6-16-2010

A Multilevel Property Hedonic Approach to Valuing Parks and Open Space

Christopher Treg
University of Vermont

A MULTILEVEL PROPERTY HEDONIC APPROACH TO VALUING PARKS AND OPEN SPACE

A Dissertation Presented

by

Treg Christopher

to

The Faculty of the Graduate College

of

The University of Vermont

In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy
Specializing in Natural Resources

January, 2010

Accepted by the Faculty of the Graduate College, the University of Vermont, in partial fulfillment of the requirements for the degree of Doctor of Philosophy specializing in Natural Resources.

Dissertation Examination Committee:

______________________ **Advisor**
Austin Troy, Ph.D.

____
Morgan Grove, Ph.D.

____
Beverley Wemple, Ph.D.

______________________ **Chairperson**
Ruth Mickey, Ph.D.

______________________ **Associate Dean, Graduate College**
Patricia A. Stokowski, Ph.D.

Date: October 15th, 2009

ABSTRACT

Many of the benefits that are generated by the natural environment are external to normal market transactions and are consequently undervalued and under-provisioned even though they substantially contribute to human welfare. One approach to valuing certain environmental goods and services is through a regression technique known as the property hedonic model. This model considers a property as a bundle of attributes where the total price of the property is decomposed into marginal, implicit prices for property-specific attributes, the context or neighborhood in which a property resides and access to environmental amenities.

The goal of this dissertation research is to estimate the value of proximity to the environmental amenities of parks and open spaces using a property hedonic model for the City of Baltimore and suburban areas of Baltimore County. While the property hedonic model has been commonly used to value environmental benefits, few of these studies have distinguished the spatial scales of neighborhood characteristics from the property-specific characteristics within a regression model. In this research, a multilevel modeling approach to the typical property hedonic model was used to model the effects of attributes at different spatial scales. This approach also allowed the effect of environmental attributes to vary across geographic space and interact with attributes across spatial scales. Such methods provide a more realistic accounting of the dynamic spatial variation of the value of environmental goods and services.

For parks in the City of Baltimore, the results of valuing proximity to parks showed a spatial dynamic not often captured in property hedonics. The overall fixed effect for distance to park was negative but insignificant. When allowed to vary by block group, the random effect for this variable indicated that only two-thirds of the 401 neighborhoods positively valued increased proximity to parks. No interactions were found to be significant for the entire study. However, for the population of block groups whose properties did positively value proximity to parks, the results of interactions with neighborhood and park characteristics showed that smaller and more open parks were valued higher than larger and more wooded parks. A high population density also increased the value for a property in close proximity to a park. Finally, properties with smaller yards placed a higher value on proximity to parks than those properties with larger yards, indicating a substitution effect.

For open space in Baltimore County, the results indicated that while higher proportions of privately-owned open space surrounding a property increased the value of that property, open space that was publicly-accessible was not significantly valued. Privately-owned open space that was potentially developable was less than half the value of the positive effect of private, open space under conservation easements or other development restrictions.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: LITERATURE REVIEW

1.1. Environmental Attributes and Valuation Methods

The natural environment provides a wide variety of goods and services that are crucial to human welfare. Ecosystem goods and services are often divided into four groups: supporting services; production services; regulation services; and cultural services (Assessment 2003; Costanza et al. 1997; de Groot et al. 2002). Supporting services refer to the primary services that are necessary for the existence of all other services such as soil formation, nutrient cycling and net primary production. Production services refer to the provision of food, fuels and raw materials and are services most readily included into market transactions. Regulation services refer to the ability of the environment to regulate climate, hydrologic and biological processes such as the sequestering of carbon to moderate global temperatures or the ability of wetlands to moderate flooding from storms. Finally, cultural services provide aesthetic and recreation opportunities as well as contributing to cultural or religious heritages.

Most of the benefits that are generated by the natural environment, also referred to as “environmental externalities”, are external to normal market transactions and consequently, are often undervalued and under-provisioned even though they greatly impact the quality of people’s lives. Capturing the monetary value of these benefits can improve human welfare because it allows for more informed choices in weighing tradeoffs between conservation and economic development. Conflicts between the natural and built environment can occur through the outright conversion of natural areas into developed ones or through the decline in quality of the natural environment from extraction or waste production related to the production and manufacture of goods. Consequently, the impacts of many environmental disamenities, or negative environmental externalities, such as air and water pollution, are over-provisioned when their costs are not captured in the market.

The focus of this dissertation research is on valuing cultural services, specifically recreation and aesthetic amenities, provided by parks and open spaces in urban and suburban environments. The non-market, economic value of the benefits of environmental amenities and disamenities can be captured through stated or revealed-preference techniques. With a stated-preference approach, such as Contingent Valuation or Contingent Choice, individuals specify their preferences (e.g. through willingness to pay) for specific changes to or combinations of environmental attributes within a hypothetical market (Mitchell and Carson 1989).

The value for environmental attributes derived from these studies may be broadly categorized as either "use" or "non-use" values. With use values, the benefits accruing to an individual are related to seeing or using the environmental amenity, such as the opportunity for recreation or providing aesthetically pleasing views. With non-use values, an individual may get satisfaction from, for example, the existence of wilderness or the ensured protection of rare and endangered species. Estimates of non-use values (also known as existence, intrinsic, preservation and passive-use values) can comprise a substantial proportion of the total economic value of ecosystem goods and services and ignoring such values can cause serious misallocations for the provision of these attributes

(Freeman 2003). For example, Carson and others (1999) found that lost non-use values resulting from the Exxon Valdez oil spill amounted to $2.8 billion dollars for the United States.

In contrast to the stated preference approach, revealed-preference techniques only capture those environmental attributes that are directly usable to the individual. With this method, use values can be inferred from information on market transactions for related goods (Freeman 2003). One such revealed preference method is known as the property hedonic model in which an individual can choose a level of use of an environmental attribute through their choice of location where they purchase a property. This type of valuation method is the focus of this dissertation research.

1.2. Background to the Property Hedonic Model

The foundation for the property hedonic model was presented by Rosen (1974) who showed the existence of a property market equilibrium where consumers and suppliers maximize their respective utility and profits by choosing to purchase and produce properties with distinct combinations of desirable attributes. While the value of each attribute is implicit and therefore not directly observed in the property transaction, the marginal economic contribution of each of these attributes to the total transaction can be estimated from a regression model using property sales data from an area with varying combinations of these attributes. Etymologically, the term “hedonics” is derived from the Greek word for pleasure, “hedonikos”. In the context of the property market, it refers to the utility or satisfaction one derives through the consumption of goods and services related to the purchase of property.

These characteristics can be broadly grouped into three categories: property-specific (including both the land and structural improvements); contextual neighborhood-specific (the socio-economic context); and environmental (locational) (Freeman 2003). Structural attributes refer to the tangible qualities of a dwelling and parcel such as lot size, house size (square footage), quality, age, number of bathrooms. Contextual attributes are those shared by a neighborhood such as crime rate, ethnicity, income and other socio-economic factors. Locational attributes refer to the proximity and/or accessibility of various land uses and buildings such as hospitals, schools, highways, industrial areas, shopping centers as well as proximity to environmental amenities such as urban green spaces of parks and golf courses. A traditional property hedonic model can be written in terms of the house price as function of a vector of these structural characteristics (S), neighborhood characteristics (N) and environmental or locational characteristics (L) (Equation 1.1).

$$P = f\,(L, S, N) \qquad \text{Eq. 1.1}$$

The price of the house is the sum of the implicit prices for the attributes that are contained in the hedonic model. The implicit marginal price of any single attribute is revealed in the regression coefficient as the additional amount that must be paid for an individual buyer to move to a housing bundle with a higher level of that characteristic, *ceteris paribus* (Freeman 2003).

1.3.1 Structural Attributes

A large amount of the variation in house price can be explained by the structural characteristics of that house, especially the number of rooms and bathrooms, the amount of floor space and the lot size.

Residential properties with greater floor space are desired by big families and buyers who can afford a better standard of living. However, house and lot size are often log transformed to represent the declining marginal value of these features. A home buyer may greatly value large amounts of floor space but the incremental value of each additional square foot between 500ft^2 and 1,000ft^2 will be greater than for the incremental difference between 1,000ft^2 and 1, 500ft^2. Research has also found that the age of a house is often negatively related to the house price. This is because *ceteris paribus*, older houses are worth less because they have obsolete fixtures and appliances, have higher annual maintenance and repair costs, and typically require more energy to heat and cool the living space (Clapp and Giaccotto 1998; Knight and Sirmans 1996). The value of a house is expected to depreciate with age but at a declining rate. This suggests that a log transformation is most suitable for this characteristic. However, other research has found that after a number of years, age will often be positively associated with house price. This may be due to unknown renovations or the “vintage effect” (Goodman and Thibodeau 1995) of older properties. For a study area with a considerable proportion of vintage houses, a quadratic transformation is preferable to a log transformation.

The quality of the construction of the house is also an important attribute in the property hedonic model. However, it is difficult to find a single objective metric for this characteristic and as a consequence most researchers tend to exclude this characteristic from their models. For the State of Maryland, where quality of construction is a measure within the state's appraisal database, some studies have found that higher quality leads to a significantly positive effect on house price that is approximately double that of the effect of average quality construction (Geoghegan 2002; Troy and Grove 2008).

1.3.2. Neighborhoods Attributes

The neighborhood where a house is located also plays an important role in the valuation of a house. The distinction between neighborhood and locational characteristics is somewhat artificial as both groups are related to the spatial context that surrounds a property. However, neighborhood attributes are often considered as such because they are "relative locational" attributes (Orford 2002), such as socio-economic characteristics, that are shared by a contiguous geographic area and cannot be disaggregated to the property level (although they may be aggregated further). The effects of "fixed locational" attributes, on the other hand, are characteristic that uniquely effect and can be measured for every individual property (Orford 2002). These unique, fixed locational attributes, such as distance to an amenity, can also be aggregated to the neighborhood level and have a separate effect from the effect at the property level. This aspect is discussed further in the section on multilevel modeling.

The quality of the public school district in which a property resides has been found to have an impact on house prices, with characteristics such as high test scores showing significant positive benefits to the house price (Brasington 1999; Clapp et al. 2007; Downes and Zabel 2002; Goodman and Thibodeau 1998; Sedgley et al. 2008).

Both actual and perceived levels of crime have been found to negatively affect property prices. Gibbons (2004) found that high levels of vandalism had a much greater, negative impact on property prices than high levels of burglary. He posited the argument that the visibility of vandalism motivated fear of crime in a community even when levels of robbery are relatively low. Lynch and Rasmussen (2001) found a significant drop in house price in areas with high levels of violent crime but no effect from high levels of burglary. They attributed this lack of significance to higher reporting rates in wealthier communities.

Other socio-economic variables, such as percent unemployment, population density, demographic distribution and median household income have been found to significantly affect house prices (Anderson and West 2006).

1.3.3 Locational Attributes

Many of the original property hedonic studies included distance to the Central Business District (CBD) as the primary and often sole measure of location. Increasing distances to the CBD were expected to reduce accessibility and increase travel costs, thereby reducing the value of a house. However, Coulson (1991) observed that prior research had great difficulty in finding that prices significantly declined with distance from the CBD. Since then, many studies have included distances to multiple employment centers (McDonald and McMillen 1990; Orford 2000; Ottensmann et al. 2008) and other measures of accessibility such as distance to highway interchanges or rail hubs (McMillen and McDonald 1998; Troy and Grove 2008). Many of these studies found distances to secondary employment centers to be significant, with their inclusion in the models improving predictions beyond those obtained by using only distance to the CBD. Des Rosiers and others (2000) estimated travel times from each property to the CBD and to highways, shopping centers, schools, and universities as a better metric for accessibility than just distances.

Other studies have focused on the relation of residential housing prices to environmental externalities in addition to measures of accessibility. Some of the environmental amenities that have been researched are discussed in the next section, while others have examined the effects of negative environmental externalities such as: hazardous waste sites (Deaton and Hoehn 2004; McCluskey 2003); landfills (Hite et al. 2001); superfund sites (Gayer et al. 2000); air pollution (Kim et al. 2003) and flood risk (Bin and Polasky 2004). There are also studies that found the externality of noise from traffic provided a significant negative effect on property values (Palmquist 1992; Wilhelmsson 2000).

1.3. Environmental Attributes and the Property Hedonic Model

The research by Ridker and Henning (1967) was one of the first property hedonic studies that attempted to use residential property values as a measure of the economic benefits that accrue to individuals from improvements to the environmental attribute of air quality. By regressing Census tract property values on a measure of sulfate air pollution, they found that a change in pollution level was significantly related to a change in property value. They argued that a change in the value for multiple, individual residences could be summed to measure the total benefits that accrue to a geographic area as a result of improvements to air quality.

In the decades that followed, the property hedonic model was used to estimate the value of numerous environmental attributes, particularly of different land uses and the changes in the quality or quantity of those areas. Some of the more recent environmental amenities that property hedonic models have attempted to estimate are: urban parks (Bolitzer and Netusil 2000; Morancho 2003; Orford 2002; Troy and Grove 2008); greenbelts (Lee and Linneman 1998); forest preserves (Garrod and Willis 1992; Thorsnes 2002; Tyrväinen and Miettinen 2000); wetlands (Mahan et al. 2001); and agriculture (Bastian 2002). The general theory and findings from this research is that undeveloped, open spaces, especially in urban built environments, provide substantial benefits to residential property values. However, the magnitude of this impact or even whether this impact is beneficial or negative varies across studies and depends upon the type and characteristics of the open space and the surrounding context of the neighborhood.

With respect to the type of urban green space, Bolitzer and Netusil (2000) found that urban parks and golf courses both have positive impacts, while cemeteries have an insignificant, negative impact on properties. Anderson and West (2006) found that urban park and golf course proximity provides a significant benefit while proximity to cemeteries has a negative impact. Smith and others (2002) found a negative effect with proximity to suburban parks and positive impact with proximity to golf courses.

Park characteristics such as size, vegetation cover and crime have also been found to affect the proximity-price relationship. Lutzenhiser and Netusil (2001) found that small, urban parks have a negative impact on nearby properties but larger, natural (wooded) areas have a positive impact for homes in Portland, Oregon. Their findings suggest that the size of the green space has an important effect on price-proximity relationship. They theorize that the benefits from large parks outweigh the negative externalities of traffic and noise that may be associated with smaller, urban parks. Anderson and West (2006) found the same, beneficial effect of size for natural areas (special parks) but a detrimental effect of size for neighborhood parks. Garrod and Willis (1992) found that parks primarily consisting of conifers have a negative impact while deciduous tree cover created a positive impact. Tyrväinen (2000) and Thorsnes (2002) found that proximity to forested views and access is beneficial to property prices although for small wooded strips, Tyrväinen found a negative impact.

Neighborhood specific factors may also affect the value of proximity to park. Anderson and West (2006) found that urban parks are generally more beneficial to nearby properties than suburban parks in the area around Minneapolis-St. Paul, Minnesota. High income and high density neighborhoods increased the value of proximity to these urban parks. Dehring and Dunse (2006) found that home buyers in high density neighborhoods favored proximity to urban parks while there was no significant effect for lower density neighborhoods. Troy and Grove (2008) found that higher levels of crime in the areas around parks reduced the positive impact of park proximity to the point where there was a negative impact of park proximity on properties near high-crime parks.

Other research has focused on values related to water such as: proximity to water bodies (Lansford and Jones 1995) changes in water quality (Leggett and Bockstael 2000; Poor et al. 2007) and river restoration from dam removal and riparian enhancement (Lewis et al. 2008; Mooney and Eisgruber 2001). Leggett and Bockstael (2000) reported that improvement to water quality at beaches in Maryland, using fecal coliform bacteria as the metric for pollution, has a significant, positive effect on property values. On the other hand, Mooney and Eisengruber (2001) found that lowering in-stream water temperature, by increasing riparian tree cover and shade, reduced nearby home values. Such discrepancies may be related to whether it is the recreational opportunities or the aesthetic amenities which are most important to nearby properties.

The most common approach to estimating the value of these environmental amenities has been to include a measure of distance from a property to the amenity as a proxy for use values of aesthetics (usually visual) or recreational opportunities provided by the environments. However, the visual aesthetics of an environmental attribute may either be experienced at the amenity location or from the view of that location from a property. It may be beneficial to include both distance and visual metrics into a property hedonic model, especially in the case of the use values for water. Benson and others (1998) examine the effect of mountain, lake and ocean views on property values in Washington State and found that both full and partial views added a significant benefit over houses at similar distances from these attributes but were without a view. Bourassa and others (2004) also found that ocean views in New Zealand provide a significant benefit over view-less properties that were the same distance from the coast. Paterson and Boyle (2002) and Muller (2009) argue that including only a measure of distance to an environmental attribute and ignoring a metric for view to that same attribute may result in a mis-specification of the model and consequently a bias in the resulting coefficient for the distance metric.

In addition to the commonly-studied distance and view metrics of environmental attributes, some researchers have measured the composition and spatial distribution of land use. Geoghehan and others (1997) found that having a relatively high amount of immediately adjacent visual and recreation amenities is considered to be beneficial by local residents but that a high proportion of open space within 1km of their property reduces the conveniences associated with developed areas such as shopping and entertainment. Acharaya and Bennett (2001) and Kestens and others (2004) reported similar results on the difference between high levels of open space in the immediate neighborhood of a property versus less-valued open areas that are a short driving-distance from a residence.

1.4. Technical Issues with the Property Hedonic Model

The property hedonic regression results is questioned by the numerous statistical issues which must be carefully considered and controlled for before a valid interpretation of regression results can be made. The model may be mis-specified as a result of the functional form of the model, collinearity among variables and omitted determinants of property value. There are also spatial issues of dependency, non-stationarity and scale that exist in a housing market composed of interrelated sub-markets.

1.4.1. Functional Form

In estimating a relationship between environmental amenities and property prices, the choice of functional form is not always clear. Rosen (1974) stressed that economic theory fails to indicate that any particular form is appropriate and a variety of functional forms have been used in the hedonic literature. A linear form assumes that an individual's preferences are linear, implying that perfect repackaging of property characteristics is possible (Freeman 2003). However, in property markets, individual house characteristics are inseparable; an individual cannot mix characteristics in any other level than is already available in each house (Garrod and Willis 1992). Also, since the price function is an equilibrium relationship determined in the marketplace by the interactions of individual buyers and sellers (Taylor 2003), the existence of a linear relationship is unlikely.

Early research tried using alternative forms such as the log-linear or double-log forms where the best form was chosen based on the goodness of fit (Freeman 2003). While the form that is chosen should ideally improve the model fit and help to satisfy important assumptions of OLS regressions, such as normally distributed residuals and homoscedasticity, this is not the main issue with choosing functional forms. The goal of finding a proper functional form is to overcome problems associated with the non-linearity that is often found in hedonic regression equations (Goodman and Thibodeau

1995). Substantively, this means that the proper functional form should be chosen so that marginal value for any given property attribute does not vary across the range of house prices.

Halvorsen and Pollakowski (1981) recommended allowing both for transformation of the dependent variable and for different transformations of each independent variable using a quadratic Box-Cox transformation (Box and Cox 1964). However, Freeman (2003) suggests this approach is more cumbersome than necessary while Palmquist (1992) recommends only transforming the dependent variable and the independent variables that are the main effects. Cropper and others (1988) suggest that in models with missing or proxy variables, a common occurrence in property hedonic studies, that simple functional forms (linear, quadratic, log-log and log-linear) or linear Box-Cox transformations are preferred over quadratic Box-Cox forms. This is because omitted variable bias will affect more coefficients in the quadratic forms than in simpler functional forms. Halstead and others (1997) note that the choice of functional form can affect both variable significance and the magnitude of the coefficients.

Since the purpose of these estimated regression functions is to generate amenity values, it may be preferable to use a relatively simple form (Freeman 2003). A log-linear form allows the marginal effect of each independent variable to vary with the level of the dependent variable. Thus, the marginal effects of independent variables change as house price varies. The double-log form, in which both the dependent variable and the main effects are transformed using the natural logarithm, may provide the most interpretable results. With this form, a coefficient is interpreted as an elasticity; the percentage change in the dependent variable given the percentage change in an explanatory variable.

A Box-Cox transformation analysis can be used to provide guidance on whether such simple forms are adequate for satisfying regression assumptions. The Box-Cox transformation of the dependent variable is shown as:

$$y(\lambda) = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \text{if } \lambda \neq 0 \\ \log y, & \text{if } \lambda = 0 \end{cases} \quad \text{Eq. 1.1}$$

With this test, the parameter, λ, is estimated through maximum likelihood to find the optimal transformation of the dependent variable. This parameter can then be tested for significant differences between the optimal value of λ and three cases of λ that correspond to simpler functional forms: a reciprocal transformation, where $\lambda=-1$; a log transformation, where $\lambda=0$; and a linear (untransformed) form, where $\lambda=1$. Some researchers have compared Box-Cox transformations with simpler functional forms and found that a log transformation of the dependent variable (semi-log) or of both dependent and independent variables (double-log) performed the best (Anderson and West 2006; Anthon et al. 2005).

1.4.2 Submarkets

Rosen’s (1974) development of the theory for the property hedonic model assumed that both supply and demand factors were mobile and elastic and that an entire city could be viewed as having a single housing market in equilibrium. Equilibrium occurs when the market settles on a hedonic price supply-demand curve that ensures households (within their budget constraints) cannot increase their utility by choosing a different property and sellers cannot increase their profits by increasing the property's price or changing its characteristics. With this assumption, the price of a property and the availability and contribution of its constituent characteristics are invariant across geographic space (Goodman and Thibodeau 1998). Since Rosen work, most researchers have found that housing markets are typically not in equilibrium and that the assumption of a single market is unrealistic except for very small study areas. The use of a single regression model for an entire city should therefore be considered inappropriate. Once a house is built, its characteristics are fixed (ignoring the potential for costly renovation) and it likely shares similar characteristics as the surrounding properties. Although more elastic than house supply, consumer preferences can also create market segmentation. People of different ethnicities, generations, incomes or social classes may desire particular combinations of property, neighborhood and locational characteristics when seeking to purchase a residence (Borjas 1998). Consumer driven segmentation can be further exaggerated through information constraints such as the tendency of the real estate agent to profile a potential buyer (racially or by income status) and selectively present properties of specific types or in specific types of neighborhoods depending on the characteristics of the potential buyers (Orford 2000).

While the existence of housing submarkets is commonly accepted, there has been little agreement on the method of identifying housing submarkets. Some researchers segment the market by sectors, that is, differentiating the housing stock by structural attributes such as dwelling type (Ekeland et al. 2002). Other researchers focus on markets based on locational-contextual attributes such as income classes or ethnicities. Orford (2000) argues that rather than distinguishing between sector-defined or context-defined submarkets, there is a joint influence of sector and context attributes that should be modeled simultaneously. Day and others (2004) used aspatial clustering methods to group properties as a function of either the housing stock/structural attributes or by neighborhood characteristics. They found that the property market was best segmented by socio-economic characteristics rather than housing stock. Even when segmenting by context rather than sector, there has been little consistency of the spatial units used to define these sub-markets. Many researchers use Census units to delineate the boundaries. Bourassa and others (2003) found that geographic, sub-market boundaries defined by real estate appraisers provided a better model than aspatial clustering techniques. Goodman and Thibodeau (1998) found that school districts were appropriate for determining sub-markets.

1.4.3. Spatial Dependency, Non-Stationarity and Scale

This issue of market segmentation raises spatial statistical concerns of spatial dependency, non-stationarity and inappropriate scale for the property hedonic model. Spatial dependency (association or correlation) refers to the likelihood that the values of observations for a particular variable become more similar with spatial proximity. Real estate agents and appraisers price and assess homes based on the value and characteristics of nearby homes, leading to further spatial dependency in house price and important characteristics such as house size (Orford 2000). Since residuals capture unexplained variation in the model, spatial error autocorrelation reveals the existence of a spatial association that has not been incorporated into the model (Paez and Scott 2004). The problem with the presence of spatial error auto-correlation in a regression model is that the statistical assumption regarding the independent distribution of errors is violated. As a consequence of these spatial dependencies (lags and error), parameter estimates will be biased and inefficient, respectively (Anselin 1988). This leads to artificially smaller standard errors and the possibility of finding a spurious significance of an effect when one does not actually exist (Type I error). Much of this dependency can be attributed to and controlled for by market segmentation and the use of multilevel models as discussed in the section below. The Moran's I statistic (Cliff and Ord 1981) can be used to test whether the final model has sufficiently accounted for spatial error autocorrelation. However, Lauridsen and Kosfeld (2006) suggest that the a high Moran's statistic may result from both strong positive spatial correlation and spatial non-stationarity. The Lagrange Multiplier test is recommended for distinguishing between these two processes (Anselin 1988; Lauridsen and Kosfeld 2006; Mueller and Loomis 2008).

Non-stationarity refers to the existence of a heterogeneous (non-constant) relationship between dependent and independent variables across geographic space (Fotheringham et al. 2002). Global approaches to hedonic modeling, such as using OLS, do not accommodate local, spatial variations in these relationships. A coefficient that is reported as insignificant within a global regression model may be the result of highly significant positive relationships cancelling out the effect of significant negative relationships in others areas.

Spatial scale is commonly viewed as being composed of two components: grain and extent (Allen and Hoekstra 1992; Turner et al. 1989). Grain is the fundamental unit of measurement or observation and is also referred to as resolution or unit of analysis while "extent" refers to the boundary of the area or system that is being studied. The use of property hedonic models to capture the value environmental attributes requires that observations come from a single housing market (Rosen 1974), which suggests that the ability to define what is an appropriate extent or boundary to a hedonic study is an issue that itself requires further research. However, the focus of this discussion is on the scale component of grain. Issues with spatial dependencies and non-stationarity are complicated by spatial scale because by simply changing the resolution it may be possible to make homogeneity out of heterogeneity and vice versa (Polsky and Easterling 2001). A classic example of this is problem was presented by Openshaw and Taylor (1979) in which the relationship between percent elderly voters and percent republican votes varied from a significant positive to a significant negative relationship depending on the resolution (and configuration) of the political boundaries used in the analysis.

This sensitivity to changes in unit size in the relationships between: the independent and dependent variable (and the issue of non-stationarity); individual observations (spatial lag); and the residuals (spatial error autocorrelation) is also known as "scalar dynamics" (Geoghegan et al. 1998). While much of the more recent research on property hedonic models has attempted to control for or explicitly model issues of spatial dependencies and non-stationarity, few have attempted to explicitly analyze the scalar dynamics of these issues because they cannot be effectively modeled with standard regression techniques.

The context in which a property resides is one example of exhibiting scalar dynamics. Both individual property attributes and their neighborhood averages can differentially affect house prices. In addition, cross-scale interactions can occur where large scale attributes such as neighborhood socio-economic characteristics mediate or constrain the preference for environmental goods, such as proximity to open space that is occurring at the individual level. This higher-level constraint is viewed as an important aspect of hierarchy theory (O'Neill et al. 1986), which suggests that modeling of different scales should occur simultaneously.

Explicitly modeling multiple scales is useful for accounting for these statistical issues as well as accounting for the change in the value of environmental attributes due to scales of human perception. While knowledge and appreciation of the environment can originate from unique occasions, much of how an individual interprets and values environmental attributes is determined by their daily activities and experiences that spatially bind the individual into “life spaces” (Reginster and Edwards 2001). This notion of “embedding” or place-based experiences will vary with a particular “audience” or consumer group (e.g. lifestyles, ethnicities, cultures, and generations) and consequently the value that is placed on a particular environmental attribute will vary with these different audiences (Hein et al. 2006).

1.5. A Multilevel Approach to Property Hedonic Models

With the multilevel approach to developing a property hedonic model, individual properties are nested within neighborhoods within a city (Brown and Uyar 2004; Gelfand et al. 2007; Goodman and Thibodeau 1998; Orford 2000, 2002). The goals of these models is to allow the simultaneous examination of the effects of group-level and individual-level variables on individual-level outcomes (the property price) while accounting for the non-independence of observations within groups (the neighborhoods).

Multilevel models were initially developed by educational researchers to examine the effects of context, such as classroom and school characteristics, on individual scholastic achievement (Goldstein 1993; Raudenbush 1991). More recently, the use of multilevel models has grown rapidly in geographic-orientated research such as relationships of socio-economic conditions and individual health (Beland et al. 2002; Diez-Roux et al. 2000; Fone and Dunstan 2006; Klassen et al. 2005; Langford et al. 1998; Moon 2003; Mujahid et al. 2007; Robert 2004; Soobader 2006; Subramanian 2001), epidemiology (Mauny et al. 2004), environmental justice (McLeod et al. 2000), criminal activity (Browning et al. 2004) and land use (Hoshino 2001; Overmars 2006; Pan 2005). This is an important development because geographical analyses are intrinsically spatial and involves the grouping of elementary units of interest (e.g. households or individuals) into higher-level spatial clusters such as neighborhoods (Orford 2000), communities (Moon 2003) or travel analysis zones (Bhat 2000).

There are several statistical and substantive reasons for explicitly modeling individual properties as belonging to neighborhoods. First, spatial dependencies (either correlation in variables or the residual) are likely to be common in property hedonic analyses since individual properties in the same neighborhood are likely to be similar in ways not fully accounted for by the property and neighborhood variables included in a single-level model (Jones and Bullen 1993). With a multilevel model, the house price and other important house characteristics found at the property level can be included at the neighborhood level to control for dependencies in house price and characteristics (spatial lag processes). In the property market, this type of dependency is known as a compositional effect whereby the neighborhood averages of individual property attributes affect the value of a individual property (Orford 2002). Multilevel models also account for the spatial error autocorrelation (dependence of the residuals) by differentiating between-individual errors from between-neighborhood errors (Orford 2000). If this dependency was not modeled, the standard errors of the independent variables would be biased downwards (underestimated), which results in spuriously significant effects (Snijders and Bosker 1999).

Second, it is possible that an average effect for a particular attribute does not represent local conditions that may occur in different areas of a study area; the issue of non-stationarity in property hedonic studies (Cho et al. 2006; Troy and Grove 2008). A possible outcome of estimating an average effect, also known as a global effect, is that contrasting relationships in different areas of the study may cancel each other out, leading to a globally insignificant estimate for that variable’s coefficient (Fotheringham et al. 2002). By considering a variable as a random effect, instead of a global fixed effect, that variable is allowed to vary across neighborhoods. While it is possible to build OLS regressions to determine this effect for each neighborhood, there will be the problem that there are insufficient observations within any given neighborhood leading to unreliable estimates. In multilevel models, random effects are constructed with Empirical Bayesian techniques which borrow strength across neighborhoods and shrink or smooth estimates for unreliable neighborhoods (those with few observations) towards the overall mean (Raudenbush and Bryk 2002). This means that the implicit prices of certain attributes (those that are considered random effects) are optimally weighted averages that combine information derived from the group itself with the mean from all other neighborhoods (Diez-Roux 2002). While unreliable submarket estimates are differentially shrunk towards the global estimate, submarkets with many properties will not be affected by this shrinkage. The Empirical Bayesian variation in slopes between neighborhoods is added to the global fitted value to reveal areas of positive and negative relationships and is one of the fundamental advantages of multilevel modeling (Subramanian 2001).

The reason for the existence of non-stationarity can then be examined by allowing that variable to interact with neighborhood-level variables (Steenbergen and Jones 2002). By specifying cross-level interactions, it is possible to determine whether the effect of a level-1 variable is conditioned or moderated by a group-level variable and is termed, "causal heterogeneity" by Western (1998). For example, neighborhood population density may change across a study area and have a direct effect on house prices. Also allowing this variable to interact with a locational variable, such as distance to park, may moderate the park-price relationship so that in areas with high population density, proximity to parks may be more highly valued than areas with a much lower population density.

Third, heterogeneity (varying relationships) between neighborhoods needs to be distinguished from the heterogeneity among individual properties. Ignoring this differentiation and modeling the behavior of interest at a single level invites the pitfalls of ecological or atomistic fallacies. Atomistic and ecological fallacies are avoided because the predictors and unexplained variation are modeled at the appropriate level (Jones and Duncan 1996). When group-level data, such as Census data, is included in an OLS model whose focus is the individual, the interpretation of the results may lead to the ecological fallacy. With this situation, data (and the processes they measure) that are collected at a broader scale are assumed to have the same importance, associations and interactions at the individual level as at the higher level. Conversely, the atomistic fallacy refers to inferences about the significance, associations and the variability of higher, group levels are based on data originating from individuals (Allen and Starr 1982; O'Neill et al. 1986). This fallacy can also refer to higher-level data that is simply an aggregation of individual-level observations (Hox 2002). For example, the association between price and house size or age at the property level may differ from the association between these variables that have been aggregated to the group level. With multilevel models, the variation in house price due to the effects of individual-level, property attributes is separated from price differences between areas that are related to neighborhood-level contextual effects. With multilevel models, the user is not forced to model at one level or the other, both (indeed several) levels are be modeled simultaneously.

Hedonic models often suffer from heteroscedasticity (unequal variation in the residuals) across neighborhoods, leading to inefficient estimates(Goodman and Thibodeau 1995). Heteroscedasticity can be caused by omitting important variables from the models or from the modeled individual property characteristics and potential interactions. For instance, the variance associated with the implicit price of floor area may be greater in larger houses than smaller houses and additionally, may be greater in homes rather than townhouses. If a neighborhood has larger houses then its variance will be greater than other neighborhoods. If that same neighborhood has a greater proportion of homes to townhouses than the average, then this non-constant variance will be further exaggerated. Multilevel models can control for heteroscedasticity in the level-1 residual by expanding the random part of the model with an additional random term for floor size. Each level-1 coefficient can be allowed to vary across neighborhoods either randomly, through the interaction with level-2 variables or through both of these options (Orford 2000).

