RESOLVING THE AGGREGATION PROBLEM THAT PLAGUES THE HEDONIC PRICING METHOD

A Thesis
Presented to
The Academic Faculty

by

Clifford Allen Lipscomb

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Public Policy

Georgia Institute of Technology

RESOLVING THE AGGREGATION PROBLEM THAT PLAGUES THE HEDONIC PRICING METHOD

Approved:

Philip Shapira, Chairman

Michael C. Farmer

Patrick S. McCarthy

Douglas Noonan

David Sawicki

Date Approved 11/18/03

DEDICATION

This paper is dedicated to two special women, without who I would not be at this point in my life. To my mother, Angie Barnette, thanks for providing me with the opportunity to become "Dr. Bubba" and to fulfill all of my academic endeavors. Without your support and contagious work ethic, none of this would be possible. And to my wife Amelia, thanks for putting up with my crazy and sometimes unforgiving work schedule. I am constantly in awe of the emotional support that you provide. I love you both very much!

EPIGRAPH

"It is a pity that econometricians have for so long neglected individual, microeconomic behavior as a fruitful field of study on the grounds that data are hard to obtain and laborious to handle."

J.S. Cramer, Econometric applications of Maximum Likelihood methods, Cambridge University Press: Cambridge, 1986.

ACKNOWLEDGMENT

I would like to acknowledge the time and effort Michael C. Farmer put into making this dissertation the most fruitful academic endeavor of my career. Michael has been a constant source of motivation and inspiration for the last five years. Michael, I value our professional relationship and non-professional friendship more than you will ever know.

TABLE OF CONTENTS

TABLE OF CONTENTS (cont'd)

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Housing hedonic studies typically assume that individuals or households are similar enough to aggregate into a single demand equation for analysis, typically relying on ordinary least squares (OLS) or some other single-line equation estimator. In these models, heterogeneity itself is managed by non-spherical disturbance corrections, typically spatial autocorrelation or heteroskedasticity in the single-line OLS estimate. This paper tests whether households in the same neighborhood can be theoretically and empirically treated as a single "type" or if households can be sorted into more than one "type." If the latter, then a single equation estimate may not be the most appropriate technique for a hedonic model that *structurally* accounts for household diversity. My technique, the seemingly unrelated regression (SUR) model, allows for more than one demand curve to represent housing demand and allows several "types" to compete over a fixed housing stock in a given residential neighborhood. As such, the SUR is the empirical translation of a theory that different household "types" can coexist in the same neighborhood. The results suggest the existence of multiple hedonic price surfaces in a single neighborhoood; and that the method developed here should be considered by professional city planners as one that should shape their often idealistic plans to incorporate the intricacies of the local real estate market.

CHAPTER 1

URBAN ECONOMICS MODELS

1.1 Overview

Just as the exurban "white flight" movement of predominantly white people occurred in the 1960s and 1970s, the 1990s ushered in a new physical movement of residential and commercial infrastructure to the central city. While still underdeveloped in many cases, urban areas are being reborn as new economic incentives encourage in-town developments. The challenges that face planners and economists include the structuring of incentives in such a way as to invite developers to enhance existing infrastructure; what hinders this process are laws, investor intransigence, the challenges of planning for one scale using data appropriate for another scale, and the ignorance of basic microeconomic theory. As these in-town developments occur, the urban landscape changes, meaning the kinds of persons attracted to the area will also change. Households will move into urban areas, for example, if they prefer urban cultural amenities and relaxations of their commuting times [Roback (1982), Cohen (1990)]. Current residents will stay or move depending on their levels of satisfaction with these urban infrastructures; their degree of tolerance for new, often transient, neighbors; their tolerance for certain levels of crime; and their desires for certain kinds of lifestyles. The model presented here seeks to find patterns in households' preferences for certain local features, such as their desires to be near public transportation and/or green space. Using this model, one can reasonably predict the impact on neighbors' dwellings' sales prices if a particular land parcel is designated for a particular use (i.e. a vacant parcel converted to open space) and if the neighborhood is assumed to remain static (i.e. no households move from or into the neighborhood). Also, one can forecast the "type" of household most likely to move into, say, a vacant house, in this neighborhood based on the dwelling characteristics most preferred by households of certain characteristics.

1.2 Neighborhood Land Uses

The household location decision involves numerous choices made simultaneously. These choices include the household's distance to work, distances to shopping and recreation opportunities, and distances to other kinds of households. A household is usually assumed to consider any externalities that may affect its quality of life at its particular location, although the time that it takes these households to fully realize the price response to these externalities varies. In one example, Kiel and McClain (1995) find evidence that *rumors* of the siting of an incinerator caused a price response in the local housing market, followed by another price response at the groundbreaking. So, while the time that it takes for a household to respond to local externalities varies depending on the information available, if the land use particulars of nearby parcels negatively affect the household, then these negative externalities (smog, foul odors from water treatment plants, excessive noise from highways, and other factors that make the parcel less attractive to buyers) will cause the household to have a lower willingness to pay for its dwelling, all else held constant. Conversely, other externalities (such as being close to open spaces, retail and commercial areas, improved access to public services, and environmental amenities), if perceived to be positive, will have an upward effect on the selling prices of these houses. In other words, externalities, both positive and negative, are reflected in the actual sales price that a household pays for a particular dwelling once they are internalized by each household. For policy analysts and local government officials, this is very important information to know about a neighborhood as it provides insight into the local tax base (which determines to a certain extent the funding potential of local governments' activities), the local amenities preferred by households (which also influences the local tax base), and how these preferences provide the foundation for targeted neighborhood redevelopment.

In this study, information on household preferences is collected from a cross-section survey of the Home Park neighborhood in Atlanta, Georgia, which will be described in more detail at the end of this chapter. The problem with cross-sectional data is that any generalized statement of household preferences of the neighborhood becomes out-of-date as parcels change from one land use designation to another or as different households move into the neighborhood. Cross-sectional data permits one to draw conclusions about the current households in the neighborhood *only* as extrapolations to other neighborhoods are not appropriate. When neighborhood planners and economists try to create "ideal" neighborhoods using the information provided by households, incremental land use changes in the Lindblom (1959) tradition most likely will not accomplish the neighborhood's economic and livability goals (Farmer 2003). To effectively push a neighborhood toward its optimal "sustainable" end, several key parcels and their most appropriate land use designations should be identified and marketed toward investors who will develop those sites in accordance with the wishes of local residents, city planners, and local government officials.[1] Key first investments may include designating open space parcels or locating well-stocked groceries, so private owners will not insist upon constructing houses on these lots. First investors must have assurances (contractual or verbal agreements) that their neighbors will reinvest in their own parcels in conformity with the neighborhood's long-range master plan. Otherwise, first investors' actions will go unmatched by current residents, thus not creating the maximum benefits for the neighborhood.

As an example, if my next-door neighbor adds $10,000 of economic value to her house with the construction of an additional bedroom and full bath, the selling price of my house (and others) is likely to increase due to the across-parcel impacts of land use attributes and dwelling features. A second example would be the increase in my sales price due to the fact that my across-the-street neighbor has a well-manicured garden that provides me joy as I leave my house to work. In both examples, even though my neighbor adds economic value to her house, she also "loses" in the sense that her proximal and surrounding neighbors are reaping benefits from the added value of her house *without necessarily adding value themselves.*[2]

Other types of housing externalities that are important to understand are those that involve the competition for space. Externality *intensive* land uses, or those that strongly impact the prices of other parcels, such as open space or key commercial sites, may be "outbid" by less economically valuable uses such as multi-family residential housing if

[1] From a policy design context, this sounds more like an incrementalist approach in the style of DeLeon (1992), who warns that incrementalism in its original conception does not have the requirement of a clear vision and statement of policy purpose.

[2] While many economists would agree that housing is not a "zero-sum game", the homeowner adding value to her dwelling increases the sales prices of "nearby" dwellings (a positive externality). The reciprocation

the owners cannot capture the residual rents attributed to their parcel neighborhood-wide.[3] In other words, if landowners cannot capture the rents from their parcel, they will be less inclined to zone that land for uses (such as open space) that bring more benefit to others than to themselves.

However, the competition for space does not necessarily suggest a coordination problem among investors. A self-interested investor, whose goal is to maximize profits from the development and sale of a particular parcel, may choose to develop a property without consulting the local neighborhood association. The potential problems with this scenario include a lack of neighborhood support for the project and the development's lack of "assimilation" into the neighborhood. On the other hand, an investor with a long-range vision that conforms to the visions of the neighborhood most likely will avoid the problems faced by the self-interested investor; this means that the coordination of investors who might be intransigent, but have the start-up capital to locate groceries and open spaces to areas that will permit the capture of rents, is contingent upon assurances by neighbors that all will develop their parcels according to an agreed upon master plan, which accounts for "the spatial dimension of coordinated simultaneous investment" (Farmer 2003). Neighborhood associations in conjunction with urban planners and other economists can address these two concerns and coordinate first investments to jump-start the envisioned neighborhood revitalization process.[4]

(other households making home improvements that increase the sales prices of other "nearby" dwellings) may not occur necessarily.

[3] The hedonic model in Chapter 3 will determine the marginal prices of being near these open spaces, brown industries, and commercial sites so one can see if externality intensive land uses are indeed outbid by less intensive uses.

[4] Overcoming investor intransigence (stubbornness or inflexibility) is not the focus of this research. But, it is a phenomenon that must be overcome for neighborhoods to be redeveloped according to their master plans.

Two implications arise from these revitalization efforts. First, any attempt to redevelop an existing urban area should consider the current state and demographic composition of the neighborhood, current land uses and laws, and an efficient way to achieve simultaneously the goals of planners, economists, local government officials, and neighborhood residents. Second, *in situ* land uses cannot rise to their highest and "best" uses in isolation; the cross externalities associated with various types of land uses and various types of neighbors are important factors that contribute to the overall value of the parcel of interest. This suggests that the coordination of investors in a systematic way no longer reinforces the land uses that impose negative externalities on neighboring parcels, but fulfills the (homeowners') goal to maximize total economic real estate value (aggregated selling prices) in an area by seizing gains from the positive externalities associated with the correct mix of land uses.

1.3 Reinvestment Obstacles

The biggest obstacles to reinvesting in the built city (and to achieving the correct mix of land uses) are the fixed costs of reinvestment (i.e. capital outlays), traffic and population congestion, air and noise pollution, other contributions to quality of life checking the benefits of urban agglomeration (Tolley 1974), the diverse characteristics of an area's residents, and investors' fears that over-improved islands of development are inconsistent with sustained growth planning schemes (Farmer 2003). These obstacles have the potential to dissuade residents and investors from supporting the land use features that may be "best" for the city. For example, a city may want to encourage more open space and reduce ambient particulate matter, but current public policy dictates that more monies be used for other services such as reducing the commuting costs of congestion.[5] This occurs because politicians and transportation authorities continue to succumb to commuters' desires for more roads instead of using these monies to make urban areas more livable or to enhance the city's capacity for mass transit opportunities (Ibid.). The decisions of these key policymakers have caused "pricing errors of the past [to be] cast in brick and asphalt, [resulting in policies that] are very expensive and have limited effectiveness" (Anas, Arnott, and Small 1998, p. 1457). Instead, if these key policymakers had used road monies to redevelop the central city to include additional green space, parks, passenger friendly mass transit, and other family friendly infrastructure, suburban families would perceive an even greater advantage to living in the central city, the result of which would be households willing to move to the urban core. But, given the past decisions of policymakers, it seems that the use of federal and state transportation funds to revitalize urban areas (and not roads) is impossible, meaning that some other fundraising method must be utilized to transform key parcels to a land use that is consistent with the neighborhood's long run master plan.[6]

Another problem that complicates the issues presented here is that all urban residents are different. Paul Samuelson (1954) argued that perfect household preference revelation would occur only in private competitive markets where consumers' choices are observed. Then, Tiebout (1956) suggested in contrast to Samuelson that households sort into different homogeneous jurisdictions within a city in a "self-organization" process, a

---

[5] The 2001 Urban Mobility Study conducted by the Texas Transportation Institute concludes that metropolitan Atlanta commuters were stuck in traffic a total of 152.5 million hours in 1999, which costs businesses and commuters approximately $2.62 billion in lost wages, lost time, vehicle depreciation, and other commuting costs.

process in which consumers reveal their true preferences for locally provided goods if an infinite number of jurisdictions is available. Even if Tiebout's assertion, that a theoretical and conceptual solution for the preference revelation problem exists, is correct, each resident's motivation for locating to a certain neighborhood or community is different. Some households choose to live in a particular jurisdiction because it has a nice view of the city skyline, or because it is close to the household members' workplaces. At a more disaggregate level, other households choose to live in a particular dwelling for its large yard and/or bay window that affords them a view of the trees and street.

Despite the fact that households are different in multiple dimensions and choose particular neighborhoods for different reasons, some dimensions might be similar enough to generalize to different classes of households. For example, Glaeser, Kahn, and Rappaport (1997) found that the rich and poor alike favor improvements in urban transportation. One of the benefits of improved urban transportation is lower auto emissions as a result of smarter intra-city logistics planning (bus and rail route coordination). Simultaneously, some other benefits of improved urban transportation planning may accrue to households with particular income levels. Kahn (1996) says that the attraction of higher income in-migrants (through improved transportation planning) supports the decline of urban manufacturing and the rise of environmentally friendly services through taxes and support of local services. In a sense, higher income households can support lower income households (i.e. blue-collar workers) if the transportation infrastructure and available amenities are conducive to the needs of higher income in-migrants.

[6] A key role that neighborhood associations could play is that of a neighborhood "bank." Responsibilities

Since these intra-city logistics improvements are tied to the developing pattern of real estate over a landscape, one can argue that a "significant portion of congestion pressures in many cities can be eased at the margin by unleashing self-interested reinvestment activity rather than by sponsoring capital intensive projects to allay pollution and traffic pressures" (Farmer, Reiss, and Ditto 2000). This suggests that the key to neighborhood transformations and urban livability is the use of economic incentives to foster reinvestment activity among investors, developers, and area residents instead of pouring additional monies into more roads. Even Richard Moe, president of the National Trust for Historic Preservation, and other Georgia historic preservationists and economic developers agree that revitalizing neighborhoods is key to curbing urban sprawl, improving quality of life, and reducing gridlocked highways (Minor 2002). These reinvestment activities will 1) attract new residents that most prefer to reside in these urban areas and 2) at the margin relieve traffic congestion, pollution, and support start-up revitalization efforts of former urban manufacturing areas. What this means for local policymakers is that a robust estimate of the economic impact of these activities is needed. This paper provides a way in which the implicit economic value of certain amenities for different classes of households within a single residential neighborhood can be estimated. Then these estimates can be used to evaluate the change in the aggregate sales prices of the houses in the neighborhood when a certain kind of change is made to the local landscape.

---

could include having a revolving low-interest loan fund for façade improvements, landscaping, and others.

1.4 Land Use Models

Revitalizing the urban landscape in accordance with consumers' preferences is a lofty goal, especially given the heterogeneity of consumers' preferences. Clark *et al.* (2002) argue that "a residential population of young professionals with more education and fewer children creates a social profile geared toward recreation and consumption concerns," which values highly "a cultural center offering diverse, sophisticated and cosmopolitan entertainment lacking elsewhere" (p. 500). Similarly, recent findings show that cities with more gay men tend to have certain types of amenities than other cities, and those amenities and gay men are some of the "strongest predictors of high-technology job growth" (Black et al. 2002 and Florida 2002, quoted in Clark *et al.* 2002). So, one can see the possibility that distinct types of citizens respond to distinct economic development strategies in a given area. In other words, a well devised economic development strategy can fail if urban planners do not install quite early the kinds of amenities appropriate for the type of human capital and skilled persons they wish to attract. This is important because the shape of the urban landscape, which attracts certain citizen types, depends on the economic development strategies employed by planners and local government officials.

At the heart of this effort to revitalize the urban landscape according to preferences are the spatial relationships between households. Firms, which locate close to common labor pools in order to take advantage of scale economies from production and to save on wages, exemplify the minimization of costs associated with production (Farmer, Reiss, and Ditto 2000). These sorts of firm-level scale economies take various forms, from a skilled work force specializing in a given industry that serves as a common labor pool for several producers of similar goods, to "a joint, or congestible, set of recreation and cultural services [that satisfy] a growing population under declining or flat marginal costs" (*Ibid.*, p. 3). In a similar vein, *households* locate close to desired amenities and employment clusters to reduce the costs (travel costs, congestion costs, and others) associated with accessing these amenities and job locations.

The role of space and externalities suggests a brief examination of the evolution of land use theories. One of the first land use models was that of von Thünen (1826), who argued that an export-oriented economy drives efficient land use and causes land to cluster into different zones (or concentric rings) around a town. In his model, the value of agricultural land was determined by transportation costs from the particular farms to the marketplace where agricultural goods were sold (Barlowe 1993). While all residents of an area were assumed to be farmers in the original model, variations of the original von Thünen model have argued that a housing band was located between the city center and agricultural land (Asami and Isard 1989, p. 514), and that high-end commercial activities would gravitate toward the city center while bulk services would locate within the housing band. Therefore, while von Thünen's model is not used explicitly today, it is worthwhile to see how it spurred other land use theories over time.

The urban planning model of Dixit and Stiglitz (1977) and Krugman (1980, 1991, 1996) adds to the theory of von Thünen by incorporating a second "sister-city" urban cluster that develops as pressure from the burgeoning population in one city forces it and the suburban perimeter to expand until a closely related product center and its manufacturing suppliers support this second cluster, forming a new city-size equilibrium. Mayer and Somerville (2000) support the Dixit-Krugman model in their discussion of residential construction; they argue that if a city experiences a population influx, "the demand for new residences increases, land and house prices rise, new construction occurs, and the city increases in size to accommodate the new residents" (p. 87). In other words, Dixit and Krugman thought that a population influx toward the city center would create a second layer of von Thünen characteristics.

Next, the Tiebout (1956) model predicts diversity among local jurisdictions in which distinct citizen types support distinct public goods packages that prevail in a given jurisdiction. Tiebout built upon previous theory by adding household characteristics to urban planning theory. Instead of the focus being on kinds of industry clusters and suburban expansion, his model focuses on how similar kinds of households relocate to the same jurisdictions because they have similar preferences for the local public goods found there.

A fourth land use model is that suggested by Alonso (1964) and Muth (1969). These models do not consider socioeconomic status or neighborhood quality as determinants of residential location. They fundamentally argue that transportation costs to the central business district (CBD) determine the spatial distribution of the land and housing prices, land consumption, and spatial residents' arrangements by income (Coulson 1991, p. 299). While this seems to be a natural extension of the von Thünen model (which in its original version says that the value of agricultural land is determined solely by transportation costs), it actually was a remarkable addition to land use theory to conceive that multiple land use layers could radiate from the CBD. Hypotheses of the original Alonso-Muth models include negative rent gradients and increased land consumption per household with increases in the distance to the CBD. As one example, Wheaton (1977) tested empirically both hypotheses: for the Alonso-Muth spatial income effect, he concluded that the "spatial income gradient in metropolitan areas would seem to be more the result of social or racial externalities [in the Tiebout tradition] and the incentives produced by municipal decentralization" than purely income effects (p. 621); for the land consumption/income effect, he concluded that "spatial bidding for land of different income groups looks almost identical" (p. 630).

