM projection than under standard representations of latitude and longitude. It allows for more accurate interpretation of spatial relationships denominated in Euclidean distance. Under the UTM projection latitude is measured in *Northings* and longitude in *Eastings*. These values increase as one moves north and east. The location of each housing unit is shown in Figure 2(a). The origin has been shifted such that the most southwest property in the sample has the coordinates $(0, 0)$. The sample is collected from a 30 km $\times$ 20 km rectangle as indicated by the minimum and maximum values

Table 1: Descriptive Statistics

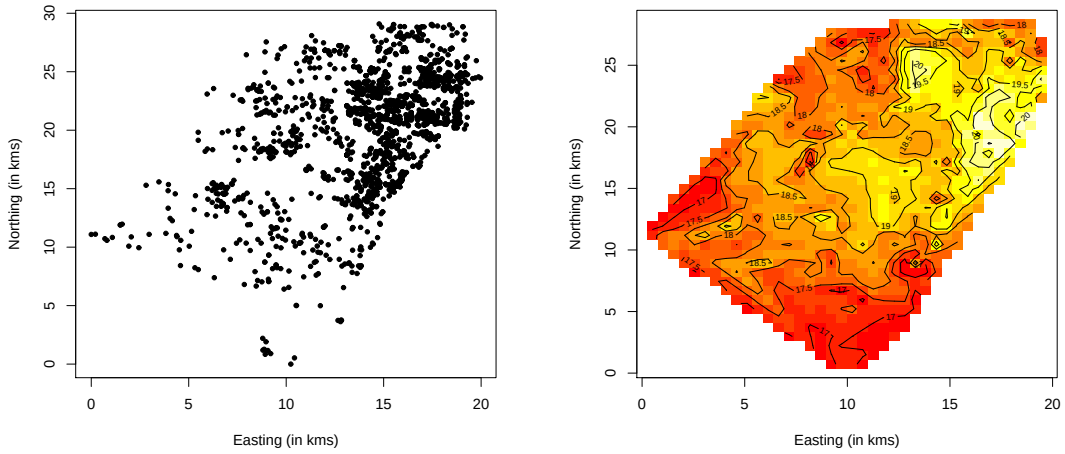
Variable	Units	Mean	Std.Dev.	Min	Max
$\ln p$	$\ln pesos$	18.95	0.77	15.88	20.77
$Northing$	km	20.19	5.12	0.00	29.09
$Easting$	km	14.21	3.46	0.00	19.97
$\#Baths$	N/A	2.84	0.83	1	4
$\#Garages$	N/A	1.08	0.83	0	2
$\ln Area$	$\ln m^2$	5.27	0.60	3.33	6.91
$Carpet$	N/A	0.61	0.49	0	1
$DiningRoom$	N/A	0.59	0.49	0	1
$24DoorKeep$	N/A	0.38	0.49	0	1
$\ln PM_{10}$	$\ln \mu g/m^3$	4.05	0.24	3.47	4.62
$\ln ZonalPark$	$\ln m$	7.04	0.80	3.91	8.42
$\ln FloodArea$	$\ln m$	7.38	1.10	3.91	8.87
$\ln 72^{nd}St$	$\ln m$	8.57	1.07	3.91	9.84
$\ln HistCent$	$\ln m$	9.02	0.95	3.91	9.78
$CrimeIndex$	homicides / 100000	3.17	5.53	0	151
$Stratum$	N/A	4.18	1.14	1	6

of $Northing$ and $Easting$. We note that the sales data is concentrated in the northeast, which is `metrocuadrado.com` concentrates its activities. Many property transactions in the south and west occur outside the purview of the formal housing market and reliable data on such transactions is unavailable.

All sales occurred between 2001 and 2006 and the prices were deflated by the Colombian Consumer Price Index, with 2005 as the base year. The price index data were obtained from Colombian Central Bank official statistics.[8] A contour map of log sales prices, $\ln p$, is shown in Figure 2(b). Note that higher prices are in the north and east. This price structure is explained by geographical and historical factors. Being naturally hilly, with a cooler climate and scenic views, north and east Bogotá have historically been favored as residential locations, and are occupied by upper income households. The south and west are more industrial and settled by lower income households. We expect $\ln p$ to increase in both location variables.

[8] Available online at `http://www.banrep.gov.co/statistics/sta_prices.htm`

Figure 2: Location and Prices of Properties



(a) Spatial Distribution

(b) Contour Plot of $\ln p$

The next seven variables in Table 1 are site-specific and / or structural. $\#Baths$, $\#Garages$ and $\ln Area$ are self-explanatory. $Carpet$, $DiningRoom$ and $24DoorKeep$ are dummy variables, set to 1 when the house has carpeting, a dining room and 24 hour doorkeeping services respectively, and 0 otherwise. Air pollution is indexed by PM_{10}, which measures the airborne concentration of suspended particulates of diameter of $10\mu g$ or less. Locations with higher PM_{10} levels are more polluted. There are 11 air monitoring stations in Bogotá, each geo-referenced. A raster of PM_{10} levels was produced through inverse distance interpolation of the emissions recorded at each station. By superimposing the raster on the property map, PM_{10} concentrations at each location were estimated. We expect $\ln p$ to increase in all site-specific variables except $\ln PM_{10}$. Pollution data were obtained from the Departmento Tècnico Administrativo del Medio Ambiente, the environmental authority in Bogotá.

The next four variables measure the distance between properties and (dis)amenities. Data on these variables were obtained from Bogotá's urban planning department. Historical Center and 72^{nd} Street are the two central business areas in Bogotá. Zonal parks are small green spaces maintained by local planning committees. Under the assumption that households prefer to live close to the city center and to green spaces we expect the corresponding

coefficients to be negative. Low-lying areas are designated as flood prone. The Flood Area variable is a proxy for elevation relative to the surrounding topography, but not absolute elevation. Properties in these areas are expected to command lower sales prices because of flood risk and because of poorer vistas implying a positive coefficient for $\ln FloodArea$. The amenities are geo-referenced and merged with properties into a single map.

The last two variables are defined at the neighborhood level. $Stratum$ is a count variable ranging from one to six and indexing average income in the neighborhood. The neighborhoods with highest mean incomes have $Stratum = 6$. It is expected to trend positively with $\ln p$. $CrimeIndex$ was compiled from neighborhood level crime data, which was geocoded by matching neighborhoods to individual properties. Under the hypothesis that households prefer to locate to low-crime neighborhoods, property prices are expected to negatively correlate to $CrimeIndex$. Crime data were obtained from the Unified System of Information on Violence and Delinquency.