Multilevel models are not the only means of addressing issues of spatial dependencies and non-stationarity. Spatial autoregression models (Anselin 2003; Dubin 1998; Paez and Scott 2004) can be utilized to compensate for the problem of biased coefficients resulting from spatial lags and inefficient standard errors from spatially autocorrelated errors. Geographically-Weighted Regression (Brunsdon et al. 1996; Fotheringham et al. 2002) can be used to examine the issue of non-stationarity. The advantage of using a multilevel approach for property hedonic models is that they can analyze variables from different scales simultaneously, while also accounting for issues of spatial dependencies and non-stationarity. Individual properties are nested within neighborhoods, with both individual and neighborhood characteristics explaining the variation in house price.

1.6. References

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

Allen, T.F.H., and Hoekstra, T.W. 1992. Toward a unified ecology. University of Columbia Press, NY, NY.

Allen, T.F.H., and Starr, T.B. 1982. Hierarchy: Perspectives for ecological complexity. The University of Chicago Press, Chicago.

Anderson, S.T., and West, S.E. 2006. Open space, residential property values, and spatial context. Reg. Sci. Urban Econ. 36: 773–789.

Anselin, L. 1988. Spatial econometrics : methods and models. Kluwer Academic Publishers London.

Anthon, S., Thorsen, B.J., and Helles, F. 2005. Urban-fringe afforestation projects and taxable hedonic values. Urban Forestry and Urban Greening 3: 79–91.

Assessment, M.E. 2003. Ecosystems and human well-being: A framework for assessment. Report of the Conceptual Framework Working Group of the Millennium Ecosystem Assessment.

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. Ecol. Econ. 40: 337-349.

Beland, F., Birch, S., and Stoddart, G. 2002. Unemployment and health: Contextual-level influences on the production of health in populations Social Science and Medicine 55(11): 2033-2052.

Benson, E.D., Hansen, J.L., Schwartz, A.L., and Smersh, G.T. 1998. Pricing residential amenities: The value of a view. Journal of Real Estate Finance and Economics 16(1): 55-73.

Bhat, C.R. 2000. A multi-level cross-classified model for discrete response variables. Transp. Res. Pt. B-Methodol. 34: 567-582.

Bin, S., and Polasky, S. 2004. Effects of flood hazards on property values: Evidence before and after Hurricane Floyd. Land Econ. 80: 490-500.

Bolitzer, B., and Netusil, N.R. 2000. The impact of open spaces on property values in Portland, Oregon. J. Environ. Manage. 59: 185-193.

Borjas, G.J. 1998. To Ghetto or Not to Ghetto: Ethnicity and Residential Segregation. Journal of Urban Economics 44: 228-253.

Bourassa, S.C., Hoesli, M., and Peng, V.S. 2003. Do housing submarkets really matter? Journal of Housing Economics 12: 12–28.

Bourassa, S.C., Hoesli, M., and Sun, J. 2004. What's in a view? Environ. Plan. A 36(8): 1427-1450.

Box, G.E.P., and Cox, D.R. 1964. An analysis of transformations. Journalof the Royal Statistical Society, Series B (Methodological) 26: 211–252.

Brasington, D.M. 1999. Which measures of school quality does the housing market value? The Journal of Real Estate Research 18(3): 395-413.

Brown, K.H., and Uyar, B. 2004. A hierarchical linear model approach for assessing the effects of house and neighborhood characteristics on housing prices. Journal of Real Estate Practice and Education 7(1): 15-23.

Browning, C.R., Feinberg, S.L., and Dietz, R.D. 2004. The paradox of social organization: Networks, collective efficacy, and violent crime in urban neighborhoods. Social Forces 83(2): 503-534.

Carson, R.T., Mitchell, R.C., Hanemann, W.M., and Kopp, R.J. 1999. A Contingent Valuation study of lost passive-use values resulting from the Exxon Valdez oil spill. Report to the Attorney General of the State of Alaska.

Clapp, J.M., A., N., and Ross, S.L. 2007. Which school attributes matter? The influence of school district performance and demographic composition on property values. Journal of Urban Economics 63: 451-466.

Clapp, J.M., and Giaccotto, C. 1998. Residential hedonic models: A rational expectations approach to age effects. Journal of Urban Economics 44(3): 415-437.

Cliff, A., and Ord, J.K. 1981. Spatial processes: Models and applications. Pion, London.

Costanza, R., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P., Van den Belt, M., D'Arge, R., De Groot, R., and Farber, S. 1997. The value of the world's ecosystem services and natural capital. Nature 387(6630): 253-260.

Coulson, N.E. 1991. Really useful tests of the mono-centric model. Land Econ. 67(3): 299-307.

Cropper, M.L., Deck, L.B., and McConnell, K.E. 1988. On the choice of functional form for hedonic price functions. The Review of Economics and Statistics 70: 668–675.

Day, B., Bateman, I., and Lake, I. 2004. Nonlinearity in hedonic price equations: An estimation strategy using model-based clustering. University of East Anglia: Centre for Social and Economic Research on the Global Environment (CSERGE).

de Groot, R.S., Wilson, M.A., and Boumans, R.M.J. 2002. A typology for the classification, description and valuation of ecosystem functions, goods and services. Ecol. Econ. 41(3): 393-408.

Deaton, B.J., and Hoehn, J.P. 2004. Hedonic analysis of hazardous waste sites in the presence of other urban disamenities. Environ. Sci. Policy 7: 499–508.

Dehring, C., and Dunse, N. 2006. Housing density and the effect of proximity to public open space in Aberdeen, Scotland. Real Estate Economics 34(4): 553 - 566.

Des Rosiers, F., Thériault, M., and Villeneuve, P. 2000. Sorting out access and neighborhood factors in hedonic price modeling. Journal of Property Investment and Finance 18(3): 291-315.

Diez-Roux, A.V. 2002. A glossary for multilevel analysis. Journal of Epidemiology and Community Health 56: 588–594.

Diez-Roux, A.V., Link, B.G., and Northridge, M.E. 2000. A multilevel analysis of income inequality and cardiovascular disease risk factors. Social Science & Medicine 50: 673-687.

Downes, T.A., and Zabel, J.E. 2002. The impact of school characteristics on house prices: Chicago 1987–1991. Journal of Urban Economics 52: 1-25.

Ekeland, I., Heckman, J.J., and Nesheim, L. 2002. Identifying hedonic models. American Economic Review 92(2): 304-309.

Fone, D.L., and Dunstan, F. 2006. Mental health, places and people: A multilevel analysis of economic inactivity and social deprivation. Health & Place 12: 332-344.

Fotheringham, A.S., Brunsdon, C., and Charlton, M.E. 2002. Geographically Weighted Regression: The analysis of spatially varying relationships. Wiley, Chichester.

Freeman, A.M. 2003. The measurement of environmental and resource values: Theory and methods. Resources for the Futures, Washington, D.C.

Garrod, G.D., and Willis, K.G. 1992. Valuing goods' characteristics: an application of the hedonic price method to environmental attributes. J. Environ. Manage. 34(1): 59-76.

Gayer, T., Hamilton, J.T., and Viscusi, W.K. 2000. Private values of risk tradeoffs at Superfund sites: Housing market evidence on learning about risk. The Review of Economics and Statistics 82(3): 439-451.

Gelfand, A.E., Banerjeeb, S., Sirmans, C.F., Tu, Y., and Ong, S.E. 2007. Multilevel modeling using spatial processes: Application to the Singapore housing market. Comput. Stat. Data Anal. 51: 3567 – 3579.

Geoghegan, J. 2002. The value of open spaces in residential land use. Land Use Pol. 19: 91-98.

Geoghegan, J., Pritchard, L., Ogneva-Himmelberger, Y., Chowdhury, R.R., Sanderson, S., and Turner, B.L. 1998. Socializing the pixel and pixelizing the social in land-use and land-cover-change. In People and pixels. Edited by D. Liverman, Moran, E., Rindfuss, R. and Stern, P. National Academy Press, Washington, DC.

Geoghegan, J., Wainger, L.A., and Bockstael, N.E. 1997. Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS. Ecol. Econ. 23(3): 251-264.

Gibbons, S. 2004. The costs of urban property crime. The Economic Journal 114: 441-463.

Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D., Thomas, S. 1993. A multilevel analysis of school examination results. Oxford Review of Education 19: 425–433.

Goodman, A.C., and Thibodeau, T.G. 1995. Age-related heteroskedasticity in hedonic house price equations. Journal of Housing Research 6(1): 25-42.

Goodman, A.C., and Thibodeau, T.G. 1998. Housing market segmentation. Journal of Housing Economics 7: 121–143.

Halstead, J.M., Bouvier, R.A., and Hansen, B.E. 1997. On the issue of functional form choice in hedonic price functions: Further evidence. Environ. Manage. 21(5): 759–765.

Hein, L., van Koppen, K., de Groot, R.S., and van Ierland, E.C. 2006. Spatial scales, stakeholders and the valuation of ecosystem services. Ecol. Econ. 57(2): 209-228.

Hite, D., Chern, W., Hitzhusen, F., and Randall, A. 2001. Property-value impacts of an environmental disamenity: The case of landfills. Journal of Real Estate Finance and Economics 22(2): 185-202.

Hoshino, S. 2001. Multilevel modeling on farmland distribution in Japan. Land Use Pol. 18(1): 75-90.

Jones, K., and Bullen, N. 1993. A multi-level analysis of the variations in domestic property prices. Urban Studies 30(8): 1409-1427.

Jones, K., and Duncan, C. 1996. People and places: the multilevel model as a general framework for the quantitative analysis of geographical data. In Spatial Analysis: Modelling in a GIS Environment Edited by P. Longley, Batty, M. John Wiley & Sons, New York. p. 389.

Kestens, Y., Theriault, M., and Des Rosiers, F. 2004. The impact of surrounding vegetation and land use on single-family house prices. Environment and Planning B: Planning and Design 31: 539-567.

Kim, C.W., Phipps, T.T., and Anselin, L. 2003. Measuring the benefits of air quality improvement: A spatial hedonic approach. J.Environ.Econ.Manage. 45(1): 24-39.

Klassen, A.C., Kulldorff, M., and Curriero, F. 2005. Geographical clustering of prostate cancer grade and stage at diagnosis, before and after adjustment for risk factors. International Journal of Health Geographics 4(1): 1-16.

Knight, J.R., and Sirmans, C.F. 1996. Depreciation, Maintenance, and Housing Prices. Journal of Housing Economics 5: 369–389.

Langford, I.H., Bentham, G., and McDonald, A. 1998. Multilevel modelling od geographically aggregated health data: A case study on malignant melanoma mortality and UV exposure in the European community. Statistics in Medicine 17: 41-57.

Lansford, N.H., and Jones, L.L. 1995. Recreational and aesthetic value of water using hedonic price analysis. J. Agric. Resour. Econ. 20(2): 341-355.

Lauridsen, J., and Kosfeld, R. 2006. A test strategy for spurious spatial regression, spatial nonstationarity, and spatial cointegration. Papers in Regional Science 85(3): 363-377.

Lee, C., and Linneman, P. 1998. Dynamics of the greenbelt amenity effect on the land market: The case of Seoul's greenbelt. Real Estate Economics 26(1): 107–129.

Leggett, C.G., and Bockstael, N.E. 2000. Evidence of the effects of water quality on residential land prices. J.Environ.Econ.Manage. 39(2): 121-144.

Lewis, L.Y., Bohlen, C., and Wilson, S. 2008. Dams, Dam removal, and river restoration: A hedonic property value analysis. Contemporary Economic Policy 26(2): 175–186.

Lutzenhiser, M., and Netusil, N.R. 2001. The effect of open spaces on a home's sale price. Contemporary Economic Policy 19(3): 291–298.

Lynch, A.K., and Rasmussen, D.W. 2001. Measuring the impact of crime on house prices. Appl. Econ. 33: 1981-1989.

Mahan, B.L., Polasky, S., and Adams, R.M. 2001. Valuing urban wetlands: A property price approach. Land Econ. 76(1): 100–113.

Mauny, F., Viel, J.F., Handschumacher, P., and Sellin, B. 2004. Multilevel modelling and malaria: a new method for an old disease. International Journal of Epidemiology 33: 1337-1344.

McCluskey, J.J., Rausser, G.C. . 2003. Hazardous waste sites and housing appreciation rates Hazardous waste sites and housing appreciation rates 45(2): 166-176.

McDonald, J.F., and McMillen, D.P. 1990. Employment subcenters and land values in a polycentric urban area: The case of Chicago. Environ. Plan. A 22(12): 1561-1574.

McLeod, H., Langford, I.H., Jones, A.P., Stedman, J.R., Day, R.J., Lorenzoni, I., and Bateman, I.J. 2000. The relationship between socio-economic indicators and air pollution in England and Wales: Implications for environmental justice. Regional Environmental Change 1(2): 78-85.

McMillen, D.P., and McDonald, J.F. 1998. Suburban subcenters and employment density in metropolitan Chicago. Journal of Urban Economics 43: 157-180.

Mitchell, R.C., and Carson, R.T. 1989. Using surveys to value public goods: The contingent valuation method. Resources for the Future, Washington, D.C.

Moon, G., Barnett, R. 2003. Spatial scale and the geography of tobacco smoking in New Zealand: A multilevel perspective New Zealand Geographer 59(2): 6-15.

Mooney, S., and Eisgruber, L.M. 2001. The Influence of riparian protection measures on residential property values: The case of the Oregon plan for salmon and watersheds. Journal of Real Estate Finance and Economics 22(2/3): 273-286.

Morancho, A.B. 2003. A hedonic valuation of urban green areas. Landscape and Urban Planning 66: 35-41.

Mueller, J.M., and Loomis, J.B. 2008. Spatial dependence in hedonic property models: Do different corrections for spatial dependence result in economically significant differences in estimated implicit prices? J. Agric. Resour. Econ. 33(2): 212-231.

Mujahid, M.S., Diez-Roux, A.V., Morenoff, J.D., and Raghunathan, T. 2007. Assessing the Measurement Properties of Neighborhood Scales: From Psychometrics to Ecometrics. American Journal of Epidemiology 165: 858 - 867.

Muller, N.Z. 2009. Using hedonic property models to value public water bodies: An analysis of specification issues. In Water Resour. Res.

O'Neill, R.V., Deangelis, D.L., Waide, J.B., and Allen, T.F.H. 1986. A hierarchical concept of ecosystems. Princeton University Press, Princeton, N.J.

Openshaw, S., and Taylor, P.J. 1979. A million or so correlation coefficients: Three experiments on the Modifiable Areal Unit Problem. In Statistical Methods in the Spatial Sciences. Edited by N. Wrigley. Pion, London. pp. 127–144.

Orford, S. 2000. Modelling spatial structures in local housing market dynamics: A multilevel perspective. Urban Studies 37(9): 1643-1671.

Orford, S. 2002. Valuing locational externalities: A GIS and multilevel modelling approach. Environment and Planning B: Planning and Design 29(1): 105-127.

Ottensmann, J.R., Payton, S., and Man, J. 2008. Urban location and housing prices within a hedonic model. Regional Analysis and Policy 38(1): 19-35.

Overmars, K.P., Verburg, P.H. 2006. Multilevel modelling of land use from field to village level in the Philippines. Agricultural Systems 89: 435-456.

Paez, A., and Scott, D.M. 2004. Spatial statistics for urban analysis: A review of techniques with examples. GeoJournal 61: 53–67.

Palmquist, R.B. 1992. Valuing localized externalities. Journal of Urban Economics 31(59-68).

Pan, W.K.Y., Bilsborrow, R.E. . 2005. The use of a multilevel statistical model to analyze factors influencing land use: A study of the Ecuadorian Amazon Global and Planetary Change 47: 232-252.

Paterson, R.W., and Boyle, K.J. 2002. Out of sight, out of mind? Using GIS to incorporate visibility in hedonic property value models. Land Econ. 78(3): 417-425.

Polsky, C., and Easterling, W.E. 2001. Adaptation to climate variability and change in the US Great Plains: A multi-scale analysis of Ricardian climate sensitivities. Agriculture, Ecosystems and Environment 85: 133–144.

Poor, P.J., Pessagnob, K.L., and Paul, R.W. 2007. Exploring the hedonic value of ambient water quality: A local watershed-based study. Ecol. Econ. 60: 797-806.

Raudenbush, S.W., and Bryk, A.S. 2002. Hierarchical Linear Models: Applications and Data Analysis Methods. Sage Publications, London.

Raudenbush, S.W., Rowan, B., King,S.J. 1991. A multilevel, multivariate model for school climate with estimation via EM algorithm and application to US high school data. Journal of Educational Statistics 16: 295-330.

Reginster, I., and Edwards, G. 2001. The concept and implementation of perceptual regions as hierarchical spatial units for evaluating environmental sensitivity. URISA Journal 13(1): 5-16.

Ridker, R.G., and Henning, J.A. 1967. The determinants of residential property values with special reference to air pollution. The Review of Economics and Statistics 49: 246-257.

Robert, S.A., Reither, E.N. 2004. A multilevel analysis of race, community disadvantage, and body mass index amongadults in the US. Social Science & Medicine 59: 2421-2434.

Rosen, S. 1974. Hedonic prices and implicit markets: product differentiation in pure competition. The Journal of Political Economy 82(1): 34–55.

Sedgley, N.H., Williams, N.A., and Derrick, F.W. 2008. The effect of educational test scores on house prices in a model with spatial dependence. Journal of Housing Economics 17: 191-200.

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

Snijders, T.A.B., and Bosker, R. 1999. Multilevel analysis. An introduction to basic and advanced multilevel modeling. Sage, Thousand Oaks, CA.

Soobader, M., Cubbin, C., Gee, G.C., Rosenbaum, A., Laurenson, J. 2006. Levels of analysis for the study of environmental health disparities. Environmental Research 102: 172-180.

Steenbergen, M.R., and Jones, B.S. 2002. Modeling multilevel data structures. Am. J. Polit. Sci. 46(1): 218-237.

Subramanian, S.V., Duncan, C., Jones, K. 2001. Multilevel perspectives on modeling census data. Environ. Plan. A 33(3): 399-417.

Thorsnes, P. 2002. The value of a suburban forest preserve: Estimates from sales of vacant residential building lots. Land Econ. 78(3): 426–441.

Troy, A., and Grove, J.M. 2008. Property values, parks, and crime: A hedonic analysis in Baltimore, MD. Landscape and Urban Planning 87: 233–245.

Turner, M.G., Dale, V.H., and Gardner, R.H. 1989. Predicting across scales: Theory development and testing. Landscape Ecology 3: 245-252.

Tyrväinen, L., and Miettinen, A. 2000. Property prices and urban forest amenities. J.Environ.Econ.Manage. 39(2): 205–223.

Western, B. 1998. Causal heterogeneity in comparative research: A Bayesian hierarchical modeling approach. Am. J. Polit. Sci. 42: 1233-1259.

Wilhelmsson, M. 2000. The impact of traffic noise on the values of single-family houses. Journal of Environmental Planning and Management 43(6): 799–815.

CHAPTER 2: A PRIMER ON MULTILEVEL, PROPERTY HEDONIC MODELING

2.1. Abstract

The property hedonic model estimates implicit prices for different structural, neighborhood and environmental attributes that are related to a property. A multilevel approach to the hedonic model accounts for the spatial effects of dependencies, non-stationarity and scale that are often not captured in typical hedonic regressions. This study provides a methodological review of multilevel modeling of property hedonics, using properties in the city of Baltimore as a practical example of this approach. Issues specific to valuing environmental attributes are also considered in this study.

2.2. Introduction

Many beneficial aspects of the environment, known as ecosystem goods and services, are external to normal market transactions and consequently, are often undervalued and under-provisioned even though they affect the quality of people's lives. Conversely, the impacts of many environmental disamenities, or negative externalities, such as air and water pollution, are over-produced when their costs are not captured in the market. The non-market, economic costs and benefits of these environmental amenities and disamenities can be captured through revealed-preference techniques such as the property hedonic model. In contrast to stated-preference techniques, property hedonic methods use market transactions to estimate the implicit use-values of environmental attributes. Use values for environmental attributes, as the name implies, pertain only to amenities such as recreation and aesthetics in contrast to other, non-use environmental benefits such as waste regulation, maintaining biodiversity and carbon sequestration (Freeman 2003).

The hedonic approach for estimating the benefits or impacts of the environment has been the focus of research in numerous locations and circumstances. Some environmental attributes that have recently been examined are: urban parks (Bolitzer and Netusil 2000; Morancho 2003; Orford 2002; Troy and Grove 2008); greenbelts (Lee and Linneman 1998); forest preserves (Garrod and Willis 1992; Thorsnes 2002; Tyrväinen and Miettinen 2000); wetlands (Mahan et al. 2001); agriculture (Bastian 2002); water quality (Leggett and Bockstael 2000; Poor et al. 2007) and river restoration (Lewis et al. 2008). Others have examined the effects of negative environmental externalities such as: hazardous waste sites (Deaton and Hoehn 2004; McCluskey 2003); landfills (Hite et al. 2001); superfund sites (Gayer et al. 2000); air pollution (Kim et al. 2003) and flood risk (Bin and Polasky 2004).

While much of this more recent research has attempted to control for or explicitly model issues of spatial scale such as spatial autocorrelation (dependency) and non-stationarity (heterogeneity in the dependent-independent variable relationship) and heteroscedasticity (systematic patterns in variance), few have attempted to account for the presence of attributes simultaneously occurring at and interacting across multiple spatial scales. Explicitly modeling multiple scales is useful for accounting for these statistical issues.

The foundation for the property hedonic model was presented by Rosen (1974) who showed the existence of a property market equilibrium where consumers and suppliers maximize their respective utility and profits by choosing to purchase and produce properties with distinct combinations of desirable attributes. While the value of each attribute is implicit and therefore not directly observed in the property transaction, the marginal economic contribution of each of these attributes to the total transaction can be estimated from a regression model using property sales data from an area with varying combinations of these attributes. These characteristics can be broadly grouped into three categories: property-specific (including both the land and structural improvements); neighborhood-specific (the socio-economic context); and locational (Freeman 2003). Structural attributes refer to the tangible qualities of a dwelling and parcel such as lot size, house size (square footage), quality of construction, age, number of bathrooms. Contextual attributes are those shared by a neighborhood such as crime rate, ethnicity, income and other socio-economic factors. Locational attributes refer to the proximity and/or accessibility of various land uses and buildings such as hospitals, schools, highways, industrial areas, shopping centers as well as proximity to environmental amenities such as urban green spaces of parks and golf courses.

However, the property hedonic model is challenged by statistical issues which complicate its implementation and interpretation. The model may be mis-specified as a result of the functional form of the model, collinearity among variables and omitted determinants of property value. There are also spatial issues of dependency, non-stationarity and scale that exist in a housing market composed of interrelated sub-markets.

Rosen's (1974) development of the property hedonic model assumed that both supply and demand factors were mobile and elastic and that an entire city could be viewed as having a single housing market in equilibrium. Equilibrium occurs when the market settles on a hedonic price supply-demand curve that ensures households (within their budget constraints) cannot increase their utility by choosing a different property and sellers cannot increase their profits by increasing the property's price or changing its characteristics. With this assumption, the price of a property and the availability and contribution of its constituent characteristics are invariant across geographic space. Since Rosen's work, most researchers have found that housing markets are typically not in equilibrium and that the assumption of a single market is unrealistic except for very small study areas. Once a house is built, its characteristics are fixed (ignoring the potential for costly renovation) and it likely shares similar characteristics as the surrounding properties. Although less locationally fixed than house supply, consumer preferences can also create market segmentation. People of different ethnicities, generations, incomes or social classes may desire particular combinations of property and locational characteristics when seeking to purchase a residence. Consumer driven segmentation can be further exaggerated through information constraints such as the tendency of the real estate agent to profile a potential buyer (racially or by income status) and selectively present properties of specific types or in specific types of neighborhoods depending on the characteristics of the potential buyers (Orford 2000).

While the existence of housing submarkets is commonly accepted, there has been little agreement on the method of segmenting areas into housing submarkets. Some researchers segment the market by sectors, that is, differentiating the housing stock by structural attributes such as dwelling type (Ekeland et al. 2002). Other researchers focus on markets based on locational-contextual attributes such as income classes or ethnicities. Orford (2000) argues that rather than distinguishing between sector-defined or context-defined submarkets, there is a joint influence of sector and context attributes that should be modeled simultaneously. Day and others (2004) used aspatial clustering methods to group properties as a function of either the housing stock/structural attributes or by neighborhood characteristics. They found that the property market was best segmented by socio-economic characteristics rather than housing stock. Even when segmenting by context rather than sector, there has been little consistency of the spatial units used to define these sub-markets. Many researchers use Census units to delineate the boundaries. Bourassa and others (2003) found that geographic, sub-market boundaries defined by real estate appraisers provided a better model than aspatial clustering techniques. Goodman and Thibodeau (1998) found that school districts were appropriate for determining sub-markets.

This issue of market segmentation raises spatial statistical concerns of spatial dependency, non-stationarity and inappropriate scale for the property hedonic model. Spatial dependency (association or correlation) refers to the likelihood that the values of observations for a particular variable become more similar with spatial proximity. Since many houses are built as part of a development, the housing stock will be very similar for neighboring houses, which leads to spatial dependencies in the regression model. Real estate agents and appraisers also price and assess homes based on the value and characteristics of nearby homes, leading to further spatial dependency in house price and important characteristics such as house size (Orford 2000). Since residuals capture unexplained variation in the model, spatial error autocorrelation reveals the existence of a spatial association that has not been incorporated into the model (Paez and Scott 2004). The problem with the presence of spatial error auto-correlation in a regression model is that the statistical assumption regarding the independent distribution of errors is violated. As a consequence of these spatial dependencies (lags and error), parameter estimates will be biased and inefficient, respectively (Anselin 1988). This leads to artificially smaller standard errors and the possibility of finding a spurious significance of an effect when one does not actually exist (Type I error). Much of this dependency can be attributed to and controlled for by market segmentation and the use of multilevel models as discussed in the section below. The Moran's I statistic (Cliff and Ord 1981) is one of several metrics that can be used to test whether the final model has sufficiently accounted for spatial error autocorrelation.

Non-stationarity refers to the existence of a heterogeneous relationship between dependent and independent variables across geographic space (Fotheringham et al. 2002). Global approaches to hedonic modeling such as using OLS, do not accommodate local, spatial variations in these relationships. A coefficient that is reported as insignificant within a global regression model may be the result of highly significant positive relationships cancelling out the effect of significant negative relationships in others areas. With respect to the effect of individual attributes on house price, it is often useful to map the existence of non-stationarity and to attempt to model the reason for its existence by allowing the coefficient to interact with other variables.

Issues with spatial dependencies and non-stationarity are further complicated by the scale at which attributes are measured or aggregated. If the values of these spatial properties change with the choice of unit used in a model, then the model exhibits scaling challenges that cannot be effectively modeled with standard regression techniques. Thus, the variance of the outcome, the relationship between the independent and dependent variable and the relationship between individual observations all may be sensitive to unit size.

With the multilevel approach, individual properties are nested within neighborhoods within a city (Brown and Uyar 2004; Gelfand et al. 2007; Goodman and Thibodeau 1998; Orford 2000, 2002). These models allow the simultaneous examination of the effects of group-level and individual-level variables on individual-level outcomes (the property price) while accounting for the non-independence of observations within groups (the neighborhoods). Multilevel models were initially developed by educational researchers to examine the effects of context, such as classroom and school characteristics, on individual scholastic achievement (Goldstein 1993; Raudenbush 1991). More recently, the use of multilevel models has grown rapidly in geographic-orientated research such as relationships of socio-economic conditions and individual health (Beland et al. 2002; Diez-Roux et al. 2000; Fone and Dunstan 2006; Klassen et al. 2005; Langford et al. 1998; Moon 2003; Mujahid et al. 2007; Robert 2004; Soobader 2006; Subramanian 2001), epidemiology (Mauny et al. 2004), environmental justice (McLeod et al. 2000), criminal activity (Browning et al. 2004) and land use (Hoshino 2001; Overmars 2006; Pan 2005).This is an important development because geographical analyses are intrinsically spatial and involves the grouping of elementary units of interest (e.g. households or individuals) into higher-level spatial geographies such as neighborhoods (Orford 2000), communities (Moon 2003) or travel analysis zones (Bhat 2000).

There are several statistical reasons for explicitly modeling this clustering of individual properties within neighborhoods. First, spatial autocorrelation of independent variables and error autocorrelation (spatial dependencies) are likely to be common in property hedonic analyses since individual properties in the same neighborhood are likely to be similar in ways not fully accounted for by the property and neighborhood variables included in a single-level model (Jones and Bullen 1993). The house price and other important house characteristics are included at the neighborhood level to control for dependencies in house price and characteristics (spatial lag processes). Multilevel models account for the spatial error autocorrelation (dependence of the residuals) by differentiating between-individual errors from between-neighborhood errors (Orford 2000). If this dependency was not modeled, the standard errors of the independent variables would be biased downwards (underestimated), which results in spuriously significant effects (Snijders and Bosker 1999).

Second, it is possible that an average effect for a particular attribute does not represent local conditions that may occur in different areas of a study area; the issue of non-stationarity in property hedonic studies (Cho et al. 2006; Troy and Grove 2008). A possible outcome of estimating an average effect, also known as a global effect, is that contrasting relationships in different areas of the study may cancel each other out, leading to a globally insignificant estimate for that variable's coefficient (Fotheringham et al. 2002). By considering a variable as a random effect, instead of a global fixed effect, that variable is allowed to vary across neighborhoods. While it is possible to build OLS regressions to determine this effect for each neighborhood, there will be the problem that there are insufficient observations within any given neighborhood leading to unreliable estimates. In multilevel models, random effects are constructed with Empirical Bayesian techniques which borrow strength across neighborhoods and shrink or smooth estimates for unreliable neighborhoods (those with few observations) towards the overall mean (Raudenbush and Bryk 2002). This means that the implicit prices of certain attributes (those that are considered random effects) are optimally weighted averages that combine information derived from the group itself with the mean from all other neighborhoods (Diez-Roux 2002). While unreliable submarket estimates are differentially shrunk towards the global estimate, submarkets with many properties will not be affected by this shrinkage. The Empirical Bayesian variation in slopes between neighborhoods is added to the global fitted value to reveal areas of positive and negative relationships and is one of the fundamental advantages of multilevel modeling (Subramanian 2001).

Third, heterogeneity (varying relationships) between neighborhoods needs to be distinguished from the heterogeneity among individual properties. Ignoring this differentiation and modeling the behavior of interest at a single level invites the pitfalls of either the ecological or atomistic fallacy. Atomistic and ecological fallacies are avoided because the predictors and unexplained variation are modeled at the appropriate level (Jones and Duncan 1996). When group-level data, such as Census data, is included in an OLS model whose focus is the individual, the interpretation of the results may lead to the ecological fallacy. With this situation, data (and the processes they measure) that are collected at a broader scale are assumed to have the same importance, associations and interactions at the individual level as at the higher level. Conversely, the atomistic fallacy refers to inferences about the significance, associations and the variability of higher, group levels are based on data originating from individuals (Allen and Starr 1982; O'Neill et al. 1986). This fallacy can also refer to higher-level data that is simply an aggregation of individual-level observations (Hox 2002). For example, the association between price and house size or age at the property level may differ from the association between these variables that have been aggregated to the group level. With multilevel models, the variation in house price due to the effects of individual-level, property attributes is separated from price differences between areas that are related to neighborhood-level contextual effects. With multilevel models, the user is not forced to model at one level or the other, both (indeed several) levels are be modeled simultaneously.

Hedonic models also often suffer from heteroscedasticity (unequal variation in the residuals) across neighborhoods, leading to inefficient estimates(Goodman and

Thibodeau 1995). Heteroscedasticity can be caused by omitting important variables from the models or from the modeled individual property characteristics and potential interactions. For instance, the variance associated with the implicit price of floor area may be greater in larger houses than smaller houses and additionally, may be greater in homes rather than townhouses. If a neighborhood has larger houses then its variance will be greater than other neighborhoods. If that same neighborhood has a greater proportion of homes to townhouses than the average, then this non-constant variance will be further exaggerated. Multilevel models can control for heteroscedasticity in the level-1 residual by expanding the random part of the model with an additional random term for floor size. Each level-1 coefficient can be allowed to vary across neighborhoods either randomly, through the interaction with level-2 variables or through both of these options (Orford 2000).

2.3. Objectives

This study reviews the steps involved in building a multilevel property hedonic model using property data from the City of Baltimore as an example. Modeling the spatial effects of dependencies, non-stationarity and scale are discussed. Statistical assumptions that are necessary for all regression models are also reviewed. I also compare OLS and multilevel models for differences in structural and socio-economic parameters, their standards errors and the model residuals. This will demonstrate how a multilevel model avoids the violation of assumptions inherent to an OLS model. Finally, I discuss some issues that are specific to the valuation of environmental attributes.

2.4. Data

Property sales and attributes were obtained from the MD Property View 2004 database, a private company which compiles sales transaction data with a property's location and structural characteristics from the state of Maryland's property-appraisal database. Transactions for the city of Baltimore for a 5-year period (1998-2002) were used in this analysis.