One complication of the negative rent gradient hypothesis is that urban submarkets may have their own rent gradients and/or rental multipliers, thus confounding (overlapping) the *market* negative rent gradient and the *market* rental multipliers.[7] Given this complication, Muth's standard assumptions of identical incomes and preferences, which Wheaton (1977) relaxed, are also relaxed in this paper. This relaxation of identical incomes is particularly important because of the high percentages of students (more than 50% in the sample) and renters in the neighborhood (approximately 65% for the sample).

As one can see, land use theories have been posited and modified over time as certain assumptions are relaxed. Most recently, Farmer (2003) integrated the Dixit-Krugman and Tiebout models into a land use model that emphasized the following: 1) housing near commercial and amenity services increase in value compared to housing away from these areas; 2) multipurpose trips are more significant than single-purpose trips (due to time constraints); and 3) pockets of agricultural land accommodate

[7] To estimate the market sales price for rental dwellings in this analysis, a rental multiplier of 120 is used for the neighborhood after consulting several local real estate agents. This multiplier is used to estimate the market sales price of a dwelling that is currently being rented. So, if a dwelling rents for $1,000 per month, I estimate that it would sell for $120,000. This is how the dependent variable for rental dwellings is determined. Also, Coy (2003), for example, uses the idea of rental multipliers to determine the financial gains of buying versus renting a dwelling, but does not explicitly state the rental multiplier for Atlanta to compare to the one used in this study.

independent centers of retail and/or amenity services *and* simultaneous convenience shopping and professional services, which may compromise opportunities for amenity and open space in the reinvestment process.

This paper focuses on the microeconomic behavior of individual households operating in a finite neighborhood space. One goal here is to devise a method that can be used to predict the degree of change in neighborhood composition (measured by the percentage of households of particular "types" that will stay or move from the neighborhood in response to a policy change). The method used to predict changes in neighborhood composition differs from typical large-scale planning methods in that the latter requires fundamental economic assumptions about individual household behavior. Planners are forced to make these assumptions about households so that they can make suggestions *at a larger scale* as to what percentage of land should be designated as open space, retail, residential, etc. My discussion of the Georgia 400 corridor (please see Chapter 6) suggests that such large-scale planning efforts and their required assumptions about household behavior may be appropriate for large areas. However, when one wishes to do neighborhood planning, the confines of a smaller space require that the planner know more about each individual parcel and the households that inhabit them.

This research continues the work of Farmer (2003) by incorporating the following features into an urban land use model that focuses on residential housing:

- Neighborhood diversity, or distinct citizen types that support neighborhood-specific and dwelling-specific characteristic packages;

- Endogenous groupings of households (defined by particular characteristics and amenities) that mitigate the aggregation problem;
- Socio-economic status and neighborhood amenities as location determinants; and
- Income diversity of neighborhood residents.

In this research, the central question to be addressed is "Do statistically aggregable subgroups of households within a single neighborhood have different marginal prices for the same local amenities and dwelling characteristics?" If so, then the full aggregation of households, or the assumption that all households in an area are analyzable using a single line equation method, masks some important distinctions between different classes of households concerning the directional impacts of the same local amenities, disamenities, and dwelling structure characteristics on the sales prices of different types of households. By separating households into different statistically aggregable "types", the empirical representation of the first-stage hedonic equation can proceed through the simultaneous analysis of several types of households, each with its own distinct set of preferences that are different from other types.

To answer this central question, the subsequent chapters consider identification within hedonic pricing models (please see Chapter 2), past research on utility theoretic frameworks (please see Chapter 3), and the different empirical estimation methods used to model the first-stage hedonic regression equation (please see Chapter 3). Also, the central question motivates new technical questions about consumer preferences. For example, do dwelling and neighborhood characteristics influence house sales prices differently for different subgroups of households? While urban planning models assume that households are similar enough to be analyzed as a single group, this research establishes a method by which subgroups of households can be differentiated from other subgroups within the same neighborhood. The results reveal price differences across statistically aggregable groups of households (please see Chapter 4). For policymakers, this means that neighborhood analyses must be conducted using more detailed (maybe non-public) data on households than available from government sources. To validate this claim, statistical exercises are conducted (please see Chapter 5) that simulate the household types that are likely to move out of or remain in the neighborhood when a green space that currently exists (Home Park green space) is complemented with an additional green space (the proposed park). Then, the data requirements for more macro-scale analyses, the policy implications of this paper, the generalizability of results and methods, some recommendations for future research, and some implications for planners and regional scientists conclude this paper (please see Chapter 6).

Finally, this paper makes progress toward a theory of household "types" for both owner-occupied and rental dwellings that is empirically verified in a single urban neighborhood, Home Park. Home Park, a neighborhood adjacent to the north part of the Georgia Institute of Technology campus, was founded in 1920 and was originally built for the workers of the Atlanta Water Works plant and steel mill. Figure 1 shows the geographic distribution of sales prices in Home Park as well as the location of several key parcels, including the Home Park green space, the proposed park, the 10th Street commercial strip, and the Meineke automobile service center. Over time, the original owners have bequeathed the properties to widowed spouses and/or their children; as a result, Home Park has a large percentage of renters (approximately 65% of all households). Of all renter households, 65% are occupied by at least one college student; of all owners, 7% have at least one college student.[8] Then, the next largest groups of households contain younger couples and older single men with at least one person employed full time. This, along with the fact that Home Park is surrounded by highways, commercial properties, and the university, which isolates it from other residential neighborhoods, makes it an interesting study area.

---

[8] This statistic illustrates two interesting facts about the neighborhood – 1) some students choose to live in owners' spare bedrooms and 2) parents purchase dwellings for their children to live in while attending college.

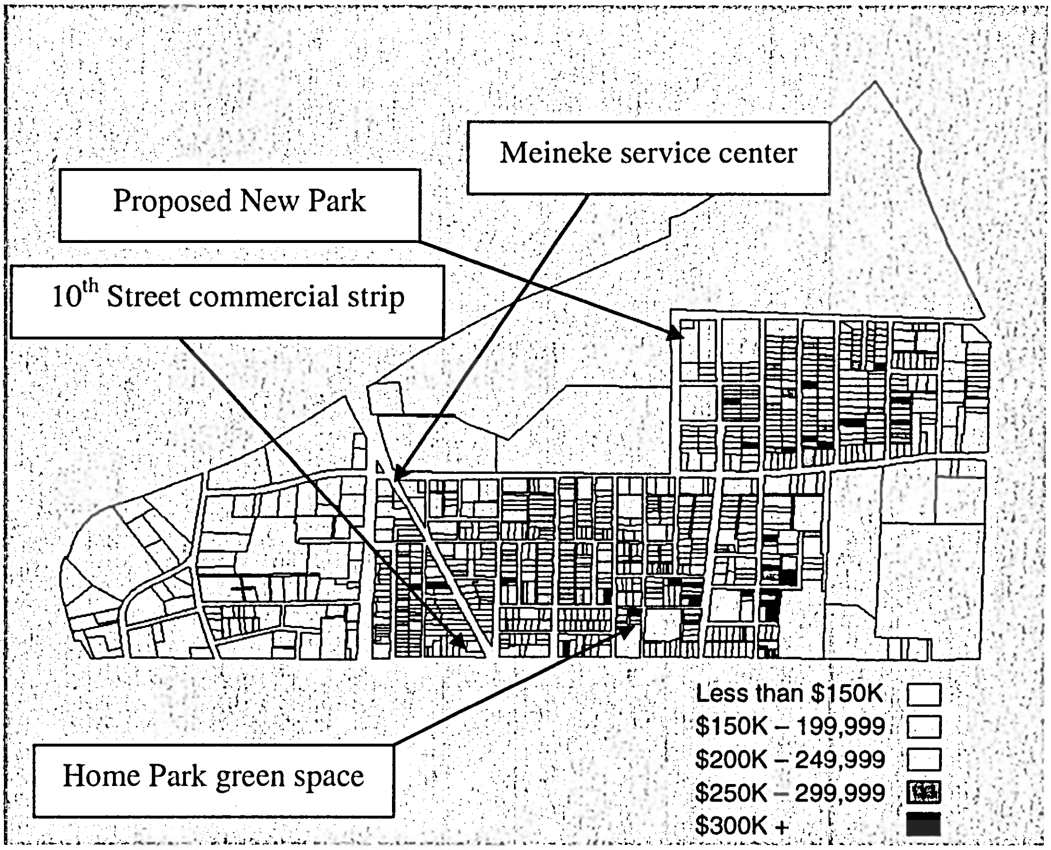


Figure 1 – Geographic Distribution of Sales Prices (Black = Highest)

Another interesting phenomenon that had an indirect effect on Home Park was the redevelopment of several nearby neighborhoods for the 1996 Summer Olympics. As millions of dollars were poured into these formerly dilapidated neighborhoods, public housing and other subsidized housing areas were rebuilt to accommodate households of all income levels. These changes have not occurred in Home Park, though, as redevelopment monies were not directed at this particular neighborhood. Yet, the changes to nearby neighborhoods did inject more confidence into Home Park real estate improvements, stabilizing the investment and resident diversity patterns noted.

CHAPTER 2

HEDONICS MODELS

In this chapter, I discuss the strengths and weaknesses associated with hedonic pricing as a method of estimating the "use" value (as opposed to the "non-use" value estimated by other forms of non-market valuation like contingent valuation) of certain dwelling characteristics and neighborhood features. I start with a history of the method, its application to housing markets, its weaknesses, and its recent applications.

2.1 The History of Hedonics Research

In the early 1900s, research that considered the selling price of a good as a function of its characteristics was novel. This technique, the hedonic pricing method, treats the price of a good as the dependent variable and the characteristics of a good as the independent variables. Using regression equations, slope coefficients (which are interpreted as unobserved, implicit prices) on each independent variable are estimated that reflect the impact of a one-unit change in the independent variable on the dependent variable, holding all other variables constant. This method is commonly used to model the demand for goods that do not have traditional economic markets (e.g. air quality or water clarity) using the prices of goods determined in other markets (e.g. the price of housing). Automobiles, water quality, trees, and houses are only a few of the goods that can be analyzed using this method of valuation.

Debates in the literature on the first use of this technique are common. Colwell and Dilmore (1999) suggest that an overlooked monograph by G.C. Haas of the Division of Agricultural Economics at the University of Minnesota Agricultural Experiment Station in 1922 was the first hedonic study. This conclusion defies those of Griliches (1961) and Goodman (1998), who argue that A.T. Court conducted the first hedonic study on automobiles in 1939. Regardless of its exact origins, Lancaster (1966) and Rosen (1974), who developed early theories of consumer behavior, popularized the hedonic method as a useful and legitimate valuation exercise.

According to Raymond Palmquist, the hedonic pricing methods are based on "the realization that some goods or factors of production are not homogeneous and can differ in numerous characteristics" (quoted in Braden and Kolstad 1991). Consumers purchase different bundles of characteristics each time a buying decision is made. When purchasing an automobile, for example, a consumer buys a specific type of engine, steering column, body style, color, and other characteristics. In this market, the manufacturer is willing to sell the automobile at a specific price based on the labor, parts, and shipping costs of the automobile, whereas the consumer is willing to pay a certain price for the characteristics of the automobile in question. When this offer and bid are in equilibrium, a transaction between buyers and sellers occurs. The intersection of multiple bid and offer curves forms the hedonic price equation.[9]

The hedonic pricing method can be best illustrated through contrasting examples of goods that do and do not have traditional economic markets. First, an example of a good that does have a traditional economic market is Nike shoes. Economists observe the quantities of shoes bought by consumers at particular prices. Then, the estimation of the demand for Nike shoes is straightforward- plot the prices and quantities on a graph.

In contrast, the estimation of the demand for a good that does not have an economic market like water clarity or air quality is not as simple. In the case of housing, which is a heterogeneous bundled good, researchers can use the hedonic pricing method to determine the demand for housing characteristics such as square footage of living space, acreage, proximity to green space, etc. Houses are listed at a price determined by the owner and based on the sellers' characteristics, the dwelling's characteristics, the dwelling's amenities, the actual selling price of similar houses in the same neighborhood, the types of land use designations surrounding the particular dwelling for sale, among other factors. But, the actual sales price of the dwelling is determined by a negotiation process that ultimately equates the purchaser's willingness to pay for the characteristics and amenities with the seller's willingness to accept a certain level of compensation for the house. This produces a buying/selling decision in the housing *market*, similar to that in the market for Nike shoes. The difference is that in the case of Nike shoes, each component or characteristic may have an explicit cost- i.e. a sole costs $5 of the total cost. For a dwelling, one does not directly observe a separate price for the fireplace in the living room, for the bay window in the kitchen, or for the bedroom on the main floor of the dwelling; regression analysis, if conducted properly, can reveal a robust estimate of the marginal implicit price for dwelling structure characteristics, certain neighborhood features, and other amenities that impact sales price.

The motivation for using the hedonic pricing method is to separate the contribution to total selling price of each of these characteristics and amenities. What is the contribution to the total selling price (marginal price), holding all other characteristics

---

[9] Please see Freeman (1993) for a nice summary of the hedonic pricing method and its limitations (p. 124-

and amenities fixed, of a bedroom? What is the marginal price of a square foot of living space? Then, once these marginal implicit prices have been calculated, researchers can estimate the demand for certain characteristics and amenities by aggregable demanders. Obviously, the derivation of marginal implicit prices differs from that for goods with true economic markets. But, hedonic marginal prices play the same role as do direct observations on prices in standard demand theory (Freeman 1993).

To conclude this section, I will briefly mention housing supply. It is clear that housing supply is not studied in the literature as much as housing demand. Smith (1976) argues that little empirical work on housing supply has been produced, citing the "difficulty of controlling simultaneously for housing quality and location amenities" as a contributing factor. For the purposes of this dissertation, the housing supply in Home Park is assumed to be fixed. This is an appropriate assumption because little new housing construction (new dwellings, not conversions or demolitions) is occurring in Home Park, as evidenced by the six building permits issued in the neighborhood since January 2000 to construct new residences (mostly duplexes). It is also appropriate because, at a basic level, an analysis of housing supply, particularly for different statistically aggregable household types that I attempt to show in this paper, suggests that different technologies might be used to construct housing for different household types. One factor that shifts supply curves is a change in the state of technology of a given economy. Since my supply-side data are very limited, and since I assume that the same technologies are used to construct housing, the estimation of the supply of housing is omitted here.

---

131).

2.2 The Hedonic Pricing Method Applied to Housing Markets

Dwellings are differentiated, bundled goods comprised of various styles, sizes, structure characteristics, and amenities. Dwelling sales prices reflect the particular combination of structure features (age of the structure, living space square footage, number of bedrooms, etc.); neighborhood characteristics (proximity to public transportation, noise levels from traffic, level of tree cover, etc.); and recreational amenities (parks, other forms of open space, presence of quality sidewalks, etc.). One of the earliest attempts to model the housing market was Muth (1960), who introduced the concept of "housing services", thereby encouraging economists to treat housing as *homogeneous*. Put another way, Muth's contribution was to use tools developed for markets with homogeneous goods to econometrically analyze housing markets.[10] However, the concept of housing services makes difficult the recovery of the marginal prices of particular housing attributes, such as a bedroom or a fireplace. Researchers after Muth treated houses as *differentiated* goods (having different characteristics and amenities/disamenities), as Lancaster (1966) first proposed and Rosen (1974) then popularized.

Traditionally, the first stage of hedonic analyses derives the marginal price surface using the selling price of a house regressed on its structural characteristics (number of rooms, baths, square footage, and others), parcel size (in acres), and nearby amenities (parks or shopping opportunities) and disamenities (landfills or industrial parks). Then, the second-stage hedonic equation uses the marginal price vector of

[10] To deal with the difficulty of treating housing units of different technological and legal characteristics as homogeneous, Muth suggests acceptance of the market's judgment "to treat as identical those units of 'housing' which command identical prices" (p. 32).

interest, say the marginal prices for the number of acres, and regresses it on the prices of relevant substitutes and complements as well as "demand shifters" such as income, race, and sex. The problem with the second-stage is that the derivation of the marginal price surface (in the first-stage) is based on a number of variables that are then used to estimate the demand for a particular characteristic, which creates an endogeneity problem. The hedonics literature is replete with articles that attempt to resolve this endogeneity problem [please see Palmquist (1984) for a good discussion of this endogeneity problem].

This paper attempts to derive better first-stage hedonic estimates. With so much emphasis on the *econometric* identification problem in the second stage, this paper's focus on the first stage addresses a *policy* concern in hedonics, that single-line estimators do not structurally capture the basic differences among households in a consistent way. These basic differences may include different willingness to pays (WTPs) for the same dwelling structure characteristics and/or local amenities. Using multiple-line estimators, I hope to pinpoint the household "types" that are better off with the construction of the new park (and who will continue to reside in the neighborhood); please see Chapter 5 for a discussion of the welfare economic impacts of this paper. Also, to reduce the problems associated with omitted variable bias, I explicitly incorporate space into the hedonic pricing model using a series of spatial variables.

2.3 Problems Inherent to the Hedonic Method

Whether one is talking about the economic valuation of clean air (Sieg *et al.* 2000; Chay and Greenstone 2000), open space (Benson *et al.* 1998; Bolitzer and Netusil 2000), or housing characteristics (Bourassa *et al.* 1999), the mitigation of the aggregation problem is a core concern. The aggregation problem is the assumption that unique consumers (households in this case) with unique preferences can be grouped together for analysis purposes. To clarify, this is different than the "identification problem that dominates discussion in the applied [hedonics] literature" described by Ekeland, Heckman, and Nesheim (2002, p. 304). The aggregation problem arises because households' unique preferences for unique bundles of goods cannot be measured "uniquely;" the hedonic identification problem concerns the second-stage hedonic equation, which I do not estimate here.

As an *extreme hypothetical example* of the aggregation problem, assume that each household, if treated as a unique buying and selling decision unto itself, requires its own hedonic regression equation to perfectly measure its marginal prices for each characteristic, neighborhood amenity, etc. In this extreme example, if regression analysis permitted the estimation of coefficients on different dwelling structure characteristics and local amenities, each household would have its marginal prices estimated from its own regression equation. Obviously, degrees of freedom limitations prevent this estimation from occurring. But, if this model could be estimated with a single observation, then the estimates obtained for each individual equation would perfectly reflect the marginal implicit prices of each dwelling characteristic and amenity *for that household that chooses to live in that dwelling*. This means that a neighborhood with approximately 800 households (like Home Park) would require a system of 800 different regression equations to be estimated simultaneously to completely capture the contribution to sales price of each unit of each characteristic, etc. The point is that this kind of simultaneous estimation is impossible statistically and provides the researcher with no indication of housing demand by a class of demanders; it only confirms the heterogeneity of household preferences over time and space regardless of the researcher's study area. In this situation, the econometric investigator learns nothing about the demand for housing at the neighborhood level.

At the other extreme, a *single* equation used to represent *all* households in a community provides the researcher with general trends in dwelling sales prices (e.g. a one percent change in dwelling square feet leads to a .5 percent change in sales price at the margin). If the researcher's question focuses on the estimation of marginal prices across a single community, then this *extreme aggregation* is appropriate. However, if the researcher's question is slightly different, say to allow the possibility that different types of households might systematically "value" these amenities and structure characteristics differently, then full aggregation only tells the researcher that all households are similar enough to analyze as a single group. Therefore, I assume that households can be allowed to sort into "types" based on their similarities in demographic characteristics, attitudes toward the neighborhood, and marginal implicit prices. Then, if household-level demographic and attitudinal data can be used to distinguish households into these "types" or aggregable subgroups, then the number of regression equations sufficient to mitigate the aggregation problem in a particular neighborhood will be less than the extreme disaggregation (one equation per household) and possibly greater than the extreme aggregation position (one equation per community or neighborhood).