4 Results and Discussion

The estimators are ranked on the basis of their accuracy, precision and overall quality. Accuracy is measured through estimator bias or distance between estimate and true value. Precision is determined from the spread or variability of the estimates: precise estimators are less variable. Quality is gauged through metrics that combine information on accuracy and precision. Inference on the performance of a method is based on the relative rankings of its parameter estimators $\hat{\boldsymbol{\theta}}$ and price prediction estimators $\hat{\boldsymbol{p}}_0$.

4.1 Estimating the Trend Component

The bias of an estimator of a trend parameter is $b(\beta, \hat{\beta}) = E(\hat{\beta}) - \beta$ where $E(\hat{\beta})$ is the expected value of the estimator $\hat{\beta}$ and $\beta \in \boldsymbol{\beta}$ is the true value of the trend parameter. In

this analysis, the true value of β is unknown, which precludes exact measurement of bias and exact inference about accuracy. However, as discussed in Section 3, we have *a priori* expectations about the sign of each β. By comparing the sign of $\hat{\beta}$ to this *a priori* expectation, approximate inference about accuracy is possible. From the subsample regressions, 38 estimates of each β are obtained for each estimation method. The plug-in estimator of $E(\hat{\beta})$ is $\bar{\beta} = \sum_{k=1}^{38} \hat{\beta}_k/38$ where $\hat{\beta}_k$ is the kth partial jackknife estimate of $\beta \in \boldsymbol{\beta}$ (Wasserman, 2004). Values of $\bar{\beta}$ corresponding to each variable in the trend, across the three methods of estimation, are reported as the unbracketed values in Table 2. Comparison of these $\bar{\beta}$s supplies inference about mean accuracy with respect to estimation of the trend $\boldsymbol{X\beta}$.

Each $\bar{\beta}$ is identical in sign and similar in magnitude across the estimation methods and has the *a priori* expected sign in all cases except $Easting$ and $\ln ZonalPark$. *A priori* $\bar{\beta}_{Easting}$ was expected to be positive on the basis of the contour map of $\ln p$ in Figure 2(b), which showed a broad upward trend in the northeast direction. However, the map also indicates pockets of high priced neighborhoods scattered throughout north Bogotá. As a result, some subsamples included high value properties located west of lower value properties, which resulted in the negative sign for $\bar{\beta}_{Easting}$. This negative effect is significant under the R and H methods, but not under the C method. Even when significant, $\bar{\beta}_{Easting}$ values are low, which implies that relocation along the East-West axis has small effect on property price. In contrast, the $\bar{\beta}_{Northing}$ values indicate that price responsiveness to movement along the North-South axis is three times higher. The positive sign on $\bar{\beta}_{\ln ZonalPark}$ is explained by anecdotal evidence that zonal parks are poorly maintained and provide negligible recreational opportunities.

For any independent variable, the sign of the estimated $\bar{\beta}$ is identical across methods. This result leads to an approximate inference that all methods are equally accurate at estimating $\boldsymbol{\beta}$ in our spatial hedonic property model. Since the true value of $\boldsymbol{\beta}$ is unknown, bias cannot be measured and exact inference and ranking on accuracy is not feasible. However, one can test to determine if $\bar{\boldsymbol{\beta}}$ estimates differ significantly across methods. Without significant differences, one may infer that all methods are equally accurate. If there are significant

Table 2: Mean, Standard Deviation and Proportion Results for $\bar{\boldsymbol{\beta}}$

Variable	Cressie IWLS (C)		Restricted ML (R)		Hierarchical (H)	
$Intercept$	16.59^{**}	[100.00]	16.57^{**}	[100.00]	17.03^{**}	[100.00]
	(0.63)		(0.56)		(0.52)	
$Northing$	0.03^{**}	[100.00]	0.03^{**}	[97.37]	0.03^{**}	[97.37]
	(0.01)		(0.00)		(0.00)	
$Easting$	-0.01	[36.84]	-0.01^{**}	[31.58]	-0.01^{**}	[34.21]
	(0.01)		(0.00)		(0.00)	
$\#Baths$	0.08^{**}	[86.84]	0.09^{**}	[97.37]	0.08^{**}	[97.37]
	(0.01)		(0.01)		(0.01)	
$\#Garages$	0.03^{**}	[73.68]	0.03^{**}	[78.95]	0.03^{**}	[73.68]
	(0.01)		(0.01)		(0.01)	
$\ln Area$	0.62^{**}	[100.00]	0.65^{**}	[100.00]	0.64^{**}	[100.00]
	(0.02)		(0.02)		(0.02)	
$Carpet$	0.01	[63.16]	0.04^{**}	[71.05]	0.03^{**}	[68.42]
	(0.02)		(0.01)		(0.01)	
$DiningRoom$	0.04^{**}	[71.05]	0.04^{**}	[73.68]	0.03^{**}	[63.16]
	(0.01)		(0.01)		(0.01)	
$24DoorKeep$	0.06^{**}	[73.68]	0.08^{**}	[86.84]	0.07^{**}	[76.32]
	(0.02)		(0.01)		(0.01)	
$\ln PM_{10}$	-0.35^{**}	[81.58]	-0.31^{**}	[78.95]	-0.32^{**}	[78.95]
	(0.07)		(0.05)		(0.06)	
$\ln ZonalPark$	0.02^{**}	[23.68]	0.04^{**}	[23.68]	0.03^{**}	[23.68]
	(0.01)		(0.01)		(0.01)	
$\ln FloodArea$	0.03^{**}	[65.79]	0.03^{**}	[68.42]	0.03^{**}	[65.79]
	(0.01)		(0.01)		(0.01)	
$\ln 72^{nd}St$	-0.01	[60.53]	-0.04^{**}	[68.42]	-0.06^{**}	[65.79]
	(0.02)		(0.02)		(0.02)	
$\ln HistCent$	-0.17^{**}	[78.95]	-0.17^{**}	[84.21]	-0.18^{**}	[84.21]
	(0.04)		(0.04)		(0.04)	
$CrimeIndex$	-0.00	[55.26]	-0.00	[57.89]	-0.00	[52.63]
	(0.00)		(0.00)		(0.00)	
$Stratum$	0.20^{**}	[100.00]	0.23^{**}	[100.00]	0.21^{**}	[100.00]
	(0.01)		(0.01)		(0.01)	