Selling prices were standardized to the year 2000 with the OFHEO (Office of Federal Housing Enterprise Oversight) housing price index for the Baltimore Metropolitan Statistical Area. This index accounted for both annual and seasonal (quarterly) fluctuations of property sales. This standardization removed the need for adding dummy variables for year and season while allowing for a sufficiently large dataset of properties that would be consistent with the 2000 Census attributes used to describe the neighborhood.

Property records were selected if they followed numerous criteria, whose values were specified within the property database: "arms-length" transactions only; single, detached homes or townhouses; total, appraised property value was within ±50% agreement of the selling price; zoning was classified as either residential or residential-commercial; and values of key variables used in the analyses were not missing. The total square footage for each house accounted for floor area for each story, excluding basements and attics. Although lot size (land area) was available in the property database, these values were recalculated within a GIS through the association of the individual transaction records with their corresponding, spatially-delineated parcels. These properties were also re-located to the center of their parcels in order to adjust for location errors in the property database. Basements and garage sizes were converted to dummy variables (presence or absence). The quality of the original construction was converted to 3 dummy variables of poor, average and high quality. The number of bathrooms and half-bathrooms were recombined into one attribute (e.g. one full bath and one half bath equals "1.5"). Following the example of Cho (2006) and Troy and Grove (2008), records with low property prices (less than $50,000 in this case) were considered as either database errors or non, arms-length transactions and were excluded from the analysis. Additionally, records with house size or lot size less than 500 ft^2 were considered to be database errors and were excluded. These and other property variables as well as their means and range are listed in Table 2.1. All of these variables, except for age, are expected to have a positive impact on property price.

Block group attributes from the Census 2000 were used as proxies for neighborhood characteristics that nearby properties shared. Median household income and percent unemployment were used to capture the relative economic status of a neighborhood. While median household income is expected to provide a positive contribution to property price, percent unemployment is expected to be negative. Median house value was expected to capture some of the spatial dependency in price that nearby properties shared. The percent of the neighborhood having a high school diploma or college degree was used as a proxy of the social status of the neighborhood. A well educated neighborhood is expected to be associated with higher property values. Population density (per hectare) was a measure of the demand for and relative scarcity of land available for development. Higher population densities are expected to drive land price (and overall property prices) higher. Mean travel time to work (in minutes) was used as a proxy for distance to the nearest employment center.

Only Census block groups with at least 5 sales transactions during the 5-year period were included in this research. This resulted in a final dataset of 13,793 properties distributed among 405 block groups.

Table 2.1. Regression Variables

Variables	Description	Min	Max	Mean
Property				
Price00[ab]	Sale price converted to Yr 2000	$50,000	$1,238,298	$107,615
EnclsFt[b]	Total area of interior (square ft.)	510	10,185	1,475
LandHa[b]	Lot size of property (hectares)	0.002	2.04	0.04
Bathnum	Number of full and half baths	1.0	10.0	1.6
Age[c]	Age of house at the time of sale	0	199	67
HouseDum	Detached home vs. townhouse	0	1	0.3
AirDum	Presence central air-conditioning	0	1	0.3
Basedum	Presence of finished basement	0	1	0.4
FireDum	Presence of a fireplace	0	1	0.2
GarDum	Presence of garage or carport	0	1	0.2
QualDumAvg	Avg. quality of construction	0	1	0.2
QualDumHigh	High quality of construction	0	1	0.1
Block group				
pUnemploy	Percent unemployment	0.1	19.3	4.9
MedValHouse	Median house value	$32,500	$596,300	$91,249
pHSDiploma	Percent with high school diploma	15.8	63.4	39.8
TravelMean	Mean, travel time to work (min)	17	53	31
PopDens[b]	Population density (per hectare)	1.3	190.2	50.9
pVacancy[b]	Percent vacant residences	1.1	65.1	9.8
MedHsInc[b]	Median household income	$12,095	$170,428	$38,858

a:dependent variable, b:natural log transformation, c:quadratic transformation

2.5. Review of Issues General to all Property Hedonic Models

2.5.1. Functional Form

In estimating a relationship between environmental amenities and property prices, the choice of functional form is not always clear. Rosen (1974) stressed that economic theory fails to indicate that any particular form is appropriate and a variety of functional forms have been used in the hedonic literature. A linear form assumes that an individual's preferences are linear, implying that perfect repackaging of property characteristics is possible (Freeman 2003). However, in property markets, individual house characteristics are inseparable; an individual cannot mix characteristics in any other level than is already available in each house (Garrod and Willis 1992). Also, since the price function is an equilibrium relationship determined in the marketplace by the interactions of individual buyers and sellers (Taylor 2003), the existence of a linear relationship is unlikely.

Early research tried using alternative forms such as the log-linear or double-log forms where the best form was chosen based on the goodness of fit (Freeman 2003). While the form that is chosen should ideally improve the model fit and help to satisfy important assumptions of OLS regressions, such as normally distributed residuals and homoscedasticity, this is not the main issue with choosing functional forms. The goal of finding a proper functional form is to overcome problems associated with the non-linearity that is often found in hedonic regression equations (Goodman and Thibodeau 1995). Substantively, this means that the proper functional form should be chosen so that marginal value for any given property attribute does not vary across the range of house prices.

Halvorsen and Pollakowski (1981) recommended allowing both for transformation of the dependent variable and for different transformations of each independent variable using a quadratic Box-Cox transformation (Box and Cox 1964). However, Freeman (2003) suggests this approach is more cumbersome than necessary while Palmquist (1992) recommends only transforming the dependent variable and the independent variables that are the main effects. Cropper and others (1988) suggest that in models with missing or proxy variables, a common occurrence in property hedonic studies, that simple functional forms (linear, quadratic, log-log and log-linear) or linear Box-Cox transformations are preferred over quadratic Box-Cox forms. This is because omitted variable bias will affect more coefficients in the quadratic forms than in simpler functional forms. Halstead and others (1997) note that the choice of functional form can affect both variable significance and the magnitude of the coefficients.

Since the purpose of these estimated regression functions is to generate amenity values, it may be preferable to use a relatively simple form (Freeman 2003). A log-linear form allows the marginal effect of each independent variable to vary with the level of the dependent variable. Thus, the marginal effects of independent variables change as house price varies. The double-log form, in which both the dependent variable and the main effects are transformed using the natural logarithm, may provide the most interpretable results. With this form, a coefficient is interpreted as an elasticity; the percentage change in the dependent variable given the percentage change in an explanatory variable.

A Box-Cox transformation analysis can be used to provide guidance on whether such simple forms are adequate for satisfying regression assumptions. The Box-Cox transformation of the dependent variable is shown as:

$$y(\lambda) = \begin{cases} \frac{y^{\lambda}-1}{\lambda}, & \text{if } \lambda \neq 0 \\ \log y, & \text{if } \lambda = 0 \end{cases} \qquad \text{Eq. 2.1}$$

With this test, the parameter, λ, is estimated through maximum likelihood to find the optimal transformation of the dependent variable. This parameter can then be tested for significant differences between the optimal value of λ and three cases of λ that correspond to commonly used functional forms: a reciprocal transformation, where λ=-1; a log transformation, where λ=0; and a linear (untransformed) form, where λ=1. Some researchers have compared Box-Cox transformations with simpler functional forms and found that a log transformation of the dependent variable (semi-log) or of both dependent and independent variables (double-log) performed the best (Anderson and West 2006; Anthon et al. 2005).

In this research, a left-hand (LHS) Box–Cox test using Stata software tested for the optimal transformation for price and found a Lambda value of -0.4. A chi-square test found this to be significantly different from zero indicating that a natural log transformation was not an optimal transformation for the dependent variable. However, for the purposes of this review and for better interpretation of results, I use the log transformation for the dependent variable.

A log transformation was also used on the continuous property variables of house size and lot size. In the hedonic literature, this is a common method to account for the (non-linear) declining marginal value of house and lot size. Houses are also expected to depreciate with age at a declining rate but after a number of years, age will often be positively associated with house price. This may be due to unknown renovations or the "vintage effect" (Goodman and Thibodeau 1995) of older properties. Therefore, a quadratic transformation is used for the age variable. The inflection point where there was a positive effect of age on price was approximately 63 years old.

2.5.2. Testing for Collinearity and Other Regression Assumptions

The regression model was then checked for collinearity within an OLS regression. A general rule of thumb is that variance inflation factors (VIF) greater than 10 are thought to be highly correlated and should be cause for further assessment before proceeding (O'Brien 2007). This research found VIF's below 4 for all variables except Age and Age-Squared. However, Shieh and Fouladi (2003) found that even in the presence of multicollinearity, for level-1 variables, the fixed-effect parameter estimates produce relatively unbiased values. Diagnostics were also used to check for the normal distribution of residuals, the existence of homogeneity of variances in the residuals, and the potential for heteroscedasticity and/or non-linear trends in the independent variable-price relationship.

2.6. Methods and Results for Building Multilevel Models

While the analysis of multilevel models can be performed with a number of statistical packages, the details of using multilevel modeling with HLM (Hierarchical Linear Modeling version 6) software are discussed below and follow the work of Raudenbush and Bryk (2002). While it is possible to model more than just two levels of hierarchical data, this review focuses on the property level and the block group (neighborhood) level to illustrate the use of multilevel models.

2.6.1. Do Neighborhoods Matter?

In the first step of building a multilevel model, I examined whether property price varied among neighborhoods. Constructing a hierarchical model, unlike a simple linear model, explains variation in the dependent variable differently at different scales. By creating a “null” or “unconditional” model, I was able to examine the baseline variation in house price at each level in the absence of explanatory variables (Raudenbush and Bryk 2002). This determines the appropriateness of defining submarkets through multilevel regression versus using a single housing market with OLS regression to model property prices in Baltimore. The variation in house price is simply decomposed into variation at the property level and variation at the neighborhood level. Mathematically, this model (known as the “null” or “unconditional” model) is described as:

$$\text{Level 1: } Y_{ij} = ß_{0j} + r_{ij}$$
$$\text{Level 2: } ß_{0j} = \gamma_{00} + U_{0j} \qquad \text{Eq. 2.2}$$

Combined, the null model is:

$$Y_{ij} = \gamma_{00} + U_{0j} + r_{ij} \qquad \text{Eq.2.3}$$

Where Y_{ij} is house i's price in neighborhood j and each neighborhood has its own intercept ($ß_{0j}$) composed of a mean price (γ_{00}) and an error term (U_{0j}) with a between-group variance (τ_{00}) that is separate from the individual-level error (r_{ij}) and its within-group variance (σ^2). It is this additional, level-2 error term that accounts for much of the non-independence between the individual observations that are nested within the level-2 units. A finding of significant between-group variance (τ_{00}) indicates that it is preferable to use multilevel regression over the simpler OLS regression. This step is statistically equivalent to conducting a one-way ANOVA for determining whether there are significant differences between groups and is determined in HLM from a chi-square test statistic (Raudenbush and Bryk 2002). Therefore, finding significant group variance indicates that house prices do vary by neighborhoods within the city as previously suggested in the discussion on market segmentation.

Table 2.2 provides the estimates for both the fixed effects and the random effects in this model. The maximum likelihood estimate of the grand mean price (γ_{00}) is 11.345 ($84,577) with a standard error of 0.020. Under the assumption of normally distributed block group errors (U_{0j}), a 95% confidence interval for the block group mean price is between 10.556 and 12.132 ($38,522-$185,691) (specified by the equation: ($\boldsymbol{\gamma}_{00}$ / $1.96(\boldsymbol{\tau}_{00})^{1/2}$). This indicates that there is a substantial range in mean price among Baltimore's block groups. The random effects section of Table 2.2 shows the decomposition of the variance into its house-level and neighborhood-level components.

The *p*-value of 0.000 shows that the variance at the neighborhood level is statistically significant at better than the 1% level of significance.

Table 2.2. Results of the Null Model

Fixed Effect	**Coef**	**SE**	**P-value**
Intercept (γ_{00})	11.345419	0.020155	0.000
Random Effects	**Variance**	**Df**	
Single/L1 (r_{ij})	0.0667		
Group/L2 (U_{0j})	0.16099	404	0.000

2.6.2. Are Properties within a Neighborhood Correlated?

This null model is also useful for examining the amount of non-independence (spatial autocorrelation) of the unexplained variation in house prices and provides one justification for using multilevel models. If the variance of the level-2 intercept is zero, there is no autocorrelation and only a single-level model is needed (Orford 2000). A single-level model, such as an OLS regression, assumes that the data does not have a hierarchical structure, that all the relevant variation is at one scale, that there is no autocorrelation and that there is a single general relationship across space (Jones and Bullen 1993).

The degree of autocorrelation in multilevel models can be examined with the intra-class correlation coefficient (ICC), which is the ratio of variation at the higher level ($\tau 00$) to the total variation of all levels ($\tau 00 + \sigma^2$) (Diez-Roux 2002). A high coefficient indicates that individual observations within a group are much more similar (spatially correlated) than individuals between neighborhoods. As discussed previously, such correlation violates the OLS assumptions of independence of residuals with the potential for spurious significance of individual variables. Thus a high coefficient will indicate that it is worthwhile to model individual properties as belonging to a particular neighborhood. However, Kreft and de Leeuw (1998) argue that even a value as low as 0.05 will indicate that it would be beneficial to use multilevel models over OLS regressions.

Using information from Table 2.2, the resulting Intraclass Correlation Coefficient (0.16/(0.16 + 0.07)) shows the proportion of variance in house prices that is attributable to differences at the neighborhood level. The ICC of 0.71 indicates that 71% of the total variance in house price is due to differences at the block group level. This suggests that there is a much greater degree of similarity for properties within a block group than between them and therefore, the use of a multilevel model can be justified.

2.6.3. What are the Important Property-Specific Variables?

After determining that neighborhoods are in fact important in accounting for variation in house price, models can be built to explain the variation at both the property and neighborhood scales. Explaining the variation at the property level requires constructing a model ("structural attributes" model) with property-specific (level-1) variables. This model is similar to a one-way ANCOVA model except that the group effect is considered to vary randomly (Raudenbush and Bryk 2002) . First, individual variables are added sequentially and a significant reduction in the Deviance Statistic (-2 Log Likelihood) is examined to determine the most efficient model in terms of minimizing complexity (number of variables) while maximizing the explanatory power of the model. The Deviance Statistic is used to compare two models that differ only in the number of variables included in the equation. In HLM, a likelihood ratio test compares the Deviance Statistics of the two models and tests whether this difference is statistically significant. If it is, this indicates that the less restrictive model (the one with more variables) fits the data significantly better than the more restrictive model. This comparison test is distributed chi-squared with degrees of freedom equal to the difference in the number of degrees of freedom between the two models (i.e., the number of variables added to the model).

Because these sub-models have a different number of fixed effects (the independent variables), this test needs to be performed with Full Maximum Likelihood estimation (MLF) rather than restricted (MLR) (Hox 2002). These two estimation methods are discussed in a later section.

A lack of finding a significant effect of a particular variable may be due to non-stationarity in the relationship with the house price. This means that a significant reduction in the Deviance Statistic after adding a variable justifies the inclusion of that variable even if it does not show a significant effect. Maximum Likelihood estimation and the Deviance Statistic are discussed at the end of section describing the multilevel steps. Mathematically, the model that includes only level-1 variables is:

$$\begin{aligned} \text{Level 1: } Y_{ij} &= ß_{0j} + ß_{1j}X_{ij} + \ldots ß_{nj}X_{ij} + r_{ij} \\ \text{Level 2: } ß_{0j} &= \gamma_{00} + U_{0j} \\ ß_{1j} &= \gamma_{10} \end{aligned} \qquad \text{Eq. 2.4}$$

The results of this estimation are shown in Table 2.3. A chi-square test comparing each successive Deviance Statistic with the previous, simpler model's Deviance Statistic showed that the increase in the model's explanatory power did not come at the cost of an overly complex model. The estimates of these fixed effects are all significant at better than the 1% level (shown in the full model in Table 2.6).

Table 2.3. Level-1 Model

Fixed Effect	**Fixed Effect**	**Fixed Effect**	**Fixed Effect**	**Fixed Effect**
Intercept	Intercept	Intercept	Intercept	Intercept
	lnEnclsFt	lnEnclsFt	lnEnclsFt	lnEnclsFt
	LandHa	LandHa	LandHa	LandHa
	Bathnum	Bathnum	Bathnum	Bathnum
		Age	Age	Age
		AgeSqd	AgeSqd	AgeSqd
			HouseDum	HouseDum
			GarDum	GarDum
			AvgQualDum	AvgQualDum
			HighQualDum	HighQualDum
				AirDum
				Basedum
				FireDum
Deviance	**Deviance**	**Deviance**	**Deviance**	**Deviance**
3481.106274	-3334.067164	-3775.719443	-9592.073827	-10165.56647
Df	**Df**	**Df**	**Df**	**Df**
2	6	8	11	15

2.6.4. What is the Proportion of Level-1 Variance Explained?

The proportion of variance explained at level-1 is an indication of how much the level-1 residual has been reduced by including these level-1 variables relative to the null model (σ^2 Null Model- σ^2Level-1 Model) / σ^2 Null Model). The outcome of 0.56 (from (0.0667- 0.02905)/ 0.0667) indicates that 56% of the variance of the between-properties sale price (level-1 variance) has been accounted for by including these property-level variables.

Table 2.4. Variances of Random Effects from the Level-1 Model

Random Effect	Variance	Df	P-Value
Single/L1 (rij)	0.02905		
Group (U0j)	0.02537	404	0.000

2.6.5. What are Important Neighborhood Characteristics?

The next step is to construct models containing only level-2 variables. This is known as a "means as outcomes" model as it is the group means (the intercept) which is predicted by the level-2 (group-level) variables. The purpose here is to determine which of the coefficients are significant and to compare the deviance statistic from several models in order to obtain the most efficient model. Mathematically, the level-2 model is:

$$\text{Level 1:} \quad Y_{ij} = ß_{0j} + r_{ij}$$
$$\text{Level 2:} \quad ß_{0j} = \gamma_{00} + \gamma_{01}W_j + .. \gamma_{0n}W_j + U_{0j} \qquad \text{Eq. 2.5}$$

So that (γ_{01}) is the effect of a level-2 variable (W_j) on the group intercept ($ß_{0j}$), which is the average sale price for a given neighborhood. The process of sequentially adding level-2 variables and comparing the resulting Deviance Statistic is the same as discussed for building the level-1 model and is not illustrated further in this article.

2.6.7. What is the Proportion of Level-2 Variance Explained?

The proportion of variance explained at second level is an indication of how much the level-2 residual has been reduced by including these level-2 variables. This compares the level-2 variance (τ00) between this model and the previous, null model (τ00 Null Model- τ00Level-2 Model) / τ00Null Model). The outcome of 0.75 (from (0.16099-0.03955)/ 0.16099) indicates that 75% of the variance of the mean neighborhood property price has been accounted for by including these neighborhood-level variables.

Table 2.5. Variances of Random Effects of the Level-2, "Means-as-Outcomes" Model

Random Effect	Variance	Df	P-Value
Single/L1 (rij)	0.0667		
Group (U0j)	0.03955	404	0.000

2.6.8. How does the Full Multilevel Model Compare to an OLS Model?

The final step is to create a full model where variables for both levels are included. This is known as an "intercepts and slopes as outcomes" model. Mathematically, this full model is:

$$\text{Level 1:} Y_{ij} = ß_{0j} + ß_{1j}X_{ij} + \ldots ß_{nj} X_{ij} + r_{ij}$$
$$\text{Level 2:} ß_{0j} = \gamma_{00} + \gamma_{01}W_j + .. \gamma_{0n}W_j + U_{0j} \qquad \text{Eq. 2.6}$$

$$ß_{1j} = \gamma_{10}$$
$$ß_{nj} = \gamma_{n0}$$

The results of the full OLS and Multilevel models are presented in Table 2.6. The property structural characteristics are all highly significant and show the expected signs. The log transformation of floor space means that this coefficient is interpreted as an elasticity, controlling for other coefficients, so that property price increases by .3% (.4% HLM) for every 1% increase in floor space. There is a semi-log relationship between price and most variables, so what is estimated is the percent change in property price with a 1-unit change of that particular variable. For lot size, there is an approximate 40% (59% HLM) increase in sale price from an additional hectare of land. There is an approximate 5% (4% HLM) increase in sale price with the addition of a half bath. Property price falls by .38% for every additional year of house age until the house becomes older than 63 years, at which point property value increases by a slight but significant amount. For the dummy variables, there is: an approximate 11% (13% HLM) increase in sale price if the building is a single, detached dwelling (vs. a townhouse); a 9% (6% HLM) increase in sale price with the presence of central air conditioning; a 3% (4% HLM) increase in sale price with the presence of a finished basement; a 1% (2% HLM) increase in sale price with the presence of a fireplace; and an approximate 4% (4% HLM) increase in sale price with the presence of a garage. There is a 40% (32% HLM) and 67% (61% HLM) increase in the sale price if the property was built with average or high quality materials, respectively, rather than poor quality materials.

Table 2.6. Property-Specific Coefficients for the Full HLM and OLS Models

Model: OLS				Model: HLM			
Variables	**Coef**	**SE**	**Sig**	**Variables**	**Coef**	**SE**	**Sig**
Property				**Property**			
lnEnclsFt	0.30290	0.0074987	0.000	lnEnclsFt	0.37433	0.00690	0.000
lnLandHa	0.39553	0.0492366	0.000	lnLandHa	0.58923	0.04420	0.000
BathNum	0.04719	0.0030855	0.000	BathNum	0.04031	0.00268	0.000
Age	-0.00334	0.0002388	0.000	Age	-0.00301	0.00023	0.000
AgeSqd	0.00002	0.0000016	0.000	AgeSqd	0.00001	2.0E-06	0.000
HouseDum	0.11209	0.0051051	0.000	HouseDum	0.12761	0.00529	0.000
AirDum	0.09424	0.0039648	0.000	AirDum	0.06115	0.00356	0.000
BaseDum	0.02575	0.0035998	0.000	BaseDum	0.03998	0.00326	0.000
FireDum	0.00730	0.0050149	0.145	FireDum	0.01704	0.00446	0.000
GarDum	0.03679	0.0044992	0.000	GarDum	0.04345	0.00408	0.000
QualDumAvg	0.40400	0.0055572	0.000	QualDumAvg	0.32136	0.00606	0.000
QualDumHigh	0.67471	0.0079183	0.000	QualDumHigh	0.61793	0.00822	0.000

Many of the estimated coefficients for the property's neighborhood characteristics are significant with the expected signs (Table 2.7). There is a 0.6% (0.7% HLM) decrease in sale price with a 1-unit increase in neighborhood (percent) unemployment; a 0.1% increase for every $10,000 increase in median house value; a 0.2% (non-significant 0.1% HLM) increase in sale price with a 1-unit increase in a neighborhood's percent of population with (at least) a high school diploma; a 0.4% (0.6% HLM) decrease for every additional minute in a neighborhood's mean travel time to work; and a 0.004% (0.005% HLM) decrease for every 1-unit increase in a neighborhood's crime index for robberies. The log transformation of the last three neighborhood characteristics means that these coefficients are interpreted as the elasticity of sales price, so that there is a property price increase of: 0.03% (0.04% HLM) for every 1% increase in population density; 0.01% (non-significant 0.004% HLM) for every 1% increase in percent vacant residences; and.13% for every 1% increase in median household income. The theory that high vacancy is a sign of a declining neighborhood suggests a negative coefficient for this variable. However, this was positive in both regressions in this research (although insignificant in HLM). This unexpected sign could be caused by omitted variable bias. Since the variables for percent vacancy and percent with high school diplomas were insignificant in HLM, they were dropped from further analyses.

Table 2.7. Neighborhood-Specific Coefficients for HLM and OLS models

Model:OLS				**Model:HLM**			
Variables	**Coef**	**SE**	**Sig**	**Variables**	**Coef**	**SE**	**Sig**
Neighborhood				**Neighborhood**			
(Constant)	7.44678	0.0883266	0.000	(Constant)	7.13554	0.24954	0.000
pUnemploy	-0.00577	0.0005827	0.000	pUnemploy	-0.00692	0.00204	0.001
MedValHouse	1.3E-06	0.0000001	0.000	MedValHouse	1.0E-06	1.0E-07	0.000
pHSDiploma	0.00236	0.0002430	0.000	pHSDiploma	0.00103	0.00089	0.251
TravelMean	-0.00399	0.0003712	0.000	TravelMean	-0.00609	0.00122	0.000
Robbery	-0.00004	0.0000050	0.000	Robbery	-0.00005	0.00002	0.010
lnPopDens	0.02692	0.0028778	0.000	lnPopDens	0.03601	0.00935	0.000
lnpVacancy	0.01394	0.0036203	0.000	lnpVacancy	0.00422	0.01235	0.733
lnMedHsInc	0.13137	0.0064201	0.000	lnMedHsInc	0.12558	0.02257	0.000

A test for spatial autocorrelation of the residuals using the global Moran's I statistic shows that there is moderate and positive correlation in the OLS model's residuals (Moran's = 0.15) while the HLM has only a slight amount (Moran's = 0.01). Upon comparison of the above results, it appears that there is little difference between the regression coefficients estimated with conventional OLS regression and multilevel regression. However, with single–level regressions, it is assumed that the observations are independently and identically distributed, even though properties are shown to be correlated. With this correlation, OLS regressions do not produce correct standard errors; therefore, HLM needs to be used as it takes the issue of correlated errors into consideration and provides more realistic and conservative statistical testing. Standard errors are larger for HLM than OLS, as HLM considers sources of errors more rigorously than OLS. Conversely, standard errors are underestimated in OLS, which can result in potentially spuriously significant effects.

2.6.9. What is the Amount of Correlation in the Full Model?

The ICC (Intraclass Correlation Coefficient) is now at .37 (from (0.01508/(0.01508 + 0.02536)), indicating that 37% of the remaining total variation in house price is due to differences between block groups after accounting for all the property-level and neighborhood-level variables in the full model (Table 2.8).

Table 2.8. Variances and ICC's for the Null and Full models

	Null Model	**Full Model**
Random Effect	Variance	Variance
Single/L1 (rij)	0.0667	0.02536
Group (U0j)	0.16099	0.01508
ICC	0.71	0.37

2.6.10. Why Allow Level-1 Variables to Randomly Vary?

The steps previously discussed only considered the level-1 attributes as fixed effects. While the intercept was allowed to randomly vary between level-2 units, the coefficients of these fixed effects remain the same across all groups in the study (i.e. a global effect). An important feature of multilevel models is to reconsider some or all of these level-1 variables as random effects. This separates a variable into a fixed component, whose resulting coefficient applies to all of these observations for this variable, and a random component that expresses each group's deviation from that global effect.

By allowing control variables to randomly vary across groups, the level-1 variance in the error term is further portioned into the variances for these random effects. This helps control for the problem of heteroscedasticity (discussed in the introduction) in the level-1 residual by attributing some of the problematic (heteroscedastistic) variation in the residual to variation in specific, independent variables such as house size and age. If these new random terms have significant variance, as indicated by a chi-square test in HLM, the model is likely to be an improvement over the model with only fixed effects. A significant reduction in the Deviance Statistic of the random coefficients model compared to the previous, fixed effects model (using MLR) indicates the added complexity (from introducing random effects) is appropriate.

The previous equation (2.6) becomes:

$$
\begin{aligned}
\text{Level 1:} Y_{ij} &= ß_{0j} + ß_{1j}X_{ij} + \ldots ß_{nj} X_{ij} + r_{ij} \\
\text{Level 2:} ß_{0j} &= \gamma_{00} + \gamma_{01}W_j + .. \gamma_{0n}W_j + U_{0j} \\
ß_{1j} &= \gamma_{10} + U_{10j} \\
ß_{nj} &= \gamma_{n0} + U_{n0j}
\end{aligned}
\qquad \text{Eq. 2.7}
$$

Table 2.9 shows the sequential addition of random coefficients and the comparison of the deviance statistic with the previous model. The chi-square tests comparing the deviances statistics showed that the improvement to the model by adding these random effects was justified. The results of the chi-square tests show that all of these variances were statistically significant and that collinearity between the variances was eliminated.

Table 2.9. Successive Models for Variances and Deviance Statistics

Rand Effect	Var	Rand Effect	Var	Rand Effect	Var
Intercept	0.01471	Intercept	0.01495	Intercept	0.01521
		lnEnclsFt	0.0491	lnEnclsFt	0.04366
				Age	0.00001
Level-1 Var.	0.02536		0.02259		0.02188
Deviance	-10211.2		-11281.1		-11477.2
Df	2		4		7
Chi-Test			Reject Ho		Reject Ho
Rand Effect	Var	Rand Effect	Var		
Intercept	0.01518	Intercept	0.01549		
lnEnclsFt	0.03983	lnEnclsFt	0.03585		
Age	0.000004	Age	0.000004		
LandHa	0.68624	LandHa	0.63009		
		BathNum	0.00181		
Level-1 Var.	0.02155		0.02111		
Deviance	-11595.7		-11708.9		
Df	11		16		
Chi-Test	Reject Ho		Reject Ho		

2.6.11. What is the Difference Between Group- and Grand-Centering?

When moving to a model containing level-1 random effects, some other diagnostics need to be performed to ensure the model's validity. HLM provides a Tau-as-correlations matrix to see whether there is any collinearity between the random coefficients. Collinearity between the variances of a predictor and the intercept term is represented in the matrix by a column (off-diagonals) of 1's. In the context of hierarchical linear models, the existence of strong correlations between level-1 variables and the intercept is a well-known cause of instability in the model (Kreft et al. 1995; Raudenbush and Bryk 2002). Such models have compromised estimates of uncertainty as well as possible bias (Gelman et al. 2007; Paccagnella 2006). One way of dealing with this problem is by centering predictors entered into the model. One type of centering, grand-mean centering, is equivalent to the type of centering normally performed in OLS regressions. However, grand-mean centering is often not appropriate because the estimate of the slope of the independent variable becomes an uninterpretable blend of within and between-group effects (Raudenbush and Bryk 2002). On the other hand, group-mean centering minimizes any confounding of individual-level effects with neighborhood-level variables (Gelman et al. 2007). With this type of centering, the previous equation (2.7) is re-specified as:

Level 1: $Y_{ij} = ß_{0j} + ß_{1j}(X_{ij}-X_j)+\ldots ß_{nj} X_{ij}+ r_{ij}$ Eq. 2.8

Where, for example, the Xij is the area of house i in neighborhood j and X_j is the average area for all houses in neighborhood j, thus expressing this coefficient as the deviation from the group average for that coefficient. The interpretation of this group-centered coefficient is thought of as a “frog-pond” effect by which individuals compare their standing relative to other members of a group (Paccagnella 2006).

Centering is also useful to control for collinearity between the variables of a quadratic, as is the case with age. Aside from statistical issues, Enders and Tofighi (2007) recommend choosing between grand- and group- mean centering depending on the theoretical questions that are to be addressed in one’s research. If the research focuses on a main effect at the first level, then group-centering is preferred. If the main effect is found at the second level, then grand-centering is preferable.

Table 2.10 shows that without group-mean centering the variances for house size and age were highly correlated with the variance of the group mean (off-diagonals close to -1). This created the situation where the variance of the group means (the level-2 intercept) becomes greatly inflated when the uncentered age or house size was allowed to randomly vary. This situation did not occur with the variables for number of bathrooms or lot size but they were group-centered anyways to account for the fact that they did not have meaningful zero points on their original scales. When group mean centering is used, the correlations between neighborhood variables and property variables, represented as off-diagonals in the Tau-as-correlations matrix, converge towards zero. The example in Table 2.11, shows the variance of the intercept becomes greatly inflated when the uncentered variable of house size is allowed to vary across neighborhoods.

Table 2.10. Tau-as-Correlations Matrices, Centered Versus Uncentered

Intercept	1.000	-0.997		Intercept	1.000	-0.816
lnEnclsFt, Uncentered	-0.997	1.000		Age, Uncentered	-0.816	1.000
Intercept	1.000	0.410		Intercept	1.000	0.228
lnEnclsFt, Group-centered	0.410	1.000		Age, Group-centered	0.228	1.000

Table 2.11. Effect of Uncentered vs. Group-Centered Random Effect on Intercept Variance

Random Effects	Variance	DF	P-value
Full Model-Random Intercept Only			
Single/L1 (rij)	0.02536		
Group (U0j)	0.01235	397	0.000
Full Model-Uncentered Random Effect of lnEnclsFt			
Single/L1 (rij)	0.02269		
Group (U0j)	2.20917	380	0.000
lnEnclsFt(U10j)	0.04358	387	0.000
Full Model-Group-mean Centered Random Effect of lnEnclsFt			
Single/L1 (rij)	0.02259		
Group (U0j)	0.01273	380	0.000
lnEnclsFt(U10j)	0.04884	387	0.000

2.6.12. Why Include an Attribute's Group Mean at Level-2?

In addition to considering certain property-level variables as random effects and group-mean centering these variables, the neighborhood means for age, house size and lot size were all added to the level-2 model. This tested whether there were compositional effects of place in addition to the contextual effects that were previously included in the full model. With respect to multilevel property hedonic models, the contextual effects -- the difference a place makes on price (e.g. neighborhood effects of unemployment, crime, median income, etc.) -- are potentially confounded with the compositional effects -- the differences produced by the housing attributes (e.g. house size, age) within each place (Orford 2000). So for example, the strongest effect of house size on price will likely occur at the property level (level 1). However, as a compositional effect, house size may have a different influence on the house price at the neighborhood level (level 2).