A second and related problem with hedonics is the "incidental parameters" problem, which says that attributes are sampled only once. First analyzed by Neyman and Scott (1948) and later applied to congressional voting by Heckman and Snyder

(1997), the incidental parameters problem involves the violation of the classical regression assumption of drawing repeated samples from the same population, which *can* be a problem in hedonics research as the "population" of houses and the attributes of households changes with each new renter or homeowner. So, analyzing a cross-sectioned sample (one time period) means that the population is held identical by the econometric investigator for analyses of that period of time. This assumption (no resident out-migration) means that any conclusions about the neighborhood are applicable only to the current residents, and that no absolute rule of how the demographic and economic compositions of this neighborhood change over time can be devised here. But, in a subsequent chapter, I show how cross-sectional data can be used to simulate the change in demographic composition and dwelling sales prices in the neighborhood if a particular parcel is zoned to a different land use designation.

In choosing a particular dwelling[11], households pay (through the selling price) an unobserved amount of money for each characteristic of the dwelling and amenities in the neighborhood. For example, if a household narrowed its search to two identical (in structure and location) houses, one with a fireplace and the other without, it would choose the house with the fireplace if it (the household) were willing to pay the difference in sales prices between the house with a fireplace and the house without a fireplace. In other words, the household's purchase of the house with a fireplace tells researchers that

---

[11] As Rosen (1974) stated, households cannot "re-package" characteristics; they cannot buy one characteristic from one house (i.e. garden size) and combine it with another characteristic from a different house (i.e. number of rooms) (Hanley and Spash 1993). If this were the case, then the first-stage hedonic price function (the selling price of a house regressed on a set of independent variables) would be linear, which means that households would have similar marginal implicit prices for a particular good.

this household's preference for fireplaces can be measured through the difference in sales prices between the otherwise identical houses. Extending this logic to the other dwelling characteristics, a household is willing to pay a certain amount of money for each feature and amenity related to that house.[12] Therefore, researchers use regression analysis to estimate these unobserved marginal prices for house features and amenities.

Other issues that complicate the hedonic pricing problem include the various functional forms of the first-stage hedonic pricing equation, the introduction of new policy variables, and different methods of identifying neighborhood effects. These issues do not complicate my discussion of the aggregation problem, but are general concerns that are addressed in the hedonics literature. Box and Cox (1964); Cropper, Deck, and McConnell (1988); Halstead, Bouvier, and Hansen (1997); and Lipscomb (2001) have employed alternative Box-Cox functional form variations to get the "best fit" of the model. Sivitanidou (1995); Helmuth, Obata, & Kassabaum et al. (1997); Morrell and Lu (2000); and Lipscomb (2001) have used interesting policy variables like noise contours and the distance to an airport in the hedonic price function. Other statistical techniques such as method of moments (Bell and Bockstael 2000), principal components analysis (Bourassa et al. 1999), simultaneous equations (Irwin and Bockstael 2001), and latent variable approaches (Arguea and Hsiao 2000) are recent additions to the hedonics literature as well.

The two main concerns of past hedonics research that will be addressed here are 1) the uses of first-stage hedonic equations without a coherent theoretical motivation and

[12] The recent incorporation of spatial variables (distances to shopping opportunities, etc.) has complicated the model specification and functional form issues involved with hedonic models (i.e. do households really

2) the ways in which neighborhood effects have been modeled in previous studies. First, several recent studies (Halstead, Bouvier, and Hansen 1997; Bolitzer and Netusil 2000; Tyrvainen and Miettinen 2000, Luttik 2000) have not used a utility theoretic framework to derive the first-stage hedonic regression equation. The motivation of the first-stage hedonic equation earlier in this chapter simultaneously complements a utility maximization model developed in Chapter 3. Second, previous studies have employed a variety of methods for defining neighborhoods (Hughes, Jr. and Turnbull 1996; Rapaport 1997; Bourassa *et al.* 1999; Immergluck 1999; Bolitzer and Netusil 2000; Sieg *et al.* 2000), using dummy variables to control for neighborhood or jurisdictional effects (Thorsnes 2000 and Sieg *et al.* 2000), and allowing the error terms to absorb neighborhood effects (Irwin and Bockstael 2002). This paper improves upon these studies by using an endogenous process to sort households into distinguishable types.

The advantage of using an *endogenous* sorting process is that the researcher does not dictate *a priori* how many household types (or subgroups) exist in the neighborhood, nor does s/he know the criteria that distinguish one type from another. For example, market segmentation (which I also call "distinguishing into household 'types'") is commonly treated in the hedonics literature according to some spatial distinction, either by school districts, cities, or other geographical boundaries. The limitation here is that geographical boundaries may not represent true distinctions in housing supply or demand; households across several school districts may enjoy the same amenities and dwelling structure characteristics but are grouped with other households based on geography instead of revealed preferences in the housing market. In this paper, I can,

---

take into account the tree cover around their dwelling when choosing where to live?; what is the according to Luc Anselin, "let the data speak" because the spatial theory is developed well (quoted in Boyce, Nijkamp, and Shefer 1991).

To conclude, the argument made here is that knowing all the limitations of the hedonic pricing method, including the role that aggregation plays in these analyses, informs my approach and allows me to extend the theory of hedonic pricing to allow for more than one type of household in a neighborhood. This extension of current theory, which involves the use of a statistically-based sorting exercise and the use of a non-standard estimator of the marginal implicit prices, is detailed in the next chapter.

---

appropriate functional form for the distance variables?).

CHAPTER 3

MODEL TO MITIGATE AGGREGATION LIMITATIONS

In this chapter, I discuss the theoretical model that allows households in a single neighborhood to be sorted into endogenous household "types." The basis of this discussion is the famous Samuelson (1954) – Tiebout (1956) debate on how efficiently household preferences are revealed by their choice of jurisdiction in which to reside. Then, I discuss the empirical translation of this theory into a sorting exercise using econometric tools.

3.1 Overview

Urban economics models frequently assume that households within a particular space are of a singular type, or similar enough in some way (usually unstated) that they can be aggregated into a single group and then analyzed using single-line regression models such as ordinary least squares (OLS) or generalized least squares (GLS). For hedonic price models designed to analyze the value of a single amenity to local real estate markets, assuming that all households are similar enough to be grouped into a single analyzable group is essential to translate real estate value changes into reasonable estimates of the impact of certain neighborhood attributes. It is also common in hedonic models to make assumptions about the error structures to compensate for the heterogeneity of preferences so that a proper GLS estimate can be made to infer efficient welfare change measurements from changes in local amenity access. These economists will be referred to as the aggregation "optimists," or those who believe generally that households are similar enough to aggregate into a single type and analyze using a single-line regression equation. These "optimists" include Sieg *et al.* (2002), Ekeland, Heckman, and Nesheim (2002), and others who treat household preference heterogeneity through error term corrections to the first-stage hedonic equation.

Next, the aggregation "pessimists" are very skeptical that household diversity can be sufficiently captured by the "one neighborhood, one type" assumption implicitly made when single-line regression models are employed. While no authors who might be "pessimists" come to mind, the theoretical argument remains that "pessimists" argue that the specification of a regression model that satisfies aggregability conditions for reliable and valid welfare analyses *from* a hedonic price model is largely unattainable. The "pessimists" appeal to strong heterogeneity, not only of household preferences, but of real estate goods as well, as each housing unit is different from any other. To them, aggregation falls on the "one house, one type" assumption. So, what can be done to appease the extremists in both camps?

Tiebout (1956) suggests, in the extreme, that each citizen-type can find the jurisdiction or community that exactly satisfies his/her preferences if the number of jurisdictions and households is infinite. His examples of jurisdictions with expenditures used to purchase private goods only and public goods only show the extreme positions on the continuum of aggregation. With an infinite number of jurisdictions, all combination bundles of local government expenditures can be made available to an infinite number of consumers able to express effective demand for particular bundles. In the original Tiebout (1956) construction, some households will be "discontented with the patterns of their community" except when the system is in equilibrium. But, with only a finite number of jurisdictions available, the decision of a particular household to reside in a particular neighborhood and a particular dwelling within that neighborhood may not be ideal. In this situation, a heterogeneous class of consumers may inhabit the same neighborhood and differentiate themselves on the intensity with which they consume different local amenities, different retail purchases, and different bundles of dwelling structural characteristics. Then, if patterns among these households can be revealed, then the common assumption of a singular type neighborhood, school district, or other classification of physical space is not sufficient to permit perfect aggregation of households for analysis purposes.

Recently, Ekeland, Heckman, and Nesheim (2002) argue that the hedonic identification problem (which I call the aggregation problem) can be mitigated *theoretically* to permit the full aggregation of households into a single type. In this paper, I argue that a *statistically* based method is also consistent with theory and can be used to determine the number of different household "types" that co-exist in a single space. This classification scheme mitigates the aggregation problem that plagues hedonic analyses by creating a model that accounts for heterogeneous preferences among households that may share a uniform array of local public services. This is accomplished by sorting households into "types" based on their demographic characteristics and their attitudes toward neighborhood features.

This sorting process presented here is not equivalent to sorting based on preference heterogeneity; it is merely a way to determine if households, through a systematic analysis of their demographic and attitudinal characteristics, are systematically different in multiple dimensions. If more than one "type" is determined by the model, then it may be possible that multiple hedonic price lines exist in the same neighborhood. A deeper probe into this question will be the subject of future research. Nevertheless, after the initial sorting process, I filter these potential differences through implicit price vectors in an iterated hedonic framework to determine which subgroups are statistically distinct based on their dwelling purchase choices.

3.2 Theoretically Justified Aggregation

The research question guides one's empirics. Past research questions in hedonics have focused on the estimation of marginal implicit prices (and sometimes the demand for certain dwelling or neighborhood attributes) across a "community." In this research, I ask a slightly different question – is it possible that each household type could have its own hedonic marginal price line if the data suggest that systematically different households live in the neighborhood? Others have been interested in across-community diversity as they assume *a priori* that all households in a particular neighborhood have homogeneous preference orderings; these include Sieg *et al.* (2000), Murty, Gulati, and Banerjee (2003), and others. While the possibility of multiple hedonic price lines in a single community is debatable (even though it seems on the surface to be a natural extension of the market segmentation argument some make at the city level), I proceed to show in the neighborhood under study that a typology of household categories is still orderly; each distinguishable group of households is aggregable, which permits the estimation of plausible within-group welfare change measures (please see Chapter 5).

It is easy to see that the choice of residence simultaneously determines the distances that each household must travel in order to access all goods, including private consumption (eggs, milk, apples, etc.) and non-private consumption (green space, national parks, etc.). Then, depending on the mode of transportation chosen by each household to access these goods, the costs of traveling to the sites where these goods can be purchased (or consumed) might be different for each household. That the household location decision simultaneously determines the distances and costs of accessing private and non-private goods suggests that different households in different geographical areas pay different prices for the same types of goods. But, in the single neighborhood modeled here, it is reasonable that (on a more macro scale) each household approximately travels the same distances to access groceries, entertainment, work, etc. However, *within* this neighborhood, the distances each household must travel to certain neighborhood features (like the Home Park green space or to the Georgia Tech campus) vary greatly, particularly when one considers the various modes of transportation available to local residents (students tend to walk to school whereas full-time employees tend to drive to work) and the time required to travel those distances. This suggests that households will react differently (i.e. have different marginal price reactions) to structural changes in the local land use (altering the land use designations of particular parcels), changes to the transportation infrastructure, and others. But, if these differences among households are similar enough in many dimensions, then it is possible that each household can be aggregated into groups with other households that have similar marginal prices.

To initially describe the aggregation problem, I start with a household utility maximization model (with a Cobb-Douglas utility specification) similar to that described in Farmer (2003) in which the arguments of the utility function are assumed to be weakly separable (the marginal rate of substitution between two arguments is independent of the quantities of all other goods).[13]

(1)

$$
\begin{aligned}
&MaxU_i(\mathbf{x}_c, \mathbf{A}_j, s_H, L) = \alpha_i \ln \mathbf{x}_c + \beta_i \ln \mathbf{A}_j + \gamma_i \ln s_H + \delta_i \ln L \\
&st \\
&\mathbf{w}_{sl}\mathbf{I}_{sl} + \mathbf{w}_{usl}\mathbf{I}_{usl} + S_i = \mathbf{p}_c \mathbf{x}_{ci} + R_H\left[s_{Hi}, \mathbf{c}(\mathbf{x}_{ci}, \mathbf{A}_{ji}, I_J), \mathbf{d}(\mathbf{x}_{ci}, \mathbf{A}_{ji}, I_J)\right] \\
&T = L_i + \sum_{J=(sl,usl)} I_J + \sum_{J=(sl,usl)} t_J d_J + \sum_{x=1}^{n} t_{\mathbf{x}} d_{\mathbf{x}i} * \mathbf{x}_{ci} + \sum_{A=1}^{m} t_{\mathbf{A}} d_{\mathbf{A}i} * \mathbf{A}_{ji} + \sum_{M=1}^{q} t_{Mi} d_{Mi} * f(\mathbf{x}_{ci}, \mathbf{A}_{ji}, I_J)
\end{aligned}
$$

This model says that households maximize utility (satisfaction and happiness) through the consumption of four items: 1) consumption other than housing space and neighborhood amenities [$\mathbf{x}_c$]; 2) cultural and recreational amenities located outside the neighborhood that are not capitalized into the local housing market [$\mathbf{A}_j$]; 3) housing space [s_H] from which households receive services (including the monetary impacts of local amenities, such as sidewalk quality and proximity to open space); and 4) leisure time [L][14].

The two main constraints in the model are income (endogenously determined in the model even though skill level is exogenously determined) and time (exogenous to the model). In the income constraint, I assume that skilled-worker households [denoted by the "sl" subscript] and unskilled-worker households [denoted by the "usl" subscript] work different amounts of time [$\mathbf{I} = (I_{sl}, I_{usl})$] for different wage rates [$\mathbf{w} = (w_{sl}, w_{usl})$]. This total household income [**wI**] is used to purchase a certain amount of housing space,

---

[13] In this model, the assumption is made that $\alpha = \beta = \gamma = \delta = 1$, which assumes increasing returns to scale.

amenities, consumables, and travel-related goods (car, gasoline, bus tickets, etc.). In the case of retired persons and students who may not earn income in the same way as other households, S_i denotes saved monies (savings from work done in previous periods, college loans, Social Security, other endowments, etc.) used to purchase consumable goods, purchase housing space, travel to various neighborhood amenities, shopping opportunities, etc., and purchase neighborhood amenities. On the right hand side of the income constraint, households pay a price $\mathbf{p}_c$ for goods consumed $\mathbf{x}_{ci}$ which include non-local, out-of-neighborhood amenities for which households pay user or entrance fees. They also pay a rental price R_H for a dwelling which is a function of square feet of housing space s_H, a vector of costs $\mathbf{c}$ associated with traveling distances $\mathbf{d}$ to consume goods, to consume within-neighborhood amenities $\mathbf{A}_{ji}$, and to work; R_H varies depending upon the intensity with which each $\mathbf{A}_{ji}$ is consumed or "accessed."

In the time constraint, each household has only 24 hours a day in which to consume leisure L_i; to work, to commute between work and home, to purchase goods, to consume amenities, and to make multiple purpose trips, respectively. $\mathbf{t}$ contains the times required to travel distances $\mathbf{d}$ at costs of $\mathbf{c}$ per unit of distance.[15]

Next, the Lagrangian formulation of the utility maximization model is differentiated with respect to each argument in the utility function as well as the shadow prices to yield the first-order conditions. Then, a rearrangement of the first-order

---

[14] Anas, Arnott, and Small (1998), in their synthesis of the urban economics literature, argue that more realistic models of urban areas require the derivation of the shadow value of time endogenously by "adding leisure and a time budget to the model" (p. 1436). This is consistent with the model presented here.

[15] For simplicity, this paper assumes that trips to consume goods other than housing and/or amenities *on the way to or on the way from work* are negligible and will not be included in either constraint in the econometric model. Also, $\mathbf{c}_i$ vary because I assume that the monetary cost per unit of distance for short trips is greater than the cost per unit of distance for longer distance trips. An example is the cost per unit of

conditions yields the Marshallian demands, which are used to derive the indirect utility function (IUF)[16]

(2)

$$V(\mathbf{p}_c, R_H[c_J, c_{\mathrm{x}}, c_{\mathrm{A}}, c_{Mi}, t_J, t_{\mathrm{x}}, t_{\mathrm{A}}, t_{Mi}, d_J, d_{\mathrm{x}}, d_{\mathrm{A}}, d_{Mi}], \mathbf{wI}; T, S)$$

Equation 2 tells me that 1) utility is not aspatial and that 2) the purchase or rental of a dwelling reflects the simultaneous determination of a household's driving/walking times to work, to shopping, to recreation, to entertainment, to health care services, and to others; noise from the nearest major road; and others. In other words, dwellings sales or rental prices reflect the many distances and thus times and costs associated with traveling those distances that are "chosen" by the household location decision.[17]

Income in this framework is not wholly exogenous; this is an important factor for non-professional urban residents in particular. Where one lives sets in motion all sorts of constraints on retail purchases, on recreation, and on available groceries. These constraints of various travel distances affect many choices; even income tends to tip the scale in favor of Tiebout-like similarity where residents, choosing from an array of similar overlapping constraints, tend to cluster vis-à-vis more extreme Samuelson-like markets where residents have very little in common beyond a few local amenities. But these households, while they may have relatively similar incomes, cannot be

distance for a single-purpose trip to the local supermarket during rush-hour traffic versus the cost for a multiple-purpose trip to the hair salon, grocery store, and the bank.

[16] The specification of the IUF is consistent with the specifications of other IUFs that include the prices of housing services and the prices of composite goods (Rapaport 1997), income (Ibid., Haab and Hicks 1997, Sieg *et al.* 2000), household and community characteristics (Rapaport 1997 and Chattopadhyay 2000), and public and private goods (Rapaport 1997 and Sieg *et al.* 2000).

[17] Individuals who maximize utility will rearrange their purchases of a particular commodity or urban amenity until the marginal rate of substitution (or the slope of an indifference curve) between a composite commodity and each desired characteristic of a dwelling is equal to the implicit price of that characteristic (Hanley and Spash 1993).

distinguished on a single dimension; this is the mistake made and realized by Rosen (1974). Therefore, to model a neighborhood that questions the common "one neighborhood, one type" assumption, I distinguish households into subgroups to segment the housing market.

Segmentation of the housing market follows Freeman (1993) who says that either supply or demand (or both) have to vary across market segments for multiple hedonic price lines to exist. Similarly, Freeman argues that so-called "barriers to mobility" that prevents arbitrage must exist so that one submarket is distinctly different from another. From the list of potential barriers that may segment a single market, Home Park has several seemingly important "barriers" in racial composition, geography, and income.

The question of how a household can be excluded from participation in other markets (and potentially purchase adequate housing at a lower cost) is paramount to the concept of segmented markets. In Home Park, two distinct ethnic enclaves (Muslims and Southeast Asians) and an emerging third (Hispanics) live close together, reflecting the importance of having a "community" in a potentially "foreign" country. This poses potential limitations on the housing stock that is available to households looking to locate to Home Park. If race is a systematic barrier that prevents otherwise qualified households from purchasing in the neighborhood, then the housing market would be segmented by race.