* and ** indicate significance at the 5% and 1% levels respectively.

differences, one may infer that accuracy differs across methods, but there is not enough information to assign a ranking of methods. We use the nonparametric Wilcoxon Signed Rank Test to determine whether each method is equally accurate at estimating $\boldsymbol{\beta}$. The Wilcoxon test is commonly used for the rank-based comparison of two matched samples because it has a high power efficiency when compared to similar tests (Hays and Winkler, 1971). The C, H and R sets of $\bar{\boldsymbol{\beta}}$ estimates are matched samples. Given three methods of estimation there are three possible pairings: $C-R$, $C-H$ and $R-H$. The Wilcoxon test is applied to each pairing and significance is determined at the 5% level. The tests indicate that there is a significant difference between $\bar{\boldsymbol{\beta}}^H$ and $\bar{\boldsymbol{\beta}}^R$, but that $\bar{\boldsymbol{\beta}}^C$ does not differ significantly from the other $\boldsymbol{\beta}$ estimates. We conclude that there are significant differences in accuracy across methods, but analysis of $\bar{\boldsymbol{\beta}}$ does not allow further inference.

The $\bar{\beta}$s enable inference about accuracy in the mean. Looking at the proportion of $\hat{\beta}_k$s (where $k = 1, \cdots, 38$) with the expected sign, on the other hand, enables inference about the likelihood that a $\hat{\beta}_k$ will have the expected sign. This allows us to rank the three methods with respect to accuracy. One may infer that the most accurate method has the greatest proportion of $\hat{\beta}_k$s with the expected sign. The most accurate method may also be considered to be the most robust to the data, in the sense that the corresponding $\hat{\beta}_k$s are most stable and least likely to change sign on account of random changes in sample composition. The proportions are reported in percentage form as the numbers in square brackets in Table 2.

Proportional accuracy of all methods is greater than 95% in the case of $Northing$, $\ln Area$ and $Stratum$.[9] The R and H methods also have a proportional accuracy of greater than 95% in the case of $\#Bath$. For almost all other variables, the proportional accuracy of each method is greater than 60%. The exceptions are $\hat{\beta}_{Easting}$ and $\hat{\beta}_{\ln ZonalPark}$, which have proportional accuracies of less than 40% across methods, and $\hat{\beta}_{CrimeIndex}$, which has a proportional accuracy of about 50%. These results are not surprising given that $\bar{\beta}_{Easting}$ and $\bar{\beta}_{\ln ZonalPark}$ also did not have the expected sign and that $\hat{\beta}_{CrimeIndex}$ is the only insignificant variable.

[9] We define 'proportional accuracy' as the proportion of $\hat{\beta}_k$s with the expected sign.

The mean proportional accuracy of each estimation method is calculated by averaging down the columns in Table 2. It is over 70% for all methods and the C method does the worst with an accuracy of 73.19%. The H method is second best with a proportional accuracy of 73.85% and the R method does best with a proportional accuracy 76.15%. Wilcoxon tests indicate that the mean proportional accuracy of the R method is significantly higher than that of the H and C methods at the 5% level. The difference in mean proportional accuracy between the C and H methods is not significant. We infer that the R method is the most statistically accurate at estimating $\hat{\boldsymbol{\beta}}$, and that there is no significant difference in accuracy between the C and H methods: $R \succ H \sim C$.

Analysis of the dispersion of $\bar{\beta}$ enables inference about its precision since, in statistics, precision is defined as the inverse of the variance (Banerjee, Carlin, and Gelfand, 2004, pg 101). Since $sd(\bar{\beta}) = V(\bar{\beta})^{1/2} = 1/\psi(\bar{\beta})^{1/2}$ where $sd(\cdot), V(\cdot)$ and $\psi(\cdot)$ signify standard deviation, variance and precision respectively, analysis of sd allows inference about precision. The lower the sd the greater the precision. The numbers in parentheses in Table 2 are the sds of the $\bar{\beta}$s for each method of estimation. The standard deviations are calculated as $sd(\bar{\beta}) = sd(\hat{\beta})/\sqrt{38}$ where $sd(\hat{\beta}) = [\sum_k (\hat{\beta}_k - \bar{\beta})^2/(k-1)]^{1/2}$.

For all 16 βs, the sd of estimates from the C method are greater than or equal to the sd of estimates from the R and H methods. Thinking of the sds as precision scores implies a set of 16 precision scores for each estimation method. Inference on any pair of methods may then be drawn by comparing the scores through the Wilcoxon test. The Wilcoxon test is applied to each pair of methods to determine if the precision of one is significantly greater than the other. Significance is determined at the 5% level. The tests show that the precision of the C method is significantly worse than the precision of the other methods. The precision of the R and H methods do not differ significantly. One may infer that the R and H methods are most precise in the estimation of $\bar{\boldsymbol{\beta}}$, the C method is the least precise. In order of precision the methods may be ranked $H \sim R \succ C$.

Inference about relative precision may also be drawn by comparing the range of every $\hat{\beta}$ estimated under the different methods. Range is defined as $r(\hat{\beta}) = \max(\hat{\beta}) - \min(\hat{\beta})$. A smaller range implies greater clustering of the $\hat{\beta}$ estimator and hence greater precision. As with the $sd(\bar{\beta})$ statistics, $r(\hat{\beta})$ statistics may also be interpreted as scores. The range of every estimator of $\beta \in \boldsymbol{\beta}$ is not reported, but Wilcoxon tests comparing ranges support the sd Wilcoxon test results. In other words, the $\hat{\beta}^C$ estimates have the biggest range and the $\hat{\beta}^H$ and $\hat{\beta}^R$ estimates have the smallest range.

We use the mean squared error of the estimator θ, $MSE(\hat{\theta}) = V(\hat{\theta}) + [b(\theta, \hat{\theta})]^2$, as a metric of its overall quality. MSE measures the difference between the estimate and the true value of the estimated parameter. A high-quality estimator, which has high accuracy and precision, will have a low MSE. This is because the MSE is the sum of the squared bias, which is inversely related to accuracy, and the variance, which is inversely related to precision. When an estimator θ is unbiased then $b(\theta, \hat{\theta}) = 0$ and $MSE(\hat{\theta}) = V(\hat{\theta})$. Since $\bar{\beta}$ is an unbiased estimator of $\beta \in \boldsymbol{\beta}$, it follows that $MSE(\bar{\beta}) = V(\bar{\beta})$. Since $V(\bar{\beta}) = sd(\bar{\beta})^2$, $MSE(\bar{\theta})$ is a monotonic transformation of $sd(\bar{\theta})$. Hence the inferences drawn on the precision of $\bar{\beta}$ are applicable to the quality of these estimators in the mean. By induction the C method supplies estimators with the significantly highest MSE. The MSEs of β estimators from the H and R methods are not significantly different from each other. One may infer that the H and R methods supplies β estimators of the highest quality and the C method supplies β estimators of the lowest quality. The preference ordering of the three methods of estimation is $H \sim R \succ C$.