The results, shown in Table 2.12, indicate that mean house size and age were both significant. Mean lot size was not found to be significant and was dropped from further analyses. There was a 0.009% increase in mean property price with each additional, square foot increase in a neighborhood's mean house size. There was also a 0.1% increase in mean property price with each additional year in a neighborhood's mean house age. This positive impact of neighborhood age is a good example of avoiding the ecological-atomistic fallacies where the relationship at one level is expected to be the same as that of another level. Property value clearly depreciates as a house ages (to a certain point) simply through deterioration of the structure from wear-and-tear. In contrast, as a neighborhood matures (as represented by mean house age), other features of the neighborhood become well developed and improve the value of an individual home. An obvious example of this is the growth of trees and vegetation improving the amenity value of the area.

Table 2.12. Neighborhood Effects of Compositional Characteristics on Mean Price

L2 Variables	Coef	SE	Sig
mnEnclsFt	0.000093	0.0000180	0.000
mnAge	0.001019	0.0003520	0.004

mnLandHa	0.305624	0.2353810	0.195

2.6.13. What is the Estimation Method Used by HLM?

The HLM multilevel approach to estimating the regression coefficients and variance components is with maximum likelihood estimation (ML) (Raudenbush and Bryk 2002).Maximum Likelihood estimators estimate the parameters of a model by providing estimates for the population values that maximize the Likelihood Function: the function that maximizes the probability of finding the sample data that has actually been found (Hox 2002). The assumption of normally distributed errors informs this likelihood function and therefore violation of this assumption should be avoided. However, if the number of level-2 units is large (greater than 30), the ML remains consistent in the estimation of the fixed effects even when the normality assumption is violated (Maas and Hox 2004). In contrast, the standard errors may become slightly biased downwards leading to a spurious significance of the effect. In HLM, robust standard errors are also reported, but may overcorrect for the violation of the normality assumption, leading to a spurious lack of significance of the fixed effect. Both Maas and Hox (2004) and Raudenbush and Bryk (2002) suggest that the robust errors should be used as a diagnostic tool where a large discrepancy between robust and normal standard errors is an indication of the violation of the normal distribution of residuals.

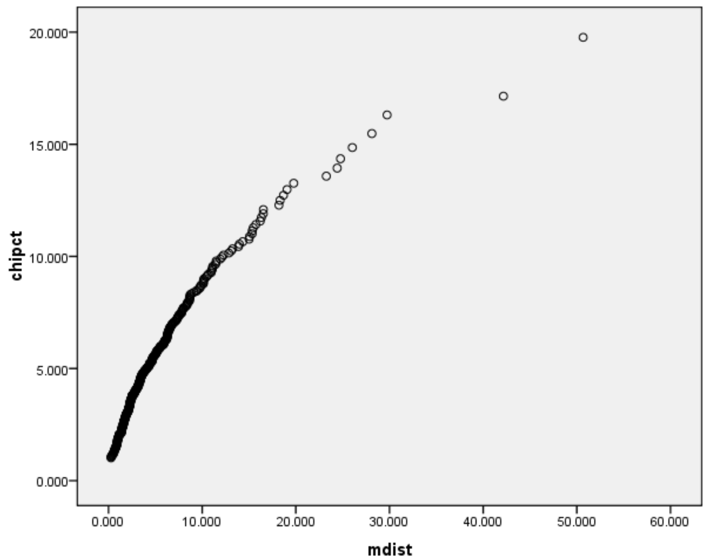
In HLM, the user has the option to choose between restricted (RML) and full (FML) maximum likelihood estimation. With FML, both the variance components and the coefficients are included in the likelihood function while with RML, only the variance components are included (Raudenbush and Bryk 2002). These two often lead to the same results particularly when there are a large number of groups. However, when there are few level-2 units or there are a large number of coefficients in the model, FML variance components tend to be biased downwards creating the potential of finding non-significant variance component. The default option of using RML in HLM corrects for situations where there are few groups. However, when building the level-1, structural model (described above in step 2), FML estimation must be used to compare the Deviance Statistic between models with successively greater number of coefficients (Hox 2002; Kreft and de Leeuw 1998). With RML, only differences in the random part between two models can be tested with the deviance statistic. Therefore, RML is used when testing whether the improved model fit (by allowing a successively greater number of coefficients to randomly vary) is worth the increase in model complexity as determined by the deviance statistic.

2.6.14. How can the Assumption of Normally Distributed Errors be Tested?

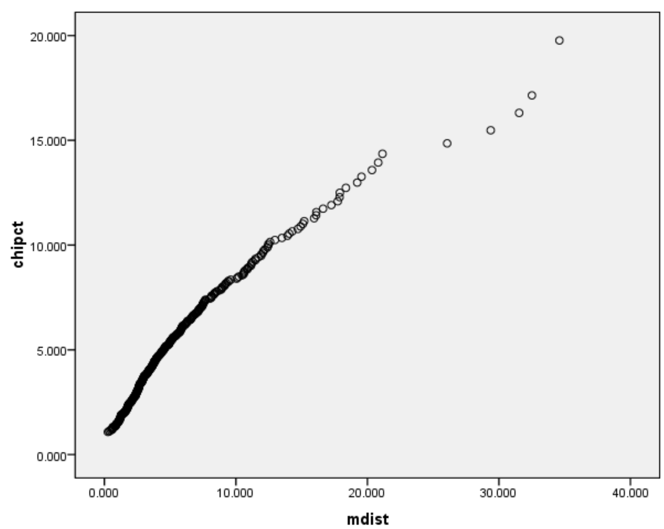
Normally distributed errors are another OLS assumption that must be satisfied in multilevel models at all levels. This assumption was visually assessed using histograms of all variances for both the log transformation of price and the Box-Cox transformation. The level-1 residuals appeared normally distributed while the level-2 residuals (intercept and independent variables) were normally distributed with a small amount of right hand skew. While Maas and Hox (2004) found that non-normal residuals at level 2 have little effect on parameter estimates, the distribution of these residuals is further examined. The residuals of the level-2 intercept were further examined visually with a Q-Q plot of the Chipct versus Mahalanobis distance (MDist). MDist is a measure of the distance of a unit's Empirical Bayesian estimate from the global, fitted value for the intercept, while Chipct is a measure of the expected values of variance from a population with a chi-square distribution (Raudenbush and Bryk 2002). If the Q-Q plot resembles a 45 degree line, the random effect for the intercept is normally distributed. Block groups where the MDist deviates dramatically from this 45 degree line are considered as outliers. Although the log transformation showed a little more deviation from normality than the Box-Cox transformation, both had the same four block groups which deviated significantly from the normal distribution for both transformations (Figure 2.1). The reason that these block groups are outliers are unknown and should be investigated further. One possibility is the presence of coding or other database errors for properties in these block groups (Raudenbush and Bryk 2002).

Figure 2.1. Chipct vs. MDist Q-Q Plots for Log and Box-Cox Transformations of Price

Price log transformed



Price Box-Cox transformed



2.6.15. What is an Optimal "Neighborhood"?

Orford (2000) and Goodman and Thibodeau (1998) suggests that the neighborhood is best thought of not as a single entity, but rather as a hierarchy of progressively more inclusive residential groupings. The inclusion of a third level, in which block groups are clustered by similar socio-economic conditions, might be a useful avenue of research that is not considered in this primer.

In addition, to the issue of including another level in multilevel models, there is also the issue of whether the block group is the most ideal unit for delineating the level-2 boundaries. In the presence of submarkets it can be difficult to segment their boundaries. With multilevel models it is necessary to determine *a priori*, geographic units for modeling these submarkets. However, models with different geographic units can be separately run and the results can be compared to each other. The goal of this segmentation is to distinguish areas where the mean price and the marginal prices of property characteristics differ markedly. Besides the block group, the level-2 units examined are: the Census tract, Baltimore city "neighborhoods" and PRIZM™ (Potential Rating index for Zip code Markets) classes.

Census tracts are the next level of aggregation in the census hierarchy. Block groups (an average of three) are perfectly nested within their respective census tract. PRIZM classes were developed by Claritas, Inc. and use factor analysis and U.S. Census data about housing, employment, education, income, ethnicity and consumer spending patterns to classify block groups into several categories. There are 3 levels of categorical resolution with this system. This research uses the 62-group resolution which further expands the characterization from Census data with information on household composition and housing characteristics. Baltimore (BNIA) Neighborhoods were developed from the Baltimore Neighborhood Indicators Alliance and are not a hierarchy within the census classification. There are about 265 BNIA Neighborhoods within the city of Baltimore.

Table 2.13 compares ICC's, residuals, stability of parameter estimates and error autocorrelation in order to find the most suitable grouping structure for a property hedonic model of Baltimore. The size of the Intraclass Correlation Coefficient (33 to 65%), which is the ratio of between-unit variation to total variation, shows that, regardless of grouping method, some manner of grouping is superior to no grouping at all. This indicates that it is important to conceptualize the hedonic model as being composed of sub-markets. A higher value ICC indicates greater agreement between properties within a neighborhood. Because these different units have different number of average observations per group it is not possible to directly compare them statistically. Smaller areas (Block Groups) have higher ICCs than larger geographic areas (PRIZM clusters). Larger geographic areas are likely to contain more heterogeneity in the characteristics being assessed, leading to less agreement among properties within a group. Therefore, Block Groups should have the largest ICC's and PRIZM clusters the smallest. However, it appears that BNA Neighborhoods provide the best agreement between properties. The results from the Moran's I coefficient indicate that grouping by Block Groups reduce spatial error autocorrelation by the greatest amount. BNA

Neighborhoods and Census Tracts also have little remaining residual dependencies. The comparison between normal and robust standard errors shows that Block Groups have the least difference in standard errors for age and house size. This indicates that the assumptions of normally distributed errors and homoscedasticity are best met by using Block Groups as the areal unit. While the BNA Neighborhoods have a higher ICC than Block Groups, the Block Groups have better statistical properties with respect to assumptions of the residuals.

Table 2.13. Comparison of Grouping Units

	Block Group			
Full Model	Variance	ICC	L1 Var Expl	l2 Var Exp
Single/L1	0.03126	0.54800463	0.57927322	0.78710257
Group	0.0379			
Moran's I	0.01			
l1 Varbs	coeff	SE	robust SE	diff
lenclft	0.411272	0.007439	0.017344	0.009905
age	-0.003191	0.00025	0.000468	0.000218
	BNA Neighborhood			
Full Model	Variance	ICC	L1 Var Expl	l2 Var Exp
Single/L1	0.03373	0.65079201	0.60106446	0.68706128
Group	0.06286			
Moran's I	0.023297			
l1 Varbs	coeff	SE	robust SE	diff
lenclft	0.412101	0.007445	0.025743	0.018298
age	-0.003875	0.000244	0.000465	0.000221
	Tract			
Full Model	Variance	ICC	L1 Var Expl	l2 Var Exp
Single/L1	0.03556	0.50999035	0.6215008	0.77448053
Group	0.03701			
Moran's I	0.03			
l1 Varbs	coeff	SE	robust SE	diff
lenclft	0.402078	0.007615	0.022181	0.014566
age	-0.003925	0.000245	0.000633	0.000388
	Prizm Clusters			
Full Model	Variance	ICC	L1 Var Expl	l2 Var Exp
Single/L1	0.04449	0.33097744	0.6531535	0.92362942
Group	0.02201			
Moran's I	0.123854			
l1 Varbs	coeff	SE	robust SE	diff
lenclft	0.366417	0.007895	0.026334	0.018439
age	-0.004738	0.000252	0.001352	0.0011

2.7. Issues When Including Environmental Attributes

2.7.1. Why Allow a Main Effect to Randomly Vary?

The main effects of the study (i.e. the research variables of main interest such as distance to an environmental attribute) can also be considered as random effects. This allows the researcher to determine: whether there is statistically significant deviation from the fixed effect across groups; the range of this deviation; and where this deviation occurs by mapping the Empirical Bayesian (EB) coefficients for these groups. By mapping these EB coefficients, non-stationarity is visualized to reveal localized patterns in this deviation. This mapping can be considered to be an EDA (exploratory data analysis) that may help to determine whether there are group-level variables that will interact with this random effect to explain the deviation from the fixed effect. Such cross-level interactions are discussed in the next section.

The Empirical Bayesian variation in slopes between neighborhoods is added to the global fitted value to reveal areas of positive and negative relationships. Modeling random coefficients to investigate the effects of place on the relation between the main effect and the dependent variable is one of the fundamental advantages of multilevel modeling (Subramanian 2001). While it is possible to build OLS regressions to determine this effect for each neighborhood, there will be the problem that there are insufficient observations within a particular neighborhood. One approach to dealing with this problem is to construct Empirical Bayes estimates which borrow strength across neighborhoods and shrink estimates for neighborhoods with few observations towards the overall mean (Raudenbush and Bryk 2002). This means that the implicit prices of certain attributes (those that are considered random effects) are optimally weighted averages that combine information derived from the group itself with the mean from neighborhoods with similar characteristics (Diez-Roux et al. 2000). Unreliable submarket estimates are differentially shrunk towards the global estimate, whereas submarkets with many properties will not be affected by this shrinkage. This pooling of information and borrowing of strength is more analogous with the definition of submarkets as being quasi-independent and functionally related (Orford 2000).

2.7.2. What are Cross-Level Interactions?

In multilevel models, interactions can occur between variables at the same level or between variables at different levels. In a cross-level interaction, a neighborhood-level variable can be used to moderate the relationship between a level-1 environmental variable and the house price. Hox (2002) recommends that if there is a significant interaction found between one of these neighborhood characteristics and the environmental coefficient, then the direct effects of that characteristic must also be included even if it is found to be insignificant.

The issue of centering is also important when interactions are included in a multilevel model. Enders and Tofighi (2007) found that use of grand mean centering (when the main effect is a level-1 variable) can artificially produce a significant effect from the interaction of either another level-1 variable or a level-2 variable (single-level or cross-level interactions) when in reality one does not exist. They conclude that group-mean centering is more appropriate when (either cross-level or same-level) interactions are the research interest.

2.7.3. What is the Purpose of a Non-Hierarchical, Cross-Classified Model?

The multilevel approach to valuing environmental amenities may be flawed in that market segmentation through common property and neighborhood characteristics might not create ideal neighborhood boundaries with respect to the valuation of park proximity. In the case of locational amenities, there is a spatial diffusion of environmental externalities which may cross the market boundaries. Diffusion occurs because the effect of an externality weakens with distance. The spatial dependency of externalities implies that neighboring properties will capture a similar impact from the externality because of their similarity in proximity to the amenity. However, these properties that share similar proximity to a local amenity may not be located in the same neighborhood. They may be located in the next, contiguous neighborhood or in a neighborhood on the other side of an amenity such as urban parks. This potential for misalignment might be corrected by considering the cross-classification of properties into both neighborhood groups and the parks to which they are closest. However, in the case of locational amenities such as those generated by a park, there is a spatial diffusion of environmental externalities which may cross the sub-market boundaries. Diffusion occurs because the effect of an externality weakens with distance. The spatial dependency of externalities implies that neighboring properties will capture a similar impact from the externality because of their similarity in proximity to the amenity. These neighboring properties may not be located in the same neighborhood defined through market segmentation. They may be located in the next, contiguous neighborhood or in a neighborhood on the other side of a park.

2.8. Conclusion

Although this primer focuses on building multilevel models with control variables commonly used in property hedonic studies, it is my intention that the information presented here will facilitate an explicit use of spatial scale into regression models for environmental economic research. Orford's (2000) paper on multilevel property hedonics is another source that details the use of this methodology. Raudenbush and Bryks' book on Hierarchical Linear Models (2002), though examining non-spatial, educational data, is an excellent source on the numerous statistical issues associated with multilevel models.

2.9. References

Allen, T.F.H., and Starr, T.B. 1982. Hierarchy: Perspectives for ecological complexity. The University of Chicago Press, Chicago.

Anderson, S.T., and West, S.E. 2006. Open space, residential property values, and spatial context. Reg. Sci. Urban Econ. **36**: 773–789.

Anselin, L. 1988. Spatial econometrics : methods and models. Kluwer Academic Publishers London.

Anthon, S., Thorsen, B.J., and Helles, F. 2005. Urban-fringe afforestation projects and taxable hedonic values. Urban Forestry and Urban Greening **3**: 79–91.

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. Ecol. Econ. **40**: 337-349.

Beland, F., Birch, S., and Stoddart, G. 2002. Unemployment and health: Contextual-level influences on the production of health in populations Social Science and Medicine **55**(11): 2033-2052.

Bhat, C.R. 2000. A multi-level cross-classified model for discrete response variables. Transp. Res. Pt. B-Methodol. **34**: 567-582.

Bin, S., and Polasky, S. 2004. Effects of flood hazards on property values: Evidence before and after Hurricane Floyd. Land Econ. **80**: 490-500.

Bolitzer, B., and Netusil, N.R. 2000. The impact of open spaces on property values in Portland, Oregon. J. Environ. Manage. **59**: 185-193.

Bourassa, S.C., Hoesli, M., and Peng, V.S. 2003. Do housing submarkets really matter? Journal of Housing Economics **12**: 12–28.

Box, G.E.P., and Cox, D.R. 1964. An analysis of transformations. Journalof the Royal Statistical Society, Series B (Methodological) **26**: 211–252.

Brown, K.H., and Uyar, B. 2004. A hierarchical linear model approach for assessing the effects of house and neighborhood characteristics on housing prices. Journal of Real Estate Practice and Education **7**(1): 15-23.

Browning, C.R., Feinberg, S.L., and Dietz, R.D. 2004. The paradox of social organization: Networks, collective efficacy, and violent crime in urban neighborhoods. Social Forces **83**(2): 503-534.

Cho, S.H., Bowker, J.M., and Park, W.M. 2006. Measuring the contribution of water and green space amenities to housing values; An application and comparison of spatially weighted hedonic models. J. Agric. Resour. Econ. **31**(3): 485-507.

Cliff, A., and Ord, J.K. 1981. Spatial processes: Models and applications. Pion, London.

Cropper, M.L., Deck, L.B., and McConnell, K.E. 1988. On the choice of functional form for hedonic price functions. The Review of Economics and Statistics **70**: 668–675.

Day, B., Bateman, I., and Lake, I. 2004. Nonlinearity in hedonic price equations: An estimation strategy using model-based clustering. University of East Anglia: Centre for Social and Economic Research on the Global Environment (CSERGE).

Deaton, B.J., and Hoehn, J.P. 2004. Hedonic analysis of hazardous waste sites in the presence of other urban disamenities. Environ. Sci. Policy **7**: 499–508.

Diez-Roux, A.V. 2002. A glossary for multilevel analysis. Journal of Epidemiology and Community Health **56**: 588–594.

Diez-Roux, A.V., Link, B.G., and Northridge, M.E. 2000. A multilevel analysis of income inequality and cardiovascular disease risk factors. Social Science & Medicine **50**: 673-687.

Ekeland, I., Heckman, J.J., and Nesheim, L. 2002. Identifying hedonic models. American Economic Review **92**(2): 304-309.

Enders, C.K., and Tofighi, D. 2007. Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. Psychol. Methods **12**(2): 121-138.

Fone, D.L., and Dunstan, F. 2006. Mental health, places and people: A multilevel analysis of economic inactivity and social deprivation. Health & Place **12**: 332-344.

Fotheringham, A.S., Brunsdon, C., and Charlton, M.E. 2002. Geographically Weighted Regression: The analysis of spatially varying relationships. Wiley, Chichester.

Freeman, A.M. 2003. The measurement of environmental and resource values: Theory and methods. Resources for the Futures, Washington, D.C.

Garrod, G.D., and Willis, K.G. 1992. Valuing goods' characteristics: an application of the hedonic price method to environmental attributes. J. Environ. Manage. **34**(1): 59-76.

Gayer, T., Hamilton, J.T., and Viscusi, W.K. 2000. Private values of risk tradeoffs at Superfund sites: Housing market evidence on learning about risk. The Review of Economics and Statistics **82**(3): 439-451.

Gelfand, A.E., Banerjeeb, S., Sirmans, C.F., Tu, Y., and Ong, S.E. 2007. Multilevel modeling using spatial processes: Application to the Singapore housing market. Comput. Stat. Data Anal. **51**: 3567 – 3579.

Gelman, A., Bafumi, J., Shor, B., and Park, D.K. 2007. Rich state, poor state, red state, blue state. What's the matter with Connecticut? Quarterly Journal of Political Science **2**: 345-367.

Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D., Thomas, S. 1993. A multilevel analysis of school examination results. Oxford Review of Education **19**: 425–433.

Goodman, A.C., and Thibodeau, T.G. 1995. Age-related heteroskedasticity in hedonic house price equations. Journal of Housing Research **6**(1): 25-42.

Goodman, A.C., and Thibodeau, T.G. 1998. Housing market segmentation. Journal of Housing Economics **7**: 121–143.

Halstead, J.M., Bouvier, R.A., and Hansen, B.E. 1997. On the issue of functional form choice in hedonic price functions: Further evidence. Environ. Manage. **21**(5): 759–765.

Hite, D., Chern, W., Hitzhusen, F., and Randall, A. 2001. Property-value impacts of an environmental disamenity: The case of landfills. Journal of Real Estate Finance and Economics **22**(2): 185-202.

Hoshino, S. 2001. Multilevel modeling on farmland distribution in Japan. Land Use Pol. **18**(1): 75-90.

Hox, J.J. 2002. Multilevel analysis: techniques and applications. Lawrence Erlbaum Associates, Mahwah, NJ.

Jones, K., and Bullen, N. 1993. A multi-level analysis of the variations in domestic property prices. Urban Studies **30**(8): 1409-1427.

Jones, K., and Duncan, C. 1996. People and places: the multilevel model as a general framework for the quantitative analysis of geographical data. *In* Spatial Analysis: Modelling in a GIS Environment *Edited by* P. Longley, Batty, M. John Wiley & Sons, New York. p. 389.

Kim, C.W., Phipps, T.T., and Anselin, L. 2003. Measuring the benefits of air quality improvement: A spatial hedonic approach. J.Environ.Econ.Manage. **45**(1): 24-39.

Klassen, A.C., Kulldorff, M., and Curriero, F. 2005. Geographical clustering of prostate cancer grade and stage at diagnosis, before and after adjustment for risk factors. International Journal of Health Geographics **4**(1): 1-16.

Kreft, I., and de Leeuw, J. 1998. Introducing multilevel modeling. Sage, Newbury Park, CA.

Kreft, I.G.G., Deleeuw, J., and Aiken, L.S. 1995. The effect of different forms of centering in hierarchical linear-models. Multivariate Behav. Res. **30**(1): 1-21.

Langford, I.H., Bentham, G., and McDonald, A. 1998. Multilevel modelling od geographically aggregated health data: A case study on malignant melanoma mortality and UV exposure in the European community. Statistics in Medicine **17**: 41-57.

Lee, C., and Linneman, P. 1998. Dynamics of the greenbelt amenity effect on the land market: The case of Seoul's greenbelt. Real Estate Economics **26**(1): 107–129.

Leggett, C.G., and Bockstael, N.E. 2000. Evidence of the effects of water quality on residential land prices. J.Environ.Econ.Manage. **39**(2): 121-144.

Lewis, L.Y., Bohlen, C., and Wilson, S. 2008. Dams, Dam removal, and river restoration: A hedonic property value analysis. Contemporary Economic Policy **26**(2): 175–186.

Maas, C.J.M., and Hox, J.J. 2004. The influence of violations of assumptions on multilevel parameter estimates and their standard errors. Comput. Stat. Data Anal. **46**: 427 – 440.

Mahan, B.L., Polasky, S., and Adams, R.M. 2001. Valuing urban wetlands: A property price approach. Land Econ. **76**(1): 100–113.

Mauny, F., Viel, J.F., Handschumacher, P., and Sellin, B. 2004. Multilevel modelling and malaria: a new method for an old disease. International Journal of Epidemiology **33**: 1337-1344.

McCluskey, J.J., Rausser, G.C. . 2003. Hazardous waste sites and housing appreciation rates Hazardous waste sites and housing appreciation rates **45**(2): 166-176.

McLeod, H., Langford, I.H., Jones, A.P., Stedman, J.R., Day, R.J., Lorenzoni, I., and Bateman, I.J. 2000. The relationship between socio-economic indicators and air pollution in England and Wales: Implications for environmental justice. Regional Environmental Change **1**(2): 78-85.

Moon, G., Barnett, R. 2003. Spatial scale and the geography of tobacco smoking in New Zealand: A multilevel perspective New Zealand Geographer **59**(2): 6-15.

Morancho, A.B. 2003. A hedonic valuation of urban green areas. Landscape and Urban Planning **66**: 35-41.

Mujahid, M.S., Diez-Roux, A.V., Morenoff, J.D., and Raghunathan, T. 2007. Assessing the Measurement Properties of Neighborhood Scales: From Psychometrics to Ecometrics. American Journal of Epidemiology **165**: 858 - 867.

O'Neilll, R.V., Deangelis, D.L., Waide, J.B., and Allen, T.F.H. 1986. A hierarchical concept of ecosystems. Princeton University Press, Princeton, N.J.

Orford, S. 2000. Modelling spatial structures in local housing market dynamics: A multilevel perspective. Urban Studies **37**(9): 1643-1671.

Orford, S. 2002. Valuing locational externalities: A GIS and multilevel modelling approach. Environment and Planning B: Planning and Design **29**(1): 105-127.

Overmars, K.P., Verburg, P.H. 2006. Multilevel modelling of land use from field to village level in the Philippines. Agricultural Systems **89**: 435-456.

Paccagnella, O. 2006. Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. Evaluation Review **30**(1): 66-85.

Paez, A., and Scott, D.M. 2004. Spatial statistics for urban analysis: A review of techniques with examples. GeoJournal **61**: 53–67.

Pan, W.K.Y., Bilsborrow, R.E. . 2005. The use of a multilevel statistical model to analyze factors influencing land use: A study of the Ecuadorian Amazon Global and Planetary Change **47**: 232-252.

Poor, P.J., Pessagnob, K.L., and Paul, R.W. 2007. Exploring the hedonic value of ambient water quality: A local watershed-based study. Ecol. Econ. **60**: 797-806.

Raudenbush, S.W., and Bryk, A.S. 2002. Hierarchical Linear Models: Applications and Data Analysis Methods. Sage Publications, London.

Raudenbush, S.W., Rowan, B., King,S.J. 1991. A multilevel, multivariate model for school climate with estimation via EM algorithm and application to US high school data. Journal of Educational Statistics **16**: 295-330.

Robert, S.A., Reither, E.N. 2004. A multilevel analysis of race, community disadvantage, and body mass index amongadults in the US. Social Science & Medicine **59**: 2421-2434.

Rosen, S. 1974. Hedonic prices and implicit markets: product differentiation in pure competition. The Journal of Political Economy **82**(1): 34–55.

Shieh, Y.Y., and Fouladi, R.T. 2003. The effect of multi-collinearity on multilevel modeling parameter estimates and standard errors. Educ. Psychol. Meas. **63**: 951-985.

Snijders, T.A.B., and Bosker, R. 1999. Multilevel analysis. An introduction to basic and advanced multilevel modeling. Sage, Thousand Oaks, CA.

Soobader, M., Cubbin, C., Gee, G.C., Rosenbaum, A., Laurenson, J. 2006. Levels of analysis for the study of environmental health disparities. Environmental Research **102**: 172-180.

Subramanian, S.V., Duncan, C., Jones, K. 2001. Multilevel perspectives on modeling census data. Environ. Plan. A **33**(3): 399-417.

Thorsnes, P. 2002. The value of a suburban forest preserve: Estimates from sales of vacant residential building lots. Land Econ. **78**(3): 426–441.

Troy, A., and Grove, J.M. 2008. Property values, parks, and crime: A hedonic analysis in Baltimore, MD. Landscape and Urban Planning **87**: 233–245.

Tyrväinen, L., and Miettinen, A. 2000. Property prices and urban forest amenities. J.Environ.Econ.Manage. **39**(2): 205–223.

CHAPTER 3: A MULTILEVEL APPROACH TO MEASURING THE CONTRIBUTION OF BALTIMORE PUBLIC PARKS TO PROPERTY VALUES

3.1. Abstract

This study measures the contribution of public parks to the value of properties in the city of Baltimore that were sold between 1998 and 2002. Multilevel models were used to define the submarkets for the city and to reveal important spatial variations for the effect of park proximity on property prices. Maps of this variation in the park-price relationship showed that only two-thirds of neighborhoods showed a positive preference for park proximity. This effect was then allowed to vary as a function of several property, neighborhood and park characteristics. The results of property-specific interactions indicate that a property's lot size provided a strong substitution effect to the beneficial impact of park proximity on property price. Smaller and more open parks also interact with park proximity to significantly increase the benefits to nearby properties (compared to larger, wooded parks). Only the neighborhood characteristic of population density was found to have a significant effect with higher densities showing an increased preference for park proximity.

3.2. Introduction

Public parks and other green spaces provide numerous benefits to an urban population, such as recreation, scenery, improved water quality, habitat for wildlife, a reduction in the "heat island effect" and much more. Most of these environmental benefits, also known as environmental externalities or “ecosystem goods and services” (ES) are external to normal market transactions and consequently, are often undervalued and under-provisioned even though they impact the quality of people’s lives. Capturing the monetary value of these benefits is important for improving people’s welfare and for urban planning issues such as zoning, development, land conservation acquisitions, property taxation and improvements to and maintenance of existing parks.

One method for capturing the benefits of parks is the property hedonic model in which an individual can choose a level of use of an environmental attribute through their choice of location where they purchase a property. However, this method only captures those benefits that are directly usable to the individual such as recreation and aesthetics (Freeman 2003). Though there are many other ecosystem services that are provided by the existence of parks these cannot be valued with the property hedonic model.

With this approach, an individual property is considered to be composed of a bundle of characteristics, each of which implicitly contributes to the price of the property. These characteristics can be broadly grouped into three categories: property-specific (including both the land and structural improvements); contextual neighborhood-specific (the socio-economic context); and environmental (locational) (Freeman 2003). Structural attributes refer to the tangible qualities of a dwelling and parcel such as lot size, house size, quality, age, number of bathrooms. Contextual attributes are those shared by a neighborhood such as crime rate, ethnicity, income and other socio-economic factors. Locational attributes refer to the proximity and/or accessibility of various land uses and buildings such as schools, highways, industrial areas, shopping centers and green spaces such as parks and golf courses.

The hedonic approach has been used to value urban parks (Cho et al. 2006; Espey and Owusu-Edusei 2001; Morancho 2003; Orford 2002; Troy and Grove 2008), a mixture of urban-suburban green spaces, such as parks, natural areas, green belts, golf courses, wetlands and cemeteries (Acharya and Bennett 2001; Anderson and West 2006; Bolitzer and Netusil 2000; Do and Grudnitski 1995; Kong et al. 2007; Lee and Linneman 1998; Lutzenhiser and Netusil 2001; Mahan et al. 2001; Tyrväinen and Miettinen 2000), or a mixture of open spaces in suburban-rural areas (Bastian et al. 2002; Garrod and Willis 1992; Geoghegan et al. 1997; Irwin 2002a; Irwin 2001; Smith et al. 2002). The general theory and findings from this research is that parks and green spaces contribute positively to property values but this benefit rapidly declines with increasing distance from the property. However, the magnitude of this impact or even whether this impact is beneficial or negative varies across studies and depends upon the type of green space, the green space characteristics, and the location of that space with respect to land use and neighborhood characteristics.

With respect to the type of urban green space, Bolitzer and Netusil (2000) found that urban parks and golf courses both have positive impacts, while cemeteries have an insignificant, negative impact on properties. Anderson and West (2006) found that urban park and golf course proximity provides a significant benefit while proximity to cemeteries has a negative impact. Smith and others (2002) found a negative effect with proximity to suburban parks and positive impact with proximity to golf courses.

Park characteristics such as size, vegetation cover and crime have also been found to affect the proximity-price relationship. Lutzenhiser and Netusil (2001) found that small, urban parks have a negative impact on nearby properties but larger, natural (wooded) areas have a positive impact for homes in Portland, Oregon. Their findings suggest that the size of the green space has an important effect on price-proximity relationship. They theorize that the benefits from large parks outweigh the negative externalities of traffic and noise that may be associated with smaller, urban parks. Anderson and West (2006) found the same, beneficial effect of size for natural areas (special parks) but a detrimental effect of size for neighborhood parks. Garrod and Willis (1992) found that parks primarily consisting of conifers have a negative impact while deciduous tree cover created a positive impact. Tyrväinen (2000) and Thorsnes (2002) found that proximity to forested views and access is beneficial to property prices although for small wooded strips, Tyrväinen found a negative impact.

Neighborhood specific factors may also affect the value of proximity to park. High income neighborhoods or neighborhoods with a high proportion families with children may place a higher value on recreation opportunities (from being proximate to parks) than low income neighborhoods or neighborhoods with a high proportion of elderly people. Anderson and West (2006) found that urban parks are generally more beneficial to nearby properties than suburban parks in the area around Minneapolis-St. Paul, Minnesota. High income and high density neighborhoods increased the value of proximity to these urban parks. Dehring and Dunse (2006) found that high density neighborhoods favored proximity to urban parks while there was no significant effect for lower density neighborhoods.

The lot size of a property may be expected to act as a substitute to the value of a park proximity (Henderson and Song 2008). However, Anderson and West (2006) found that lot size was positively correlated with the value of proximity to parks, suggesting that lot size is a complement to rather than a substitute for parks.