However, this is not likely the only barrier that segments the housing market. Other barriers include geography and income. From discussions with neighborhood residents, there seems to be a mental barrier that separates those who live north of 14th Street (North Home Park) from those who live south of 14th Street (South Home Park).

As an example, those who live in North Home Park and walk their pets do not seem to cross into South Home Park on a consistent basis. As I explain in Chapter 5, this phenomenon seems to favor the conversion of a vacant parcel in Home Park to a green space opportunity for North Home Park residents. And, for those who choose to cross the 14th Street "barrier," two green space opportunities may be available.

As for income as a barrier that suggests market segmentation, renters typically have a certain kind of dwelling in mind given their budget constraints. So, to argue generally that renters and owners have access to the same kinds of dwellings is misguided. While it is possible for some renters and owners to be "in the market" for the same kinds of dwellings, income is likely the main barrier that determines the kinds of dwellings available to these households. Given these barriers that suggest market segmentation and the IUF as specified, the possibility exists that each household may face a different implicit marginal price for each good, including common market goods like food and clothing. Then, if households exhibit similar implicit prices for the same dwelling characteristics and local amenities, it may be possible to specify more than one hedonic price surface in a single neighborhood.

So, what if a city planner assumes Tiebout-like homogeneity in her work at the neighborhood level (which translates to $V_{neighborhood}$ for full aggregation) when actually households can be distinguished into different subgroups that are more homogeneous than the entire sample of households (somewhere between $V_{neighborhood}$ and V_i)? This is the essence of the aggregation problem: how can unique households, which have unique preferences for unique housing bundles, be modeled in a manner that implicitly assumes that all households are similar enough in some way to be perfectly aggregated into a single group? The answer may be "by just assuming that IUFs are identical for each household in a given 'type' if my sorting process tells me so." If this is the case, then the complexity of the housing purchase decision as a complex hedonic good lends support to both aggregation theories: perfect neighborhood aggregability a la Tiebout (1956) at one end of the continuum requires the researcher to have full information on households, their purchasing behaviors, their travel habits, etc.; conversely, impenetrable disaggregation a la Samuelson (1954) at the other end requires little information as all households are different and must be modeled differently. Figure 2 illustrates the role that information plays in these various degrees of aggregation.

Extreme Samuelson (1954)	Extreme Tiebout (1956)
No functional form problems	Large functional form problems
No information learned	Full information learned
No household aggregability	Full household aggregability

Figure 2: The Role of Information in Household Aggregation

3.3 Identifying Household "Types"

3.3.1 Data Requirements and Estimators

In the sorting process to be described later, households only face the exact same set of marginal prices if they are sorted into the same type. If this occurs, then a single type will emerge from the sorting process. While it is possible for all households to face the same marginal prices for all independent variables in the hedonic regression model, my intuition tells me that households that systematically differ in multiple dimensions will have different marginal prices for the majority of variables in the model.

The next step, then, is to devise a method that permits one to use the detailed data obtained from a household survey to aggregate households into "types," allowing for the possibility that all households *could* sort into a single type that requires single-line equation analysis (for example, OLS or GLS models). Here I describe the process used to find a common ground between these two extremes of the aggregability continuum so that households can be sorted into statistically aggregable "types."

Location theory tells me that households will locate to jurisdictions that have the best set of amenities and public goods among all alternative jurisdictions; this location decision, though, does not guarantee a perfect expression of preference orderings over all goods and amenities. But, if "the provision of local public goods mirror[s] that of an efficient private market", then households sort efficiently into jurisdictions (Tiebout 1956).[18] This suggests that when households have more neighborhoods from which to choose, they can choose simultaneously the dwelling and neighborhood that provide the best bundle of characteristics, services, features, amenities, etc. over all others.

Housing hedonic studies that seek to estimate marginal implicit prices for a "community" generally have no concerns about aggregation, as they are only interested in marginal prices across a set of households regardless of their differences in tastes, which are handled econometrically by corrections for heteroskedasticity or spatial

---

[18] Goodman (1978) found significant differences in the mean attribute levels and mean attribute effects across a single metropolitan area (New Haven, Connecticut Metropolitan Statistical Area), which supports the existence of spatially segmented housing markets within a single area in the Tiebout tradition.

weights matrix approaches. However, if I assume that households in a given study area (neighborhood, school district, tax district, etc.) are *not* similar enough to fully aggregate into a single group for analysis purposes, what am I supposed to do? Forego my theory to satisfy those researchers who ask a different question? The answer is to identify the groups of households, identify the data requirements for aggregating households, and choose an appropriate estimator.

First, to avoid the problems associated with *a priori* defined groups, household "types" are defined endogenously through a sorting process. Defining types endogenously permits households to sort themselves into groups based on the data instead of arbitrary classification strategies by the researcher(s). Second, the data required for this exercise are much more specific than those required for earlier hedonic studies that assume that households have homogeneous preferences and/or that they live in the same neighborhood or school district. The data required to distinguish households include households' demographic characteristics and attitudes toward certain neighborhood features (data that Tiebout said could distinguish one jurisdiction from another), which are used to support the claim that the location of a particular type of household to a particular jurisdiction (or neighborhood) does not mean automatically that all households in that neighborhood can be reasonably described by a single rubric, nor that all households can be aggregated into a single citizen-type (or "consumer-voter" according to Tiebout) for analysis purposes.

This study has relatively more detailed demographic and attitudinal information than any other study that I found. The information is enough to permit the aggregation of subgroups, but not enough to observe the implicit time and distance variables that would allow different types to be fully aggregated, *even if* all utility functions obeyed log-linear substitutability. The data include variables that are important to the household location decision, such as dwelling square footage, the number of bedrooms, the number of acres, etc., as well as the distances to various local amenities, schools, grocery stores, and shopping opportunities. If one remembers that the household location decision affects the implicit prices of all consumable goods, how one spends her leisure time, the costs of enjoying various amenities, etc., she might argue that households with similar implicit prices for these goods and amenities will cluster together or share the approximate same neighborhood space. Most likely, though, households choose to live in a particular neighborhood for the amenities found there, the proximity to one's workplace, and for the neighbors who are "more like them" than in other neighborhoods. So, within a single neighborhood, my intuition tells me that it is reasonable to expect that households with similar demographic compositions are likely to have similar attitudes toward the same neighborhood amenities and similar preferences for leisure time, goods consumed, local amenities, and housing space. While I have no formal test for this assertion, my intuition would be disproved by a result that suggests the prevalence of a *single* household type in the neighborhood. In any case, the coefficient estimates derived from a regression model that is the empirical translation of a theoretically consistent market segmentation exercise for the Home Park housing market will be efficient.

Third, once household types have been defined and data requirements identified, one must choose an appropriate estimator. In the case of a single household type, OLS (or some GLS correction for heteroskedasticity) will be the appropriate estimator. In the case of multiple types, the location of these different types to individual neighborhood parcels suggests competition among them for the finite space in the neighborhood. This competition between household types for various parcels is emblematic of the Seemingly Unrelated Regression (SUR) estimator identified by Zellner (1962). The SUR estimator is $\hat{\beta}^*(\hat{\Sigma}^*)=(\mathbf{X}^{*\prime}\hat{\Sigma}^{*-1}\mathbf{X}^*)^{-1}\mathbf{X}^{*\prime}\hat{\Sigma}^{*-1}\mathbf{y}^*$, where $\hat{\beta}^*$ is the estimated coefficient vector, $\hat{\Sigma}^*$ is the estimated error variance-covariance matrix, $\mathbf{X}^*$ is the matrix of independent variables, and $\mathbf{y}^*$ is the vector of observations on the dependent variable. Feasible GLS estimation proceeds as the error variance-covariance matrix $\hat{\Sigma}^*$ is estimated using OLS and then used as a weight to obtain the coefficient vector β^* (Greene 2003, p. 342). Then, to determine whether the SUR estimator realizes more efficiency gains than equation-by-equation OLS, the off-diagonal elements of the error variance-covariance matrix are tested for their statistical difference from zero with a Breusch-Pagan test which, if significant, means that the off-diagonal elements are indeed non-zero and suggests that the SUR estimator is more efficient than OLS. Without a lot of information about the interrelationships between types, SUR has the advantage of correcting for an unmodelled relationship between equation estimates, which makes it the most efficient estimator of the two. In fact, as the correlations between the independent variables get stronger, the SUR estimator provides ever more efficient estimates than equation-by-equation OLS.

Two examples illustrate the uses of SUR. The first example appears in Zellner (1962) and is refined by Kmenta (1986). In a monopolistically competitive market of differentiated goods, the supply decision of one firm is conditioned on the supply movements of other distinct and imperfect substitutes. The same technology factors may affect production at each firm, but the individual firm's choice to accelerate production requires that the movements in the other firms be tracked simultaneously. Without the full *system* estimation of the demand for each different product similar to that described in Intriligator (1978), or a key substitution parameter provided exogenously (Krugman and Smith 1994), the relationship between output decision choices does not require (or meet the rank conditions for) a full simultaneous systems estimation, yet the competitive conditions between the purchasers mean that these seemingly independent choices are indeed related.

A second example is food demand systems where weakly separable groups of food items compete for scarce budget resources. For example, fish, pork, beef, and chicken serve as good types in a protein group that is considered weakly separable from, say, food staples. Each demand equation appears independent as it responds to its own price, to total expenditures over the group, and to the cross-prices of close substitutes (other food items in the group). But, the unexplained portion of each regression equation is related to the unexplained portion of every other equation. Similarly, in the housing market modeled here, the household types active in the market individually demand a set of attributes that makes a given space attractive. Yet as types bid for the same space, competition dictates that the unexplained portion of revealed choice to purchase a given parcel at a given price is related to the unexplained portion of revealed choices by the other active consumer types in that market. So, the competition for space among different household types indicates that the SUR method of system estimation is appropriate for this exercise.

To my knowledge, this is the first paper that uses detailed household level data to identify statistically aggregable subgroups of households from the sample and then to estimate a marginal price schedule for each household type.[19] The data on Home Park residents show that very different kinds of households (in terms of income, why they moved to the area, etc.) live in the neighborhood and share a bundle of attributes that each values differently. So, I argue that different types of households can reside in the same jurisdiction, neighborhood, or other defined spatial area, each with a certain set of preferences for certain neighborhood features that entice them to locate to the neighborhood. To determine whether different household types exist in the same neighborhood space, a sorting process must be used to distinguish one "type" of household from another. I do not claim *a priori* knowledge of the number of household types that are statistically aggregable in this neighborhood, so I limit the amount of bias introduced into this process by iteratively estimating the number of endogenous household types in the neighborhood.

3.3.2 The Iterative Household Sorting Process

The challenge to endogenize the definition of a household type flows from the aggregation problem itself. Rosen (1974) holds in the limit that households cluster or organize into distinct entities of *one element only*. Even if many households have similar preferences and choose to purchase relatively similar houses, the *vector* of characteristics that induce the unique household to purchase its individual dwelling is unique to that household. Unlike uniform market goods, each dwelling is distinct and will, if sold, go to

[19] The model presented by Hoyt and Rosenthal (1997) is close to what we model here. Their data from the American Housing Survey, however, do not permit them to determine whether households sort efficiently at the community level, but only at the street level. Despite this, their sample is comprised of clusters around particular houses, not of households across a single contiguous space.

the household offering the highest bid for the particular bundle of characteristics possessed by that dwelling. If a class of buyers that shares many common preferences pursues an array of similar houses, then sorting household preferences (to the particular houses best suited to those preferences) becomes a restatement of the aggregation problem itself, as *similar* households with *similar* preferences pursue *similar* bundles of housing characteristics. In this way, the problem of full aggregation can be reduced if households can be sorted into types.

To start the household sorting process, I assume that groups of households in Home Park have similar enough characteristics and attitudes to be aggregated into discrete and finitely numbered "types". Just as Tiebout (1956) proposes and Goodman (1978) empirically supports that homogeneous households would sort themselves into jurisdictions based on sets of public goods, I argue at the intra-neighborhood level that different types of statistically aggregable households can locate into a single neighborhood, which can be perceived to contradict Tiebout (1956). I speculate that some households will value positively a local amenity viewed by others as negatively related to sales price, and vice versa, but that as a bundle, enough similarities exist to sort households into statistically aggregable types. I assume that aggregation problems do not exist within types, implying common demand functions among households of a particular type.

To begin, I have a large set of potential variables (75 of them) to explain sales price variation, the ultimate goal of this exercise. Given the sample size (n = 400), the use of all 75 variables to explain sales price variation violates the minimum sample size requirements determined by Griffiths *et al.* (2002). Plus, I do not have a strong theoretical reason for including the demographic and attitudinal variables as explanatory variables for sales price. So, to initiate the sorting process, I use principal components analysis (PCA) to reduce the dimensionality of the data set and identify "groups" of households with similarities in demographic characteristics and attitudes.[20] PCA creates factors (subsets of a larger set of variables) of the demographic and attitudinal variables found in Table 1 that explain significant amounts of the correlation between variables, or "clustering" among demographic and attitudinal variables.[21] Formally, the first principal component y_1 is a linear combination of the set of variables $\mathbf{x_i}$ $\left(y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1p}x_p = \sum_{i=1}^{p} a_{1i}x_i \right)$ "such that the variance of the created factor is maximized given the constraint that the sum of the squared weights is equal to one" $\left(\sum_{i=1}^{p} a_{1i}^2 = 1 \right)$ (Dunteman 1989). Otherwise, the variance $\left(\sum_{i=1}^{p} \sum_{j=1}^{p} a_i a_j \sigma_{ij} \right)$ of the linear composite $\left(\sum_{i=1}^{p} a_i x_i \right)$, where σ_{ij} is the covariance between the i^{th} and j^{th} variables, is made artificially large by the selection of large weights (Ibid.). In other words, PCA locates the weight vector (a) that maximizes the variance of the linear composite given

---

[20] Other options to initiate the sorting process include factor analysis (also known as cluster analysis to some) or logit models. Factor analysis has a key underlying assumption "that the i^{th} variable in the variable set x_i can be expressed as a linear combination of hypothetical unobservable common factors plus a factor unique to that variable" (Dunteman 1989, p. 56). This, coupled with the fact that principal components analysis focuses on the decomposition of the total variance of a set of variables, is why factor analysis is not used. As for logit models, usually employed in the analysis of time-series data, the dependent variable would be "types" (yes, no – are you classified as a Type A household?), which would be updated through the sorting process *instead* of households sorting based on similarities in implicit prices for dwelling structure characteristics and neighborhood features.

[21] PCA is used regularly in other literatures (e.g. see Huettman and Diamond [2001] for an example of PCA used to classify bird species by characteristics of birds). Generally, though, economists believe the use of this technique is too arbitrary. Yet a search of EconLit on 3/5/2003 found 162 articles that use the technique, 50 of them since Jan. 2000.

the constraint that all squared weights must equal one. In this way I explain the most amount of covariance or correlation between the variables using the fewest number of components.

Table 1: Variable List

Demographic and attitudinal variables (from housing survey) used in the PCA:
Total Household Income (IV); multinomial
Race of the respondent (IV); dichotomous (0=non-white, 1=white)
Number of adults in the household (IV); interval
Number of children in the household (IV); interval
Education level of the respondent (IV); multinomial
Age of the survey respondent (IV); continuous
Sex of the survey respondent (IV); dichotomous (0=female, 1=male)
Is the house owner-occupied? (IV); dichotomous (0=no, 1=yes)
Is some part of the house rented? (IV); dichotomous (0=no, 1=yes)
"What particular features attracted you to Home Park?" *(Please see Appendix A for the complete list of questions.)*

Dwelling Characteristics (from housing survey and Metropolitan Atlanta Multiple Listing Service) used in the SUR:
Rental/Selling price of house (DV); continuous in unadjusted dollars
Square footage of the house (IV); continuous in square feet
Number of bedrooms (IV); discrete
Number of baths (IV); discrete
Year of last sale (IV); discrete;
Age of the structure (IV); continuous in years
Is the dwelling at or above street level? (IV); dichotomous

Spatial features (from GIS) used in the SUR:
Road network distance to nearest brown industry (IV); continuous in meters
Road network distance to child care facility (IV); continuous in meters
Road network distance to nearest church (IV); continuous in meters
Road network distance to Home Park, Piedmont Park) and playground (IV); continuous in meters
Road network distance to Piedmont Park (IV); continuous in meters
Road network distance to nearest elementary school (IV); continuous in meters
Road network distance to local convenience and ethnic grocery stores (IV); continuous in meters
Road network distance to nearest retail opportunity (IV); continuous in meters
Road network distance to nearest commercial opportunity (IV); continuous in meters
Road network distance to nearest public transportation bus stop (IV); continuous in meters
Road network distance to Georgia Tech (IV); continuous in meters
Road network distance to Feminist Women's Center (IV); continuous in meters

Adjacency Variables (Yes or No dichotomous variables created from GIS) used in the SUR:
"Do you live adjacent to a renter?"
"Do you live adjacent to a homeowner?"
"Do you live adjacent to a student?"
"Do you live adjacent to a college graduate?"
"Do you live adjacent to a household that wishes to move in the next two years?"
"Do you live adjacent to a household that wishes not to move in the next two years?"
"Do you live adjacent to a household of a different race than your own?"
"Do you live adjacent to a household that has made home improvements in the last two years?"
"Do you live adjacent to a household that chose Home Park for its proximity to Georgia Tech?"

My *a priori* contention is that households with similar demographic characteristics and similar attitudes toward the same neighborhood features are likely to have similar implicit prices for the same neighborhood amenities, which means that implicit prices will be more efficient than those estimated via OLS given that the households in each type are statistically aggregable and the true prices differ between types' markets. Groups determined using PCA allow the observations to sort into groups with other households of demographic and attitudinal "similarity," as will become apparent throughout the sorting process. So, from the PCA, 24 principal components or factors f_i are returned, meaning that 24 different orthogonal linear combinations of the variables explain correlation clusters among the set of variables.[22] Included in these variables are some demand shift variables like income and the number of children; some choice variables like education level and whether or not to be a renter or an owner; and others. One of the caveats of using all of these variables is that the results might differ by using the non-choice variables only. Preliminary runs of the PCA model indicate that the number of demographic and attitudinal variables used caused some but not discouraging variation in the eigenvalues of the top eight factors.

The choice of the number of components to use is debatable. The concern is that having too many groups of households prevents us from obtaining useful information with such a small sample size (n = 400). Using the Kaiser (1960) criterion, all principal components with an eigenvalue, λ, greater than or equal to one would be retained; using the Jolliffe (1972) criterion, factors with $\lambda \geq .7$ would be retained; and using the Cattell

[22] An equal number of components is created as there are variables. If a variable v_1 can be expressed as a linear combination of other variables (v_2, v_3, v_4) in the data set, only then is the dimensionality reduced by the number of variables on the righthand side (3).