4.2 Estimating the Error Component

Since all the error coefficient estimates $\hat{\boldsymbol{\alpha}}$ are positive one cannot draw inference about accuracy by studying patterns in sign. Instead we draw inference by individually analyzing the estimates of $\boldsymbol{\alpha}$ across methods. The plug-in estimator of $E(\hat{\alpha})$ is $\bar{\alpha} = \sum \hat{\alpha}_k / 38$ where

$\hat{\alpha}_k$ is the kth partial jackknife estimate of $\alpha \in \boldsymbol{\alpha} = \{\tau^2, \sigma^2, \phi\}$. The $\bar{\alpha}$s and their standard deviations are reported in Table 3. Standard deviations are in parentheses. The nugget τ^2

Table 3: Mean and Standard Deviation Results for $\bar{\boldsymbol{\alpha}}$

Variable	Cressie IWLS	Restricted ML	Hierarchical
Nugget, τ^2	0.03**	0.03**	0.04**
	(0.01)	(0.01)	(0.00)
Partial Sill, σ^2	6.45	0.34**	0.11**
	(3.58)	(0.12)	(0.00)
Range, ϕ	1399.80*	18.68*	0.48**
	(660.27)	(8.90)	(0.00)

* and ** indicate significance at the 5% and 1% levels respectively.

captures idiosyncratic and microscale variations. Across the three procedures, $\bar{\tau}^2$ values are similar, but Wilcoxon tests show that $\bar{\tau}^2_H$ is significantly greater than the C and R estimates of τ^2. We cannot make inferences about the relative accuracy of the methods because the true value of τ^2 is unknown.

When spatial correlation decays exponentially over distance, the partial sill σ^2 captures the random unexplained variation in price between two properties some distance $\boldsymbol{h}$ apart. Consider a property with mean attributes and sold at mean price US$77,915.[10] Given $\bar{\sigma}^2_C = 6.45$, an identical property 1 km away might randomly differ in value by 100% after controlling for size, location and other effects through the trend. This property may randomly cost nothing or twice as much. The corresponding σ^2 estimates for the R and H methods imply unexplained price fluctuations of 27.4% and 1.2% respectively. The C and R methods imply extremely high, unexplained price instability in the property market, which is unlikely. We infer that the C and R methods supply biased estimators of σ^2 and that the H method does not.

The range parameter ϕ defines the distance over which spatial correlation effects may be discerned. The effective range of spatial dependence, which is the distance at which the correlation drops to 0.05, is about $-\log(0.05)/\phi$ (Finley, Banerjee, and Carlin, 2007). The

[10]The exchange rate used was US$1 = 2179 Colombian pesos, valid on 06/30/2009

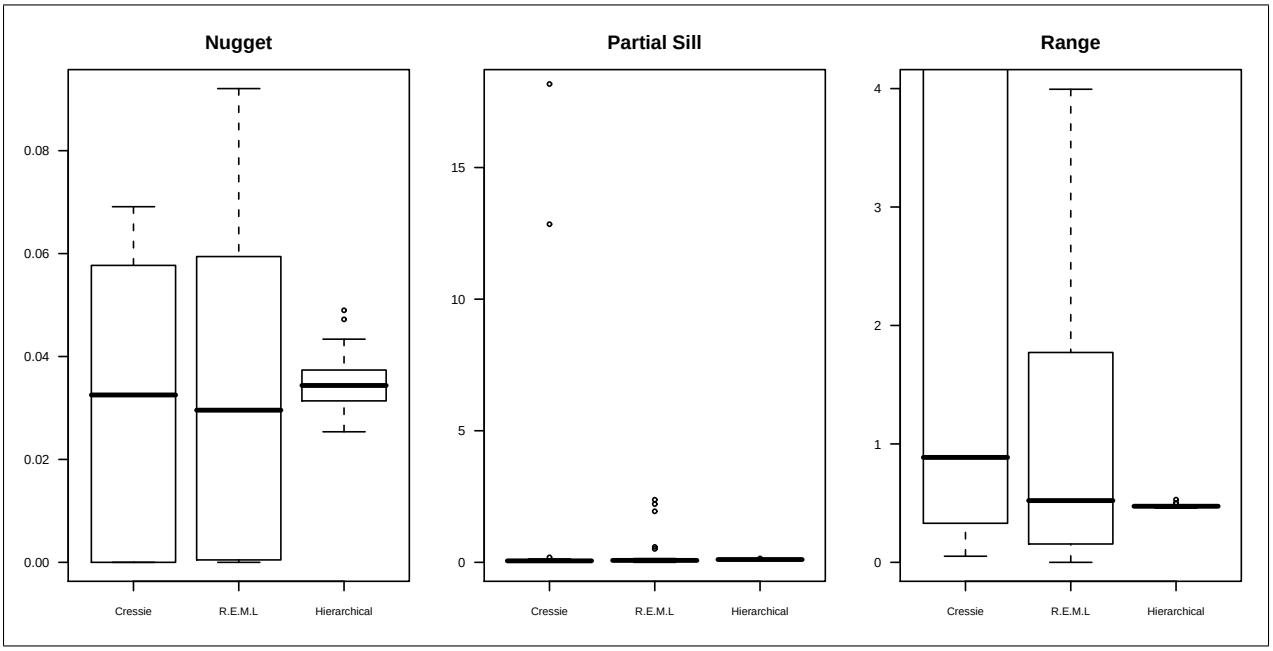
mean effective ranges are 2 meters for the C method, 161 m for the R method and 6.29 km for the H method. With effective ranges of 2 m or 161 m, the hedonic model becomes approximately non-spatial given the geographical scale of the Bogotá property market. The variogram analysis on the other hand implied significant spatial effects. We infer that the C and R methods supply biased and inaccurate range estimates and that the C method is the more inaccurate of the two. The mean range estimated by the H is more plausible and we infer that the H method supplies reasonably accurate estimates of ϕ.