Proximity to public parks is generally expected to provide a positive contribution to property values. This is primarily a function of the visual and recreation amenities a park provides. There may also be indirect benefits which are captured by park proximity such as the absence of negative externalities (for example noise, traffic, pollution) associated with other land uses such as industrial areas or shopping centers. However, a park's positive externalities may be mitigated or overwhelmed by negative externalities associated with the park, which may reduce or even overwhelm the positive benefits to the local property values. Examples of such negative impacts are the potential for increased car traffic and noise (especially near a park's parking lot or entrances), noise from team sports and providing a refuge for criminals or homeless people. Troy and Grove (2008) found that higher levels of crime in the areas around parks reduced the positive impact of park proximity to the point where there was a negative impact of park proximity on properties near high-crime parks.

These negative impacts may have different distance –decay rates than the decay rate of the park's positive externalities. Espey and Owusu-Edusei (2001) found a positive impact with proximity to small or medium urban parks (less than 9 hectares) in

Greenville, South Carolina for distances between 100 to 500 meters but a negative impact for properties less than 100 meters from a park.

The interaction of these contextual elements with the park's distance variable can be expected to create a localized and spatially-varying price-proximity relationship even within the geographic area of a single city. These elements can generally be grouped into characteristics related to individual properties, socio-economic or vegetation characteristics of the neighborhood and park-specific characteristics.

The foundation for the property hedonic model was presented by Rosen (1974) who showed the existence of a property market equilibrium where consumers and suppliers maximize their respective utility and profits by choosing to purchase and produce properties with distinct combinations of desirable attributes. Although this approach has been used to address a wide variety of environmentally-related issues in since Rosen's work, there are numerous statistical and econometric issues that need to be accounted for before a valid interpretation of regression results can be made. The model may be mis-specified as a result of the functional form of the model, collinearity among variables and omitted determinants of property value. There are also spatial issues of dependency, non-stationarity and scale that exist in a housing market composed of interrelated sub-markets.

Rosen's (1974) development of the property hedonic model assumed that both supply and demand factors were mobile and elastic and that an entire city could be viewed as having a single housing market in equilibrium. Equilibrium occurs when the market settles on a hedonic price supply-demand curve that ensures households (within their budget constraints) cannot increase their utility by choosing a different property and sellers cannot increase their profits by increasing the property's price or changing its characteristics. With this assumption, the price of a property and the availability and contribution of its constituent characteristics are invariant across geographic space.

Since Rosen's work, researchers have found that housing markets are typically not in equilibrium and that the assumption of a single market is unrealistic except for very small study areas (Bourassa et al. 2003; Day et al. 2004; Ekeland et al. 2002; Goodman and Thibodeau 1998, 2003; Orford 2000). With property hedonic models there are also spatial statistical concerns of spatial dependency, non-stationarity and inappropriate scales of analysis. Spatial dependency (association or lags) refers to the likelihood that the values of observations for a particular variable are more similar for observation in close spatial proximity to each other. An example of this spatial dependency in the housing market is the compositional effect of neighboring property characteristics (e.g. house age, size and value) influencing the selling price of an individual residence (Orford 2000). Spatial error autocorrelation, on the other hand, refers to the existence of spatial associations that have not been incorporated into the regression model (Paez and Scott 2004). The problem with the presence of spatial error autocorrelation in a regression model is that the statistical assumption regarding the independent distribution of errors is violated. As a consequence of these two types of spatial dependencies (lags and error), parameter estimates will be biased and inefficient, respectively (Anselin 1988). Inefficient standard errors leads to the possibility of finding a spurious significance of an effect when one does not actually exist (Type I error).

Non-stationarity refers to the existence of a heterogeneous relationship between dependent and independent variables across geographic space (Fotheringham et al. 2002). Global approaches to hedonic modeling such as using OLS do not accommodate local, spatial variations in these relationships. A coefficient that is reported as insignificant within a global regression model may be the result of highly significant localized relationships cancelling out the effect of significant negative relationships in others areas. With respect to the effect of locational amenities on house price, it is often useful to determine the existence of non-stationarity and to attempt to model the reason for its existence by allowing the coefficient to interact with other variables.

Issues with spatial dependencies and non-stationarity are further complicated by the scale at which attributes are measured or aggregated. If the values of these spatial properties change with the choice of unit used in a model, then the model exhibits scaling challenges that cannot be effectively modeled with standard regression techniques. Thus, the variance of the outcome, the relationship between the independent and dependent variable and the relationship between individual observations all may be sensitive to unit size.

This research uses a multilevel modeling framework for addressing the challenges of the property hedonic model. With the multilevel approach, individual properties are nested within submarkets (neighborhoods) within a city (Brown and Uyar 2004; Gelfand et al. 2007; Goodman and Thibodeau 1998; Orford 2000, 2002). These models allow the simultaneous examination of the effects of group-level and individual-level variables on individual-level outcomes (the property price) while accounting for the non-independence of observations within groups (the neighborhoods).

There are several statistical and substantive reasons for explicitly modeling individual properties as belonging to neighborhoods. First, spatial dependencies (either correlation in variables or the residual) are likely to be common in property hedonic analyses since individual properties in the same neighborhood are likely to be similar in ways not fully accounted for by the property and neighborhood variables included in a single-level model (Jones and Bullen 1993). With a multilevel model, the house price and other important house characteristics found at the property level can be included at the neighborhood level to control for dependencies in house price and characteristics (spatial lag processes). In the property market, this type of dependency is known as a compositional effect whereby the neighborhood averages of individual property attributes affect the value of a individual property (Orford 2002). Multilevel models also account for the spatial error autocorrelation (dependence of the residuals) by differentiating between-individual errors from between-neighborhood errors (Orford 2000). If this dependency was not modeled, the standard errors of the independent variables would be biased downwards (underestimated), which results in spuriously significant effects (Snijders and Bosker 1999).

Second, it is possible that an average effect for a particular attribute does not represent local conditions that may occur in different areas of a study area; the issue of non-stationarity in property hedonic studies (Cho et al. 2006; Troy and Grove 2008). A possible outcome of estimating an average effect, also known as a global effect, is that contrasting relationships in different areas of the study may cancel each other out, leading to a globally insignificant estimate for that variable's coefficient (Fotheringham et al. 2002). By considering a variable as a random effect, instead of a global fixed effect, that variable is allowed to vary across neighborhoods. While it is possible to build OLS regressions to determine this effect for each neighborhood, there will be the problem that there are insufficient observations within any given neighborhood leading to unreliable estimates. In multilevel models, random effects are constructed with Empirical Bayesian techniques which borrow strength across neighborhoods and shrink or smooth estimates for unreliable neighborhoods (those with few observations) towards the overall mean (Raudenbush and Bryk 2002). This means that the implicit prices of certain attributes (those that are considered random effects) are optimally weighted averages that combine information derived from the group itself with the mean from all other neighborhoods (Diez-Roux 2002). While unreliable submarket estimates are differentially shrunk towards the global estimate, submarkets with many properties will not be affected by this shrinkage. The Empirical Bayesian variation in slopes between neighborhoods is added to the global fitted value to reveal areas of positive and negative relationships and is one of the fundamental advantages of multilevel modeling (Subramanian 2001).

Third, heterogeneity (varying relationships) between neighborhoods needs to be distinguished from the heterogeneity among individual properties. Ignoring this differentiation and modeling the behavior of interest at a single level invites the pitfalls of either the ecological or atomistic fallacy. Atomistic and ecological fallacies are avoided because the predictors and unexplained variation are modeled at the appropriate level

(Jones and Duncan 1996). When group-level data, such as Census data, is included in an OLS model whose focus is the individual, the interpretation of the results may lead to the atomistic fallacy. With this situation, data (and the processes they measure) that are collected at a broader scale are assumed to have the same importance, associations and interactions at the individual level as at the higher level. With multilevel models, the variation in house price due to the effects of individual-level, property attributes is separated from price differences between areas (contextual effects), while also avoiding the risk of ecological fallacies by using disaggregate data. Corresponding to this is the ecological fallacy where inferences about the significance, associations and the variability of higher, group levels are based on data originating from individuals (Allen and Starr 1982; O'Neilll et al. 1986). This fallacy can also refer to higher-level data that is simply an aggregation of individual-level observations (Hox 2002). For example, the association between price and house size or age at the property level may differ from the association between these variables that have been aggregated to the group level. With multilevel models, the variation in house price due to the effects of individual-level, property attributes is separated from price differences between areas that are related to neighborhood-level contextual effects. With multilevel models, the user is not forced to model at one level or the other, both (indeed several) levels are be modeled simultaneously.

Hedonic models also often suffer from heteroscedasticity (unequal variation in the residuals) across neighborhoods, leading to inefficient estimates (Conway et al. 2008). Heteroscedasticity can be caused by omitting important variables from the models or mis-specified individual property characteristics and potential interactions. For instance, the variance associated with the implicit price of floor area may be greater in larger houses than smaller houses and additionally, may be greater in homes rather than townhouses. If a neighborhood has larger houses then its variance will be greater than other neighborhoods. If that same neighborhood has a greater proportion of homes to townhouses than the average, then this non-constant variance will be further exaggerated. Multilevel models can control for heteroscedasticity in the level-1 residual by expanding the random part of the model with an additional random term for floor size. Each level-1 coefficient can be allowed to vary across neighborhoods either randomly, through the interaction with level-2 variables or through both of these options (Orford 2000).

3.3. Objectives

This research uses multilevel property hedonic models to explore the value of park proximity and the contextual characteristics that may mediate this value. The effect of proximity to park on property values is first allowed to vary across neighborhoods as a random effect. As a random effect, the park coefficient is separated into a fixed component, whose resulting coefficient applies to all of these observations for this variable, and a random component that expresses each group’s deviation from that global effect. This deviation is then mapped to determine where and by how much the effect of park proximity varies across the city.

I then attempt to explain this variation by creating interactions with a number of property-level, neighborhood and park characteristics. The property specific characteristic of lot size is expected to act as a substitute for nearby parks and therefore diminish the value of proximity. Concerning neighborhood characteristics, I expect that the value of recreation and aesthetic opportunities will be greater for densely-populated and higher income neighborhoods. Households in high crime areas may be reluctant to utilize public spaces so that the value of park proximity will be less for these areas. With respect to park characteristics, high-crime parks should have a strong negative effect on proximity to parks. I expect that size and openness of a park will positively impact the value of park proximity.

3.4. Data

Property sales and attributes were obtained from the MD Property View 2004 database, a private company which compiles sales transaction data with a property's location and structural characteristics from the state of Maryland's property-appraisal database. Transactions for the city of Baltimore for a 5-year period (1998-2002) were used in this analysis.

Selling prices were standardized to the year 2000 with the OFHEO (Office of Federal Housing Enterprise Oversight) housing price index for the Baltimore Metropolitan Statistical Area. This index accounted for both annual and seasonal (quarterly) fluctuations of property sales. This standardization removed the need for adding dummy variables for year and season while allowing for a sufficiently large dataset of properties that would be consistent with the 2000 Census attributes used to describe the neighborhood.

Property records were selected if they followed numerous criteria, whose values were specified within the property database: “arms-length” transactions only; single, detached homes or townhouses; total, appraised property value was within ±50% agreement of the selling price; zoning was classified as either residential or residential-commercial; and values of key variables used in the analyses were not missing. The total square footage for each house accounted for floor area for each story, excluding basements and attics. Although lot size (land area of the property) was available in the property database, these values were recalculated within a GIS by associating individual transaction records with their corresponding, spatially-delineated parcels. These properties were also re-located to the center of their parcels in order to adjust for location errors in the property database. Finished basements and garage sizes were converted to dummy variables (presence or absence). The quality of the original construction was converted to 3 dummy variables of poor, average and high quality. The number of bathrooms and half-bathrooms were recombined into one attribute (e.g. one full bath and one half bath equals “1.5”). Following the example of Cho (2006) and Troy and Grove (2008), records with low property prices (less than $50,000 in this case) were considered as either database errors or non, arms-length transactions and were excluded from the analysis. Additionally, records with house and lot sizes less than 500 ft^2 were considered to be database errors and were excluded. These and other property variables as well as their means and range are listed in Table 3.1. All of these variables, except for age, are expected to have a positive impact on property price.

Block group attributes from the 2000 Census were used as proxies for neighborhood characteristics that nearby properties shared. Median household income and percent unemployment were used to capture the relative economic status of a neighborhood. While median household income is expected to provide a positive contribution to property price, percent unemployment is expected to be negative. Median house value was expected to capture some of the spatial dependency in price that nearby properties shared. The percent of neighborhood residents having a high school diploma or college degree was used as a proxy of the social status of the neighborhood. A well educated neighborhood is expected to be associated with higher property values. Population density (per hectare) was a measure of the demand for and relative scarcity of land available for development. Higher population densities are expected to drive land price (and overall property prices) higher. Mean travel time to work (in minutes) was used as a proxy for distance to the nearest employment center.

The robbery index was a proxy for the social decay of a neighborhood. This variable was obtained from Tetrad, Inc.'s Crime database for Baltimore. The various crime statistics compiled in this database were initially reported in the FBI's Uniform Crime Report. These reports were grouped for the years 1996 through 2003 and standardized through a number of procedures described at the company's website. The final crime indices were calculated by Tetrad, Inc using a modeling procedure that adjusted for socio-economic variables and reported the results for any given block group as a number relative to the national total. With the national average index of 100, Baltimore's mean robbery index of over 700 indicates that it has a much higher level of crime, which can affect both house prices and the benefits of park proximity. The robbery index was the only crime statistic used in this research as it was felt that this would be one of the more visible and frequent crime indices that would affect property values in relation to both neighborhoods and nearby parks.

Two locational dummy variables were also included in this research to control for noise-related, negative externalities. If a property was less than 50m from the edge of a major road or highway (Census Feature Class Codes A11-A39) then it was expected that the noise-related externalities would have a negative impact on house prices that would outweigh the convenience of increased accessibility (to work, shopping, etc.). This dataset was derived from the GDT Inc.'s (Geographic Data Technologies) Dynamap™ road network. Similarly, a property less than 50m from an area zoned as commercial or industrial would be expected to be negatively impacted due to noise-externalities. Commercial and industrial sites were determined from a combination of Maryland's Land Use Land Cover 2000 dataset and the USGS 2001 National Land Cover Database.

Only Census block groups with at least 5 sales transactions during the 5-year period were included in this research. This resulted in a final dataset of 13,633 properties distributed among 401 block groups.

Table 3.1. Description of Regression Variables

Variables	Description	Min	Max	Mean
Property				
Price00[ab]	Sale price converted to Yr 2000	\$50,000	\$1,238,298	\$107,615
EnclsFt[b]	Total floor space (square feet)	510	10,185	1,475
LandM[b]	Lot size (square meters)	47	20,369	412
Bathnum	Number of full and half bathrooms	1.0	10.0	1.6
Age[c]	Age of house at the time of sale	0	199	67
HouseDum	Detached home vs. townhouse	0	1	0.3
AirDum	Presence of central air-conditioning	0	1	0.3
Basedum	Presence of basement	0	1	0.4
FireDum	Presence of a fireplace	0	1	0.2
GarDum	Presence of garage or carport	0	1	0.2
QualDumAvg	Avg. quality of original construction	0	1	0.2
QualDumHigh	High quality of original construction	0	1	0.1
Block group				
pUnemploy	Percent unemployment	0.1	19.3	4.9
MedValHouse	Median house value	\$32,500	\$596,300	\$91,249
pHSDiploma	Percent with a high school diploma	15.8	63.4	39.8
TravelMean	Mean, travel time to work (min)	17	53	31
Robbery	Crime risk index for robbery	24	1707	756
PopDens	Population density (per hectare)	1.3	190.2	50.9
MedHsInc	Median household income	\$12,095	\$170,428	\$38,858
Location				
ComIndDum	Within 50m of Comm./Ind. LULC	0	1	0.1
MjRDDum	Within 50m of a major road/hwy	0	1	0.2

a:dependent variable, b:natural log transformation, c:quadratic transformation

Parks data were obtained primarily from several data sources: official city parks from Baltimore's Department of Recreation and Parks; large open spaces created by the Parks and People Foundation where public green spaces were adjacent to official parks were re-designated as parks; and Baltimore County parks that were within a kilometer of the city boundary. The park boundaries were modified in a GIS to match parcel boundaries and aerial photos. Many of the features designated as official parks were actually very small slivers of open space associated with roads, intersections and medians. Following previous research by Troy and Grove (2008) and Bolitzer and Netusil (2000) open space less than 2 hectares were excluded from this research. The majority of these “parks” that were less than 2 hectares were actually just undevelopable open spaces such as grassy medians on boulevards or at the intersections of transportation arteries. Other large “parks” that were primarily built environments (e.g. sport complexes and parking lots) were also excluded from this research. Parks were further characterized within a GIS by describing the percent grass (i.e. open) and wooded, the area of the park (in hectares) and the robbery index for block groups to which that park belonged. Since a park often crossed multiple block group boundaries, the robbery index for individual parks was calculated from the area-weighted average of their block group values. While the level of park crime is expected to affect (interact with) the park-price relationship, the level of neighborhood crime is expected to directly affect the price of the property. Depending on the locations of both a property and its nearest park, the values of the two robbery indices may be very different from each other. The further a house is from the nearest park the less likely they belong to the same block group. A correlation test for the potential collinearity between these two crime variables indicated only a 0.25 Pearson’s Correlation Coefficient.

Distance to park was calculated in a GIS as both a Euclidean distance from a property’s point location to the nearest park boundary and a network distance defined by distance travelled along the road network (excluding highways) from the property to the edge of the nearest park. While network distance was the primary variable considered in this research, this value was replaced for properties less than 30m Euclidean distance from a park. It was felt that these properties were immediately adjacent to park boundaries and the network distance would be a larger and less accurate representation of their proximity. This replacement affected approximately 200 properties. The park distance variable was also log transformed, which suggests declining marginal impacts of park proximity with increasing distance to the park (the distance-decay effect) (Bin and Polasky 2004; Cho et al. 2006; Mahan et al. 2001; Orford 2002; Troy and Grove 2008). That is, the impact of increasing the distance to park from 100 to 200m will be greater than the impact from 500 to 600m. The actual form of the decay rate can be tested through model calibration as in Orford (2002) but was not performed in this research. With the general expectation of a park providing positive benefits to nearby properties, the resulting coefficient will be negative. Figure 3.1 shows the distribution of parks and block groups used for this research.

Figure 3.1. Map of Baltimore's Research Block Groups and Parks

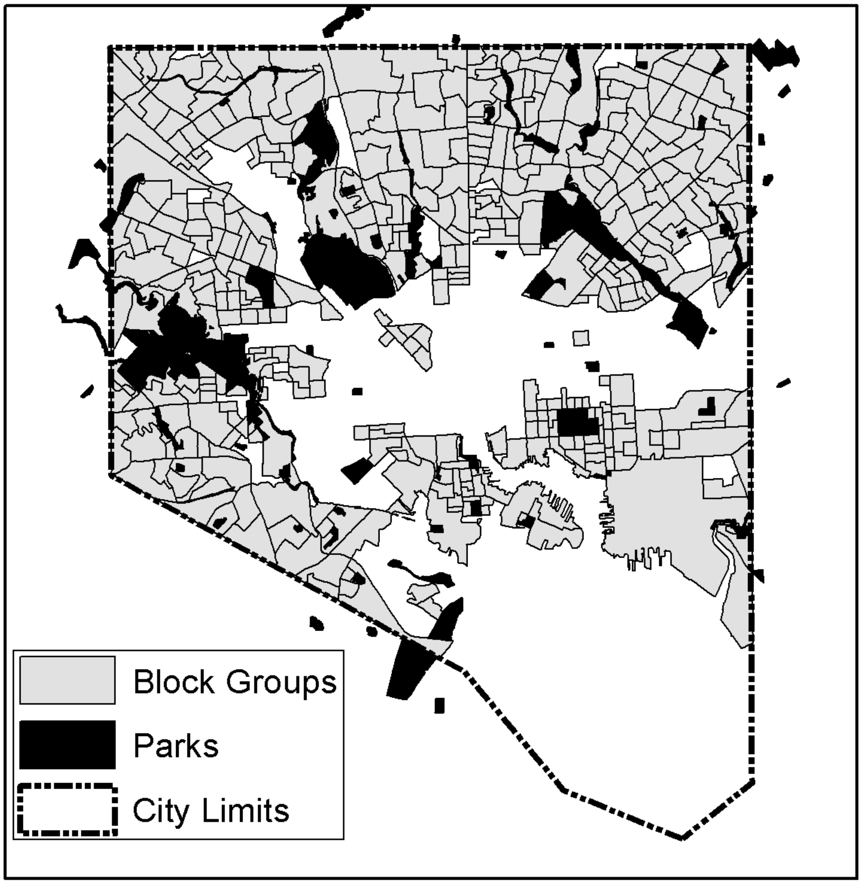


Table 3.2. Park Characteristics

Park Variables	Description	Min	Max	Mean
DPark	Distance to nearest park (meters)	1	2265	640
SzParkHa	Area of park (hectares)	2	380	55
pOpenPark	Percent without tree cover	0	100	58
RobPark	Crime Risk Index for parks	3	1544	680

3.5. Methods

In estimating a relationship between environmental amenities and property prices, the choice of functional form is not always clear. Rosen (1974) stressed that economic theory fails to indicate that any particular form is appropriate. Consequently, a variety of functional forms have been used in the hedonic literature. A linear form assumes that an individual's preferences are linear, implying that perfect repackaging of property characteristics is possible (Freeman 2003). However, in property markets, individual house characteristics are inseparable; an individual cannot mix characteristics in any other level than is already available in each house (Garrod and Willis 1992).

While the form that is chosen should ideally improve the model fit and help to satisfy important assumptions of OLS regressions, such as normally distributed residuals and homoscedasticity, this is not the main issue with choosing the proper functional form. The goal of finding a proper functional form is to overcome problems associated with the non-linearity that is often found in hedonic regression equations (Goodman and Thibodeau 1995). Substantively, this means that the proper functional form should be chosen so that marginal value for any given property attribute does not vary across the range of house prices.

A Box-Cox transformation analysis can be used to provide guidance on whether such simple forms are adequate for satisfying regression assumptions. The Box-Cox transformation of the dependent variable is shown as:

$$y(\lambda) = \begin{cases} \frac{y^{\lambda}-1}{\lambda}, & \text{if } \lambda \neq 0 \\ \log y, & \text{if } \lambda = 0 \end{cases} \quad \text{Eq. 3.1}$$

With this test, the parameter, λ, is estimated through maximum likelihood to find the optimal transformation of the dependent variable. This parameter can then be tested for significant differences between the optimal value of λ and three cases of λ that correspond to simpler functional forms: a reciprocal transformation, where λ=-1; a log transformation, where λ=0; and a linear (untransformed) form, where λ=1.

In this research, a left-hand (LHS) Box–Cox test found that the optimal transformation for price was indicated by Lambda value of -0.4. A chi-square test found this to be significantly different from zero indicating that a natural log transformation was not an optimal transformation for the dependent variable. However, results are reported for both of these transformations. Since the purpose of these estimated regression functions is to generate amenity values, it may be preferable to use a relatively simple form (Freeman 2003). The double-log form, in which both the dependent variable and the main effects are transformed using the natural logarithm, may provide the most interpretable results. With this form, a coefficient is interpreted as an elasticity; the percentage change in the dependent variable given the percentage change in an explanatory variable.

A log transformation was also used on the continuous property variables of house size and lot size. In the hedonic literature, this is a common method to account for the (non-linear) declining marginal value of house and lot size. Houses are also expected to depreciate with age at a declining rate but after a number of years, age will often be positively associated with house price. This may be due to unknown renovations or the "vintage effect" (Goodman and Thibodeau 1995) of older properties. Therefore, a quadratic transformation is used for the age variable. The inflection point where there was a positive effect of age on price was approximately 63 years old.

The regression model was then checked for collinearity within an OLS regression. A general rule of thumb is that variance inflation factors (VIF) greater than 10 are thought to be highly correlated and should be cause for further assessment before proceeding (O'Brien 2007). This research found VIF's below 4 for all variables except Age and Age-Squared. However, Shieh and Fouladi (2003) found that even in the presence of multicollinearity, for level-1 variables, the fixed-effect parameter estimates produce relatively unbiased values. Diagnostics were also used to check for the normal distribution of residuals, the existence of homogeneity of variances in the residuals, and the potential for heteroscedasticity and/or non-linear trends in the independent variable-price relationship.

While the analysis of multilevel models can be performed with a number of statistical packages, the details of using multilevel modeling with HLM (Hierarchical Linear Modeling) software are discussed below and follow the work of Raudenbush and Bryk (2002). A traditional property hedonic function (Equation 3.2) under the assumption of a single market can be written in terms of a vector of structural characteristics (S_i), neighborhood characteristics (N_i) and environmental or land use characteristic (L_i) and includes a single error term (r_i).

$$Y_i = S_i + N_i + L_i + r_i \qquad \text{Eq. 3.2}$$

With multilevel models, the property market is assumed to be composed of sub-markets where within-place property attributes (S) at the first level are separated from between-place, neighborhood characteristics (N) at the second level. The intercept of the level-1 equation becomes the dependent variable of the level-2 equation. The error term is also expanded so that there is unexplained variation at both levels (Equation 3.3).

$$\begin{aligned} &\text{Level 1: } Y_{ij} = ß_{0j} + ß_{1j}S_{ij} + \ldots ß_{nj}S_{ij} + r_{ij} \\ &\text{Level 2: } ß_{0j} = \gamma_{00} + \gamma_{01}N_j + .. \gamma_{0n}N_j + U_{0j} \end{aligned} \qquad \text{Eq. 3.3}$$

Where *Yij* is house *i*'s price in neighborhood *j* and each neighborhood has its own intercept ($ß_{0j}$) composed of a mean price (γ_{00}) and an error term (U_{0j}) with a between-group variance (τ_{00}) that is separate from the individual-level error (r_{ij}) and its within-group variance(σ^2). Since the model contains more than one error term, it cannot be estimated using OLS regression. Instead, an iterative maximum likelihood procedure is used (Raudenbush and Bryk 2002).

Property hedonic models often suffer from the effects of heteroscedasticity, or non-constant variance in the residuals. The presence of heteroscedasticity can create a downward bias to standard errors, which may create a spurious significance of a coefficient (Type I error). Multilevel models help mitigate this problem by separating the error term into two levels (Orford 2000). In addition, attributes that contribute most to the problem of heteroscedasticity in property hedonic models, such as age of dwelling

(Goodman and Thibodeau 1998), can be allowed to randomly vary across level-2 units (Equation 3.3). This distributes some of the unexplained variation expressed by the level-1 error term to variation specific to a random effect variable and helps to reduce the problem of heteroscedasticity that may be found in the residuals (Raudenbush and Bryk 2002). The presence of heteroscedasticity can be tested in HLM and if significant heteroscedasticity remains, then it is recommended that robust, "Huber-corrected" standard errors be used instead (Poor et al. 2007; White 1980).

In order to remove the problem of heteroscedasticity in this research, the structural variables of age, bathroom, house size and lot size were allowed to vary across block groups. However, the results from the HLM test for heteroscedasticity rejects the null hypothesis of homogeneity of level-1 variances. This indicates that heteroscedasticity still exists in the model and as such, only robust standard errors are reported in this research.

$$\begin{aligned} \text{Level 1: } Y_{ij} &= ß_{0j} + ß_{1j}S_{ij} + \ldots ß_{nj}S_{ij} + r_{ij} \\ \text{Level 2: } ß_{0j} &= \gamma_{00} + \gamma_{01}N_j + .. \gamma_{0n}N_j + \gamma_{0n}S_j + U_{0j} \qquad \text{Eq. 3.4} \\ ß_{1j} &= \gamma_{10} + U_{1j} \end{aligned}$$

Collinearity between the variances of the random effects of independent variables and the variance of the intercept term is a well-known cause of instability in the model (Kreft et al. 1995; Raudenbush and Bryk 2002). Such models have compromised estimates of uncertainty as well as possible bias (Gelman et al. 2007; Paccagnella 2006). One way of dealing with this problem is by group-mean centering those predictors entered into the model as random effects (Gelman et al. 2007; Raudenbush and Bryk 2002) (Equation 3.5). An examination of the Tau-As-Correlations matrix (Raudenbush and Bryk 2002) both before and after centering these random effects showed that group-mean centering removed this collinearity between the intercept variance and the variance of the independent variables that were modeled as random effects.

$$\begin{aligned}
\text{Level 1: } Y_{ij} &= ß_{0j} + ß_{1j}(S_{ij}-S_j)+\ldots ß_{nj} S_{ij} + r_{ij} \\
\text{Level 2: } ß_{0j} &= \gamma_{00} + \gamma_{01}N_j + .. \gamma_{0n}N_j + \gamma_{0n}S_j + U_{0j} \\
ß_{1j} &= \gamma_{10} + U_{1j}
\end{aligned} \qquad \text{Eq. 3.5}$$

In addition to considering certain property-level variables as random effects and group-mean centering these variables, the neighborhood means for age, house size and lot size were all added to the level-2 model. This tested whether there were compositional effects of place in addition to the contextual effects that were previously included in the full model. With respect to multilevel property hedonic models, the contextual effects -- the difference a place makes on price (e.g. neighborhood effects of unemployment, crime, median income, etc.) -- are potentially confounded with the compositional effects -- the differences produced by the housing attributes (e.g. house size, age) within each place (Orford 2000). So for example, the strongest effect of house size on price will likely occur at the property level (level 1). However, as a compositional effect, house size may have a different influence on the house price at the neighborhood level (level 2).

With multilevel property hedonics, spatial autocorrelation can be treated as the norm since individual houses in the same sub-market are likely to be more similar, in some way, than houses drawn from the entire housing market at random. Hence, autocorrelation is to be expected in hierarchical data, and the multilevel approach exploits this dependence to derive improved estimates, while the standard errors of the estimates are adjusted to take into account the autocorrelation (Goldstein 2003; Orford 2000). A test for spatial autocorrelation of the residuals using the global Moran's "I" statistic (Moran 1948) shows that there is a significant but slight amount of positive correlation remaining in the multilevel model's level-1 residuals (Moran's = 0.15) after including all regression variables except for land use. This is half the spatial autocorrelation that exists for the same model under OLS regression (Moran's I=0.3).

The variation of the park distance-price relationship across Baltimore is modeled by re-estimating the regression model with park distance as a random effect. The Empirical Bayesian variation in slopes between neighborhoods is added to the global fitted value to reveal areas of positive and negative relationships. Modeling random coefficients to investigate the effects of place on the relation between the main effect and the dependent variable is one of the fundamental advantages of multilevel modeling (Subramanian 2001). While it is possible to build OLS regressions to determine this effect for each neighborhood, there will be the problem that there are insufficient observations within a particular neighborhood. One approach to dealing with this problem is to construct Empirical Bayesian estimates which borrow strength across neighborhoods and shrink estimates for neighborhoods with few observations towards the overall mean (Raudenbush 2002). This means that the implicit prices of certain attributes (those that are considered random effects) are optimally weighted averages that combine information derived from the group itself with the mean from neighborhoods with similar characteristics (Diez-Roux et al. 2000). Unreliable submarket estimates are differentially shrunk towards the global estimate, whereas submarkets with many properties will not be affected by this shrinkage. This pooling of information and borrowing of strength is more analogous with the definition of submarkets as being quasi-independent and functionally related (Orford 2000).

In multilevel models, interactions can occur between variables at the same level or between variables at different levels. In this research, the interest is on whether there is a mediating effect of certain property, park and neighborhood characteristics on the park proximity-price relationship. Following Hox (2002), if there is a significant interaction found between one of these characteristics and the park coefficient, then the direct effects of that characteristic must also be included even if it is found to be insignificant. There is also the decision on whether to group-mean center the park variable. Enders and Tofighi (2007) found that use of grand mean centering can artificially produce a significant effect from the interaction between variables (single-level or cross-level) when in reality one does not exist. They conclude that group-mean centering is more appropriate when (either cross-level or same-level) interactions are the research interest.

3.6. Results

The results of the multilevel models are presented in Table 3.3. The property structural characteristics are all highly significant (at the 99% confidence level) and show the expected signs. The log transformation of floor space and lot size means that these coefficients are interpreted as the elasticity of sales price, controlling for other coefficients, so that property price increases by 0.29% and 0.06% for every 1% increase in floor space and lot size, respectively. There is a semi-log relationship between price and most variables, so what is estimated is the percent change in property price with a 1-unit change of that particular variable. There is an approximate 4% increase in sale price with the addition of a half bath. Property price falls by .28% for every additional year of house age until the house becomes older than 63 years, at which point property value increases by a slight but significant amount. For the dummy variables, there is: an approximate 11% increase in sale price if the building is a single, detached dwelling (vs. a townhouse); a 6% increase in sale price with the presence of central air conditioning; a 4% increase in sale price with the presence of a basement; a 3% increase in sale price with the presence of a fireplace; and a 4% increase in sale price with the presence of a garage. There is a 33% and 59% increase in the sale price if the property was built with average or high quality materials, respectively, rather than poor quality materials.