(1966) criterion, a "scree" graph of the eigenvalues is used to determine arbitrarily the "steepness" of the line that runs between two graphed eigenvalues, with "steeper" lines corresponding to factors that are retained. To select the number of factors to retain, I use the eigenvalues and eigenvectors of the correlation matrix of demographic and attitudinal variables along with Figure 3, a "scree" graph that displays the eigenvalues for the first 14 principal components. Since eight factors have eigenvalues greater than one and explain 65.4% of the variance of the total set of variables, *and* since the additional variance explained by each subsequent factor tends to dissipate between components 7 and 10, I feel confident in drawing a cut-off point at $\lambda \geq 1$. These principal components and their factor loadings are detailed in Tables 2a and 2b.[23]

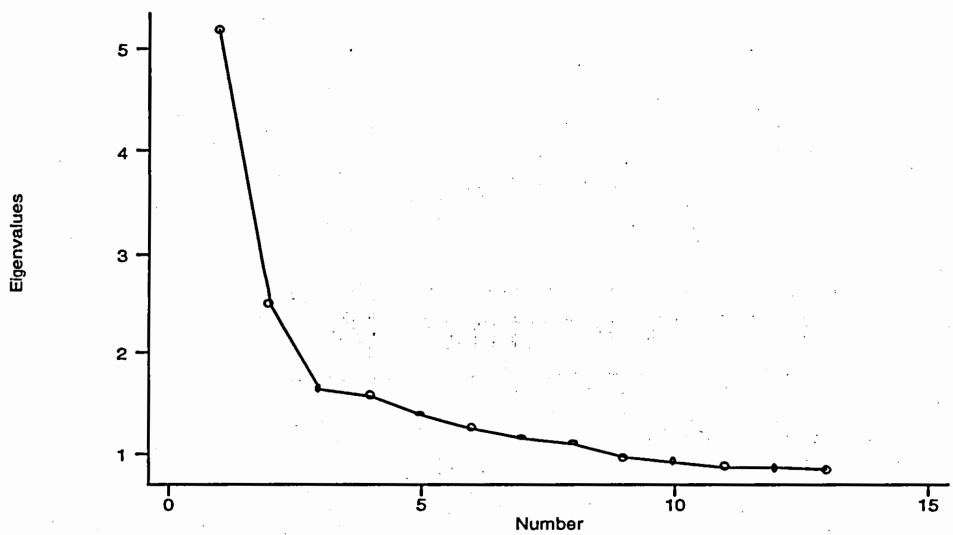


Figure 3 – Graph of Eigenvalues

[23] Please note that the λ used here to model eigenvalues is not the same λ used to illustrate the shadow prices in the utility maximization model.

Table 2a – Statistics for the Principal Components

Component	Eigenvalue	Proportion of Variance Explained	Cumulative Variance Explained	Mean (Standard Deviation)	R^2 value in SUR model
1	*5.17*	*.215*	*.215*	*6.24e-10 (2.27)*	*.07*
2	*2.46*	*.102*	*.318*	*-2.39e-09 (1.57)*	*.15*
3	*1.63*	*.067*	*.386*	*-1.53e-09 (1.27)*	*.07*
4	*1.56*	*.065*	*.451*	*-1.16e-09 (1.25)*	*.08*
5	*1.37*	*.057*	*.508*	*-1.08e-09 (1.17)*	*.12*
6	*1.24*	*.052*	*.560*	*-1.95e-09 (1.11)*	*.05*
7	*1.14*	*.047*	*.608*	*3.03e-09 (1.07)*	*.11*
8	*1.09*	*.045*	*.654*	*8.50e-10 (1.04)*	*.08*

Table 2b – Eigenvectors for Top 8 Principal Components

	1	2	*3*	*4*	*5*	*6*	7	*8*
Renter/Owner	*.33*	*.02*	*.08*	*.06*	*.05*	*.19*	*.18*	*.08*
#Adults	*-.27*	*.04*	*-.04*	*-.08*	*.33*	*.39*	*-.01*	*.09*
#Children	*.09*	*-.11*	*.27*	*.06*	*.17*	*.19*	*-.30*	*-.07*
#Adults Work	*-.17*	*-.12*	*-.03*	*-.05*	*.47*	*.42*	*-.00*	*.19*
Respondent's Age	*.32*	*.30*	*.02*	*.03*	*.10*	*-.09*	*-.07*	*.06*
Respondent's Sex	*-.02*	*-.06*	*.11*	*.14*	*.21*	*-.12*	*.16*	*.54*
Dwelling Tenure	*.23*	*.41*	*.01*	*.04*	*.04*	*.09*	*-.06*	*.01*
Race	*.21*	*-.00*	*-.24*	*-.14*	*-.20*	*.17*	*-.15*	*.19*
Employed Full Time	*.29*	*-.35*	*-.01*	*-.02*	*.00*	*.05*	*-.10*	*-.01*
Student	*-.38*	*.10*	*-.00*	*.01*	*-.04*	*.00*	*.08*	*-.04*
Retired	*.14*	*.40*	*.04*	*.08*	*.08*	*-.16*	*.05*	*.18*
Education Level	*.02*	*-.27*	*.27*	*.20*	*.27*	*-.28*	*.10*	*-.08*
Income	*.32*	*-.26*	*.10*	*.10*	*.08*	*-.00*	*.09*	*.11*
Crime	*.00*	*-.00*	*.35*	*-.64*	*-.06*	*-.10*	*.23*	*.17*
Environment	*.06*	*-.11*	*-.61*	*.00*	*.24*	*-.11*	*.05*	*-.23*
Education Issue	*-.10*	*-.07*	*.18*	*.59*	*-.31*	*.15*	*-.17*	*.18*
Social Security	*.07*	*.32*	*.07*	*.10*	*.27*	*.12*	*-.22*	*-.24*
Politics	*-.00*	*-.09*	*-.04*	*.11*	*.35*	*-.37*	*.11*	*-.00*
Close to GT	*-.34*	*.02*	*.12*	*.06*	*-.05*	*-.05*	*.01*	*-.01*
Investment Potential	*.21*	*-.02*	*-.06*	*.08*	*-.00*	*.34*	*.45*	*.03*
Tree Cover	*-.02*	*-.03*	*-.35*	*.02*	*.01*	*-.22*	*-.26*	*.48*
Near Park	*.04*	*-.11*	*-.10*	*.17*	*-.15*	*.10*	*.38*	*-.24*
Near MARTA	*-.02*	*.23*	*.10*	*.13*	*.18*	*-.14*	*.16*	*-.17*
Near Work	*.13*	*-.22*	*.18*	*-.15*	*.09*	*.04*	*-.40*	*-.21*

Using these eight factors, each household is assigned to the group that best correlates the demographic and attitudinal variables for household j to the factor f_i. The process assigns each household into a group G based on the following statistic: $G = i$ for $Min\left|\frac{f_i - \bar{f}_i}{s_i}\right|$, where $i = 1, 2, \ldots, 8$. The statistic is the absolute value of the normalized difference between the actual and the mean factor scores for each observation, where s_i is the standard deviation from the mean. This initial assignment is the starting point for further sorting into a typology of hedonic demands.

However, the fact that eight different significant orthogonal principal components emerge tells us nothing more than eight different linear combinations of the demographic and attitudinal variables explain significant amounts of the covariance among these variables. As a second step used to determine the number of household "types" that live in Home Park, the additional information that I have on households (sales price, structural characteristics, local amenities, etc.) is used in a seemingly unrelated regression (SUR) model that accounts for information asymmetry (households looking to locate to Home Park compete for the same space, but may not have full information on local amenities or dwellings for rent or sale, which might reduce their perceived feasible set of alternative housing sites), demand shocks that affect all households regardless of type, and correlated error terms between the household types.

The aggregation problem alluded to earlier is not resolved simply by explaining the covariance among a set of demographic and attitudinal variables. The goal here is to explain dwelling sales price variation among the groups of households in an alternative way. To accomplish that goal, households are allowed to sort into "groups" via PCA and then iteratively re-sort into "types" based on the SUR model, so that households with similar beta coefficient vectors (which reflect similar marginal prices for dwelling characteristics and neighborhood features) are sorted into the same type. What this suggests is a segmentation of the Home Park housing market in such a way that multiple hedonic price surfaces exist in a single neighborhood. Given the previous discussion on the competition for space among households, it seems possible that a given space, say a city, can accommodate different types of households. Then, the random selection of a neighborhood from that city will reveal different sets of households that pay different amounts implicitly for certain amenities provided publicly by the local government.

But, Tiebout (1956) homogeneity suggests that these households will be similar, but gives no indication of what "similar" means – does it mean similar in demographic terms, or similar in attitudes? It is likely that these households consume different amenities in the area with different intensities – some will walk their pets everyday, utilizing the city streets, sidewalks, and green space in the process, whereas others may recreate in the green space once a month. The point is that his assumption of homogeneity at the jurisdictional level never can be formally tested, just like I do not have enough detailed data to determine the intensity with which households consume local amenities. So, to extend theory, some assumptions must be made regarding the meaning of homogeneity and similarity. This paper assumes that "similar" households will have similar demographic characteristics, have similar attitudes toward certain neighborhood features, and have similar levels of consumption intensity of local amenities. If this occurs, then a first attempt to tease out the degree of similarity (in multiple dimensions) among a set of households in a particular neighborhood can be accomplished.

Some of the requirements for segmented markets across multiple cities include different households in different markets (Palmquist 1984, Freeman 1993). This, along with a utility restriction (Ekeland, Heckman, and Nesheim 2002) or a segmentation of the cost function (Freeman 1995), are the usual requirements for the identification of the amenity demand function. In this paper, I assume that multiple markets exist (given the "barriers" mentioned earlier in this chapter) and can be identified in a separate analysis before the first-stage hedonic regression model is conducted. This assumption seems plausible due to several characteristics of the Home Park neighborhood. First, a high percentage of renters (who also happen to be local university students) reside in the neighborhood. Due to absentee homeowners and a prevalence of apartment units in the neighborhood, the stock of rental housing is actually greater than the stock available to homeowners. Also, it is unlikely that renters will be able to afford the same kinds of dwellings as homeowners, who tend to have households of larger sizes than renters that require more living space and more bedrooms.

Second, related to the renter/owner distinction, income is widely different among households. Income, a kind of liquidity constraint, prevents renters from purchasing the same kinds of dwellings as homeowners. Third, two well known ethnic enclaves exist in the Home Park neighborhood – a Muslim enclave in the central section and a Southeast Asian enclave on the eastern section – and prefer to live near other households "like" them. Finally, zoning requirements imposed by the City of Atlanta limit the size of parcels within the city limits. In a sense, PCA seems to be good mechanism by which the multiple differences between households are managed and seems to reflect the consolidation of all of these different "barriers" into a coherent estimation of household "groups" and then "types".

Ekeland, Heckman, and Nesheim (2002) are quick to say, however, that theoretically consistent estimation does not presume segmented markets in the absence of a "source of variation that makes price functions differ when preferences, technology, and the distribution of tastes and technology are common across markets" (p. 307). A formal derivation of how a single market could attract multiple household types is beyond the scope of this paper. In fact, no other paper in the hedonics literature has put forth a formal proof of the existence of multiple hedonic price surfaces in a city, much less a single neighborhood. Yet, space constraints imposed by the built environment, coupled with the existing housing stock that caters to students and families, the old and young, etc. provide enough motivation to informally ask the question if multiple household types can exist in a single neighborhood.

Using the eight groups determined by PCA, an eight-line SUR model is run with the log of selling prices of houses (from the housing survey) in a particular group (predicted by PCA) regressed on the same set of independent variables, which includes structural characteristics of the dwellings (**S**), distances to various amenities such as open spaces (**D**), and adjacency variables **A** (please see Table 1). Again, this particular model is chosen because of the likelihood that the error terms from the system of equations (one equation per household group estimated from the PCA) are correlated; alternatively, this model is appropriate because small changes to a particular housing bundle may cause a different type of household to occupy the dwelling. Otherwise, one might naively run equation-by-equation OLS models on each type separately (see Kmenta [1986] for the proof that OLS provides inconsistent coefficient estimates in the presence of correlated error terms across equations). The formal SUR model (with eight equations) is

$$(3)\qquad \begin{bmatrix} \ln SP_i^1 \\ \ln SP_i^2 \\ \vdots \\ \ln SP_i^8 \end{bmatrix} = \begin{bmatrix} \mathbf{S}_1+\mathbf{D}_1+\mathbf{A}_1 & \cdots & \cdots \\ \vdots & \mathbf{S}_2+\mathbf{D}_2+\mathbf{A}_2 & \vdots \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \mathbf{S}_8+\mathbf{D}_8+\mathbf{A}_8 \end{bmatrix} \begin{bmatrix} \beta_1,\delta_1,\psi_1 \\ \beta_2,\delta_2,\psi_2 \\ \vdots \\ \beta_8,\delta_8,\psi_8 \end{bmatrix} + \begin{bmatrix} \varepsilon_i^1 \\ \varepsilon_i^2 \\ \vdots \\ \varepsilon_i^8 \end{bmatrix}$$

Here, β, δ, and ψ are estimated parameters of **S, D,** and **A**, respectively. Also, the expected value of the error variance-covariance matrix $E(\varepsilon_i \varepsilon_k') = \sigma_{ik}\mathbf{I}_T$, where ε_i is the error term from the i^{th} equation, σ_{ik} is the covariance of the error terms of the i^{th} and k^{th} equations, and $\mathbf{I}_T$ is an identity matrix of order 400 x 400. Greene's (2003) proof, that equation-by-equation OLS can be used to estimate the SUR model when the same independent variables are used for each equation, rests on the assumption that the researcher has additional information on the relationship between each type of household that can be used as "weights", or using the variables to approximate the estimated variance-covariance matrix of residuals. Therefore, a fully efficient GLS weight of an OLS equation can approximate asymptotically the efficient SUR weights. Barring this information, the SUR framework corrects for the relationships between the three equation estimates in an efficient and theoretically consistent manner as a way to account for cross-equation correlation of the error terms (where small changes to a housing bundle or in information may cause a different type of household to occupy a particular dwelling).

Once the initial beta coefficient vectors have been estimated in the eight-line model, the process re-sorts *each* observation into the household group category whose beta coefficient vector best predicts the sales price for that observation. This process sorts households into "types" based on the minimization of the absolute value of the difference in the predicted and actual values of the dependent variable (natural log of sales price) from the SUR model, normalized by the standard error of the linear prediction for each specified equation (which can also be thought of as "the standard error of the predicted expected value or mean for the observation's covariate pattern" [Stata Reference Manual S-Z, p. 157); or $Min \left| \left(\ln \hat{P}_i - \ln P_i \right) / \sigma_i \right|$, $i = 1 \ldots 8$.[24] In essence, this statistic calculates the absolute normalized difference between the beta coefficient vector (and corresponding predicted dwelling sales price) associated with a particular group of households determined by PCA and the self-reported sales price of dwellings. The importance of rescaling the absolute deviation statistic for each observation is analogous to the importance that a *t* statistic plays for a regression coefficient of .0001, which might be perceived as "insignificant" without normalizing it by the standard error.

It seems possible that the highest eigenvalues could become the basis for household "types." If this occurs, then the PCA "groups" will translate directly into SUR "types" and I will see high R^2 values in the initial rounds of the SUR model. But, if these PCA "groups" do not have similar implicit prices for dwelling structure characteristics and amenities, then I might expect to see very low R^2 values. Also, while it seems possible that groups with large standard errors will fit "better," this is not the case; the

[24] I realize that the absolute value notation may distort the calculation of this statistic, as a negative number subjected to absolute value will be different in natural log terms than a number that is positive before being subjected to an absolute value transformation, a problem of nonlinearity. However, I am concerned with

largest standard errors are associated with the largest eigenvalues from the PCA. But, the two PCAs with the largest eigenvalues do not become final household "types", which does not support that statistics with large standard errors will fit "better." Please see Table 2a for statistics on these principal components.

As an example of the process, say that a household is initially sorted into Type A based on the principal components. Then, after estimating the SUR model, say that same household's dwelling sales price is "best" explained by the beta coefficient vector of Type B households. As additional information (structural characteristics, etc.) is used in the SUR process, households are permitted to re-sort into the types that best explain their sales price so that the marginal implicit prices that emerge from the final SUR model are consistent and more efficient than those obtained via OLS, a good proof of which can be found in Kmenta (1986). After two iterations, the model met the convergence criteria (stop iterations when less than 1 percent of the households re-sort into a different type). Table 3 shows the sorting process after the PCA stage and the SUR stage. The SUR do-loop process sorted all but four households into the three final groups used to explain sales price variation. These four households were re-sorted into one of the final three types. One interesting trend in Table 3 is that the groups from PCA are ordered from highest to lowest eigenvalues. It is interesting that the highest eigenvalues, which by definition explain the largest amount of variance in the demographic and attitudinal variables, do not become the "types" after the do-loop process. This suggests that the principal components that explain the most demographic and attitudinal variance do not

---

raw differences between the predicted and actual sales prices, not necessarily the direction of those differences. In future research, this is another area that will be explored.

explain sales price variation well; these components only show similarities in households from a demographic and attitudinal perspective that are used to obtain the initial baseline beta vectors used in the updated sorting process.

Table 3: Sorting Process in Various Stages

	Types After SUR Iterative Process			
Groups from PCA	*A*	*B*	*C*	*Total*
1	4	14	19	*37*
2	4	13	11	*28*
3	19	29	16	*64*
4	16	50	20	*86*
5	10	18	19	*47*
6	8	27	17	*52*
7	2	12	28	*42*
8	7	21	16	*44*
Total	*70*	*184*	*146*	*400*

CHAPTER 4

DATA AND RESULTS

In this chapter, I describe the limitations of publicly available data that spurned the administration of a housing survey to the residents of Home Park (please see Appendix A). Then, I describe the quality of the data and proceed to estimate a model that explains sales price while correcting for heteroskedasticity and potential spatial effects.

4.1 Data Acquisition, Quality, and Description

Existing secondary data sources [including the 2000 Census and Metropolitan Atlanta Multiple Listing Service (MLS)] do not provide the information required to adequately complete a spatial housing hedonic exercise at the household level. Data acquired from the MLS include tax parcel identification numbers; house values used for taxing purposes (assessed values); house square footage; the size (in acres) of each parcel; date of construction; the last sales price; and the date of the most recent sale. To complete the hedonics exercise, additional data were collected by 1) the administration of a survey to the residents of Home Park and 2) the construction of spatial variables using ArcView GIS (geographic information system). First, a housing survey (please see Appendix A) was administered to each dwelling in the neighborhood in August 2002 to September 2002, from which the demographic and attitudinal variables were collected. Only one survey per household was recorded (the first one I received), as some residents received more than one survey. I received 273 completed surveys after Round 1 (including the test round), another 127 surveys in Round 2, and the final 20 surveys (used for the out-of-sample comparison) in Round 3. Then I obtained structure characteristics data for dwellings in Home Park from the MLS through special internet access granted to me in October 2002. But, with an ever greater emphasis placed on the explicit incorporation of space into hedonics models, I used resources in the Georgia Tech Center for GIS (Subrahmanyam Muthukumar in particular) to generate the road network distances to various features, such as parks and commercial opportunities. The explicit incorporation of space follows a current trend in the literature (McLeod 1984; Kulshreshtha and Gillies 1993; Benson et al. 1998; Irwin and Bockstael 2001 and 2002; Paterson and Boyle 2002) to incorporate *more* GIS data into hedonic property value models.