The C and R methods supply biased mean estimates for two of the three α parameters. The bias in the mean R estimates is less than in the mean C estimates. The H method supplies plausible and, by inference, unbiased mean estimates of $\boldsymbol{\alpha}$. We infer that if accuracy of the $\boldsymbol{\alpha}$ estimators is desirable then the preference ordering over methods is $H \succ R \succ C$.

Comparison of the standard deviation of $\boldsymbol{\alpha}$ across methods implies large differences in precision. The H method supplies the most precise estimates and the C method the least precise. In the case of τ^2, $sd(\bar{\tau}^2_H) \approx sd(\bar{\tau}^2_C)/5 \approx sd(\bar{\tau}^2_R)/5$. In the case of σ^2, $sd(\bar{\sigma}^2_H) \approx sd(\bar{\sigma}^2_R)/90 \approx sd(\bar{\sigma}^2_C)/2550$. And finally, in the case of ϕ, $sd(\bar{\phi}_H) \approx sd(\bar{\phi}_R)/4450 \approx sd(\bar{\phi}_C)/330,000$. These substantial differences in the sd of $\bar{\boldsymbol{\alpha}}$ across methods suggest that $\bar{\boldsymbol{\alpha}}^H$ is the most precise estimator of $\boldsymbol{\alpha}$ and that $\bar{\boldsymbol{\alpha}}_C$ and $\bar{\boldsymbol{\alpha}}_R$ are the least and second least precise estimators of $\boldsymbol{\alpha}$. With respect to precision in estimation of $\boldsymbol{\alpha}$ the estimation methods are ranked $H \succ R \succ C$.

The huge disparity in the precision of $\bar{\boldsymbol{\alpha}}$s across procedures and the superior precision of the H procedure are made clear through the box-and-whisker plots in Figure 3. The rectangular boxes mark the interquartile range, the thick horizontal line in the boxes mark the median and the whiskers extend out to the most extreme estimate that is no more than 1.5 times the interquartile range from the box. The dots beyond the whiskers mark outliers. The ranges of the C and R estimates of $\boldsymbol{\alpha}$ are much higher than the ranges of the H estimates, supplying further evidence of imprecision. For all error parameters, note that the median values are similar across methods, but the C and R methods have outliers with extremely high values,

Figure 3: Box-and-Whisker Plots of $\bar{\boldsymbol{\alpha}}$



which push the mean estimates upward. This is especially evident with the $\bar{\sigma}^2$ estimates.

Inference on the quality of the $\boldsymbol{\alpha}$ estimates across methods is identical to the inference on the precision of the $\boldsymbol{\alpha}$ estimates because $\hat{\alpha} = \bar{\alpha}$ and the mean-squared errors of $\boldsymbol{\alpha}$ are identical to their variance. By induction, the H method supplies the highest quality estimators of α and the C method supplied the poorest quality estimators. If the quality of $\boldsymbol{\alpha}$ estimators is important then preferences over the methods of estimation will be $H \succ R \succ C$.

4.3 Price Predictions

Unlike with the model coefficients $\boldsymbol{\theta}$, the exact accuracy of price predictions can be calculated because the sales prices of the properties in the prediction data set are known. Prediction bias is $b(p, \hat{p}) = E(\hat{p}) - p$, where $\hat{p}$ is the estimator of house price $p \in \boldsymbol{p}_0$. The plug-in estimator of $E(\hat{p})$ is the mean prediction $\bar{p} = \sum_{k=1}^{38} \hat{p}_k/38$, where $\hat{p}_k$ is the kth prediction estimate of p. Hence estimated prediction bias is $\hat{b}(p, \hat{p}) = \bar{p} - p$. The smaller the estimated

bias, the more accurate the prediction.

Mean bias $\bar{p}$ is reported as the un-bracketed numbers in Table 4. In the mean, the C method supplied price predictions with the lowest mean bias at seven locations and the highest mean bias at four location. Price predictions supplied by the H method had the lowest and highest mean bias at four and six locations respectively. The corresponding numbers for the R method are four and five. We treat absolute mean bias as a score and apply the Wilcoxon Signed Rank test to determine if differences in mean bias across methods are significant. The tests show that the differences in bias when predicting house price are not significant at the 5% level.

However there is a spatial pattern to the biases. Comparing Figures 1 and 2(a), one notes that locations A to I are sited in areas with a high density of sample units and locations J to O are spatial outliers in low density areas. Wilcoxon tests indicate that prediction bias among non-outliers is significantly less when using the C method as compared to when using the R and H methods. Significance is calculated at the 10% level. We conclude that the methods are ranked $C \succ R \sim H$ when accurate prediction of non-outliers is the criterion. There is no significant difference in prediction bias among outliers when using the R and H methods. Since all methods are equally accurate when predicting prices at outlying locations, but the C method is best for prediction in high density areas, we infer that with respect to prediction accuracy the methods are ranked $C \succ H \sim R$.

In Table 4 we see that prediction biases at locations I and L are highest. This holds true for all methods. I is not an outlier but L is, which indicates that spatial effects might not be to blame for the poor prediction. The data shows that the property at I has an unusually high price and the property at L has an unusually low price. The model specification was unable to capture the price idiosyncrasy at these locations. Instead, the model predicted prices that tended towards the mean, implying a large downward bias at I and a large upward bias at L.

Table 4: Mean Prediction Bias, Standard Deviation and Estimated MSE

Location	Cressie IWLS (C)		Restricted ML (R)		Hierarchical (H)	
A	0.00	[0.03]	0.03	[0.01]	0.03	[0.01]
	(0.03)		(0.02)		(0.02)	
B	0.05	[0.04]	0.04	[0.02]	0.05*	[0.02]
	(0.03)		(0.02)		(0.02)	
C	0.36**	[0.16]	0.36**	[0.15]	0.36**	[0.15]
	(0.03)		(0.02)		(0.02)	
D	−0.07*	[0.03]	−0.07**	[0.02]	−0.05*	[0.02]
	(0.03)		(0.02)		(0.02)	
E	0.00	[0.03]	0.01	[0.01]	0.02	[0.02]
	(0.03)		(0.02)		(0.02)	
F	0.36**	[0.17]	0.38**	[0.17]	0.37**	[0.16]
	(0.03)		(0.02)		(0.02)	
G	−0.34**	[0.16]	−0.38**	[0.17]	−0.38**	[0.18]
	(0.03)		(0.03)		(0.03)	
H	−0.08**	[0.03]	−0.13**	[0.03]	−0.10**	[0.02]
	(0.03)		(0.02)		(0.02)	
I	−0.47**	[0.25]	−0.53**	[0.29]	−0.52**	[0.28]
	(0.03)		(0.02)		(0.02)	
J	0.14**	[0.05]	0.10**	[0.02]	0.14**	[0.05]
	(0.03)		(0.02)		(0.03)	
K	0.16**	[0.06]	0.15**	[0.05]	0.15**	[0.05]
	(0.03)		(0.03)		(0.03)	
L	0.61**	[0.40]	0.62**	[0.39]	0.60**	[0.39]
	(0.03)		(0.02)		(0.02)	
M	-0.05	[0.05]	-0.04	[0.04]	-0.03	[0.05]
	(0.04)		(0.03)		(0.03)	
N	0.13**	[0.06]	0.11**	[0.04]	0.12**	[0.04]
	(0.03)		(0.03)		(0.03)	
O	−0.20**	[0.07]	−0.25**	[0.07]	−0.25**	[0.08]
	(0.03)		(0.02)		(0.02)	