Table 3.3. Regression Results

Model 1: Log transformed Price				Model 2: Box-Cox Price $((Price^{-0.4} - 1)/-0.4)$			
Variables	Coef.	S.E.	Sig.	Variables	Coef.	S.E.	Sig.
Property				**Property**			
HouseDum	0.105757	0.009656	0.000	HouseDum	0.001151	0.000094	0.000
AirDum	0.056331	0.004245	0.000	AirDum	0.000570	0.000042	0.000
BaseDum	0.039594	0.003690	0.000	BaseDum	0.000410	0.000036	0.000
FireDum	0.029402	0.005391	0.000	FireDum	0.000345	0.000049	0.000
GarDum	0.039586	0.004912	0.000	GarDum	0.000391	0.000049	0.000
QualDumAvg	0.328124	0.014903	0.000	QualDumAvg	0.003363	0.000153	0.000
QualDumHigh	0.588582	0.019238	0.000	QualDumHigh	0.005485	0.000189	0.000
BathNum	0.028686	0.003728	0.000	BathNum	0.000260	0.000036	0.000
Age	-0.002842	0.000439	0.000	Age	-0.000030	0.000004	0.000
AgeSqd	0.000010	0.000003	0.001	AgeSqd	0.000000	0.000000	0.000
lnEnclsFt	0.288438	0.012723	0.000	lnEnclsFt	0.002856	0.000114	0.000
lnLandM	0.062850	0.005689	0.000	lnLandM	0.000635	0.000056	0.000
Block Group				**Block Group**			
Intercept	10.791893	0.058301	0.000	Intercept	2.467887	0.000633	0.000
MedHsInc	0.000001	0.000001	0.588	MedHsInc	0.000000	0.000000	0.701
pUnemploy	-0.008082	0.001783	0.000	pUnemploy	-0.000098	0.000019	0.000
MedValHouse	0.000002	0.000000	0.000	MedValHouse	0.000000	0.000000	0.001
pHSDiploma	0.001803	0.000748	0.017	pHSDiploma	0.000025	0.000008	0.002
TravelMean	-0.003041	0.001054	0.005	TravelMean	-0.000041	0.000011	0.000
Robbery	-0.000040	0.000015	0.006	Robbery	-0.000001	0.000000	0.001
PopDens	0.000444	0.000221	0.044	PopDens	0.000004	0.000002	0.068
mnEnclsFt	0.000092	0.000026	0.001	mnEnclsFt	0.000001	0.000000	0.000
mnAge	0.001264	0.000407	0.002	mnAge	0.000012	0.000004	0.008
mnLandM	0.000028	0.000030	0.343	mnLandM	0.000000	0.000000	0.297
Location				**Location**			
ComIndDum	-0.013940	0.006199	0.024	ComIndDum	-0.000162	0.000063	0.010
MjRdDum	-0.025466	0.005845	0.000	MjRdDum	-0.000232	0.000057	0.000
Main Effect				**Main Effect**			
lnDistToPark	-0.003336	0.002718	0.220	lnDistToPark	-0.000047	0.000028	0.091

Although many of the estimated coefficients for the property's neighborhood characteristics are significant (at the 90% confidence level) with the expected signs, median household income was found to be insignificant and excluded from further analyses. There is a 0.8% decrease in sale price with a 1-unit increase in neighborhood (percent) unemployment; a 0.2% increase for every $10,000 increase in median house value; a 0.2% increase in sale price with a 1-unit increase in a neighborhood's percent of population with (at least) a high school diploma; a 0.3% decrease for every additional minute in a neighborhood's mean travel time to work; a 0.004% decrease for every 1-unit increase in a neighborhood's crime index for robberies; and a 0.04% increase for every 1-unit increase in population density.

The results from including block group averages for house size, lot size and age indicate that mean house size and age were both positive and significant. Mean lot size was not found to be significant and was dropped from further analyses. There was a 0.009% increase in mean property price with each additional, square foot increase in a neighborhood's mean house size. There was also a 0.1% increase in mean property price with each additional year in a neighborhood's mean house age.

The estimated coefficient for distance to park, although negative as expected, is non-significant in the first model (log transformed price) while significant (at the 90% confidence level) with a Box-Cox transformation of the dependent variable. With the first model, there is an elasticity between the park coefficient and price (i.e. both are log transformed) and the result can be interpreted as a 0.003% decrease in price with every 1% increase in distance.

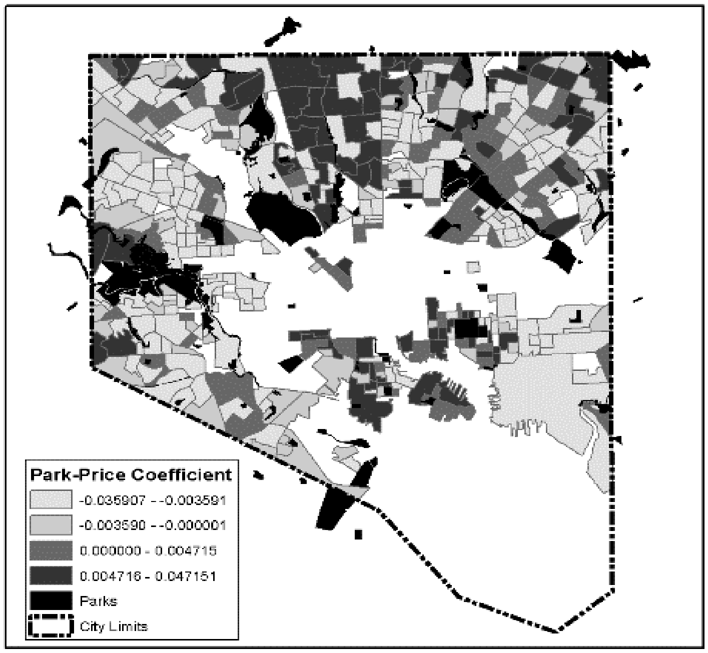
The variance component for the random park coefficient (as well as for the other random effects) indicates that there is significant variation in the slopes across neighborhoods (Table 3.4). The maps of the variation in the park coefficient by block group (Figure 3.2) do not appear to show any consistent spatial pattern to this variation.

Table 3.4. Variance Components for Model 1

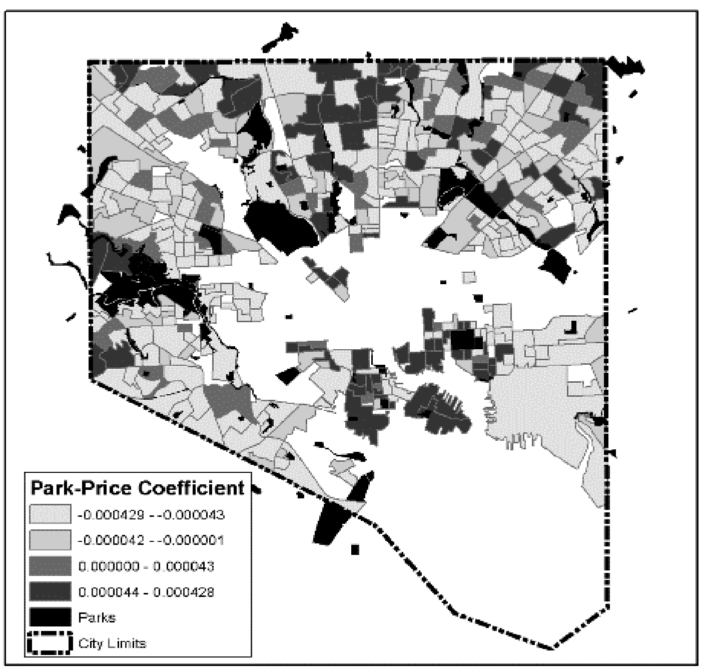
Random Components	SD	Variance	df	Chi-Square	Sig.
Level-1, r	0.14357	0.02061			
Intercept, U_0	0.11575	0.01340	362	9437.9	0.000
BathNum, U_8	0.04107	0.00169	372	561.0	0.000
Age, U_9	0.00183	0.00000	372	705.3	0.000
lnEnclsFt, U_{11}	0.16581	0.02749	372	771.6	0.000
lnLandM, U_{12}	0.04783	0.00229	372	497.7	0.000
lnDistToPark, U_{15}	0.02354	0.00055	372	568.8	0.000

Figure 3.2. Maps of the Variation of the Park Coefficient Across Block Groups

Model 1: Log transformed Price



Model 2: Box-Cox Price ($(\text{Price}^{-0.4} - 1)/-0.4$)



The effect of park proximity on property price is further modeled with interactions between the park variable and several property, park or neighborhood characteristics. None of these interactions were found to be significant for the entire model. However, it is possible that the variables that mediate the park effect may be different between properties that positively value proximity to parks and those that negatively value this proximity. Therefore, the dataset is split into those two population groups and reexamined for interactions. The dataset of properties for neighborhoods that positively value proximity to parks contains 8,667 properties distributed among 283 block groups and comprises approximately two-thirds of the original dataset. The dataset of properties for neighborhoods that do not appear to value proximity to parks contains 4,966 properties distributed among 118 block groups. None of these interactions are significant for the population of properties in neighborhoods that value being farther away from parks. The results of significant interactions for those properties that positively value proximity to parks are shown in Table 3.5.

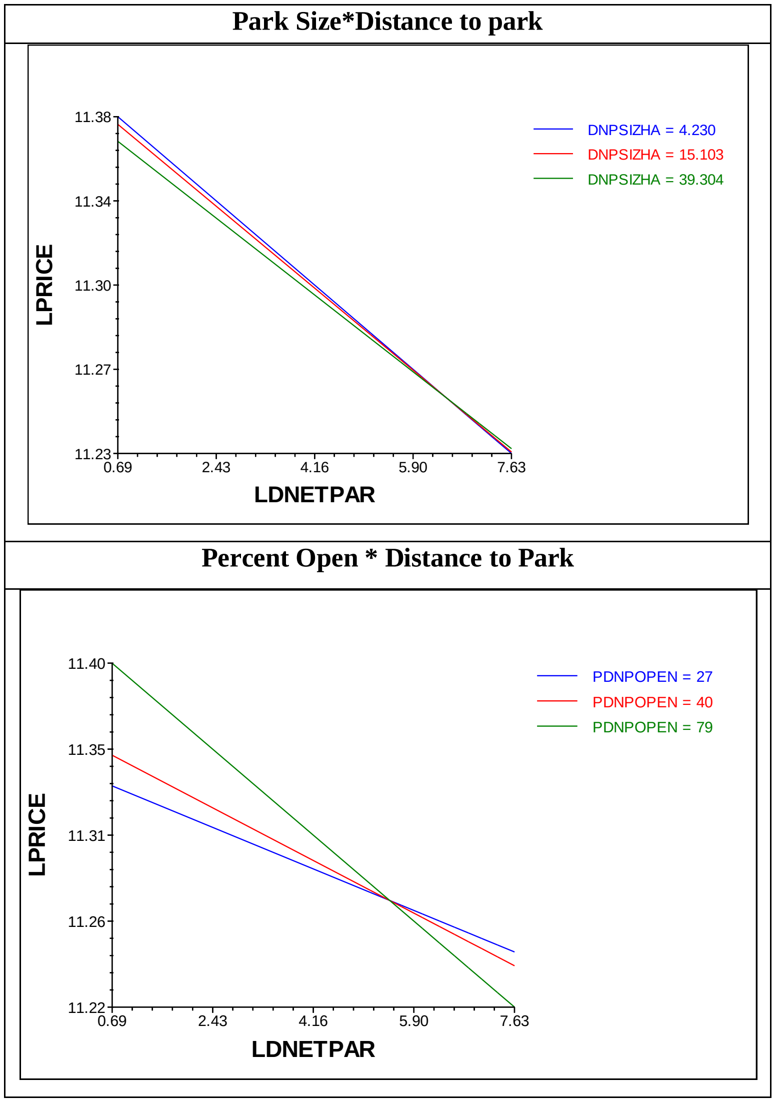
Table 3.5. Significant Interactions for Properties in Block Groups that Positively Value Proximity to Parks

Model 3: Log transformed Price				Model 4: Box-Cox Price $((Price^{-0.4} - 1)/-0.4)$			
<u>Variables</u>	<u>Coef.</u>	<u>S.E.</u>	<u>Sig.</u>	<u>Variables</u>	<u>Coef.</u>	<u>S.E.</u>	<u>Sig.</u>
Main Effect				**Main Effect**			
DistPark	-0.017612	0.002740	0.000	DistPark	-0.000195	0.000027	0.000
Property				**Property**			
lnLandM	0.031809	0.011997	0.009	lnLandM	0.000209	0.000129	0.107
lnLandM *lnDPark	0.008091	0.003169	0.011	lnLandM *lnDPark	0.000082	0.000033	0.013
Park				**Park**			
ParkSize	-0.000258	0.000142	0.068	ParkSize	-0.000002	0.000002	0.176
ParkSize* lnDPark	0.000048	0.000023	0.032	ParkSize* lnDPark	0.000000	0.000000	0.074
pOpen	0.000615	0.000640	0.337	pOpen	0.000003	0.000006	0.608
pOpen* lnDPark	-0.000162	0.000096	0.090	pOpen* lnDPark	-0.000001	0.000001	0.269
pWod	-0.000792	0.000649	0.223	pWood	-0.000005	0.000007	0.456
pWood* lnDPark	0.000179	0.000097	0.065	pWood* lnDPark	0.000001	0.000001	0.203
Block Group				**Block Group**			
PopDens	0.000104	0.000235	0.659	PopDens	0.000000	0.000003	0.851
PopDens* lnDPark	-0.000186	0.000090	0.039	PopDens* lnDPark	-0.000002	0.000001	0.053
mnLand	0.000038	0.000032	0.245	mnLand	0.000001	0.000000	0.138
mnLand* lnDPark	0.000013	0.000008	0.101	mnLand* lnDPark	0.000000	0.000000	0.089

For Model 3, the interaction of a property's lot size (lnLandM*lnDPark) is positive and significant at the 95% confidence level, indicating that as lot size increases there is a decline in the value of being in close proximity to a park.

Park characteristics are also expected to have an effect on the park-price relationship. For Model 3, the results of the direct effect of park size on property price are negative and significant at the 90% confidence level. A 1-hectare increase in park size decreases the value of properties by 0.03%. The interaction with park distance is positive and significant, indicating that there is a 0.005% increase in property price with every 1-unit increase in the product of the logged park distance and park size. The results of the direct effect of percent openness of a park (i.e. the proportion that is not wooded) on property price are positive and significant at the 95% confidence level. A 1-unit increase in percent openness increases the value of properties by 0.06%. The interaction with park distance is negative and significant, indicating that there is a 0.02% decrease in property price with every 1-unit increase in the product of the logged distance and percent openness (Figure 3.3). The level of crime, as indicated by the robbery index, was not found to provide a significant interaction effect with the park variable.

Figure 3.3. Interactions with Park Size and Openness for Model 3



Neighborhood characteristics are also expected to have an effect on the park-price relationship. However, only the neighborhood's population density and mean lot size were found to have significant interactions with park variable. For Model 3, the results of the direct effect of population density have a negative and insignificant effect on price while the effect is positive and significant in Models 1 and 2. The interaction with park distance is negative and significant, indicating that there is a 0.02% decrease in property price with every 1-unit increase in the product of the logged park distance and population distance. The results of the interaction with park distance and the average lot size for a neighborhood are similar to the results found in the interaction with individual lot sizes.

3.7. Discussion

The inclusion of an average block group for age, house size and lot size tested whether there were compositional effects of place in addition to the neighborhood's contextual effects that were previously included. This positive impact of neighborhood age is a good example of the avoiding the ecological-atomistic fallacies where the relationship at one level is expected to be the same as that of another level. Property value clearly depreciates as a house ages (to a certain point) simply through deterioration of the structure from wear-and-tear. In contrast, as a neighborhood matures (as represented by mean house age), other features of the neighborhood become well developed and improve the value of an individual home. An obvious example of this is the growth of trees and vegetation improving the amenity value of the area.

The non-significant effect of the level-1 park coefficient may be due to local variation (non-stationarity) in the price-park distance relationship. A strongly negative park-price relationship in some areas of the city may negate the existence of a significant and positive relationship across the city. This possibility is examined by allowing the park coefficient's slopes to vary across neighborhoods. The maps of the variation in park distance suggest that the global estimate of the park coefficient may underestimate or overestimate the amenity value of parks in particular neighborhoods by a substantial margin although there does not seem to be a clear spatial trend to this variation. The degree of local heterogeneity found in this research suggests that the many previous studies that have found a positive, global effect of park proximity may not be adequately describing the complex spatial dynamics involved in property markets and the valuation of environmental amenities.

The park coefficient was also allowed to interact with property, neighborhood and park characteristics to determine possible reasons for this dramatic variation in the park-price relationship. The positive interaction of a property's lot size with the park coefficient indicates that as lot size increases there is a decline in the value of being in close proximity to a park. This suggests that a large yard acts a substitute for the recreational and aesthetic opportunities provided by parks. While this substitution effect was expected, it is in contrast to the complementary effect found by Anderson and West (2006).

The results of park characteristics of size and percent open space suggest that smaller and more open parks command a higher value for nearby properties than those larger, wooded parks maintained in a more natural state. This is in contrast to the findings of Bolitzer and Netusil (2000) who found that larger, natural parks in Portland, Oregon provided a greater contribution to the value of a house. One possible explanation is that Baltimore residents have a heightened level of fear of being victims of crime because most crime indices for Baltimore are much higher than the national average. Larger, natural parks can buffer criminals from law enforcement and neighborhood watch groups who can better monitor conditions in small, open parks. While the interaction with neighborhood and park crime was found to be insignificant, this study only examined the crime index for robbery. It is possible that including other crimes such as rape and murder may show a more significant interaction with proximity to parks.

While most neighborhood interactions were found to be insignificant, population density provided a negative and significant effect. This suggests that properties found in neighborhoods with the highest population densities command the largest premium for park proximity. Such results are expected and are inversely correlated with the effect of large lot sizes; a densely populated neighborhood will have fewer private open spaces, thus increasing the value of local parks.

Unfortunately, none of these interactions examined in this study could significantly explain why a third of the neighborhoods had properties that negatively that valued close proximity to parks. As this is in contrast to the many studies that found a positive global effect for park proximity, further research is necessary to determine whether this is an anomaly related to unidentified characteristics of the City of Baltimore or whether this is a result of multilevel modeling that explicitly captures the spatial variation in the valuation of environmental amenities.

The multilevel approach used in this research may be flawed in that market segmentation through common property and neighborhood characteristics might not create ideal neighborhood boundaries especially concerning the valuation of park proximity. The spatial nature of the environmental amenities from parks implies that neighboring properties will capture a similar impact from the externality because of their similarity in proximity to the amenity. However, these properties that share similar proximity to a local amenity may not be located in the same neighborhood. They may be located in the next, contiguous neighborhood or in a neighborhood on the other side of a park. This potential for misalignment might be corrected by considering the cross-classification of properties into both neighborhood groups and the parks to which they are closest.

Orford (2000) and Goodman and Thibodeau (1998) suggests that the neighborhood is best thought of not as a single entity, but rather as a hierarchy of progressively more inclusive residential groupings. Under this reasoning, the inclusion of a third level such as the Tract or PRIZM cluster might have created different results than those found.

3.8. Conclusion

This research uses multilevel modeling to the hedonic approach of estimating the effect of proximity to parks on sales price in Baltimore, Maryland. The multilevel approach first modeled the neighborhoods or submarkets of the city's housing market as a function of property-level and socio-economic, neighborhood-level factors. The effect of parks was then allowed to vary by neighborhoods and mapped to reveal the complex spatial variation found with this variable. These maps showed that only two-thirds of the neighborhoods examined in this research positively value being in close proximity to parks. The remaining one-third of the neighborhoods show a preference for being more distant from parks.

Separating these neighborhoods into two datasets, I then tested whether the effects of park proximity were moderated by property, park and neighborhood characteristics. The results indicated that larger lots at both individual and average neighborhood levels acted as strong substitutes for parks. Park characteristics also seem to have some effect on the park-price relationship. Smaller and more open parks increased the value of nearby parks. Only the neighborhood characteristic of population density seemed to affect the park-price relationship. The value for being in close proximity to parks increased with increasing population densities. However, these interaction effects were only significant for the population of properties in neighborhoods that showed a preference for being close to parks. The reason that the remaining one-third of properties did not value park proximity was not explained by the interactions examined in this research.

While the property hedonic model can be used to estimate the value of some ecosystem goods and services, it is important to remember that the method provides only a limited measure of total economic benefits. Parks provide many services in addition to the amenities of view and recreation. Parks also provide recreation to visitors as well as to the local residents. This research only studied the effects of park proximity on home and townhouse prices. The potential, property hedonic benefits to condominiums, apartments, and businesses was not observed. For these reasons, estimates from hedonic house price models will generally under-represent the total, non-market value of the ecosystem goods and services that are provided by parks.

3.9. References

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

Allen, T.F.H., and Starr, T.B. 1982. Hierarchy: Perspectives for ecological complexity. The University of Chicago Press, Chicago.

Anderson, S.T., and West, S.E. 2006. Open space, residential property values, and spatial context. Reg. Sci. Urban Econ. **36**: 773–789.

Anselin, L. 1988. Spatial econometrics : methods and models. Kluwer Academic Publishers London.

Bastian, C.T., McLeod, D.M., Germino, M.J., Reiners, W.A., and Blasko, B.J. 2002. Environmental amenities and agricultural land values: a hedonic model using geographic information systems data. Ecol. Econ. **40**: 337-349.

Bin, S., and Polasky, S. 2004. Effects of flood hazards on property values: Evidence before and after Hurricane Floyd. Land Econ. **80**: 490-500.

Bolitzer, B., and Netusil, N.R. 2000. The impact of open spaces on property values in Portland, Oregon. J. Environ. Manage. **59**: 185-193.

Bourassa, S.C., Hoesli, M., and Peng, V.S. 2003. Do housing submarkets really matter? Journal of Housing Economics **12**: 12–28.

Brown, K.H., and Uyar, B. 2004. A hierarchical linear model approach for assessing the effects of house and neighborhood characteristics on housing prices. Journal of Real Estate Practice and Education **7**(1): 15-23.

Cho, S.H., Bowker, J.M., and Park, W.M. 2006. Measuring the contribution of water and green space amenities to housing values; An application and comparison of spatially weighted hedonic models. J. Agric. Resour. Econ. **31**(3): 485-507.

Conway, D., Li, C.Q., Wolch, J., Kahle, C., and Jerrett, M. 2008. A spatial autocorrelation approach for examining the effects of urban greenspace on residential property values. *In* Journal of Real Estate Finance and Economics.

Day, B., Bateman, I., and Lake, I. 2004. Nonlinearity in hedonic price equations: An estimation strategy using model-based clustering. University of East Anglia: Centre for Social and Economic Research on the Global Environment (CSERGE).

Dehring, C., and Dunse, N. 2006. Housing density and the effect of proximity to public open space in Aberdeen, Scotland. Real Estate Economics **34**(4): 553 - 566.

Diez-Roux, A.V. 2002. A glossary for multilevel analysis. Journal of Epidemiology and Community Health **56**: 588–594.

Diez-Roux, A.V., Link, B.G., and Northridge, M.E. 2000. A multilevel analysis of income inequality and cardiovascular disease risk factors. Social Science & Medicine **50**: 673-687.

Do, A.Q., and Grudnitski, G. 1995. Golf courses and residential house prices: an empirical examination. Journal of Real Estate Finance and Economics **10**: 261–270.

Ekeland, I., Heckman, J.J., and Nesheim, L. 2002. Identifying hedonic models. American Economic Review **92**(2): 304-309.

Enders, C.K., and Tofighi, D. 2007. Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. Psychol. Methods **12**(2): 121-138.

Espey, M., and Owusu-Edusei, K. 2001. Neighborhood parks and residential property values in Greenville, South Carolina. Journal of Agricultural and Applied Economics **33**(3): 487-492.

Fotheringham, A.S., Brunsdon, C., and Charlton, M.E. 2002. Geographically Weighted Regression: The analysis of spatially varying relationships. Wiley, Chichester.

Freeman, A.M. 2003. The measurement of environmental and resource values: Theory and methods. Resources for the Futures, Washington, D.C.

Garrod, G.D., and Willis, K.G. 1992. Valuing goods' characteristics: an application of the hedonic price method to environmental attributes. J. Environ. Manage. **34**(1): 59-76.

Gelfand, A.E., Banerjeeb, S., Sirmans, C.F., Tu, Y., and Ong, S.E. 2007. Multilevel modeling using spatial processes: Application to the Singapore housing market. Comput. Stat. Data Anal. **51**: 3567 – 3579.

Gelman, A., Bafumi, J., Shor, B., and Park, D.K. 2007. Rich state, poor state, red state, blue state. What's the matter with Connecticut? Quarterly Journal of Political Science **2**: 345-367.

Geoghegan, J., Wainger, L.A., and Bockstael, N.E. 1997. Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS. Ecol. Econ. **23**(3): 251-264.

Goldstein, H. 2003. Multilevel statistical models. Oxford University Press, NY, NY.

Goodman, A.C., and Thibodeau, T.G. 1995. Age-related heteroskedasticity in hedonic house price equations. Journal of Housing Research **6**(1): 25-42.

Goodman, A.C., and Thibodeau, T.G. 1998. Housing market segmentation. Journal of Housing Economics **7**: 121–143.

Goodman, A.C., and Thibodeau, T.G. 2003. Housing market segmentation and hedonic prediction accuracy. Journal of Housing Economics **12**: 181–201.

Henderson, K.K., and Song, Y. 2008. Can nearby open spaces substitute for the size of a property owner's private yard? International Journal of Housing Markets and Analysis **1**(2): 147 - 165.

Hox, J.J. 2002. Multilevel analysis: techniques and applications. Lawrence Erlbaum Associates, Mahwah, NJ.

Irwin, E.G. 2002. The effects of open space on residential property values. Land Econ. **78**(4): 465-480.

Irwin, E.G., Bockstael, N.E. 2001. The problem of identifying land use spillovers: Measuring the effects of open space on residential property values. Am. J. Agr. Econ. **83**(3): 698–704.

Jones, K., and Bullen, N. 1993. A multi-level analysis of the variations in domestic property prices. Urban Studies **30**(8): 1409-1427.

Jones, K., and Duncan, C. 1996. People and places: the multilevel model as a general framework for the quantitative analysis of geographical data. *In* Spatial Analysis: Modelling in a GIS Environment *Edited by* P. Longley, Batty, M. John Wiley & Sons, New York. p. 389.

Kong, F., Yin, H., and N., N. 2007. Using GIS and landscape metrics in the hedonic price modeling of the amenity value of urban green space: A case study in Jinan City, China. Landscape and Urban Planning **79**: 240–252.

Kreft, I.G.G., Deleeuw, J., and Aiken, L.S. 1995. The effect of different forms of centering in hierarchical linear-models. Multivariate Behav. Res. **30**(1): 1-21.

Lee, C., and Linneman, P. 1998. Dynamics of the greenbelt amenity effect on the land market: The case of Seoul's greenbelt. Real Estate Economics **26**(1): 107–129.

Lutzenhiser, M., and Netusil, N.R. 2001. The effect of open spaces on a home's sale price. Contemporary Economic Policy **19**(3): 291–298.

Mahan, B.L., Polasky, S., and Adams, R.M. 2001. Valuing urban wetlands: A property price approach. Land Econ. **76**(1): 100–113.

Moran, P.A.P. 1948. The interpretation of statistical maps. Journal of the Royal Statistical Society: Series B **10**: 243-251.

Morancho, A.B. 2003. A hedonic valuation of urban green areas. Landscape and Urban Planning **66**: 35-41.

O'Brien, R.M. 2007. A caution regarding rules of thumb for variance inflation factors. Quality & Quantity **41**: 673–690.

O'Neilll, R.V., Deangelis, D.L., Waide, J.B., and Allen, T.F.H. 1986. A hierarchical concept of ecosystems. Princeton University Press, Princeton, N.J.

Orford, S. 2000. Modelling spatial structures in local housing market dynamics: A multilevel perspective. Urban Studies **37**(9): 1643-1671.

Orford, S. 2002. Valuing locational externalities: A GIS and multilevel modelling approach. Environment and Planning B: Planning and Design **29**(1): 105-127.

Paccagnella, O. 2006. Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. Evaluation Review **30**(1): 66-85.

Paez, A., and Scott, D.M. 2004. Spatial statistics for urban analysis: A review of techniques with examples. GeoJournal **61**: 53–67.

Poor, P.J., Pessagnob, K.L., and Paul, R.W. 2007. Exploring the hedonic value of ambient water quality: A local watershed-based study. Ecol. Econ. **60**: 797-806.

Raudenbush, S.W., and Bryk, A.S. 2002. Hierarchical Linear Models: Applications and Data Analysis Methods. Sage Publications, London.

Rosen, S. 1974. Hedonic prices and implicit markets: product differentiation in pure competition. The Journal of Political Economy **82**(1): 34–55.

Shieh, Y.Y., and Fouladi, R.T. 2003. The effect of multi-collinearity on multilevel modeling parameter estimates and standard errors. Educ. Psychol. Meas. **63**: 951-985.

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

Snijders, T.A.B., and Bosker, R. 1999. Multilevel analysis. An introduction to basic and advanced multilevel modeling. Sage, Thousand Oaks, CA.

Subramanian, S.V., Duncan, C., Jones, K. 2001. Multilevel perspectives on modeling census data. Environ. Plan. A **33**(3): 399-417.

Thorsnes, P. 2002. The value of a suburban forest preserve: Estimates from sales of vacant residential building lots. Land Econ. **78**(3): 426–441.

Troy, A., and Grove, J.M. 2008. Property values, parks, and crime: A hedonic analysis in Baltimore, MD. Landscape and Urban Planning **87**: 233–245.

Tyrväinen, L., and Miettinen, A. 2000. Property prices and urban forest amenities. J.Environ.Econ.Manage. **39**(2): 205–223.

White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica **48**(4): 817-838.

CHAPTER 4: A MULTILEVEL PROPERTY HEDONIC APPROACH TO MEASURING THE CONTRIBUTION OF OPEN SPACE TO PROPERTY VALUES IN SUBURBAN AREAS OF BALTIMORE COUNTY

4.1 Abstract

Undeveloped open spaces, both private and public, provide numerous social and ecological benefits that are not fully apparent in the property-land market. Further development of such areas may result in welfare losses. This study uses multilevel models to estimate the contribution of open space to the value of properties in the suburban areas of Baltimore County that were sold between 1998 and 2000. The area of three different types of open space: private-conserved; private-developed and public, were measured within 100m,500m and 1000m distances of an individual property and for the block group in which each property resides. The results show that, after controlling for structural and neighborhood characteristics, privately-owned open space significantly increase property values, while public open space provides a lesser and statistically insignificant benefit. In addition, private open space that is permanently preserved provides a higher premium than potentially developable open space. The difference between private conserved versus potentially-developable open space suggest that residents negatively value the uncertainty about the future status of the latter, with the potential for higher levels of negative externalities than currently exist. The large difference in benefit between private and public open space suggests that negative externalities of public open space may outweigh its positive recreational or aesthetic

value. The implications for open space funding suggest that purchasing the development rights, through conservation easements might be preferable to outright public acquisition of undeveloped land.

4.2 Introduction

Open spaces provide numerous benefits to the local population in a suburban setting, including recreation opportunities, habitat and scenic amenity, improvement of water quality, mitigation of flooding, preservation of agricultural and forestry jobs and more. Most of the environmental benefits related to open space are external to normal market transactions and consequently are often undervalued and under-provisioned even though they enhance the quality of people's lives. Capturing the monetary value of these benefits is important for improving individual and social welfare and for urban planning issues such as zoning, development, land conservation acquisitions, property taxation and improvements to and maintenance of existing open spaces.

One approach for estimating these non-market environmental values uses a revealed-preference technique, known as the property hedonic model. With this approach, an individual property is considered to be composed of a bundle of characteristics, each of which implicitly contributes to the price of the property. These characteristics can be broadly grouped into three categories: property-specific (including both the land and structural improvements); contextual neighborhood-specific (the socio-economic context); and locational (Freeman 2003). The composition of open space and competing land uses are important location characteristics that can have a significant impact on property values.

The hedonic approach for estimating the impact of open spaces on property value has been researched in numerous locations and circumstances. Much of this previous research has focused on how distance to a specific type of open space affected the property price. In general, the recreational opportunities and visual amenities provide positive but declining benefits with increasing distance between open space and local residences. Some of the various types of open space that have been examined are: urban parks (Bolitzer and Netusil 2000; Morancho 2003; Orford 2002; Troy and Grove 2008), golf courses (Do and Grudnitski 1995; Lutzenhiser and Netusil 2001), greenbelts (Lee and Linneman 1998), forest preserves (Garrod and Willis 1992; Thorsnes 2002; Tyrväinen and Miettinen 2000), wetlands (Mahan et al. 2001) and agriculture (Bastian 2002). Many of these types are permanently undeveloped, publicly owned and publicly accessible. The emphasis in these studies has been to estimate the value of accessibility, often using Euclidean distance as a proxy for the accessibility to the closest open space. However, the composition (amount) of all types of land uses surrounding a property will likely have important effects that are not fully captured by distance metrics. For example, finding the distance to the closest park ignores the existence of other parks that are slightly further from the residence.

The value of these various and competing land uses may also be tied to the issue of spatial scale. The potential amenities or negative impacts from the presence of nearby open spaces on residences will depend on the spatial extent of the externalities of various land uses that spillover onto a property. Rather than defining the neighborhood at one scale, it would be better to interpret this space at multiple scales that may be hierarchical or overlapping depending on the particular variable of interest. For the suburban region surrounding Baltimore, one would expect that a diversity of land types would be valuable to a residence because it increases the number of nearby destinations such as working, shopping and recreation. However, in the immediate neighborhood of a home, mixed uses may be less desirable because it may increase the level of negative externalities. It would therefore seem important to capture this variation in impacts by examining the effects at multiple scales (hierarchical and/or non-hierarchical) within a property hedonic model.