Another important research area is the role that amenities play in the household location decision. Generally, households will choose a location that is "near" cultural amenities, entertainment and recreation opportunities, and others. Clark et al. (2002) argue that "a residential population of young professionals with more education and fewer children creates a social profile geared toward recreation and consumption concerns," which values highly "a cultural center offering diverse, sophisticated and cosmopolitan entertainment lacking elsewhere" (p. 500). Similarly, recent findings show that cities with more gay men tend to have more amenities than other cities; and that amenities and gay men are some of the strongest predictors of high-technology job growth (Black et al. 2002; Florida 2002). This is particularly important for the City of Atlanta, which has a relatively large gay population, high-technology job growth, and many families with relatively fewer children. For these reasons, the impacts of local amenities on the household location decision process are incorporated into the empirical analysis via the attitudinal variables (please see Table 1). In addition to attitudinal variables, the other variables used for this study can be classified as dwelling features, demographic variables, spatial variables, and neighborhood features.

The response rate for this study was 51.2 percent (400/781). Of those who responded, 64.5 percent are renters and 35.5 percent are owners. The descriptive statistics for the full sample (Table 4a) show that the average household in Home Park has 2.6 adults and 0.13 children (only 33 households in the sample, or 8.2 percent, have any children). The average survey respondent is 33 years old (the median age is 28), not surprising given the prevalence of college students in the neighborhood clustered in the 20-25 age bracket and the percentage of retired persons in the neighborhood (5 percent). 65 percent of respondents to the survey were male. The average resident has lived in Home Park slightly longer than five years, is white, has at least an undergraduate degree, and makes between $25,000 and $34,999 per year. Approximately 34 percent, 26 percent, 32 percent, and 7 percent of Home Park residents feel that the most important issue besides national security is crime, the environment, education, and Social Security, respectively. This variable limits the responses only to one of these four options for survey response rate reasons; so these possible answers may not necessarily be the single most important issue to these households besides national security. Overwhelmingly, Home Park residents (57.7 percent of them) chose to live in the area because of its proximity to Georgia Tech. Also, more residents chose their particular dwellings based on the number of bedrooms (38 percent) than any other factor. Finally, 50.7 percent of homeowners have made at least $10,000 in improvements to their homes over the last three years. This in part accounts for the average market price of homeowners' dwellings ($223,917).

Block level Census data (for Census Tract 10, Block Group 2 and Census Tract 6, Block Group 1 that comprise Home Park and a few streets outside the neighborhood that are comprised of mostly commercial and retail enterprises) tell me that 82.2 percent of individuals are renters, that 29.3 percent of residents are between the ages of 15 and 24, that 40.6 percent of residents are between the ages of 25 and 34, that 59.1 percent of residents are white, and that 29.4 percent of residents are Asian. In general, the Census block group data suggest that I may have oversampled the owners in Home Park (who are older on average than renters), the white persons in the neighborhood (as the white roommates may have completed the survey even though they have roommates of other races), and older households (which corresponds with oversampling the owners). This may have occurred because of language barriers, particularly with the Muslim households that do not understand the English language very well, and the possibility that younger students may not see the merit in completing the survey.

In general, I believe these statistics have remained fairly stable since the 1996 Summer Olympic Games, when a large influx of dollars was directed at nearby neighborhoods and Georgia Tech. Also, with the debates on school quality raging in Georgia and throughout the United States, I should point out that all Home Park households are located in the same school district. So, school quality does not vary and would add nothing to the analysis.

Table 4a: Descriptive Statistics on Full Sample of Survey Respondents

Variable (400 obs.)	Mean	Standard Deviation
Rent/Own (Renter = 0, Owner = 1)	.35	.47
Current Market Value of Dwelling	148115	69400
High School Graduate (Yes = 1)	.09	.29
*College Graduate (Yes = 1)**	.66	.47
# of Adults in Household	2.62	1.51
# of Children in Household	.13	.51
Age of Survey Respondent	33.44	14.44
Sex (Female = 0, Male = 1)	.65	.47
Dwelling Tenure	5.09	10.21
Race (Non-white = 0, white = 1)	.72	.44
Employed Full Time (Yes = 1)	.43	.49
Student (No = 0, Yes = 1)	.45	.49
Undergraduate Student (Yes = 1)	.23	.42
Graduate Student (Yes = 1)	.19	.39
Retired (No = 0, Yes = 1)	.05	.23
Dwelling Year of Last Sale	1992	6.59
Politically Conservative	.21	.40
Politically Moderate	.47	.50
Politically Liberal	.31	.46

* Note that "High School Graduate" denotes only those survey respondents who have a high school education. If the question were the percentage of households with at least a high school education, the answer would be approximately 80 percent.

4.2 Results

To reiterate, principal components analysis (PCA) performed on demographic and attitudinal variables created factors (categorizations of residents) that were used to sort households into types. These baseline types were used as guides for the sorting process that used the SUR iterative model. Then, after the SUR iterative process, which had two iterations, households sorted into three types (named A, B, and C from this point forward), the descriptive statistics of which can be found in Table 4b. Also, Tables 4c and 4d display summary statistics for each household type by student status and by owner/renter status.

Table 4b: Descriptive Statistics by Type

	Type A (70 observations)		Type B (184 observations)		Type C (146 observations)	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Rent/Own (Renter= 0 Owner= 1)	.18	.39	.26	.44	.55	.49
Current Market Value of Dwelling	126582	61120	134926	61173	175060	74680
High School Graduate (Yes = 1)	.14	.35	.07	.26	.09	.29
*College Graduate (Yes = 1)**	.47	.50	.61	.48	.80	.40
Age of Survey Respondent	31.43	14.90	32.37	14.35	35.74	14.12
Sex (Female = 0, Male = 1)	.63	.47	.64	.47	.67	.46
Dwelling Tenure	4.02	8.21	5.34	11.62	5.28	9.16
Race (Non-white = 0, white = 1)	.85	.35	.73	.44	.65	.47
Employed Full Time (Yes = 1)	.34	.47	.36	.48	.54	.49
Student (No = 0, Yes = 1)	.50	.50	.52	.50	.33	.47
Undergraduate Student (Yes = 1)	.35	.48	.28	.45	.10	.30
Graduate Student (Yes = 1)	.11	.32	.22	.41	.20	.40
Retired (No = 0, Yes = 1)	.07	.25	.04	.20	.06	.25
Dwelling Year of Last Sale	1992	6.30	1991	6.88	1993	6.30
Politically Conservative	.28	.45	.21	.41	.17	.37
Politically Moderate	.40	.49	.47	.50	.50	.50
Politically Liberal	.31	.46	.30	.46	.32	.46
Dwelling Square Footage	1299	474	1322	577	1361	496
Number of Beds	2.72	.86	2.15	.58	2.50	.77
Number of Baths	1.32	.50	1.26	.49	1.39	.68
Number of Acres	.17	.09	.22	.19	.25	.45
Distance to Nearest Commercial	528	169	596	216	477	169
Distance to Home Park green space	446	289	545	234	556	280
Distance to Piedmont Park	2528	205	2514	244	2451	185
Distance to Georgia Tech	1293	196	1285	238	1314	210

Table 4c: Selected Means by Type and Student Status

	Type A (70 observations)		Type B (184 observations)		Type C (146 observations)	
Student? (# of observations)	**Yes (35)**	**No (35)**	**Yes (96)**	**No (88)**	**Yes (49)**	**No (97)**
Rent/Own (Renter= 0, Owner= 1)	.05	.31	.02	.52	.12	.77
Current Market Value of Dwelling	113485	139679	112406	159494	122946	201386
High School Graduate (Yes = 1)	.00	.28	.00	.16	.00	.14
College Graduate (Yes = 1)	.34	.60	.48	.76	.71	.84
Age of Survey Respondent	23.00	39.86	23.69	41.85	25.67	40.82
Sex (Female = 0, Male = 1)	.66	.61	.64	.65	.73	.64
Dwelling Tenure	.98	7.07	1.14	9.92	1.65	7.12
Race (Non-white = 0, white = 1)	.82	.88	.57	.91	.24	.86
Employed Full Time (Yes = 1)	.00	.68	.00	.77	.00	.82
Retired (No = 0, Yes = 1)	.00	.14	.00	.09	.00	.10
Dwelling Year of Last Sale	1992	1992	1991	1991	1992	1993
Politically Conservative	.22	.34	.18	.23	.18	.16
Politically Moderate	.45	.34	.53	.42	.42	.54
Politically Liberal	.31	.31	.28	.34	.38	.28

Table 4d: Selected Means by Type and Renter/Owner Status

	Type A (70 observations)		Type B (184 observations)		Type C (146 observations)	
Rent/Own Status? (# of obs.)	**Renter (57)**	**Owner (13)**	**Renter (136)**	**Owner (48)**	**Renter (65)**	**Owner (81)**
Current Market Value of Dwelling	102314	232991	106411	215718	109938	227319
High School Graduate (Yes = 1)	.12	.23	.02	.22	.06	.12
*College Graduate (Yes = 1)**	.43	.61	.60	.66	.75	.84
Age of Survey Respondent	27.79	47.38	27.31	46.72	28.10	41.86
Sex (Female = 0, Male = 1)	.59	.84	.67	.57	.75	.61
Dwelling Tenure	1.93	13.19	1.58	15.98	1.73	8.14
Race (Non-white = 0, white = 1)	.84	.92	.64	.97	.38	.87
Employed Full Time (Yes = 1)	.31	.46	.26	.66	.29	.75
Student (No = 0, Yes = 1)	.56	.15	.67	.04	.66	.08
Undergraduate Student (Yes = 1)	.43	.07	.38	.02	.18	.03
Graduate Student (Yes = 1)	.13	.07	.29	.02	.38	.05
Retired (No = 0, Yes = 1)	.03	.23	.007	.14	.01	.11
Dwelling Year of Last Sale	1992	1991	1991	1991	1990	1995
Politically Conservative	.28	.30	.18	.29	.15	.18
Politically Moderate	.42	.30	.5	.41	.44	.55
Politically Liberal	.29	.38	.31	.29	.40	.25

The three types of households predicted by the model are distinct, as one might expect from a non-clustering, self-organization process in a particular neighborhood. According to Table 4b, Type A households are generally white male student renters who have lived in the area an average of four years. Type B households on average are mostly renters comprised of white males in their early to middle 30s who have lived in the area for an average of five years. Finally, Type C households on average are owners that also are the most racially diverse group of the three types. Generally speaking, though, Type C households are mostly white (despite the racial diversity), employed full-time, and, as expected, have the relatively highest education levels and dwelling sales prices of Home Park residents (since they are primarily established homeowners). Note that the sorting on demographic characteristics and attitudinal variables was used to provide a starting point for the SUR iterative process instead of the atheoretical random assignment of households to groups. Despite this, there may be a slight interpretation problem here, as types of households were determined based on the "best" fitting hedonic price lines, which implies that prices and housing attributes are the primary sorting criteria. Nonetheless, these are the general characteristics of the three final types of households that emerge from the SUR iterative process.

As everyone knows, starting points matter; this may be viewed as a fundamental limitation of the approach used in this paper. In any case, the goal is to get beta coefficient estimates that reasonably describe an aggregable group of households, defined by demographic characteristics and attitudes toward certain neighborhood features. The possibility exists that households in other PCA determined groups may not be classified with other households that have similar implicit prices for the same dwelling structure characteristics. This is the reason that I re-sort households into the group whose corresponding beta coefficient vector best minimizes the normalized difference in actual and predicted sales prices. This method assures me that the households within each of the three final subgroups have identical (though nonlinear) implicit prices for each of the independent variables.

An interesting statistic I see here, moving from Type A to Type C households, is the change in the kind of student that sorts into each type. Type A households, 50 percent of which is students, contains 35 percent undergraduate and 11 percent graduate students, respectively. Then, moving to Type C, I see that the percentage of undergraduate students falls and the percentage of graduate students rises. Considering the average tenures of each type, this result is consistent with the fact that graduate students tend to take longer to finish their degrees than undergraduates. This trend is supported by contingency tables that sort households by type and student status (please see Table 4c).

Another way to look at the demographic differences is to distinguish renters from homeowners. Table 4d shows me that 59 percent of renter survey respondents in Type A were male, compared with 67 percent and 75 percent in Types B and C, respectively. Also, 84 percent of Type A renter respondents were white compared with 64 percent and 38 percent in Types B and C, respectively. Finally, 31 percent, 26 percent, and 29 percent of Types A, B, and C renter respondents, respectively, were employed full-time; and the trends concerning the type compositions of undergraduate and graduate students also hold when statistics are examined by whether a household rents or owns its dwelling. These statistics verify the trends at local universities- the majority of Home Park renters are non-white, male Georgia Tech graduate students (mostly Asian) who consider their graduate assistantships to be part-time work. Also, the full-time employment rates illustrate the relatively difficult financial situations that undergraduates living off-campus (especially out-of-state students) endure compared to graduate students, who generally have more financial support through fellowships and assistantships (either research or teaching-oriented).

Next, after using several joint F tests to eliminate insignificant explanatory variables for the natural log of sales price, a parsimonious regression model is determined. The coefficient estimates from the parsimonious SUR model are displayed in Table 5.[25]

[25] The variables used in the SUR iterative process but omitted from the final model include the year the dwelling was constructed, the distances to two various churches in Home Park, the distance to the local mosque, the distance to the four within-neighborhood grocery stores, a dummy variable that answers the question "Are you living adjacent to someone who wishes to move out of the neighborhood in the next two years?", the distance to Centennial Elementary School (that services Home Park elementary school students), the distance to the Feminist Women's Center, a dummy variable representing dwellings north of 14th Street, the distance to the Georgia Tech Student Athletic Center (currently under construction), and several interaction variables. These variables were subjected to F tests that tested the restriction that all of the coefficients on these variables *across all three types* were jointly equal to zero. Finally, a variable that distinguishes renters from owners was not included as an explanatory variable for sales price because it was used in the PCA; the use of that variable in both regressions would not be appropriate.

Table 5: Final Iteration of SUR Estimates (double log specification)
[Dependent Variable: Natural log of Self-Reported Sales Price]

	Type A	Type B	Type C
	Coefficient (Std. Error)	*Coefficient (Std. Error)*	*Coefficient (Std. Error)*
Constant	4.78 (9.90)	5.50 (11.04)	2.19 (9.60)
Structural Variables			
Natural log of Dwelling Square Feet	-1.28** (.65)	.56 (.77)	.92 (.72)
Natural log of Number of Beds	3.60*** (.68)	-4.93*** (.81)	1.40* (.76)
Natural log of Number of Baths	-1.18* (.69)	-1.17 (.81)	2.45*** (.77)
Natural log of Number of Acres	-.19 (.34)	-.64 (.40)	.72* (.38)
Condominium (yes = 1)	---	---	-.17* (.09)
Spatial Variables			
Ln Dist to 14th St. Commercial Site	-.40 (1.48)	1.60 (1.63)	-1.78 (1.44)
Ln Dist to Piedmont Park	.44 (.37)	-1.85*** (.43)	1.37*** (.41)
Ln Dist to Meineke Service Station	1.85** (.74)	-3.97*** (.86)	2.17*** (.79)
Ln Dist to Georgia Tech	-1.76*** (.68)	4.00*** (.78)	-2.52*** (.73)
*Ln Dist to Home Park Green Space*HP <200m*	.50*** (.10)	---	---
*Ln Dist to Home Park Green Space*HP >200m*South Home Park*	.11 (.10)	.40*** (.11)	-.02 (.09)
*Ln Dist to Home Park Green Space*North Home Park*	---	.50*** (.10)	---
Is Dwelling Above Street Level?	-1.14** (.45)	-.57 (.53)	1.84*** (.50)
Located on State Street?	1.22 (.96)	-6.23*** (1.13)	5.00*** (1.07)
*Above Street Level*State Street*	-.40 (1.40)	5.42*** (1.66)	-5.27*** (1.55)
Adjacency Variables			
Lives Adjacent to a Renter	1.74*** (.67)	-2.29*** (.79)	.24 (.74)
Lives Adjacent to Undergrad Student	1.25** (.56)	1.39** (.67)	-2.65*** (.63)
Lives Adjacent to a College Educated person	-.19 (.63)	-.82 (.75)	1.17* (.71)
Lives Adjacent to a Person of a Different Race Than You	-1.92*** (.47)	-2.69*** (.55)	4.58*** (.52)
Lives Adjacent to a Dwelling that made Home Improvements	.69 (.48)	-3.98*** (.57)	3.61*** (.53)
Lives Adjacent to Someone who Lives in Home Park Because it is Near Georgia Tech	1.47*** (.57)	2.61*** (.67)	-4.13*** (.63)

Table 5 continued

Other Statistics			
χ^2	112.87 (p = .0000)	245.56 (p = .0000)	312.12 (p = .0000)
Root MSE	3.88	4.59	4.34
R^2	.23	.38	.43

*= Significant at .10 level; ** = Significant at .05 level; *** = Significant at .01 level

Knowing the characteristics of the three types of households provides me a more rich interpretation of the first-stage hedonic regression coefficients in Table 5.[26] I rescaled each distance variable to be equal to the maximum distance of any dwelling in Home Park from the landmark less the actual distance to each landmark. All variables were rescaled so that the farthest dwelling from each landmark would have a distance value of zero (which forces the price effect of a particular landmark to zero at the neighborhood boundaries); so, being located exactly at the same location as the landmark would give that dwelling a distance equal to the maximum distance to that landmark of the farthest dwelling. In terms of interpretation, positive coefficients on distance variables indicate that dwelling sales prices increase at the margin as a household gets

[26] The hedonics literature is clear on the use of dwelling structural characteristics and distance variables as explanatory variables of sales prices, but vaguer on the proper use of adjacency variables and attitudes toward certain neighborhood amenities and features. So, in the Tiebout spirit, I hypothesize that attitudinal variables can be used to sort households into types, but cannot be used effectively to explain sales price variation. Indeed, when attitudinal variables were included in the SUR model, very few were significant predictors of sales price variation. In particular, the responses to the question regarding the most important issue facing the country (crime, environment, education, or social security) show that different household types view social issues differently, evidenced by the significant and changing signs on these coefficients.

variables indicate that dwelling sales prices increase at the margin as a household gets closer to the landmark, or equivalently if the landmark expands in size by one meter on a four sides. Related to this, the specification of the model represented in Table 5 focuses on the effects of the distance to the Home Park green space variable, as these marginal prices will be used to estimate the impact on aggregated sales prices of a change in each household's distance to the park as well as identifying the dwellings that will be occupied by different household types if a particular vacant parcel is converted to a second green space for the neighborhood (please see Chapter 5).

Immediately, one can see that different independent variables explain significant amounts of sales price variation for different types of households. But, the signs are different, which suggests that the neighborhood might be treated as having different household types. The R^2 statistics show that 23 percent, 38 percent, and 43 percent of the variation in sales prices for Type A, B, and C households, respectively, is explained by the model.

For the dwelling structure variables, I would expect that households, regardless of type, would have positive marginal prices; that they would be willing to pay higher prices for more living space, additional beds, additional bathrooms, and additional acreage. Looking at the dwelling structure coefficients, I see several that are negative for Types A and B. A literal interpretation of the coefficient on the natural log of living space (square footage) means that a household at the margin is willing to pay 1.28 percent less in sales price for each additional percentage change in square feet of living space. Besides being counterintuitive, it is not consistent with the goals of any rent-seeking landlord, to provide more living space at lower prices. But, these coefficients *do* suggest that maybe different types of households are purchasing/renting their dwellings for particular structural characteristics. For Type A households, I observe a positive marginal price for the number of beds; for Type B households, for living space; and for Type C, living space, beds, baths, and acres. This suggests that Type A households may not have a positive WTP for other dwelling features except the number of bedrooms. Since I know from the physical addresses of Type A households that they are generally student renters who live in two bedroom, one bath apartment units, that they have a positive WTP for bedrooms is not surprising. I see a similar occurrence for Type B households, who have a positive WTP for living space. This result makes sense because Type B households tend to be student renters who live in apartment units or houses with more non-bedroom living space than Type A households. Finally, for Type C households, the dwelling structure coefficient estimates are significant at the .10 level and support the theory that additional living space, bedrooms, baths, or acres will command a higher sales price at the margin. Also, Type C households that live in condominiums, as expected, have a lower sales price at the margin than non-condominium residents, most likely due to the fewer acres that they own. Overall, the estimates for dwelling structure characteristics, with the exception of Type C estimates, are not consistent with economic theory but provide much insight into what particular types of households are willing to pay for particular structure characteristics in the housing market.