* and ** indicate significant bias at the 5% and 1% levels respectively.

The standard deviations of price predictions are used to compare the precision of the different methods. The standard deviations are calculated as $sd(\ln \bar{p}) = sd(\ln \hat{p})/\sqrt{38}$ where $sd(\hat{p}) = \sqrt{\sum_k (\hat{p}_k - \bar{p})^2/38}$. The sd are reported as the numbers in Table 4 in parentheses. The R method supplies the most precise predictions at all locations while the C method supplies the least precise predictions at 14 of the 15 locations; the sole exception being J. Wilcoxon tests indicate significant differences in precision across methods at the 5% level. From the most precise to the least precise, the methods are ranked $R \succ H \succ C$. Identical results on significant differences in precision hold when looking at only non-outliers (at locations A to I) or at only outliers (at locations J to O). Once again significance is at the 5% level, except when the C and H methods are compared. Here the the p-value is 6%.

Prediction quality is gauged through the Estimated Mean Squared Error $EMSE(\hat{p}) = V(\hat{p}) + \hat{b}(p, \hat{p})^2$. The lower the $EMSE$ the better the estimator. The $EMSE$ results are reported as the numbers in square brackets in Table 4. The R method supplies mean predictions with the lowest $EMSE$ at 9 locations and with the highest $EMSE$ at 2 locations. The corresponding numbers for the C method are 3 and 10, and for the H method are 3 and 3. The Wilcoxon tests indicate that the C method supplies predictions of the lowest quality. The p-values in the tests comparing the C and R methods and the C and H methods are 0.03 and 0.06 respectively. There is no significant difference in prediction quality between the R and H methods. The three methods are ranked $H \sim R \succ C$. These results are not corroborated when we look only at the non-outlier set, where no significant differences are found in prediction quality. When comparing prediction quality with respect to outliers, C performs worse than the H and R methods, but at 10% levels of significance. The H and R methods always perform equivalently. We infer that the prediction quality of each method is statistically similar when predicting at locations with a high density of observations, but the quality of the C method suffers when making predictions at outlying locations.

5 Conclusion

The rankings of the three methods of estimation with respect to the accuracy, precision and quality of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\ln \boldsymbol{p}$ estimates are summarized in Table 5. The best performer is assigned a rank of one and the worst a rank of three. When two or more methods perform equivalently then they share the mean rank. For example, since the Cressie IWLS and Hierarchical methods are equally accurate when estimating $\boldsymbol{\beta}$, but less accurate than the Restricted ML method, they are jointly assigned a rank of 2.5, which is the mean of ranks 2.0 and 3.0.

The Cressie IWLS estimates of $\ln \boldsymbol{p}$ are most accurate, which validates the suggestion in Cressie (1993) that the $IWLS$ approach, when used with the Cressie Loss function (2), will supply the most accurate prediction estimates. However, the Cressie IWLS method performs poorly on all other metrics. It has the lowest precision and quality of the three methods and is also relatively inaccurate when predicting the model parameters. The rankings indicate

Table 5: Relative Performance of Methods of Estimation

		C	R	H
Accuracy	$\boldsymbol{\beta}$	2.5	1.0	2.5
	$\boldsymbol{\alpha}$	3.0	2.0	1.0
	$\ln \boldsymbol{p}$	1.0	2.5	2.5
Precision	$\boldsymbol{\beta}$	3.0	1.5	1.5
	$\boldsymbol{\alpha}$	3.0	2.0	1.0
	$\ln \boldsymbol{p}$	3.0	1.0	2.0
Quality	$\boldsymbol{\beta}$	3.0	1.5	1.5
	$\boldsymbol{\alpha}$	3.0	2.0	1.0
	$\ln \boldsymbol{p}$	3.0	1.5	1.5

that the Hierarchical method is most suited to estimating the error structure of a spatial hedonic model since its estimates of $\boldsymbol{\alpha}$ are most accurate and precise. Overall it supplies the best quality estimates of $\boldsymbol{\theta} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}\}$ due to high precision, but its accuracy is less than that of the Restricted ML method.

In most studies, the accuracy, precision and quality of the $\boldsymbol{\beta}$ estimates is most important to economists. In such cases, our results suggest that the analyst should apply the Restricted ML method, which has the highest ranking on all three counts. The Hierarchical method has equal precision and overall quality, but is not quite as accurate. The results also favor use of the Restricted ML method for prediction since its mean ranking across accuracy, precision and quality of $\ln p$ estimates is lowest.

The results are suggestive, but not conclusive with regard to the overall merits of the three methods of estimation. The suitable choice is, to some degree, contingent on the research question. If overall accuracy of the $\boldsymbol{\theta}$ and $\ln \boldsymbol{p}$ estimates is paramount when estimating a spatial hedonic model, then the Restricted ML and Hierarchical methods would be preferred to the Cressie IWLS since mean accuracy ranks are 2.7 for the Cressie IWLS, 1.8 for the Restricted ML and 2.0 for the Hierarchical method. If overall precision is important, then once again the Restricted ML and Hierarchical methods (with mean ranks of 1.5 and 1.5) are preferred to the Cressie IWLS method (with mean rank of 3.0). With respect to overall quality, the three methods have mean ranks of 3.0, 1.7 and 1.3. If overall quality of estimation, inclusive of predictive and estimative capabilities, is important then our ranking favors the Hierarchical method.