Geoghehan and others (1997) researched the issue of the effects of multiple spatial scales of certain types of open space on property values. The authors theorized that the amount of open space within 100m of a property would have a different impact on property values than the amount of open space within 1km of the same property. They find that within the 100m buffer, the proportion of open space positively impacts property values, but within a 1-km buffer this variable negatively influences land prices. Their findings suggest that having a relatively high amount of immediately adjacent visual and recreation amenities is considered to be beneficial by local residents but that a high proportion of open space within 1km of their property reduces the conveniences associated with developed areas such as shopping and entertainment.

Acharaya and Bennett (2001) also research this differential effect of spatial scales on the impact open space on property values, using 1/4-mile and 1 mile buffers to distinguish between visual and neighborhood impacts within a walk-able distance from the home. They found that proportion of open space for both buffers generated a positive effect on house price. Including a variable for the percentage squared (that is, a quadratic form), they find that an increase in the percentage of open space around a house increases the value of the property but at a decreasing rate (the coefficient on the squared term is negative).

Kestens and other (2004) also examine the effects of land use at several spatial scales. They used buffers of 40m and 100m to represent the visual space surrounding a property. Their 500m buffer represented the walking space around the property and their 1km buffer represented the overall effect of a neighborhood. They theorize that visual impacts would be the most important to an individual residence and therefore the benefits of nearby open space and natural vegetation would be greatest in the immediate area surrounding a property.

In addition to the issue of spatial scale, the type of open space used in a hedonic study will affect the estimated coefficients. Open space in previous research has been described by cover type (e.g. deciduous vs. coniferous forest), land use (e.g. pasture vs. park), ownership (e.g. private vs. public), accessibility (e.g. open to the public vs. semi-public open vs. public no-trespass) and the potential for development (e.g. land conserved in perpetuity vs. potentially developable). Smith and others (2002) classify this last element of open space as being “fixed” (permanent) or “adjustable” (potentially developable). Because fixed open space is somewhat permanent, the effect on properties should be consistent across time (controlling for shifts in cultural values, socio-economic conditions and the property market). However, the price effect of adjustable open space (those areas that could potentially be developed) will be less clear because different residents will have different expectations about the future use of this land. Potential property purchasers may expect that this open space will be developed in the near future, thus diminishing the value that this open space provided to a property in the first place. Therefore, the amenity value of adjustable open space should be less than fixed open space but will vary according to buyers' knowledge and expectations, which are dependent on the local development conditions for an area. If there is visual evidence of nearby open areas currently being developed into residences then a buyer is likely more aware of the potential for conversion of open space nearby their prospective homes.

Geoghegan (2002) researched the differential effects of developable vs. permanently-conserved open space and found that permanent open space was three times more valuable than developable open space. This research used only one 1600m buffer distance from each home interpreted as a 20 minute walk from one's home. Irwin (2002b) also distinguishes between nearby, developable agricultural lands from those with conservation easements. The author includes the areal proportions of ten types of land uses within 400 m of a home and examines the marginal value of these land uses on the log transformed price of the property. She finds that conserved agricultural land and public open spaces were positive, cropland was positive but insignificant and forests were negative.

This study draws on previous research concerning the value of preserved and developable open space at multiple spatial scales. Proportions of different types of open space are measured at four spatial scales. These scales are defined by 100m, 500m and 1km distance buffers around each individual property as well as the census block group in which each property is located. The 100m distance was chosen to represent the immediate visual and auditory space that an individual perceives while on their property. The 500m distance represents the area of frequent walking distances from a residence, while the 1km buffer approximated a larger activity space that an individual frequently travels through during their daily routines. Open space is separated into private lands that are developable, private lands that are restricted from development, and a combination of publicly-owned and publicly-accessible lands such as parks, golf courses, cemeteries and school athletic fields. These open spaces are not likely to be owned by the individual properties that are examined in this research but are nonetheless expected to provide differential benefits to these properties depending on the type of open space and the scale at which they are examined. The size of the coefficients for public lands and private, conserved lands are expected to be greater than the coefficient for developable open spaces.

The foundation for the property hedonic model was presented by Rosen (1974) who showed the existence of a property market equilibrium where consumers and suppliers maximize their respective utility and profits by choosing to purchase and produce properties with distinct combinations of desirable attributes. Although this approach has been used to address a wide variety of environmentally-related issues in since Rosen's work, there are numerous statistical and econometric issues that are not fully accounted for. Rosen's (1974) development of the property hedonic model assumed that both supply and demand factors were mobile and elastic and that an entire city could be viewed as having a single housing market in equilibrium. Equilibrium occurs when the market settles on a hedonic price supply-demand curve that ensures households (within their budget constraints) cannot increase their utility by choosing a different property and sellers cannot increase their profits by increasing the property's price or changing its characteristics. With this assumption, the price of a property and the availability and contribution of its constituent characteristics are invariant across geographic space.

Since Rosen's work, most researchers have found that housing markets are typically not in equilibrium and that the assumption of a single market is unrealistic except for very small study areas (Bourassa et al. 2003; Day et al. 2004; Ekeland et al. 2002; Goodman and Thibodeau 1998, 2003; Orford 2000). With property hedonic models there are also spatial statistical concerns of spatial dependency, non-stationarity and inappropriate scales of analysis. Spatial dependency (association or lags) refers to the likelihood that the values of observations for a particular variable are more similar for observation in close spatial proximity to each other. An example of this spatial dependency in the housing market is the compositional effect of neighboring property characteristics (e.g. house age, size and value) influencing the selling price of an individual residence (Orford 2000). Spatial error autocorrelation, on the other hand, refers to the existence of spatial associations that have not been incorporated into the regression model (Paez and Scott 2004). The problem with the presence of spatial error autocorrelation in a regression model is that the statistical assumption regarding the independent distribution of errors is violated. As a consequence of these two types of spatial dependencies (lags and error), parameter estimates will be biased and inefficient, respectively (Anselin 1988). Inefficient standard errors leads to the possibility of finding a spurious significance of an effect when one does not actually exist (Type I error).

Non-stationarity refers to the existence of a heterogeneous relationship between dependent and independent variables across geographic space (Fotheringham et al. 2002). Global approaches to hedonic modeling such as using OLS do not accommodate local, spatial variations in these relationships. A coefficient that is reported as insignificant within a global regression model may be the result of highly significant positive relationships cancelling out the effect of significant negative relationships in others areas. With respect to the effect of locational amenities on house price, it is often useful to determine the existence of non-stationarity and to attempt to model the reason for its existence by allowing the coefficient to interact with other variables.

Issues with spatial dependencies and non-stationarity are further complicated by the scale at which attributes are measured or aggregated. If the values of these spatial properties change with the choice of unit used in a model, then the model exhibits scaling challenges that cannot be effectively modeled with standard regression techniques. Thus, the variance of the outcome, the relationship between the independent and dependent variable and the relationship between individual observations all may be sensitive to unit size.

This research uses a multilevel modeling framework for addressing the challenges of the property hedonic model. With this approach, individual properties are nested within neighborhoods (Brown and Uyar 2004; Gelfand et al. 2007; Goodman and

Thibodeau 1998; Orford 2000, 2002). These models allow the simultaneous examination of the effects of group-level and individual-level variables on individual-level outcomes (the property price) while accounting for the non-independence of observations within the neighborhoods. Multilevel models also account for the spatial error autocorrelation (dependence of the residuals) by differentiating between-individual errors from between-neighborhood errors (Orford 2000). If this dependency was not modeled, the standard errors of the independent variables would be biased downwards (underestimated), which results in spuriously significant effects (Snijders and Bosker 1999). Multilevel models also allow independent variables to vary across geographic space. Deviations from the global relationship between price and the variable can be mapped to determine the magnitude and location of this non-stationarity in the relationship. Each level-1 coefficient can be allowed to vary across neighborhoods either randomly, through the interaction with level-2 variables or through both of these options (Orford 2000).

4.3. Objectives

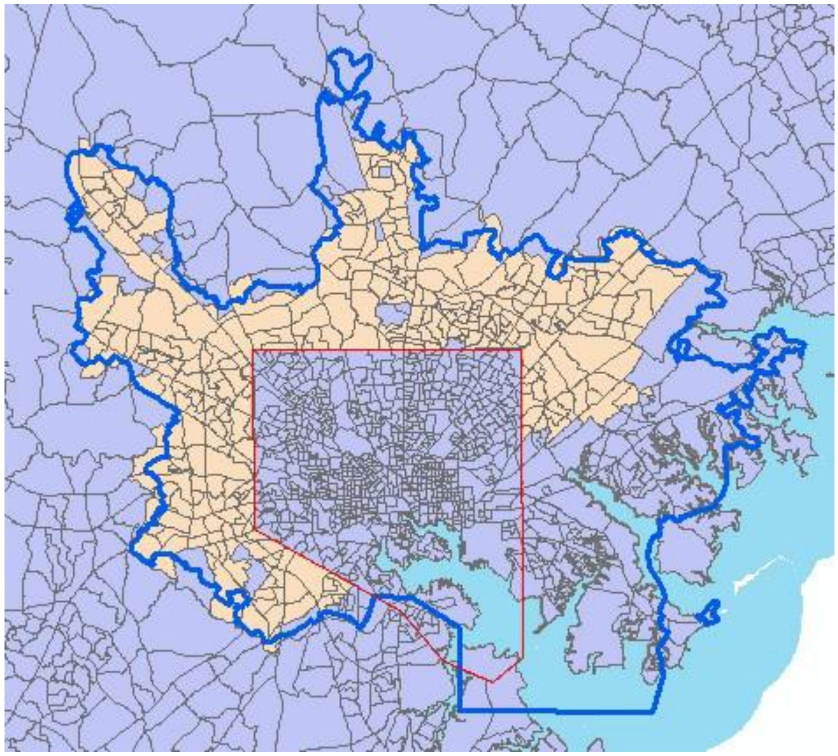
This research focuses on measuring the value of open space that contributes to the price of nearby residential properties while controlling for residential and commercial land uses as well as structural and neighborhood characteristics. Specifically, I examine the effect on property value of proximity to privately conserved, privately developable and public open spaces at four spatial scales: 100m, 500m and 1km distances from the individual residence and the block group in which the property is located.

4.4. Data

The study area for this hedonic model contains those areas of Baltimore County surrounding the city of Baltimore and extending to an urban boundary defined by the extent of the urban sewer infrastructure. Using this urban sewer boundary as the study extent helps to distinguish between the housing markets for rural properties and suburban properties and is considered a de facto urban growth boundary for the Baltimore metropolitan area. These markets differ with respect to zoning, types of properties and the consumers of these properties. Due to the inaccuracy with the actual boundary, properties were included if at least 50% of the area of their corresponding block group fell within the urban boundary. Properties located within block groups whose center was within 2km of the coast of Chesapeake Bay were also excluded. Coastal neighborhoods were excluded because of the potential for confounding the effects of the value of open land with the value of proximity to water.

Property sales and attributes were obtained from the MD Property View 2004 database, a private company which compiles sales transaction data with a property's geographic location, lot size and structural characteristics from the state of Maryland's property-appraisal database. Sales transactions for a 3-year period (1998-2000) were used in this analysis. This created a dataset with 12,196 properties distributed among 295 block groups. The properties and boundaries used in this research are shown in Figure 4.1.

Figure 4.1. Study Area in Baltimore County



Selling prices were standardized to the year 2000 with the OFHEO (Office of Federal Housing Enterprise Oversight) housing price index for the Baltimore Metropolitan Statistical Area. This index accounted for both annual and seasonal (quarterly) fluctuations of property sales. This standardization removed the need for adding dummy variables for year and season while allowing for a sufficiently large dataset of properties that would be consistent with the 2000 Census attributes used to describe the neighborhood.

Property records were selected if they followed numerous criteria: “Arm’s length” transactions only; single, detached homes or townhouses; appraised, total property value was within ±50% agreement of the selling price; zoning was classified as either residential or residential-commercial; and values of key variables used in the analyses were not missing. The total square footage for each house accounted for floor area for each story, excluding basements and attics. Basements and garages were converted to dummy variables (presence or absence) rather than using the area of the features because of the numerous omissions and errors within the database. The field denoting quality of construction, initially specified with nines codes ranging from “low cost” to “luxurious plus”, was converted to 3 dummy variables of poor, average and high quality. The number of bathrooms and half-bathrooms were recombined into one attribute (e.g. one full bath and one half bath equals “1.5”). Following the example of Cho (2006) and Troy and Grove (2008), records with low property prices (less than $50,000) were considered as either database errors or non,-arm’s length transactions and were excluded from the analysis. Approximately 0.5% of properties were excluded from further research for this reason. Additionally, a few records with house size or lot size less than 500 ft^2 were considered to be database errors and were excluded. These and other property variables as well as their means and range are listed in Table 4.1. All of these variables, except for age of structure, are expected to have a positive impact on property price.

A large number of Census attributes, obtained at the block group level from the 2000 Census, were available for use in this research. These were used as proxies for neighborhood characteristics that nearby properties shared. The percent of the neighborhood with a bachelor’s degree or higher was used as a proxy of the social status of the neighborhood. A well educated neighborhood is expected to be associated with higher property values. Population density (per hectare) was a measure of the demand for and relative scarcity of land available for development. Higher population densities are expected to drive land price (and overall property prices) higher. Percent vacancy is included as an indicator of property market conditions for the neighborhood. A higher vacancy rate is expected to have a negative impact on property prices. Mean travel time to work (in minutes) was used as a proxy for distance to the nearest employment center. As increased commuting time represents increased opportunity cost of time and travel expenses (e.g. fuel), an increase in this variable is expected to reduce the value of a property. Conversely, wealthy people tend to live further from downtown Baltimore, where they can access higher-quality amenities (e.g. better schools, public services, cleaner environment) and avoid urban blight externalities such as crime. Thus the variable for distance to downtown Baltimore is also included and is expected that greater distances will have a positive association with property prices. Median house value was expected to capture some of the spatial dependency in price that nearby properties shared.

In addition, the property-level variables of age, house size and lot size were group-mean centered and their block groups means were added to the model at the neighborhood level. This group-mean centering approach minimizes the confounding of individual-level effects of structure with neighborhood-level, compositional effects (Gelman et al. 2007; Orford 2000). This compositional effect of place is a form of spatial dependency where the individual attributes (the level-1 independent variables) of all the properties in a neighborhood have a combined influence on the value of an individual house separate from the effects of the individual attributes for that house.

Table 4.1. Property and Neighborhood Regression Variables

Variables	Description	Min	Max	Mean
Property				
Price00[ab]	Sale price converted to Yr 2000	\$50,062	\$1,754,793	\$147,730
EnclsFt [b]	Total area of interior (ft^2)	561	8,492	1,556
LandM [b]	Lot size of property (m^2)	63	19,950	740
Bathnum	Number of full and half bathrooms	1	8.5	2.0
Age [b]	Age of house at the time of sale	0	201	31
HouseDum	Detached home vs. townhouse	0	1	0.5
Basedum	Presence of basement	0	1	0.6
GarDum	Presence of garage or carport	0	1	0.3
QualDumAvg	Avg. quality of construction	0	1	0.6
QualDumHigh	High quality of construction	0	1	0.1
Block Group				
MedValHouse	Median house value	\$76,100	\$422,100	\$135,366
pBachDiploma	Percent with a college degree	0	45	20
TravelMean	Mean travel time to work	17	41	29
PopDens	Population density per hectare	0.3	103.6	20.6
DistBaltKm	Distance to downtown Baltimore	5	27	13
pVacancy	Percent vacant houses	0.4	15.2	3.8
mnAge	Mean age of residences	1	89	39
mnEnclsFt	Mean house size (ft^2) of residences	1,509	884	3,086
mnLandM	Mean lot size (m^2) of properties	190	5,170	894

a: dependent variable b:natural log transformed in regressions

The land use/land cover data for this research was derived from a number of sources. The public open space variable was a combination of parks, natural areas, golf courses, cemeteries, public school athletic fields and open areas of college campuses. Officially designated parks, natural areas and publicly-owned golf courses were obtained from county, state and federal agencies. Other parks, cemeteries, golf courses and public school athletic fields were identified and delineated from the USGS Geographic Names Information System (GNIS), parcel boundaries and aerial photos. All of these were combined into the “public” category because they all allowed some sort of public access and were maintained as manicured, landscaped or natural open spaces. Only open spaces that were greater than 2 hectares were included in this category. Approximately 20% of the total land for the study area was composed of public open space.

Forests were mapped (1997) into nine categories ranging in degrees of conservation by the Baltimore County Department of Environmental Protection and Resource Management (Appendix x). These categories were: publicly-owned forests, which were subsequently added to the public open space category; private forests with no development restrictions; and private forests that were under conservation easements, in resource conservation areas or were within 100 feet of a stream and unlikely to be developed due to zoning and environmental regulations. Other forests identified by the USGS 2001 National Land Cover Database (NLCD01) that were not already included in the forests described above were added to the private, developable open space category. Areas of agricultural open space were defined by a combination of the NLCD01 and the Maryland Department of Planning's 2000 Land use/Land Cover (MDLULC00) datasets for agricultural and pastoral land with a minimum mapping unit of 10 acres. These agricultural areas were added to the private, developable open space category. Approximately 5% and 15% of the total land for the study area was composed of agricultural and private forests, respectively.

Commercial lands were defined from the NLCD's high-density class that was located within the MDLULC02 commercial areas. Low-medium density residential lands were combined from NLCD01 low and medium-density developed areas and MDLULC00 residential areas that were not already considered as open space from the previous data processing. Residential land use was considered last because of the high amount of classification error associated with these NLCD01 categories (Irwin et al. 2006). Approximately 5% and 38% of the total land for the study area was composed of commercial and low/medium-density residential, respectively.

Figure 4.2. Land Use Classes

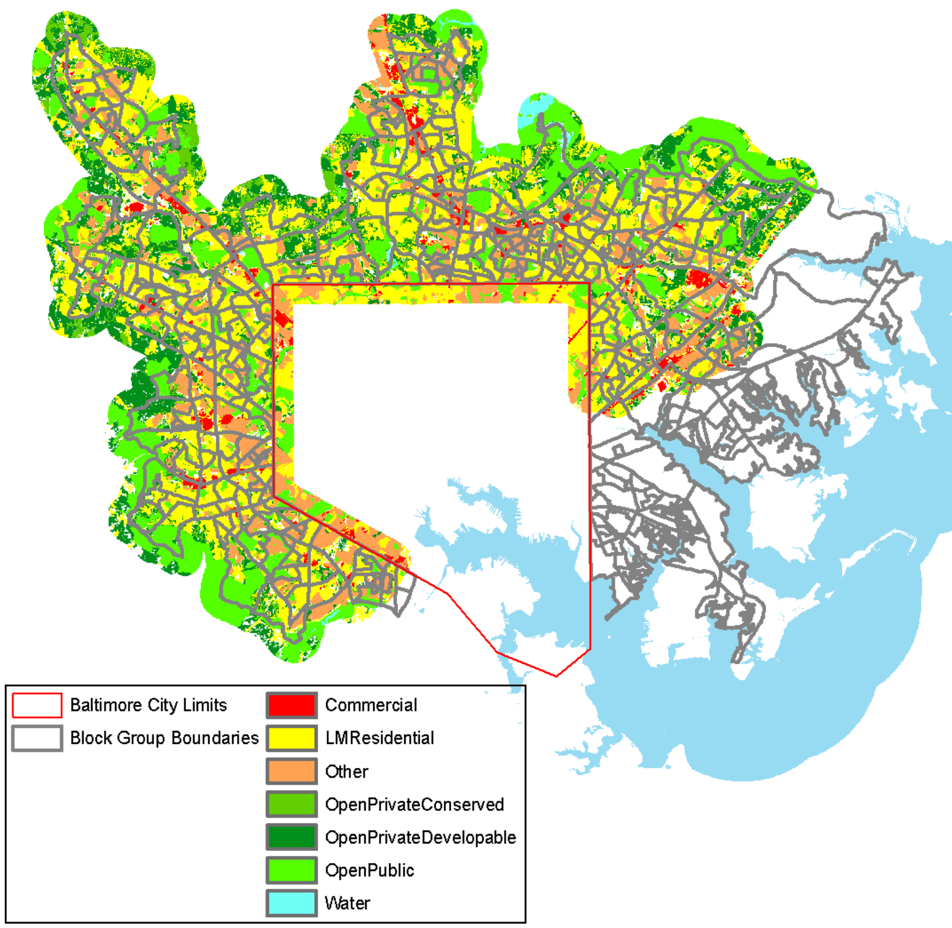


Table 4.2. Land Use Regression Variables

Variables	Description	Min	Max	Mean
100m				
pPrvCon	Private open space, conserved	0	0.54	0.01
pPrvDev	Private open space, developable	0	1	0.07
pPublic	Public open space	0	1	0.03
pLMRes	Low-med density residential	0	1	0.51
pComm	Commercial area	0	0.95	0.06
pPrvOpen	Private open spaces combined	0	1	0.06
500m				
pPrvCon	Private open space, conserved	0	0.4	0.02
pPrvDev	Private open space, developable	0	0.71	0.09
pPublic	Public open space	0	0.79	0.1
pLMRes	Low-med density residential	0	1	0.46
pComm	Commercial area	0	0.5	0.03
pPrvOpen	Private open spaces combined	0	0.76	0.1
1km				
pPrvCon	Private open space, conserved	0	0.27	0.02
pPrvDev	Private open space, developable	0	0.54	0.1
pPublic	Public open space	0	0.76	0.13
pLMRes	Low-med density residential	0	0.89	0.43
Commercial	Commercial area	0	0.26	0.04
pPrvOpen	Private open spaces combined	0	0.73	0.12
Block Group				
pPrvCon	Private open space, conserved	0	0.19	0.02
pPrvDev	Private open space, developable	0	0.43	0.06
pPublic	Public open space	0	0.63	0.11
pLMRes	Low-med density residential	0	1	0.48
Commercial	Commercial area	0	0.55	0.05
pPrvOpen	Private open spaces combined	0	0.49	0.08

4.5. Methods

In estimating a relationship between environmental amenities and property prices, the choice of functional form is not always clear. Rosen (1974) stressed that economic theory fails to indicate that any particular form is appropriate. Consequently, a variety of functional forms have been used in the hedonic literature. A linear form assumes that an individual's preferences are linear, implying that perfect repackaging of property characteristics is possible (Freeman 2003). However, in property markets, individual house characteristics are inseparable; an individual cannot mix characteristics in any other level than is already available in each house (Garrod and Willis 1992).

While the form that is chosen should ideally improve the model fit and help to satisfy important assumptions of OLS regressions, such as normally distributed residuals and homoscedasticity, this is not the main issue with choosing the proper functional form. The goal of finding a proper functional form is to overcome problems associated with the non-linearity that is often found in hedonic regression equations (Goodman and Thibodeau 1995). Substantively, this means that the proper functional form should be chosen so that marginal value for any given property attribute does not vary across the range of house prices.

A Box-Cox transformation analysis can be used to provide guidance on whether such simple forms are adequate for satisfying regression assumptions. The Box-Cox transformation of the dependent variable is shown as:

$$y(\lambda) = \begin{cases} \frac{y^{\lambda}-1}{\lambda}, & \text{if } \lambda \neq 0 \\ \log y, & \text{if } \lambda = 0 \end{cases} \qquad \text{Eq. 4.1}$$

With this test, the parameter, λ, is estimated through maximum likelihood to find the optimal transformation of the dependent variable. This parameter can then be tested for significant differences between the optimal value of λ and three cases of λ that correspond to simpler functional forms: a reciprocal transformation, where λ=-1; a log transformation, where λ=0; and a linear (untransformed) form, where λ=1.

In this research, a left-hand (LHS) Box–Cox test found that the optimal transformation for price was indicated by Lambda value of -0.018. A chi-square test did not find this to be significantly different from a lambda value of zero, indicating that a natural log transformation was suitable for the dependent variable. The regression residuals and the dependent variable were found to be normally distributed after using this transformation for property price.

A log transformation was also used on the continuous property variables of house size, lot size and age. In the hedonic literature, this is a common method to account for the (non-linear) declining value of each additional increment of house and lot size. Houses are also expected to depreciate with age at a declining rate. While the addition of a quadratic term for age often helps to account for unknown renovations or the “vintage effect” of older properties (Goodman and Thibodeau 1995), this was not included in this study.

The regression model was then checked for collinearity within an OLS regression. A general rule of thumb is that variance inflation factors (VIF) greater than 10 are thought to be highly correlated and should be cause for further assessment before proceeding (O'Brien 2007). This research found VIF’s below 4 for all property and neighborhood variables. Because the 3 buffer distances used for the land use variables were not exclusive of each other (e.g. the 500m buffer included the information from the

100m buffer), collinearity diagnostics showed a moderate amount of collinearity between some of the same type of land use at different buffer distances. For this reason, regressions and results for land use variables are performed and reported separately for each of these buffer distances. Diagnostics were also used to check for the normal distribution of residuals, the existence of homogeneity of variances in the residuals, and the potential for heteroscedasticity and/or non-linear trends in the independent variable-price relationship.

While the analysis of multilevel models can be performed with a number of statistical packages, the details of using multilevel modeling with HLM (Hierarchical Linear Modeling) software are discussed below and follow the work of Raudenbush and Bryk (2002). A traditional property hedonic function (Equation 4.2) under the assumption of a single market can be written in terms of a vector of structural characteristics (S_i), neighborhood characteristics (N_i) and environmental or land use characteristic (L_i) and includes a single error term (r_i).

$$Y_i = S_i + N_i + L_i + r_i \qquad \text{Eq. 4.2}$$

With multilevel models, the property market is assumed to be composed of sub-markets where within-place property attributes (S) at level 1 are separated from between-place, neighborhood characteristics (N) at level 2. The intercept of the level-1 equation becomes the dependent variable of the level-2 equation. The error term is also expanded so that there is unexplained variation at both levels (Equation 4.3).

$$\text{Level 1: } Y_{ij} = ß_{0j} + ß_{1j}S_{ij} + \ldots ß_{nj} S_{ij} + r_{ij}$$
$$\text{Level 2: } ß_{0j} = \gamma_{00} + \gamma_{01}N_j + .. \gamma_{nj}N_{nj} + U_{0j} \qquad \text{Eq. 4.3}$$

The level-1 structural attributes such as house size, lot size and age may be group-mean centered and the group average included at level 2 to account for the compositional effects of place as described previously. In addition, level-1 attributes can be allowed to vary across submarkets (Equation 4.4). Since the model contains more than one error term, it cannot be estimated using OLS regression. Instead, an iterative maximum likelihood procedure is used (Raudenbush and Bryk 2002).

$$\text{Level 1: } Y_{ij} = ß_{0j} + ß_{1j}(S_{ij} - S_j) + \ldots ß_{nj} S_{ij} + r_{ij}$$
$$\text{Level 2: } ß_{0j} = \gamma_{00} + \gamma_{01}N_j + .. \gamma_{0n}N_j + \gamma_{0n}S_j + U_{0j} \qquad \text{Eq. 4.4}$$
$$ß_{1j} = \gamma_{10}$$

With multilevel property hedonics, spatial autocorrelation can be treated as the norm since individual houses in the same sub-market are likely to be more similar, in some way, than houses drawn from the entire housing market at random. Hence, autocorrelation is to be expected in hierarchical data, and the multilevel approach exploits this dependence to derive improved estimates, while the standard errors of the estimates are adjusted to take into account the autocorrelation (Goldstein 2003; Orford 2000). A test for spatial autocorrelation of the residuals using the global Moran's "I" statistic (Moran 1948) shows that there is a significant but slight amount of positive correlation remaining in the multilevel model's level-1 residuals (Moran's = 0.15) after including all regression variables except for land use. This is half the spatial autocorrelation that exists for the same model under OLS regression (Moran's I=0.3).

Property hedonic models often suffer from the effects of heteroscedasticity, or non-constant variance in the residuals. The presence of heteroscedasticity can create a downward bias to standard errors, which may create a spurious significance of a coefficient (Type I error). Multilevel models help mitigate this problem by separating the error term into two levels (Orford 2000). In addition, attributes that contribute most to the problem of heteroscedasticity in property hedonic models, such as age of dwelling (Goodman and Thibodeau 1998), can be allowed to randomly vary across level-2 units. This distributes some of the unexplained variation expressed by the level-1 error term to variation specific to a random effect variable and helps to reduce the problem of heteroscedasticity that may be found in the residuals (Raudenbush and Bryk 2002). The presence of heteroscedasticity can be tested in HLM and if significant heteroscedasticity remains, then it is recommended that robust, "Huber-corrected" standard errors be used instead (Poor et al. 2007; White 1980).

In order to remove the problem of heteroscedasticity in this research, the structural variables of age, bathroom, house size and lot size were allowed to vary across block groups. However, the results from the HLM test for heteroscedasticity rejects the null hypothesis of homogeneity of level-1 variances. This indicates that heteroscedasticity still exists in the model and as such, only robust standard errors are reported in this research.

4.6. Results

The first model examines structural and neighborhood variables without including variables related to the main effects of land use. The property structural characteristics are all highly significant and show the expected signs (Table 4.3). The log transformation of floor space, lot size and age means that these coefficient are interpreted as the elasticity of sales price, controlling for other coefficients, so that property price increases by approximately 0.4% and 0.1% and decreases by 0.1% for every 1% increase in floor space, lot size and age, respectively.

There is a semi-log relationship between price and most variables, so coefficients can be interpreted as the percent change in property price with a 1-unit change of that particular variable, all else constant. There is an approximate 4% increase in sale price with the addition of a half bath. For the dummy variables, there is: an approximate 14% increase in sale price if the building is a single, detached dwelling (vs. a townhouse); a 6% increase in sale price with the presence of a basement; and a 6% increase in sale price with the presence of a garage. There are approximate 8% and 26% increases in the sale price if the property was built with average or high quality materials, respectively, rather than poor quality materials.

Concerning neighborhood characteristics, there is: a 2% increase for every $10,000 increase in median house value; a 0.5% increase in sale price with a 1-unit increase in a neighborhood's percent of population with a Bachelor's degree or higher; and a 0.7% decrease in sale price with a 1-unit increase in a neighborhood's percent vacant houses; and an insignificant decrease in sale price with increasing population density.

Concerning the location characteristics, there is a 0.3% decrease in sale price with every 1-minute increase in a neighborhood's mean travel time to work while there is a 0.4% increase in sale price with every 1km increase in distance to downtown Baltimore.

The block group's mean age and house size of the properties indicates that there is a compositional effect of neighboring houses on the price of a single home. A neighborhood with large houses will increase the price of a single home while an older neighborhood reduces the price of a single home. The block group variable for lot size was not found to be significant.

Table 4.3. Regression Results

Variables	Coef	SE	Sig
Property Level			
lnEnclsFt	0.401	0.011	0.000
lnLandM	0.081	0.005	0.000
BathNum	0.041	0.003	0.000
lnAge	-0.078	0.004	0.000
HouseDum	0.148	0.013	0.000
BaseDum	0.063	0.003	0.000
GarDum	0.056	0.005	0.000
QualDumAvg	0.083	0.010	0.000
QualDumHigh	0.258	0.022	0.000
Block Group			
(Constant)	10.932	0.057	0.000
MedValHouse	2.000E-06	0.000E+00	0.000
pBachDegree	0.005	0.001	0.000
pVacancy	-0.007	0.002	0.002
TravelMean	-0.003	0.001	0.019
PopDens	-6.170E-04	3.840E-04	0.109
DistBaltKm	0.004	0.001	0.004
mnAge	-0.002	3.680E-04	0.000
mnEnclsFt	2.840E-04	3.300E-05	0.000
mnLandM	-1.200E-05	1.100E-05	0.277

The next set of models examines the main effects of land use from each of the four spatial scales of neighborhood that were considered: 100m; 500m; 1km; and the block group. All the variables in the previous model, except for the non-significant population density and mean lot size, were included in all subsequent models. Table 4.4 presents the results of the models for each of these four spatial scales.

Concerning the results of both privately-conserved and developable, open space, the estimated coefficients are positive and significant across all spatial scales except for the block group. The coefficient for privately-conserved space is substantially greater than developable open space across all four spatial scales. The differential benefit between these two types increases from being almost twice as beneficial at the 100m buffer, to 4 times as beneficial at the 500m buffer, to 7 times as beneficial at the 1km buffer. Public open space does not have a significant effect on property value in this research.

Table 4.4. Regression Results for Land Use Variables

Land Variables	Coef	SE(robust)	Sig	Land Variables	Coef	SE(robust)	Sig
100m Buffer				**1km Buffer**			
PrvCon	0.122	0.044	0.006	PrvCon	0.960	0.299	0.002
PrvDev	0.065	0.016	0.000	PrvDev	0.134	0.076	0.080
PubCon	0.030	0.022	0.173	PubCon	0.046	0.058	0.426
Residential	0.023	0.009	0.006	Residential	0.047	0.042	0.265
Commercial	-0.074	0.033	0.026	Commercial	0.091	0.086	0.288
PrvOpen	0.074	0.015	0.000	PrvOpen	0.259	0.031	0.000
500m Buffer				**Block Group**			
PrvCon	0.570	0.241	0.018	PrvCon	0.497	0.304	0.103
PrvDev	0.136	0.051	0.008	PrvDev	0.018	0.076	0.814
PubCon	0.028	0.036	0.393	PubCon	-0.005	0.037	0.897
Residential	0.035	0.020	0.081	Residential	-0.030	0.023	0.184
Commercial	-0.070	0.039	0.074	Commercial	0.041	0.064	0.520
PrvOpen	0.203	0.039	0.000	PrvOpen	0.102	0.053	0.058

4.7. Discussion

These results suggest that there is a preference for conserved areas over developable open spaces, which indicates that individuals recognize the potential for these areas to be developed in the near future.