For the coefficients on the spatial variables, one can see that different variables significantly explain sales price variation across the three types. But, the signs of the significant coefficient estimates are intriguing. There is a fairly equal mix of positive and negative signs on the significant coefficient estimates in the SUR model. The variable

"Above Street Level * State_St" equals one if a household lives in a dwelling that is above street level *and* is located on State Street, a major access road between Home Park north of 14^{th} Street and 10^{th} Street. So, the significant and positive coefficient on this variable for Type B households tells me that these households have higher sales or rental prices if they are above State Street level, all else held constant. This is interesting given the coefficients on "Above Street Level" and "Located on State Street?", which suggest that all other households *except* those that are located above State Street level have lower sales prices at the margin. This might occur because of the proximity to the Home Park green space of these State Street houses that are above street level, which afford these households a better view of the street and (in some cases) the park.

Using the same coefficient for Type C households, I cannot definitely say what the effect of topography is on State Street sales prices. One interpretation could be that Type B dwellings have higher sales prices due to the owners that sorted into this type and that the negative coefficient for Type C households is due to the renters that sorted into Type C. The other, more general interpretation is that Type B (Type C) households on average have higher (lower) sales prices if they live on State Street and are located above street level, all else held constant. Then, this coefficient is insignificantly negative for Type A households.

Regarding the other spatial variables, the natural log of the distance to the local Meineke automobile repair shop has significantly positive, negative, and positive effects on the selling prices of Types A, B, and C households, respectively. This illustrates the more detailed estimates of different household types that one can get when markets are segmented. In addition, one can see that household types exhibit changing signs on all of the other significant distance variables (distance to the center of Georgia Tech, distance to the 14th Street commercial strip, and distance to Piedmont Park). Regarding the coefficient magnitudes, nothing highly unusual piques my interest. The coefficient on distance to Georgia Tech for Type B households can be strictly interpreted as the following: a 1 percent decrease in the distance to Georgia Tech leads to a 4 percent increase in sales price at the margin. This seems to be a little high, so I would be careful to generalize this particular result to a different neighborhood. Regardless, this empirical evidence shows the information learned by using the SUR model as the first-stage hedonic estimator – that different types have different WTPs for the same local amenities. The SUR estimator itself, I posit, is responsible for the changing coefficient signs. This is likely given that the SUR iterative process described in Chapter 3 essentially provides each observation from each regression equation the opportunity to re-sort to the group whose coefficient vector best predicts its sales price. This movement of observations from one group to another in the iterative process

Next, the adjacency variables are defined using street addresses. Since this is a sample of households, it is unrealistic to expect that all survey responses would come from adjacent households. To remedy this, each household was assigned a "nearest neighbor" on the left and right sides of the dwelling (looking out the front door); in the case of apartments in a complex, the nearest apartment unit was used as the neighbor.[27] While these variables are problematic, particularly for the dwellings that do not have an

[27] Using this method, one dwelling had a neighbor that was fifty street numbers away (approximately one-tenth of a mile). Otherwise, more than half of the observations had an adjacent dwelling as a neighbor and the other half averaged about fifteen street numbers away from their nearest neighbor on either side.

immediate nearest neighbor, I use them to estimate the impacts of certain non-structural attributes of the households on their "neighboring" dwelling sales prices.

The coefficient estimates show me that Type A, B, and C households have higher, lower, and higher sales prices if they live adjacent to other renters, respectively. This relationship between sales price and renter adjacency for Type A households reflects the implicit value that renters are willing to pay to live next to other renters (who happen to be mostly students). A significant number of survey respondents (4 percent) wrote in that they chose their particular dwelling because of nearby friends, which suggests that this percentage is most likely higher, as the survey instrument did not provide "friends" as an option. For Type B households, who tend to be more graduate students and young homeowners, the negative and significant coefficient might signal the desire not to be near renters, who happen to be mostly students, who tend to socialize at night and on the weekends. For Type C households, which are mostly owners, this positive coefficient signals the higher sales prices associated with mostly owners being near renters. This might indicate a desire to be near younger renters for crime protection reasons or for generally accessibility reasons (help with mowing the lawn, to watch the house during the day, etc.) Also, being adjacent to a renter could be interpreted as a proxy for subneighborhood diversity, or the interaction of generally older and younger folks.

The results also tell me that Type A and B households have significantly higher dwelling market values if they live adjacent to undergraduate students, as I would expect given that Type A and B households are mostly student renters. In preliminary regressions, being adjacent to graduate students did not significantly affect selling prices, most likely due to the fact that graduate students exhibit characteristics more like homeowners than renters. Also, being adjacent to a dwelling that has at least one college graduate is a significant and negative predictor for sales prices for Type C households. This, along with the coefficient estimates on the "adjacent to a college educated person" variable, suggests that owners generally prefer to be near college educated households, and they may be willing to pay higher sales prices for that situation than to be near undergraduate students.

Then, having neighbors of different races than one's own lowers selling prices for Type A and B households, all else held constant. It is likely that this variable is significant because of the fact that there are two geographically-defined Asian enclaves in the neighborhood. The first is the Southeast Asian enclave in the east central part of Home Park along Westshire Place; the other enclave, the majority of which are Muslim, are located along Ethel Street in central Home Park near the Muslim school and mosque. If the households that are sorted into Type B, for example, are mostly from the streets known to be part of the Asian enclave, which traditionally have physically smaller, lower priced dwellings than other sections of Home Park, then their selling prices might be lower if they are located near other Asians. In this case, the physical addresses of Type B households show that a significant percentage of households near both Asian enclaves were sorted into Type B. This suggests that a definitive statement regarding the inverse relationship between neighbors' races and one's dwelling market value cannot be made. However, one can interpret the significant and positive coefficient for all households as their average value of racial diversity holding all else constant.

Next, being adjacent to owners who have made improvements to their dwellings insignificantly increases the dwelling market value of Type A households, significantly decreases the dwelling market values for Type B households, and significantly increases the market value of Type C households, all else held constant. To homeowners (particularly shown by the significantly positive coefficient for Type C households), having neighbors make home improvements simultaneously increases their dwelling's market value. This empirically supports the statement in Chapter 1 about housing externalities, that making home improvements at one dwelling increases the market value of neighbors' dwellings by some amount. In contrast, Type B households have significantly lower sales prices if their neighbors made home improvements, which contradicts the assertion of positive housing improvement externalities. However, without a series of higher-order spatial lag variables that examines the home improvement activities of the neighbors' neighbors, etc., I cannot discern the extent to which these positive and negative externalities dissipate.

Finally, Type A and B households have positive valuations of being adjacent to households that located in Home Park because of its proximity to Georgia Tech. Type C households (who are mostly owners), in contrast, view adjacency to households that located in Home Park because of its proximity to Georgia Tech as a negative impact on their dwelling sales prices, all else held constant. First, the positive impact on sales price for Type A and B households is apparent, as these types are comprised mostly of student renters (and researchers) who attend (work for) Georgia Tech. Type C households (mostly owners), it seems, view Georgia Tech-attracted neighbors negatively, as several survey respondents expressed in the open-response question in the survey. This coefficient for Type C households is likely to be picking up similar effects captured in the coefficient on adjacency to undergraduate students.

At first glance, the explanatory power of the SUR model is not very high compared to other housing hedonics models, some of which report R^2 values in the .90 to .96 range. The primary difference between other hedonics research and this research is that others have used OLS or GLS models for the first-stage hedonic equation. So, it is misleading to argue that other models are "better" solely based on explanatory power. To compare the differences between the two models, Table 6 shows the pooled OLS estimates (with all independent variables from the three-equation SUR model) and the estimates by household type (using the same model specifications as in Table 5). One can see that the majority of explanatory variables are insignificant predictors of dwelling sales prices in the pooled model and especially in the OLS models by household type. This might occur because the independent variables were eliminated through joint F tests in an SUR model setting as well as high collinearity among the distance variables.

It is possible that the elimination of insignificant variables starting with an OLS framework would result in the same model specification as the SUR. However, the theory for each estimator is different: SUR is used to model the competition between types for a limited amount of space whereas no competition can be structurally modeled using the OLS model, unless one counts discrete type variables (are you a Type A household?) as "structually modeling competition between types." In any case, as a statistical validation of the theory of multiple types, the Breusch-Pagan (1980) statistic in Table 5 indicates that the variance-covariance matrix of the error terms is not diagonal (i.e. the off-diagonal elements of the error covariance matrix are non-zero). As Greene (1993) points out, "the greater the correlation of the disturbances, the greater the efficiency gain accruing to GLS" (p. 489). So, this means that the use of an OLS estimator, while consistent, does not provide as efficient estimates as the SUR system of equations. In terms of unbiasedness, according to Kakwani (1967), the feasible GLS estimator of the beta coefficient vector is unbiased regardless of which divisor is used in the consistent estimation of error variance-covariance matrix.[28] Third, when two of the three discrete "Type" variables are included in the pooled OLS model in Table 6 with the hypothesis that at least one type would be a significant predictor of sales price, I find that systematic differences in those households' prices are captured, evidenced by t = -3.61 for Type A and t = -2.51 for Type B households. These Type variables explain additional variation in sales prices (i.e. price changes in response to *type differences*), which indicate some systematic variation in households that OLS does not capture.

To conclude this section, I want to compare the estimates between Tables 5 and 6. From econometrics I know that the (feasible) SUR estimator differs from the OLS estimator by a weighting mechanism, the estimated error variance-covariance matrix, which is used as a weight to obtain the coefficient vector; this is why the SUR coefficients and standard errors in Table 5 are larger than those in Table 6. However, theoretically, the SUR is more appropriate to analyze multiple household types than OLS, as I have demonstrated through the correlation matrix of the SUR residuals as well as the significant Breusch-Pagan (1980) test. If I could conclude from the Breusch-Pagan test that the off-diagonal elements of the error covariance matrix were not significantly different than zero, then equation-by-equation OLS would be the appropriate estimator, as the gain in efficiency from SUR is negligible. These results, along with the similar

[28] The Stata "sureg" command uses the number of sample observations (n = 400 in this case) as the divisor in computing the covariance matrix for the residuals; this estimator is "asymptotically justified" (p. 156).

coefficient magnitudes to those already in the hedonics literature, assure me that the SUR estimates are correct and accurately estimate the effects on sales price of marginal changes in neighborhood features and dwelling structural characteristics.

Table 6: OLS Regression Models (Pooled and by Type)

	Pooled OLS (400 obs.)	Type A (70 obs.)	Type B (184 obs.)	Type C (146 obs.)
	Coefficient (Std. Error)	*Coefficient (Std. Error)*	*Coefficient (Std. Error)*	*Coefficient (Std. Error)*
Constant	8.35*** (1.46)	5.30 (5.97	8.65*** (1.46)	12.33*** (1.53)
Structural Variables				
Natural log of Dwelling Square Feet	.24*** (.06)	.21 (.18)	.39*** (.08)	-.07 (.11)
Natural log of Number of Beds	.09 (.06)	.17 (.19)	.11 (.09)	-.03 (.13)
Natural log of Number of Baths	.08 (.06)	-.17 (.18)	-.05 (.10)	.13 (.11)
Natural log of Number of Acres	-.12*** (.03)	.28* (.16)	-.23*** (.04)	-.08 (.05)
Condominium (yes = 1)	-.16* (.09)	---	---	-.37*** (.13)
Spatial Variables				
Ln Dist to 14th St. Commercial Site	.33 (.27)	1.42 (.97)	.03 (.23)	.32* (.18)
Ln Dist to Piedmont Park	-.02 (.03)	-.44 (.28)	-.01 (.03)	-.29 (.18)
Ln Dist to Meineke Service Station	-.14 (.09)	-.79 (.50)	-.05 (.10)	-.48** (.19)
Ln Dist to Georgia Tech	.17 (.13)	.49 (.60)	.03 (.08)	.49** (.20)
*Ln Dist to Home Park Green Space*HP <200m*	-.13 (.16)	.005 (.10)	---	---
*Ln Dist to Home Park Green Space*HP >200m*South Home Park*	-.14 (.16)	-.02 (.09)	-.02 (.03)	-.02* (.01)
*Ln Dist to Home Park Green Space*North Home Park*	-.11 (.15)	---	-.005 (.03)	---
Is Dwelling Above Street Level?	.11** (.04)	.13 (.14)	.07 (.05)	.04 (.07)
Located on State Street?	.04 (.09)	.25 (.24)	-.01 (.15)	-.43** (.18)
*Above Street Level*State Street*	-.22 (.13)	-.69** (.32)	-.28 (.23)	.30 (.24)
Adjacency Variables				
Lives Adjacent to a Renter	-.31*** (.06)	.19 (.34)	-.37*** (.07)	-.26** (.12)
Lives Adjacent to Undergrad Student	-.02 (.05)	.12 (.13)	.004 (.07)	-.14 (.15)

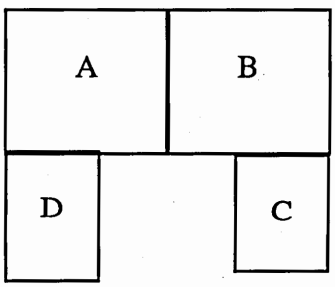
Table 6 continued

Lives Adjacent to a College Educated person	.14** (.06)	.01 (.16)	.13* (.07)	-.15 (.17)
Lives Adjacent to a Person of a Different Race Than You	-.04 (.04)	-.02 (.16)	-.12* (.06)	-.15* (.17)
Lives Adjacent to a Dwelling that made Home Improvements	.31*** (.04)	.38** (.16)	.28*** (.07)	.36*** (.07)
Lives Adjacent to Someone who Lives in Home Park Because it is Near Georgia Tech	-.05 (.05)	-.33 (.24)	.06 (.08)	-.000 (..08)
Other Statistics				
F	9.32 (p = .0000)	1.21 (p = .2900)	7.12 (p = .0000)	5.33 (p = .0000)
Root MSE	.37	.42	.33	.34
Adjusted R^2	.30	.05	.38	.36

*= Significant at .10 level; ** = Significant at .05 level; *** = Significant at .01 level

4.3 Limitations of Spatial Weights Matrix Approaches

The inclusion of spatial variables in the SUR model does not guarantee that *all* of the possible spatial effects that may influence sales prices in urban neighborhoods are accounted for. Therefore, one must check for the presence of additional spatial dependence that has not been captured by the spatial variables already in the model. In the literature on spatial econometrics, the most common approach used to account for unobserved spatial effects is the use of a spatial weights matrix, which is used to specify the "neighbors" of each unit of observation. Figure 4 shows the physical structure and resulting weights matrix **W*** (which is then typically row-standardized, called **W** from here on) for a hypothetical neighborhood.



translates into $\mathbf{W}^* = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$

Figure 4: Illustration of Spatial Weights Matrix

Two types of models emerge from the use of a spatial weights matrix. First, spatial lag models, which are roughly analogous to lags in time-series estimation, permit the use of one's "neighbors" in a spatial weights matrix, to lend weight to observations that are spatially more proximate to the observation of interest than others. Traditionally, these models take the form of $y = \rho Wy + X\beta + \varepsilon$, where y is an $n \times 1$ vector of observations on the dependent variable, W is an *n x n* row-standardized spatial weights matrix that formalizes the structure of neighborhood influences, ρ is the spatial autoregressive lag parameter to be estimated, X is an *n x k* matrix of observations on the exogenous variables, β is the *k x 1* regression coefficient vector, and ε is a *n x 1* vector of random error terms. Then, after making the relevant simplifications, the spatial lag model simplifies to $y = (I - \rho W)^{-1} X\beta + (I - \rho W)^{-1} \varepsilon$, where $(I - \rho W)^{-1}$ is called the "spatial multiplier." This shows that the dependent variable y_i "is determined by the error terms at all locations in the system, and not just the error at *i*. This *simultaneity* [italics in original] makes the spatially lagged Wy variable endogenous, which necessitates specialized estimation techniques, such as maximum likelihood estimation or instrumental variables approaches…" (Anselin 2002, p. 249).

Second, spatial error models incorporate the spatial weights matrix differently than the spatial lag models. Whereas in the spatial lag model the weights matrix was used to introduce a spatial lag of the dependent variable, a spatial error model introduces spatial dependence through the error terms. According to Kim, Phipps, and Anselin (2003), the spatial error model is "based on the assumption that there is one or more omitted variables in the hedonic price equation and that the omitted variable(s) vary spatially. Due to this spatial pattern in the omitted variables, the error term of the hedonic price equation tends to be spatially autocorrelated." Empirically, the spatial error model is $y = \mathbf{X}\beta + \varepsilon$, where $\varepsilon = \lambda \mathbf{W}\varepsilon + \mu$. Here, λ is the spatial autoregressive coefficient and μ is an independently and identically distributed (i.i.d.) vector of error terms. This model, which is a special case of a regression equation with a non-spherical error variance-covariance matrix, provides consistent, unbiased, but inefficient OLS estimates.

Therefore, estimation must be based on either maximum likelihood or a generalized moments approach in the Kelejian and Prucha (1999) style.

As one can see, the conceptualization of the spatial weights matrix is very important in spatial econometrics models, as it determines the extent to which neighbors influence the sales prices of dwellings. But, the determination of "who is my neighbor" with a simple relationship (yes or no), represented by zeros and ones in the spatial weights matrix, does not capture the actual influences between "neighboring" households in terms of actual characteristics, actual attitudes, and others that exist between them. With the effort spent in this analysis to obtain the number of spatial and attitudinal variables that I have at the household level, simply arguing that a 0,1 based spatial weights matrix can capture the spillovers or interdependencies between households seems to waste the valuable data collected. Instead of asking "Are you my neighbor, yes or no?" to describe "neighborliness," I ask the question "Are you my neighbor? If so, then what is your square footage? What part of your sales price is not explained by the model?" So, instead of zeros and ones describing my influence on my neighbors, the actual values of dwelling structure characteristics, etc. are used to measure the influence of neighbors.

The theory developed in Chapter 3 suggests that subgroup identification (the three types of households) does not necessarily eliminate the need to use a spatial weights matrix or an error term correction to handle household heterogeneity. Instead of relying on a spatial weights matrix to determine neighbors' effects on parcels of interest, I argue that a GLS-like correction to the error terms is possible through an analysis of the residuals from the SUR model described in Table 5. The goal is to correct for heteroskedasticity in the error terms, not household heterogeneity. In this correction, I assume that the characteristics of neighboring parcels and dwellings, as well as some of the attitudinal information I collected on households, can be used to explain the unexplained variation in sales price from the SUR model described in Table 5. The next section describes this alternative estimator.