This paper is a preliminary attempt at comparing different estimation frameworks in a spatial hedonic context. Our assessment of the accuracy, precision and quality of parameter estimates provides a signal with regard to the overall reliability of marginal implicit price estimates of housing attributes and neighborhood amenities, contingent on estimation method. Yet, given the paucity of empirical analyses of different frameworks in a spatial context, we hope that this research will contribute to the broader discourse on methodology in spatial statistics as applied to economic problems. There are many ways in which this work can be extended. Most obviously, the results can be made more robust by increasing the number of n-dimensional realizations from the data set. Further testing of the different procedures is always possible through the design of more metrics. Other estimation approaches can also

be tested.

References

Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Studies in Operational Regional Science, Dordrecht: Kluwer Academic Publishers.

Assunção, R., J. Potter, and S. Cavenaghi. 2002. "A Bayesian Space Varying Parameter Model applied to estimating Fertility Schedules." *Statistics in Medicine* 21:2057–2075.

Banerjee, S., B. Carlin, and A. Gelfand. 2004. *Hierarchical Modeling and Analysis for Spatial Data*. Monographs in Statistics and Applied Probability 101, Chapman & Hall / CRC.

Banerjee, S., M. Wall, and B. Carlin. 2003. "Frailty Modeling for Spatially Correlated Survival Data, with Application to Infant Mortality in Minnesota." *Biostatistics* 4:123–142.

Baranzini, A., and J. Ramirez. 2005. "Paying for Quietness: The Impact of Noise on Geneva Rents." *Urban Studies* 42:633–646.

Ben-Jamma, F., M. Marino, and H. Loaiciga. 1995. "Sampling Design for Contaminant Distribution in Lake Sediments." *Journal of Water Resources Planning and Management* 121:71–79.

Bernardo, J., and A. Smith. 1994. *Bayesian Theory*. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons.

Beron, K., Y. Hanson, J. Murdoch, and M. Thayer. 2004. *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Berlin: Springer-Verlag, chap. Hedonic Price Functions and Spatial Dependence: Implications for the Demnd for Urban Air Quality. Advances in Spatial Science, pp. 267–281.

Brasington, D., and D. Hite. 2005. "Demand for environmental quality: A Spatial Hedonic Analysis." *Regional Science and Urban Economics* 35:57–82.

Chay, K.Y., and M. Greenstone. 2005. "Does Air Quality Matter? Evidence from the Housing Market." *Journal of Political Economy* 113:376–424.

Clapp, J., H.J. Kim, and A. Gelfand. 2002. "Predicting Spatial Patterns of House Prices using LPR and Bayesian Smoothing." *Real Estate Economics* 30:505–532.

Cohen, J., and C. Coughlin. 2007. "Spatial Hedonic Models of Airport Noise, Proximity and House Prices." Unpublished, Working Paper 2006-026C, Federal Reserve Bank of St. Louis.

Cressie, N. 1985. "Fitting Variogram Models by Weighted Least Squares." *Journal of the International Association for Mathematical Geology* 17:563–586.

—. 1993. *Statistics for Spatial Data*, revised ed. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons.

Diggle, P., and P. Ribeiro. 2002. "Bayesian Inference in Gaussian Model-Based Geostatistics." *Geographical and Environmental Modelling* 6:129–146.

Finley, A., S. Banerjee, and B. Carlin. 2007. "spBayes: An R Package for Univariate and Multivariate Hierarchical Point-referenced Spatial Models." *Journal of Statistical Software* 19.

Fisher, R. 1935. *The Design of Experiments*. Edinburgh: Oliver and Boyd.

—. 1925. *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.

Freeman, A. 2003. *The Measurement of Environmental and Resource Values, Theory and Methods*, 2nd ed. Resources for the Future.

Gelfand, A., M. Ecker, J. Knight, and C. Sirmans. 2004. "The Dynamics of Location in House Price." *Journal of Real Estate Finance and Economics* 29:149–166.

Gelman, A., G. Roberts, and W. Gilks. 1996. *Bayesian Statistics 5*, Oxford: Oxford University Press, chap. Efficient Metropolis Jumping Rules. pp. 599–607.

Geweke, J. 1992. *Bayesian Statistics 4*, Oxford, UK: Clarendon Press, chap. Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments.

Ghosh, G., and F. Carriazo-Osorio. 2007. "Bayesian and Frequentist Approaches to Hedonic Modeling in a Geo-Statistical Framework." Portland, OR, Selected Paper presented at the American Agricultural Economics Association Annual Meetings.

Gotway, C., and N. Cressie. 1993. "Improved Multivariate Prediction under a General Linear Model." *Journal of Multivariate Analysis* 45.

Greene, W. 2003. *Econometric Analysis*, 5th ed. Pearson Education.

Griffith, D. 1988. *Advanced Spatial Statistics*, vol. 12 of *Advanced Studies in Theoretical and Applied Econometrics*. Boston: Kluwer Academic Publishers.

Gujarati, D. 2003. *Basic Econometrics*, 4th ed. McGraw-Hill.

Haining, R. 1978. "A Spatial Model for High Plains Agriculture." *Annals of the Association of American Geographers* 68:493–504.

Hays, W., and R. Winkler. 1971. *Statistics: Probability, Inference and Decision*. International Series in Decision Processes, New York: Holt, Rinehart and Winston, Inc.

Heidelberger, P., and P. Welch. 1983. "Simulation Run Length Control in the Presence of an Initial Transient." *Operations Research* 31:1109–1144.

Jeffreys, H. 1961. *Theory of Probability*, 3rd ed. Oxford, UK: Oxford University Press.

Kim, C., T. Phipps, and L. Anselin. 2003. "Measuring the Benefits of Air Quality Improvements: A Spatial Hedonic Approach." *Journal of Environmental Economics and Management* 45:24–39.

Leggett, C., and N. Bockstael. 2000. "Evidence of the Effects of Water Quality on Residential Land Prices." *Journal of Environmental Economics and Management* 39:121–144.

Mardia, K., and R. Marshall. 1984. "Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression." *Biometrika* 76:289–295.

Nelson, A., J. Genereux, and M. Genereux. 1992. "Price Effects of Landfills on House Values." *Land Economics* 68:359–365.

Neyman, J., and E. Pearson. 1933. "On the Problem of the Most Efficient Tests of Statistical Hypotheses." *Philosophical Transactions of the Royal Society A* 231:289–337.

Palmquist, R. 1999. *Handbook of Environmental and Resource Economics*, Cheltenham, UK: Edward Elgar, chap. Hedonic Models. pp. 765–776.