The insignificant effect of public open space seems contrary to expected theory. It is reasonable to expect that the negative externalities of public open space (e.g. increased traffic and crowd noises) may compete with the benefits (e.g. recreation and aesthetics) when these areas are in close proximity to a residence, resulting in an insignificant or even negative effect. However at the block group or 1km level, it would be expected that the convenience of having nearby recreation opportunities would provide a positive and significant benefit to homes. This unexpected result may be due to the inclusion of cemeteries, golf courses and athletic fields into the category of open space.

The effect of commercial and residential land across these scales is consistent with the expectations. In the immediate area (100m or 500m from a home), residents dislike the negative externalities associated with commercial areas. However, there is likely a distance-decay effect for these negative externalities so that at 1km, the positive aspects of commercial areas (e.g. the convenience of shopping) begin to outweigh the negative aspects. Conversely, the benefits of having more residential areas become insignificant at 1km and negative but insignificant at the block group. This suggests that the inconvenience of increased travel time to work and shopping are competing with the benefits of being surrounded by other homes.

These results show that the effect of different types of land use surrounding an individual residence, changes with the spatial scale at which these land uses are measured. This may be due to the different scales at which individual perceive various positive and negative externalities. Individuals are likely to value being surrounded by other residences or open space in the immediate visual and auditory distances from their homes. At larger scales, there appears to be a greater preference for increasing the opportunities associated with an individual's daily activities.

4.8. Conclusion

Both the type of open space and the scale at which the amount of open space is measured can influence the outcome of a property hedonic model. Concerning the type of open space, both conserved and potentially developable private lands appear to hold more value to residents than public open space. Public open space was not found to provide a significant benefit for two possible reasons. First, this category included a mix of open space types from actual public parks to golf courses, cemeteries and school athletic fields. Some of these types, particularly cemeteries, may be negatively valued to a degree that overwhelms the benefits from other types of open space, particularly parks. Second, the negative externalities potentially associated with these features, such as increased crowd noise and traffic may overwhelm the aesthetic and recreation benefits of these places. However, the much larger coefficients of private open space compared to public space suggest that it may be the absence of negative externalities associated with developed areas rather than specific open space amenities, such as recreation, that are most valuable to local residents. As local and county governments develop policy initiatives to preserve open space in the suburban area around Baltimore, this research suggests that purchasing the development rights, through conservation easements might be preferable to outright public acquisition of undeveloped land.

Concerning the difference between private conserved areas and potentially developable areas, it appears that residents do perceive the potential for increased negative externalities from the conversion of open space to developed areas. Given this recognition of potential development it would seem that the price differential between these types would be greatest for the immediate areas around a house where the negative externalities of development would be greatest. However, it may be that residents also recognize the existence of zoning laws that likely prevent disparate development from occupying nearby areas. Given the greater likelihood of this open space in the immediate area of a residence being developed into other residences, rather than commercial or industrial uses would mean that the differential between conserved and developable open space might be less in the immediate area of a residence.

While the property hedonic model can be used to estimate the value of some ecosystem goods and services associated with open space it is important to remember that this method provides only a limited measure of total economic benefits. Public open spaces provide recreation to visitors and other residents as well as to the local property owners. The potential, property hedonic benefits to condominiums, apartments, and businesses was not observed. Open spaces, both public and private, also provide numerous non-use benefits to the entire population of the area such as protecting water quality through storage, filtration and mitigation of runoff and providing habitat for local wildlife. For these reasons, estimates from hedonic house price models will generally under-represent the total non-market value of these open spaces.

4.9. References

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

Anselin, L. 1988. Spatial econometrics : methods and models. Kluwer Academic Publishers London.

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. Ecol. Econ. **40**: 337-349.

Bolitzer, B., and Netusil, N.R. 2000. The impact of open spaces on property values in Portland, Oregon. J. Environ. Manage. **59**: 185-193.

Bourassa, S.C., Hoesli, M., and Peng, V.S. 2003. Do housing submarkets really matter? Journal of Housing Economics **12**: 12–28.

Brown, K.H., and Uyar, B. 2004. A hierarchical linear model approach for assessing the effects of house and neighborhood characteristics on housing prices. Journal of Real Estate Practice and Education **7**(1): 15-23.

Cho, S.H., Bowker, J.M., and Park, W.M. 2006. Measuring the contribution of water and green space amenities to housing values; An application and comparison of spatially weighted hedonic models. J. Agric. Resour. Econ. **31**(3): 485-507.

Day, B., Bateman, I., and Lake, I. 2004. Nonlinearity in hedonic price equations: An estimation strategy using model-based clustering. University of East Anglia: Centre for Social and Economic Research on the Global Environment (CSERGE).

Do, A.Q., and Grudnitski, G. 1995. Golf courses and residential house prices: an empirical examination. Journal of Real Estate Finance and Economics **10**: 261–270.

Ekeland, I., Heckman, J.J., and Nesheim, L. 2002. Identifying hedonic models. American Economic Review **92**(2): 304-309.

Fotheringham, A.S., Brunsdon, C., and Charlton, M.E. 2002. Geographically Weighted Regression: The analysis of spatially varying relationships. Wiley, Chichester.

Freeman, A.M. 2003. The measurement of environmental and resource values: Theory and methods. Resources for the Futures, Washington, D.C.

Garrod, G.D., and Willis, K.G. 1992. Valuing goods' characteristics: an application of the hedonic price method to environmental attributes. J. Environ. Manage. **34**(1): 59-76.

Gelfand, A.E., Banerjeeb, S., Sirmans, C.F., Tu, Y., and Ong, S.E. 2007. Multilevel modeling using spatial processes: Application to the Singapore housing market. Comput. Stat. Data Anal. **51**: 3567 – 3579.

Gelman, A., Bafumi, J., Shor, B., and Park, D.K. 2007. Rich state, poor state, red state, blue state. What's the matter with Connecticut? Quarterly Journal of Political Science **2**: 345-367.

Geoghegan, J. 2002. The value of open spaces in residential land use. Land Use Pol. **19**: 91-98.

Geoghegan, J., Wainger, L.A., and Bockstael, N.E. 1997. Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS. Ecol. Econ. **23**(3): 251-264.

Goldstein, H. 2003. Multilevel statistical models. Oxford University Press, NY, NY.

Goodman, A.C., and Thibodeau, T.G. 1995. Age-related heteroskedasticity in hedonic house price equations. Journal of Housing Research **6**(1): 25-42.

Goodman, A.C., and Thibodeau, T.G. 1998. Housing market segmentation. Journal of Housing Economics **7**: 121–143.

Goodman, A.C., and Thibodeau, T.G. 2003. Housing market segmentation and hedonic prediction accuracy. Journal of Housing Economics **12**: 181–201.

Irwin, E.G. 2002. The effects of open space on residential property values. Land Econ. **78**(4): 465–480.

Irwin, E.G., Bockstael, N.E., and Cho, H.J. 2006. Measuring and modeling urban sprawl: Data, scale and spatial dependencies. *In* Regional Science Association Meetings, Toronto, Canada. pp. 1-35.

Kestens, Y., Theriault, M., and Des Rosiers, F. 2004. The impact of surrounding vegetation and land use on single-family house prices. Environment and Planning B: Planning and Design **31**: 539-567.

Lee, C., and Linneman, P. 1998. Dynamics of the greenbelt amenity effect on the land market: The case of Seoul's greenbelt. Real Estate Economics **26**(1): 107–129.

Lutzenhiser, M., and Netusil, N.R. 2001. The effect of open spaces on a home's sale price. Contemporary Economic Policy **19**(3): 291–298.

Mahan, B.L., Polasky, S., and Adams, R.M. 2001. Valuing urban wetlands: A property price approach. Land Econ. **76**(1): 100–113.

Moran, P.A.P. 1948. The interpretation of statistical maps. Journal of the Royal Statistical Society: Series B **10**: 243-251.

Morancho, A.B. 2003. A hedonic valuation of urban green areas. Landscape and Urban Planning **66**: 35-41.

O'Brien, R.M. 2007. A caution regarding rules of thumb for variance inflation factors. Quality & Quantity **41**: 673–690.

Orford, S. 2000. Modelling spatial structures in local housing market dynamics: A multilevel perspective. Urban Studies **37**(9): 1643-1671.

Orford, S. 2002. Valuing locational externalities: A GIS and multilevel modelling approach. Environment and Planning B: Planning and Design **29**(1): 105-127.

Paez, A., and Scott, D.M. 2004. Spatial statistics for urban analysis: A review of techniques with examples. GeoJournal **61**: 53–67.

Poor, P.J., Pessagnob, K.L., and Paul, R.W. 2007. Exploring the hedonic value of ambient water quality: A local watershed-based study. Ecol. Econ. **60**: 797-806.

Raudenbush, S.W., and Bryk, A.S. 2002. Hierarchical Linear Models: Applications and Data Analysis Methods. Sage Publications, London.

Rosen, S. 1974. Hedonic prices and implicit markets: product differentiation in pure competition. The Journal of Political Economy **82**(1): 34–55.

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

Snijders, T.A.B., and Bosker, R. 1999. Multilevel analysis. An introduction to basic and advanced multilevel modeling. Sage, Thousand Oaks, CA.

Thorsnes, P. 2002. The value of a suburban forest preserve: Estimates from sales of vacant residential building lots. Land Econ. **78**(3): 426–441.

Troy, A., and Grove, J.M. 2008. Property values, parks, and crime: A hedonic analysis in Baltimore, MD. Landscape and Urban Planning **87**: 233–245.

Tyrväinen, L., and Miettinen, A. 2000. Property prices and urban forest amenities. J.Environ.Econ.Manage. **39**(2): 205–223.

White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica **48**(4): 817-838.

CHAPTER 5: COMPREHENSIVE BIBLIOGRAPHY

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

Allen, T.F.H., and Hoekstra, T.W. 1992. Toward a unified ecology. University of Columbia Press, NY, NY.

Allen, T.F.H., and Starr, T.B. 1982. Hierarchy: Perspectives for ecological complexity. The University of Chicago Press, Chicago.

Anderson, S.T., and West, S.E. 2006. Open space, residential property values, and spatial context. Reg. Sci. Urban Econ. **36**: 773–789.

Anselin, L. 1988. Spatial econometrics : methods and models. Kluwer Academic Publishers London.

Anselin, L. 2003. Spatial externalities, spatial multipliers and spatial econometrics. Int. Reg. Sci. Rev. **26**(2): 153–166.

Anthon, S., Thorsen, B.J., and Helles, F. 2005. Urban-fringe afforestation projects and taxable hedonic values. Urban Forestry and Urban Greening **3**: 79–91.

Assessment, M.E. 2003. Ecosystems and human well-being: A framework for assessment. Report of the Conceptual Framework Working Group of the Millennium Ecosystem Assessment.

Bastian, C.T., McLeod, D.M., Germino, M.J., Reiners, W.A., and Blasko, B.J. 2002. Environmental amenities and agricultural land values: a hedonic model using geographic information systems data. Ecol. Econ. **40**: 337-349.

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. Ecol. Econ. **40**: 337-349.

Beland, F., Birch, S., and Stoddart, G. 2002. Unemployment and health: Contextual-level influences on the production of health in populations Social Science and Medicine **55**(11): 2033-2052.

Benson, E.D., Hansen, J.L., Schwartz, A.L., and Smersh, G.T. 1998. Pricing residential amenities: The value of a view. Journal of Real Estate Finance and Economics **16**(1): 55-73.

Bhat, C.R. 2000. A multi-level cross-classified model for discrete response variables. Transp. Res. Pt. B-Methodol. **34**: 567-582.

Bin, S., and Polasky, S. 2004. Effects of flood hazards on property values: Evidence before and after Hurricane Floyd. Land Econ. **80**: 490-500.

Bolitzer, B., and Netusil, N.R. 2000. The impact of open spaces on property values in Portland, Oregon. J. Environ. Manage. **59**: 185-193.

Borjas, G.J. 1998. To Ghetto or Not to Ghetto: Ethnicity and Residential Segregation. Journal of Urban Economics **44**: 228-253.

Bourassa, S.C., Hoesli, M., and Peng, V.S. 2003. Do housing submarkets really matter? Journal of Housing Economics **12**: 12–28.

Bourassa, S.C., Hoesli, M., and Sun, J. 2004. What's in a view? Environ. Plan. A **36**(8): 1427-1450.

Box, G.E.P., and Cox, D.R. 1964. An analysis of transformations. Journalof the Royal Statistical Society, Series B (Methodological) **26**: 211–252.

Brasington, D.M. 1999. Which measures of school quality does the housing market value? The Journal of Real Estate Research **18**(3): 395-413.

Brown, K.H., and Uyar, B. 2004. A hierarchical linear model approach for assessing the effects of house and neighborhood characteristics on housing prices. Journal of Real Estate Practice and Education **7**(1): 15-23.

Browning, C.R., Feinberg, S.L., and Dietz, R.D. 2004. The paradox of social organization: Networks, collective efficacy, and violent crime in urban neighborhoods. Social Forces **83**(2): 503-534.

Brunsdon, C., Fotheringham, A.S., and Charlton, M.E. 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. Geographical Analysis **28**: 281–298.

Carson, R.T., Mitchell, R.C., Hanemann, W.M., and Kopp, R.J. 1999. A Contingent Valuation study of lost passive-use values resulting from the Exxon Valdez oil spill. Report to the Attorney General of the State of Alaska.

Cho, S.H., Bowker, J.M., and Park, W.M. 2006. Measuring the contribution of water and green space amenities to housing values; An application and comparison of spatially weighted hedonic models. J. Agric. Resour. Econ. **31**(3): 485-507.

Clapp, J.M., A., N., and Ross, S.L. 2007. Which school attributes matter? The influence of school district performance and demographic composition on property values. Journal of Urban Economics **63**: 451-466.

Clapp, J.M., and Giaccotto, C. 1998. Residential hedonic models: A rational expectations approach to age effects. Journal of Urban Economics **44**(3): 415-437.

Cliff, A., and Ord, J.K. 1981. Spatial processes: Models and applications. Pion, London.

Conway, D., Li, C.Q., Wolch, J., Kahle, C., and Jerrett, M. 2008. A spatial autocorrelation approach for examining the effects of urban greenspace on residential property values. *In* Journal of Real Estate Finance and Economics.

Costanza, R., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R.V., Paruelo, J., Raskin, R.G., Sutton, P., Van den Belt, M., D'Arge, R., De Groot, R., and Farber, S. 1997. The value of the world's ecosystem services and natural capital. Nature **387**(6630): 253-260.

Coulson, N.E. 1991. Really useful tests of the mono-centric model. Land Econ. **67**(3): 299-307.

Cropper, M.L., Deck, L.B., and McConnell, K.E. 1988. On the choice of functional form for hedonic price functions. The Review of Economics and Statistics **70**: 668–675.

Day, B., Bateman, I., and Lake, I. 2004. Nonlinearity in hedonic price equations: An estimation strategy using model-based clustering. University of East Anglia: Centre for Social and Economic Research on the Global Environment (CSERGE).

de Groot, R.S., Wilson, M.A., and Boumans, R.M.J. 2002. A typology for the classification, description and valuation of ecosystem functions, goods and services. Ecol. Econ. **41**(3): 393-408.

Deaton, B.J., and Hoehn, J.P. 2004. Hedonic analysis of hazardous waste sites in the presence of other urban disamenities. Environ. Sci. Policy **7**: 499–508.

Dehring, C., and Dunse, N. 2006. Housing density and the effect of proximity to public open space in Aberdeen, Scotland. Real Estate Economics **34**(4): 553 - 566.

Des Rosiers, F., Thériault, M., and Villeneuve, P. 2000. Sorting out access and neighborhood factors in hedonic price modeling. Journal of Property Investment and Finance **18**(3): 291-315.

Diez-Roux, A.V. 2002. A glossary for multilevel analysis. Journal of Epidemiology and Community Health **56**: 588–594.

Diez-Roux, A.V., Link, B.G., and Northridge, M.E. 2000. A multilevel analysis of income inequality and cardiovascular disease risk factors. Social Science & Medicine **50**: 673-687.

Do, A.Q., and Grudnitski, G. 1995. Golf courses and residential house prices: an empirical examination. Journal of Real Estate Finance and Economics **10**: 261–270.

Downes, T.A., and Zabel, J.E. 2002. The impact of school characteristics on house prices: Chicago 1987–1991. Journal of Urban Economics **52**: 1-25.

Dubin, R.A. 1998. Spatial autocorrelation: A primer. Journal of Housing Economics **7**: 304–327.

Ekeland, I., Heckman, J.J., and Nesheim, L. 2002. Identifying hedonic models. American Economic Review **92**(2): 304-309.

Enders, C.K., and Tofighi, D. 2007. Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. Psychol. Methods **12**(2): 121-138.

Espey, M., and Owusu-Edusei, K. 2001. Neighborhood parks and residential property values in Greenville, South Carolina. Journal of Agricultural and Applied Economics **33**(3): 487-492.

Fone, D.L., and Dunstan, F. 2006. Mental health, places and people: A multilevel analysis of economic inactivity and social deprivation. Health & Place **12**: 332-344.

Fotheringham, A.S., Brunsdon, C., and Charlton, M.E. 2002. Geographically Weighted Regression: The analysis of spatially varying relationships. Wiley, Chichester.

Freeman, A.M. 2003. The measurement of environmental and resource values: Theory and methods. Resources for the Futures, Washington, D.C.

Garrod, G.D., and Willis, K.G. 1992. Valuing goods' characteristics: an application of the hedonic price method to environmental attributes. J. Environ. Manage. **34**(1): 59-76.

Gayer, T., Hamilton, J.T., and Viscusi, W.K. 2000. Private values of risk tradeoffs at Superfund sites: Housing market evidence on learning about risk. The Review of Economics and Statistics **82**(3): 439-451.

Gelfand, A.E., Banerjeeb, S., Sirmans, C.F., Tu, Y., and Ong, S.E. 2007. Multilevel modeling using spatial processes: Application to the Singapore housing market. Comput. Stat. Data Anal. **51**: 3567 – 3579.

Gelman, A., Bafumi, J., Shor, B., and Park, D.K. 2007. Rich state, poor state, red state, blue state. What's the matter with Connecticut? Quarterly Journal of Political Science **2**: 345-367.

Geoghegan, J. 2002. The value of open spaces in residential land use. Land Use Pol. **19**: 91-98.

Geoghegan, J., Pritchard, L., Ogneva-Himmelberger, Y., Chowdhury, R.R., Sanderson, S., and Turner, B.L. 1998. Socializing the pixel and pixelizing the social in land-use and land-cover-change. *In* People and pixels. *Edited by* D. Liverman, Moran, E., Rindfuss, R. and Stern, P. National Academy Press, Washington, DC.

Geoghegan, J., Wainger, L.A., and Bockstael, N.E. 1997. Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS. Ecol. Econ. **23**(3): 251-264.

Gibbons, S. 2004. The costs of urban property crime. The Economic Journal **114**: 441-463.

Goldstein, H. 2003. Multilevel statistical models. Oxford University Press, NY, NY.

Goldstein, H., Rasbash, J., Yang, M., Woodhouse, G., Pan, H., Nuttall, D., Thomas, S. 1993. A multilevel analysis of school examination results. Oxford Review of Education **19**: 425–433.

Goodman, A.C., and Thibodeau, T.G. 1995. Age-related heteroskedasticity in hedonic house price equations. Journal of Housing Research **6**(1): 25-42.

Goodman, A.C., and Thibodeau, T.G. 1998. Housing market segmentation. Journal of Housing Economics **7**: 121–143.

Goodman, A.C., and Thibodeau, T.G. 2003. Housing market segmentation and hedonic prediction accuracy. Journal of Housing Economics **12**: 181–201.

Halstead, J.M., Bouvier, R.A., and Hansen, B.E. 1997. On the issue of functional form choice in hedonic price functions. Environ. Manage. **21**(5): 759–765.

Halvorsen, R., and Pollakowski, H.O. 1981. Choice of functional form for hedonic price equations. Journal of Urban Economics **10**(1): 37-49.

Hein, L., van Koppen, K., de Groot, R.S., and van Ierland, E.C. 2006. Spatial scales, stakeholders and the valuation of ecosystem services. Ecol. Econ. **57**(2): 209-228.

Henderson, K.K., and Song, Y. 2008. Can nearby open spaces substitute for the size of a property owner's private yard? International Journal of Housing Markets and Analysis **1**(2): 147 - 165.

Hite, D., Chern, W., Hitzhusen, F., and Randall, A. 2001. Property-value impacts of an environmental disamenity: The case of landfills. Journal of Real Estate Finance and Economics **22**(2): 185-202.

Hoshino, S. 2001. Multilevel modeling on farmland distribution in Japan. Land Use Pol. **18**(1): 75-90.

Hox, J.J. 2002. Multilevel analysis: techniques and applications. Lawrence Erlbaum Associates, Mahwah, NJ.

Irwin, E.G. 2002a. The effects of open space on residential property values. Land Econ. **78**(4): 465-480.

Irwin, E.G. 2002b. The effects of open space on residential property values. Land Econ. **78**(4): 465–480.

Irwin, E.G., Bockstael, N.E., and Cho, H.J. 2006. Measuring and modeling urban sprawl: Data, scale and spatial dependencies. *In* Regional Science Association Meetings, Toronto, Canada. pp. 1-35.

Irwin, E.G., Bockstael, N.E. 2001. The problem of identifying land use spillovers: Measuring the effects of open space on residential property values. Am. J. Agr. Econ. **83**(3): 698–704.

Jones, K., and Bullen, N. 1993. A multi-level analysis of the variations in domestic property prices. Urban Studies **30**(8): 1409-1427.

Jones, K., and Duncan, C. 1996. People and places: the multilevel model as a general framework for the quantitative analysis of geographical data. *In* Spatial Analysis: Modelling in a GIS Environment *Edited by* P. Longley, Batty, M. John Wiley & Sons, New York. p. 389.

Kestens, Y., Theriault, M., and Des Rosiers, F. 2004. The impact of surrounding vegetation and land use on single-family house prices. Environment and Planning B: Planning and Design **31**: 539-567.

Kim, C.W., Phipps, T.T., and Anselin, L. 2003. Measuring the benefits of air quality improvement: A spatial hedonic approach. J.Environ.Econ.Manage. **45**(1): 24-39.

Klassen, A.C., Kulldorff, M., and Curriero, F. 2005. Geographical clustering of prostate cancer grade and stage at diagnosis, before and after adjustment for risk factors. International Journal of Health Geographics **4**(1): 1-16.

Knight, J.R., and Sirmans, C.F. 1996. Depreciation, Maintenance, and Housing Prices. Journal of Housing Economics **5**: 369–389.

Kong, F., Yin, H., and N., N. 2007. Using GIS and landscape metrics in the hedonic price modeling of the amenity value of urban green space: A case study in Jinan City, China. Landscape and Urban Planning **79**: 240–252.

Kreft, I., and de Leeuw, J. 1998. Introducing multilevel modeling. Sage, Newbury Park, CA.

Kreft, I.G.G., Deleeuw, J., and Aiken, L.S. 1995. The effect of different forms of centering in hierarchical linear-models. Multivariate Behav. Res. **30**(1): 1-21.

Langford, I.H., Bentham, G., and McDonald, A. 1998. Multilevel modelling od geographically aggregated health data: A case study on malignant melanoma mortality and UV exposure in the European community. Statistics in Medicine **17**: 41-57.

Lansford, N.H., and Jones, L.L. 1995. Recreational and aesthetic value of water using hedonic price analysis. J. Agric. Resour. Econ. **20**(2): 341-355.

Lauridsen, J., and Kosfeld, R. 2006. A test strategy for spurious spatial regression, spatial nonstationarity, and spatial cointegration. Papers in Regional Science **85**(3): 363-377.

Lee, C., and Linneman, P. 1998. Dynamics of the greenbelt amenity effect on the land market: The case of Seoul's greenbelt. Real Estate Economics **26**(1): 107–129.

Leggett, C.G., and Bockstael, N.E. 2000. Evidence of the effects of water quality on residential land prices. J.Environ.Econ.Manage. **39**(2): 121-144.

Lewis, L.Y., Bohlen, C., and Wilson, S. 2008. Dams, Dam removal, and river restoration: A hedonic property value analysis. Contemporary Economic Policy **26**(2): 175–186.

Lutzenhiser, M., and Netusil, N.R. 2001. The effect of open spaces on a home's sale price. Contemporary Economic Policy **19**(3): 291–298.

Lynch, A.K., and Rasmussen, D.W. 2001. Measuring the impact of crime on house prices. Appl. Econ. **33**: 1981-1989.

Maas, C.J.M., and Hox, J.J. 2004. The influence of violations of assumptions on multilevel parameter estimates and their standard errors. Comput. Stat. Data Anal. **46**: 427 – 440.

Mahan, B.L., Polasky, S., and Adams, R.M. 2001. Valuing urban wetlands: A property price approach. Land Econ. **76**(1): 100–113.

Mauny, F., Viel, J.F., Handschumacher, P., and Sellin, B. 2004. Multilevel modelling and malaria: a new method for an old disease. International Journal of Epidemiology **33**: 1337-1344.

McCluskey, J.J., Rausser, G.C. . 2003. Hazardous waste sites and housing appreciation rates Hazardous waste sites and housing appreciation rates **45**(2): 166-176.

McDonald, J.F., and McMillen, D.P. 1990. Employment subcenters and land values in a polycentric urban area: The case of Chicago. Environ. Plan. A **22**(12): 1561-1574.

McLeod, H., Langford, I.H., Jones, A.P., Stedman, J.R., Day, R.J., Lorenzoni, I., and Bateman, I.J. 2000. The relationship between socio-economic indicators and air pollution in England and Wales: Implications for environmental justice. Regional Environmental Change **1**(2): 78-85.

McMillen, D.P., and McDonald, J.F. 1998. Suburban subcenters and employment density in metropolitan Chicago. Journal of Urban Economics **43**: 157-180.

Mitchell, R.C., and Carson, R.T. 1989. Using surveys to value public goods: The contingent valuation method. Resources for the Future, Washington, D.C.

Moon, G., Barnett, R. 2003. Spatial scale and the geography of tobacco smoking in New Zealand: A multilevel perspective New Zealand Geographer **59**(2): 6-15.

Mooney, S., and Eisgruber, L.M. 2001. The Influence of riparian protection measures on residential property values: The case of the Oregon plan for salmon and watersheds. Journal of Real Estate Finance and Economics **22**(2/3): 273-286.

Moran, P.A.P. 1948. The interpretation of statistical maps. Journal of the Royal Statistical Society: Series B **10**: 243-251.

Morancho, A.B. 2003. A hedonic valuation of urban green areas. Landscape and Urban Planning **66**: 35-41.

Mueller, J.M., and Loomis, J.B. 2008. Spatial dependence in hedonic property models: Do different corrections for spatial dependence result in economically significant differences in estimated implicit prices? J. Agric. Resour. Econ. **33**(2): 212-231.

Mujahid, M.S., Diez-Roux, A.V., Morenoff, J.D., and Raghunathan, T. 2007. Assessing the Measurement Properties of Neighborhood Scales: From Psychometrics to Ecometrics. American Journal of Epidemiology **165**: 858 - 867.

Muller, N.Z. 2009. Using hedonic property models to value public water bodies: An analysis of specification issues. *In* Water Resour. Res.

O'Brien, R.M. 2007. A caution regarding rules of thumb for variance inflation factors. Quality & Quantity **41**: 673–690.

O'Neill, R.V., Deangelis, D.L., Waide, J.B., and Allen, T.F.H. 1986. A hierarchical concept of ecosystems. Princeton University Press, Princeton, N.J.

O'Neilll, R.V., Deangelis, D.L., Waide, J.B., and Allen, T.F.H. 1986. A hierarchical concept of ecosystems. Princeton University Press, Princeton, N.J.

Openshaw, S., and Taylor, P.J. 1979. A million or so correlation coefficients: Three experiments on the Modifiable Areal Unit Problem. *In* Statistical Methods in the Spatial Sciences. *Edited by* N. Wrigley. Pion, London. pp. 127–144.

Orford, S. 2000. Modelling spatial structures in local housing market dynamics: A multilevel perspective. Urban Studies **37**(9): 1643-1671.

Orford, S. 2002. Valuing locational externalities: A GIS and multilevel modelling approach. Environment and Planning B: Planning and Design **29**(1): 105-127.

Ottensmann, J.R., Payton, S., and Man, J. 2008. Urban location and housing prices within a hedonic model. Regional Analysis and Policy **38**(1): 19-35.

Overmars, K.P., Verburg, P.H. 2006. Multilevel modelling of land use from field to village level in the Philippines. Agricultural Systems **89**: 435-456.

Paccagnella, O. 2006. Centering or not centering in multilevel models? The role of the group mean and the assessment of group effects. Evaluation Review **30**(1): 66-85.

Paez, A., and Scott, D.M. 2004. Spatial statistics for urban analysis: A review of techniques with examples. GeoJournal **61**: 53–67.

Palmquist, R.B. 1992. Valuing localized externalities. Journal of Urban Economics **31**(59-68).

Pan, W.K.Y., Bilsborrow, R.E. . 2005. The use of a multilevel statistical model to analyze factors influencing land use: A study of the Ecuadorian Amazon Global and Planetary Change **47**: 232-252.

Paterson, R.W., and Boyle, K.J. 2002. Out of sight, out of mind? Using GIS to incorporate visibility in hedonic property value models. Land Econ. **78**(3): 417-425.

Polsky, C., and Easterling, W.E. 2001. Adaptation to climate variability and change in the US Great Plains: A multi-scale analysis of Ricardian climate sensitivities. Agriculture, Ecosystems and Environment **85**: 133–144.

Poor, P.J., Pessagnob, K.L., and Paul, R.W. 2007. Exploring the hedonic value of ambient water quality: A local watershed-based study. Ecol. Econ. **60**: 797-806.

Raudenbush, S.W., and Bryk, A.S. 2002. Hierarchical Linear Models: Applications and Data Analysis Methods. Sage Publications, London.

Raudenbush, S.W., Bryk, A.S. . 2002. Hierarchical Linear Models: Applications and Data Analysis Methods. Sage Publications, London.

Raudenbush, S.W., Rowan, B., King,S.J. 1991. A multilevel, multivariate model for school climate with estimation via EM algorithm and application to US high school data. Journal of Educational Statistics **16**: 295-330.

Reginster, I., and Edwards, G. 2001. The concept and implementation of perceptual regions as hierarchical spatial units for evaluating environmental sensitivity. URISA Journal **13**(1): 5-16.

Ridker, R.G., and Henning, J.A. 1967. The determinants of residential property values with special reference to air pollution. The Review of Economics and Statistics **49**: 246-257.

Robert, S.A., Reither, E.N. 2004. A multilevel analysis of race, community disadvantage, and body mass index amongadults in the US. Social Science & Medicine **59**: 2421-2434.

Rosen, S. 1974. Hedonic prices and implicit markets: product differentiation in pure competition. The Journal of Political Economy **82**(1): 34–55.

Sedgley, N.H., Williams, N.A., and Derrick, F.W. 2008. The effect of educational test scores on house prices in a model with spatial dependence. Journal of Housing Economics **17**: 191-200.

Shieh, Y.Y., and Fouladi, R.T. 2003. The effect of multi-collinearity on multilevel modeling parameter estimates and standard errors. Educ. Psychol. Meas. **63**: 951-985.

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

Snijders, T.A.B., and Bosker, R. 1999. Multilevel analysis. An introduction to basic and advanced multilevel modeling. Sage, Thousand Oaks, CA.

Soobader, M., Cubbin, C., Gee, G.C., Rosenbaum, A., Laurenson, J. 2006. Levels of analysis for the study of environmental health disparities. Environmental Research **102**: 172-180.

Steenbergen, M.R., and Jones, B.S. 2002. Modeling multilevel data structures. Am. J. Polit. Sci. **46**(1): 218-237.

Subramanian, S.V., Duncan, C., Jones, K. 2001. Multilevel perspectives on modeling census data. Environ. Plan. A **33**(3): 399-417.

Taylor, L.O. 2003. The hedonic method. *In* A primer on non-market valuation. *Edited by* P.A. Champ, K.J. Boyle, and T.C. Brown. Kluwer, Dordrecht, the Netherlands. p. 592.

Thorsnes, P. 2002. The value of a suburban forest preserve: Estimates from sales of vacant residential building lots. Land Econ. **78**(3): 426–441.

Troy, A., and Grove, J.M. 2008. Property values, parks, and crime: A hedonic analysis in Baltimore, MD. Landscape and Urban Planning **87**: 233–245.

Turner, M.G., Dale, V.H., and Gardner, R.H. 1989. Predicting across scales: Theory development and testing. Landscape Ecology **3**: 245-252.

Tyrväinen, L., and Miettinen, A. 2000. Property prices and urban forest amenities. J.Environ.Econ.Manage. **39**(2): 205–223.

Western, B. 1998. Causal heterogeneity in comparative research: A Bayesian hierarchical modeling approach. Am. J. Polit. Sci. **42**: 1233-1259.

White, H. 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica **48**(4): 817-838.

Wilhelmsson, M. 2000. The impact of traffic noise on the values of single-family houses. Journal of Environmental Planning and Management **43**(6): 799–815.