Palmquist, R., F. Roka, and T. Vukina. 1997. "Hog Operations, Environmental Effects and Residential Property Values." *Land Economics* 73:114–124.

Patterson, H., and R. Thompson. 1971. "Recovery of Interblock Information when Block Sizes are Unequal." *Biometrika* 58:545–554.

Poor, P., K. Boyle, L. Taylor, and R. Bouchard. 2001. "Objective versus Subjective Measures of Water Clarity in Hedonic Property Value Models." *Land Economics* 77:482–493.

Pope, J. 2005. "Information, Attention and Revealed Preference: Limits to the Hedonic Model." Unpublished, Job Market Paper, North Carolina State University.

Ready, R., and C. Abdalla. 2005. "The Amenity and Disamenity Impacts of Agriculture: Estimates from a Hedonic Pricing Model." *American Journal of Agricultural Economics* 87:314–326.

Ridker, R., and J. Henning. 1967. "The Determinants of Residential Property Values with Special Reference to Air Pollution." *Review of Economics and Statistics* 49:246–225.

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

Schabenberger, O., and C. Gotway. 2005. *Statistical Methods for Spatial Data Analysis*. Texts in Statistical Science, Boca Raton, FL: Chapman and Hall / CRC.

Smith, V., and T. Deyak. 1975. "Measuring the impact of Air Pollution on Property Values." *Journal of Regional Science* 15:277–288.

Stein, M. 1988. "Asymptotically Efficient Prediction of a Random Field with a Misspecified Covariance Function." *Annals of Statistics* 16:55–63.

Taylor, L. 2003. *A Primer on Nonmarket Valuation*, Kluwer Academic Publishers, chap. The Hedonic Method. The Economics of Non-Market Goods and Resources, pp. 331–393.

Tiebout, C. 1956. "A Pure Theory of Local Expenditure." *Journal of Political Economy* 64:416–424.

Tobler, W. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region." *Economic Geography* 46:234–240.

Upton, G., and B. Fingleton. 1985. *Spatial Data Analysis by Example: Point Pattern and Quantitative Data*, vol. I of *Wiley Series in Probability and Mathematical Statistics*. Chichester, UK: John Wiley & Sons.

Waller, L., and C. Gotway. 2004. *Applied Spatial Statistics for Public Health Data*. New York, NY: Wiley.

Wasserman, L. 2004. *All of Statistics: A Concise Course in Statistical Inference*. Springer Texts in Statistics, New York, NY: Springer-Verlag New York Inc.

E.ON Energy Research Center

fcn Institute for Future Energy Consumer Needs and Behavior

RWTH AACHEN UNIVERSITY

List of FCN Working Papers

2009

Madlener R., Mathar T. (2009). Development Trends and Economics of Concentrating Solar Power Generation Technologies: A Comparative Analysis, FCN Working Paper No. 1/2009, Institute for Future Energy Consumer Needs and Behavior, RWTH Aachen University, November.

Madlener R., Latz J. (2009). Centralized and Integrated Decentralized Compressed Air Energy Storage for Enhanced Grid Integration of Wind Power, FCN Working Paper No. 2/2009, Institute for Future Energy Consumer Needs and Behavior, RWTH Aachen University, November.

Kraemer C., Madlener R. (2009). Using Fuzzy Real Options Valuation for Assessing Investments in NGCC and CCS Energy Conversion Technology, FCN Working Paper No. 3/2009, Institute for Future Energy Consumer Needs and Behavior, RWTH Aachen University, November.

Westner G., Madlener R. (2009). Development of Cogeneration in Germany: A Dynamic Portfolio Analysis Based on the New Regulatory Framework, FCN Working Paper No. 4/2009, Institute for Future Energy Consumer Needs and Behavior, RWTH Aachen University, November.

Westner G., Madlener R. (2009). The Benefit of Regional Diversification of Cogeneration Investments in Europe: A Mean-Variance Portfolio Analysis, FCN Working Paper No. 5/2009, Institute for Future Energy Consumer Needs and Behavior, RWTH Aachen University, November.

Lohwasser R., Madlener R. (2009). Simulation of the European Electricity Market and CCS Development with the HECTOR Model, FCN Working Paper No. 6/2009, Institute for Future Energy Consumer Needs and Behavior, RWTH Aachen University, November.

Lohwasser R., Madlener R. (2009). Impact of CCS on the Economics of Coal-Fired Power Plants – Why Investment Costs Do and Efficiency Doesn't Matter, FCN Working Paper No. 7/2009, Institute for Future Energy Consumer Needs and Behavior, RWTH Aachen University, November.

Holtermann T., Madlener R. (2009). Assessment of the Technological Development and Economic Potential of Photobioreactors, FCN Working Paper No. 8/2009, Institute for Future Energy Consumer Needs and Behavior, RWTH Aachen University, November.

Ghosh G., Carriazo F. (2009). A Comparison of Three Methods of Estimation in the Context of Spatial Modeling, FCN Working Paper No. 9/2009, Institute for Future Energy Consumer Needs and Behavior, RWTH Aachen University, November.

2008

Madlener R., Gao W., Neustadt I., Zweifel P. (2008). Promoting Renewable Electricity Generation in Imperfect Markets: Price vs. Quantity Policies, FCN Working Paper No. 1/2008, Institute for Future Energy Consumer Needs and Behavior, RWTH Aachen University, July (revised May 2009).

Madlener R., Wenk C. (2008). Efficient Investment Portfolios for the Swiss Electricity Supply Sector, FCN Working Paper No. 2/2008, Institute for Future Energy Consumer Needs and Behavior, RWTH Aachen University, August.

Omann I., Kowalski K., Bohunovsky L., Madlener R., Stagl S. (2008). The Influence of Social Preferences on Multi-Criteria Evaluation of Energy Scenarios, FCN Working Paper No. 3/2008, Institute for Future Energy Consumer Needs and Behavior, RWTH Aachen University, August.

Bernstein R., Madlener R. (2008). The Impact of Disaggregated ICT Capital on Electricity Intensity of Production: Econometric Analysis of Major European Industries, FCN Working Paper No. 4/2008, Institute for Future Energy Consumer Needs and Behavior, RWTH Aachen University, September.

Erber G., Madlener R. (2008). Impact of ICT and Human Skills on the European Financial Intermediation Sector, FCN Working Paper No. 5/2008, Institute for Future Energy Consumer Needs and Behavior, RWTH Aachen University, September.