4.4 Spatial Dependence Checks and Efficient Spatial Estimator

To check for additional spatial dependence in the error terms from the SUR model, I set up a structural model (simultaneous equations model) to determine the appropriate weights that can be used in a White-style (1980) heteroskedasticity correction of the SUR model estimated previously.[30] To accomplish this, I follow the tradition of Guilkey and Schmidt (1973), who devise an efficient estimator for SUR models with vector autoregressive errors that uses a transformation matrix **R** to weight both sides of the equation. While their original article was intended for time-series data, Guilkey and Schmidt's method is fairly translatable to cross-sectional data where one wishes to model the simultaneous dependencies of error terms in a type of *spatial* conditional heteroskedasticity error approach.[31] In this structural model, I expect that at least one of

---

[30] Generally, autocorrelation is a problem in time-series data while heteroskedasticity is a problem in cross-sectional data. For autocorrelation and heteroskedasticity corrections, Kmenta (1986, p. 296) suggests that the residuals and residuals squared, respectively, be used in the calculation of the weights to be employed in a weighted least squares approach similar to White (1980). However, a White test (as described in Gujarati) suggests that no pure heteroskedasticity exists in the model. So, I fit different forms of the residuals before settling on the residuals squared as the appropriate dependent variables in the structural model.

[31] Other spatial lag variables were tested and rejected as insignificant predictors of the squared residual system. These variables included the nearest neighbor's living space, nearest neighbor's distances to various neighborhood attributes, and others. Including these characteristics of nearest neighbors seems like a spatial weights matrix approach. The difference is that these variables are included as explanatory variables directly, whereas the use of this information in a spatial weights matrix would also include the

the unexplained variations (residuals) of the nearest households of all three types to a particular households will be a significant predictor of the residuals of a parcel of interest. If this occurs, this will demonstrate that the residuals of "neighbors" regardless of type are indeed related, suggesting that additional spatial dependence can be captured using the structural model. Then, the incorporation of this additional information into the transformation matrix **R** will result in more efficient estimates of the marginal prices for each household type.

To test that no additional spatial dependence is present in the residuals from the basic SUR model, I model six simultaneous equations in the style of Guilkey and Schmidt (1973) and Hendry (1971)[32]:

(7a) $\varepsilon_{1i}^2 = \alpha_0 + \alpha_1 * \varepsilon_{1i,ADJ}^2 + \alpha_2 * \varepsilon_{2i,ADJ}^2 + \alpha_3 * \varepsilon_{3i,ADJ}^2 + \alpha_4 * \mathbf{Z} + \mu_{1i}$

(7b) $\varepsilon_{2i}^2 = \alpha_5 + \alpha_6 * \varepsilon_{1i,ADJ}^2 + \alpha_7 * \varepsilon_{2i,ADJ}^2 + \alpha_8 * \varepsilon_{3i,ADJ}^2 + \alpha_9 * \mathbf{Z} + \mu_{2i}$

(7c) $\varepsilon_{3i}^2 = \alpha_{10} + \alpha_{11} * \varepsilon_{1i,ADJ}^2 + \alpha_{12} * \varepsilon_{2i,ADJ}^2 + \alpha_{13} * \varepsilon_{3i,ADJ}^2 + \alpha_{14} * \mathbf{Z} + \mu_{3i}$

(7d) ln (distance to Muslim school) =

$\gamma_0 + \gamma_1 * \varepsilon_{1i,ADJ}^2 + \gamma_2 * \varepsilon_{2i,ADJ}^2 + \gamma_3 * \varepsilon_{3i,ADJ}^2 + \gamma_4 * \varepsilon_{1i}^2 + \gamma_5 * \varepsilon_{2i}^2 + \gamma_6 * \varepsilon_{3i}^2 + \mu_{4i}$

(7e) ln (distance to 14th Street commercial strip) =

$\varpi_0 + \varpi_1 * \varepsilon_{1i,ADJ}^2 + \varpi_2 * \varepsilon_{2i,ADJ}^2 + \varpi_3 * \varepsilon_{3i,ADJ}^2 + \varpi_4 * \varepsilon_{1i}^2 + \varpi_5 * \varepsilon_{2i}^2 + \varpi_6 * \varepsilon_{3i}^2 + \mu_{5i}$

(7f) ln (distance to Home Park green space) =

$\phi_0 + \phi_1 * \varepsilon_{1i,ADJ}^2 + \phi_2 * \varepsilon_{2i,ADJ}^2 + \phi_3 * \varepsilon_{3i,ADJ}^2 + \phi_4 * \varepsilon_{1i}^2 + \phi_5 * \varepsilon_{2i}^2 + \phi_6 * \varepsilon_{3i}^2 + \mu_{6i}$

---

multiplication of the matrix **W** times the error term in the spatial multiplier. The implications of this approach versus the spatial weights matrix approach will be the subject of future research.

[32] Simultaneous equation and SUR models are similar statistically. The reasons for using simultaneous equations here are to model the endogeneity of the dependent variables in equations 7d-7f in a more straightforward way, to assume zero correlations among the error terms unlike SUR which assumes

In this model, which maintains the market segmentation of households Types A, B, and C (equations 7a, 7b, and 7c)[33], **Z**= [Natural log of distance to Muslim school; Natural log of distance to Home Park; yes/no – Are you politically conservative?; yes/no – Did you choose to live in Home Park because it is near public transportation?; yes/no – Did you purchase/rent your particular dwelling because of its windows?; and yes/no – Did you purchase/rent your particular dwelling because of its proximity to entertainment opportunities?]. Also, $\varepsilon_{1i}^2 \ldots \varepsilon_{3i}^2$ are the squared residuals from the basic SUR model, corresponding to Types A, B, and C, respectively; $\varepsilon_{1i,\,ADJ}{}^2 \ldots \varepsilon_{3i,\,ADJ}{}^2$ are the squared residuals of the closest neighboring households of Types A, B, and C;[34] **Z** are vectors of spatial and attitudinal variables that are hypothesized to explain the squared residuals (determined through a series of F tests where each spatial and attitudinal variable was tested for joint significance across the three equations at the .10 level); and $\mu_{1i \ldots 3i}$ are independently and identically distributed error terms with no assumed correlation between them.[35] I hypothesize in lines a-c that the squared residuals can be explained by the squared residuals of the neighboring households as well as certain spatial variables.

---

correlation among the error terms, and to get consistent estimates of the parameters a la Guilkey and Schmidt (1973).

[33] Remember that the determination of household types A, B, and C was not a completely aspatial process, as distance variables were used in the SUR iterative process that generated the types.

[34] Most spatial econometrics models define contiguous parcels based on vertex adjacency, side adjacency, or both (called rook, bishop, and queen adjacency, respectively). Here, I assume that households on either side (left or right) or across the street (straight or diagonally) are considered "neighbors." I assume that backyard neighbors have no influence on the households on other streets. While this might seem problematic in practice, other variables considered to explain sales price, like whether a household perceives crime as a problem in the neighborhood, were insignificant in earlier runs of the SUR model.

[35] The structural model uses the residuals from the SUR model reported in Table 5. Earlier SUR iterations used distance and other spatial variables to account for the influence of space in the neighborhood. The role of the structural model correction is to test for any additional spatial dependence that was not captured in the original SUR model and to correct for any endogeneity that might arise due to the location and development of the Home Park green space, the Muslim day school, or the 14th Street commercial strip.

Then, lines d-f are used to model the other potentially endogenous land uses in Home Park through the distance to the Muslim day school, distance to the 14th Street commercial strip, and distance to the Home Park green space. These three variables are treated endogenously in the residual system because they are not predetermined factors in the neighborhood. The Muslim day school was constructed in reaction to the high concentration of Muslim residents in the neighborhood, which then caused another influx of Muslim residents; the commercial strip was built as residents moved into the neighborhood and required nearby restaurants and shopping opportunities; and Home Park (green space) was built as a place for local residents to walk their pets and enjoy recreational opportunities without having to travel several miles to other green spaces like Piedmont Park. Therefore, since the exact locations of these three neighborhood features were not predetermined before the construction of the neighborhood in 1920, they are modeled as variables in the residual equations. As an example, suppose that a Type A household lives in an area surrounded by households of other types. The first line in Equation 7 says that the unexplained sales price variation of this Type A household is dependent on the unexplained sales price variation of the adjacent households of all types, including other Type A households, and on other variables (**Z**) that are hypothesized to influence the residuals.

The main concern with the structural model is to make sure that order and rank conditions are satisfied, meaning that each equation in the model can be identified. The literature on the identification problem tells me that the order condition is the first concern econometrically. Harvey (1990) says that "Fortunately, the order condition is usually sufficient to ensure identifiability, and although it is important to be aware of the rank condition, a failure to verify it will rarely result in disaster" (p. 328). So, the fact that the model depicted in Equation 7 is exactly identified in equations a-c and overidentified in equations d-f adequately satisfies the order condition.

The first round of results provides little evidence for the endogeneity of the dependent variables in Equations 7d, 7e, and 7f. Using a Hausman (1978) exogeneity test, I discover that I have weak instruments for the dependent variables in Equations 7d-7f, mostly due to high multicollinearity between the independent variables in lines d-f. But, as expected, the predicted values from lines a-c that are included in equations d-f in the simultaneous equations model are significantly different from zero, suggesting that the dependent variables from the other equations are endogenous. But, since Hausman tests reveal the weakness of the instruments for the dependent variables in Equations 7d-7f, I choose to eliminate lines d-f from the structural model, with the residual estimation via 3SLS to include only the first three lines of the original structural model, where each of the independent variables is treated as exogenous to the system. Another reason that this is appropriate is because the modeling of the distances to the Muslim school, 14th Street commercial strip, and the Home Park green space, when a functional form and specification were found that gave me positive R^2 values for all six equations, did not improve the estimates of the first three lines of the structural model in a likelihood ratio (LR) test. The final residual model estimates are reported in Table 7.

Table 7: Residual Analysis in a Structural Model

	Dependent Variables		
	Type A Residuals²	*Type B Residuals²*	*Type C Residuals²*
Constant	46.21*** (11.61)	10.12 (12.91)	31.24*** (10.91)
Independent Variables			
Residuals² of Nearest Type A	.35*** (.07)	.02 (.11)	.06 (.09)
Ln Dist to Home Park	-5.62*** (1.86)	2.15 (2.07)	-1.72 (1.75)
Conservative Politically?	5.56** (2.77)	2.54 (3.08)	-2.30 (2.60)
Near Public Transportation?	8.94** (3.65)	-4.11 (4.06)	-1.74 (3.43)
Windows?	-5.53* (3.16)	-8.96** (3.51)	-5.17* (2.96)
Near Entertainment Opportunities?	1.94 (2.78)	-4.82 (3.10)	-3.26 (2.61)
Other Statistics			
χ^2	42.06 (p = .00)	11.59 (p = .07)	8.21 (p = .22)
Root MSE	22.43	24.95	21.08
R^2	.09	.02	.02
Durbin-Wu-Hausman tests of endogeneity#	F (3,390) = 3.23 (p = .02)	F (3,390) = .97 (p = .40)	F (3,390) = 3.55 (p = .01)

*= Significant at .10 level; ** = Significant at .05 level; *** = Significant at .01 level

In this final version of the model (squared residuals of the nearest Type B and C households were also included in preliminary runs of the structural model), Durbin-Wu-Hausman tests of endogeneity for the first three lines of the structural model include the following explanatory variables in an OLS model: the original independent variables shown in Table 7 plus the saved residuals from an auxiliary 3SLS regression model of only the last three lines of the structural model. Then, I test the joint significance of the three predicted residual variables; all three jointly are significantly different from zero, meaning that an OLS regression of the SUR residuals on the adjacent residuals, etc., gives inconsistent estimates, meaning that either maximum likelihood or generalized moments estimation is required. Similar tests of endogeneity for the last three lines of the structural model, which are reported in Table 7, show the joint significance of the coefficients on the residuals from the Muslim equation (7d) and coefficients on the residuals from the commercial equation (7e) in two of the three equations, suggesting that OLS provides inconsistent estimates.

Since 3SLS and SUR provide identical estimates in this case, lines a-c in Equation 7 are treated as seemingly unrelated. The Breusch-Pagan test statistic of 304 (probability = .0000) tells us that the error term covariance matrix of seemingly unrelated equations is not diagonal (i.e. there are non-zero elements in the off-diagonal elements), which indicates significant dependence among the error terms in this regression; indeed the "residuals of the residuals" are dependent on each other. Treating equations a-c as seemingly unrelated, I get coefficient estimates and other statistics in Table 7. Now, the question is which functional form should be used to analyze the residuals. Using the likelihood ratio (LR) test as the basis for this decision, I start with the full set of independent variables described in Equation 7 and test the explanatory power of different functional form specifications, settling on those variables reported in Table 7.

In Table 7, I see that this choice of independent variables explains 13 percent, 3 percent, and 6 percent of the variation in the squared residuals for Type A, B, and C households, respectively. The basis for choosing this functional form was to increase the explanatory power of the model for Type A households, as it had the weakest explanatory power of the three types in Table 5. I see that the independent variables significantly explain variation in the residuals of all three types even though the focus was on Type A residuals. So, in the spirit of Guilkey and Schmidt (1973), the square root of the predicted values of the dependent variables will be used as weights in the original SUR model as a White-style heteroskedasticity correction. So, the transformation matrix **R** takes the form $\left(\sqrt{\hat{\varepsilon}^2_{ji}}\right)^{-1}$, where j = 1 to 3, and will be used to weight both sides of the SUR model as a GLS correction for heteroskedasticity.

I should note here that these "spatial" weights are based on an earlier iterative sorting process that is not weighted, at least not different from the "weighting" done by the feasible GLS estimator. By controlling for nonspherical errors in this way and not earlier in the sorting process itself, I may have carried some bias into the structural model by not accounting for spatial interdependencies in earlier sorting stages. However, as I mentioned before, the spatial information I had on individual households (distance variables, whether a dwelling was above street level, etc.) was used to sort households into types. I am not sure if the results would differ greatly if the information on "neighbors" were used earlier in the process.

Nonetheless, the weighted SUR estimates are reported in Table 8. Compared to Table 5, I see that the standard error estimates have dropped across the model. Also, no significant coefficient in Table 5 maintains significance *and* changes signs in Table 8, which shows the ability of the residual structural model to correct the standard errors (a second GLS correction if you will that adds efficiency without affecting the consistency of estimates) but not change the directions of influence of the independent variables on sales prices. Additional evidence, such as the generally higher χ^2 statistics and lower Root Mean Square Errors, indicates the strength of the model and efficiency of the estimates.

From these results, one can see that non-spherical disturbances associated with spatial autocorrelation do not appear. And, only a modest form of heteroskedasticity appears, which suggests that the model summarized in Table 5 is an efficient representation of the characteristics that describe sales prices in this neighborhood. All of the spatial weighting that occurs in an Anselin-like process designed to correct for spatial autocorrelation is left up to the researcher's discretion, making it possible for the researcher to miss subtle relationships between various local landmarks and dwelling sales prices or to miss the form of the heteroskedasticity correction that is built into the SUR estimator. In this case, the gain in efficiency in the coefficient estimates through the simultaneous equations model followed a very exhaustive non-spherical disturbance correction exploration.

Table 8: Weighted SUR Estimates (double log specification)

	Type A	Type B	Type C
	Coefficient (Std. Error)	*Coefficient (Std. Error)*	*Coefficient (Std. Error)*
Constant	-.01 (.02)	-.01 (.02)	-.06 (.04)
Structural Variables			
Natural log of Dwelling Square Feet	-.12 (.39)	1.49*** (.57)	.30 (.59)
Natural log of Number of Beds	5.60*** (.50)	-6.19*** (.68)	1.01 (.70)
Natural log of Number of Baths	-1.25** (.49)	-1.43** (.63)	2.27*** (.69)
Natural log of Number of Acres	-.05 (.23)	-.87** (.35)	.79** (.33)
Condominium (yes = 1)	---	---	.25 (.44)
Spatial Variables			
Ln Dist to 14th St. Commercial Site	-1.13** (.50)	2.46*** (.89)	-1.01 (.84)
Ln Dist to Piedmont Park	.50** (.23)	-2.60*** (.47)	1.94*** (.41)
Ln Dist to Meineke Service Station	3.30*** (.49)	-5.34*** (.78)	1.95*** (.71)
Ln Dist to Georgia Tech	-3.20*** (.47)	4.78*** (.71)	-2.32*** (.68)
*Ln Dist to Home Park Green Space*HP <200m*	.81*** (.16)	---	---
*Ln Dist to Home Park Green Space*HP >200m*South Home Park*	.15** (.06)	.74*** (.09)	-.09 (.07)
*Ln Dist to Home Park Green Space*North Home Park*	---	.86*** (.09)	---
Is Dwelling Above Street Level?	-.44 (.28)	-.74* (.42)	1.50*** (.45)
Located on State Street?	-.42 (.69)	-4.55*** (.97)	4.57*** (.98)
*Above Street Level*State Street*	-.37 (1.43)	5.69*** (1.33)	-5.22*** (1.48)
Adjacency Variables			
Lives Adjacent to a Renter	1.12** (.45)	-2.04*** (.65)	.34 (.66)
Lives Adjacent to Undergrad Student	2.66*** (.38)	.01 (.58)	-2.26*** (.56)
Lives Adjacent to a College Educated person	-.69 (.45)	-1.00* (.58)	1.53** (.61)
Lives Adjacent to a Person of a Different Race Than You	-1.71*** (.31)	-3.38*** (.46)	5.16*** (.48)
Lives Adjacent to a Dwelling that made Home Improvements	.72** (.29)	-4.05*** (.45)	3.63*** (.47)
Lives Adjacent to Someone who Lives in Home Park Because it is Near Georgia Tech	1.19*** (.33)	3.62*** (.52)	-4.90*** (.54)

Table 8 continued

Other Statistics			
χ^2	446.11 (p = .00)	753.04 (p = .00)	534.81 (p = .00)
Root MSE	.30	.22	.23
Adjusted R^2	.32	.51	.50

*= Significant at .10 level; ** = Significant at .05 level; *** = Significant at .01 level

4.5 Sorting Process Checks

In the original SUR iteration process that allowed households to re-sort, approximately 50 independent variables were used to explain sales price variation. Now, to check how well the sorting process works, I use the 19 explanatory variables in each equation of the unweighted SUR models (instead of the 50 explanatory variables from the original iteration) in an *alternative* run of the iterative process used to determine the number of household types that emerge when the best functional form of the model is used. Remember that this household sorting process has two parts: the PCA run on demographic and attitudinal variables and the SUR iteration process that allows households to re-sort into the groups that best explain the sales price of particular dwellings. Here, I take the PCA run as given and start at the SUR iteration process. The only difference in this second run is that the variables used to explain sales price variation are only the 19 unweighted variables from the model represented in Table 5, not the full set of 50 explanatory variables used the first time. This process is identical to our first run that gave me three distinct household types *except* that only the variables from the model described in Table 5 are used as the re-sorting criteria after the PCA.

My *a priori* expectation is that even though the model described in Table 5 is parsimonious, it may not be the model that most accurately sorts households into types, as many explanatory variables are omitted in the parsimonious model. While I can think of no formal way to test this expectation, preliminary runs of the model suggest that small changes in the number of variables used to sort households have major impacts on the outcome. The results, in fact, show that the 19 variables from the unweighted model, when run through the SUR iteration, yield two primary household types (48 and 352 observations, respectively).[36] This outcome shows me that, with fewer explanatory variables, the nuanced differences that make these types *seemingly unrelated* are muted. In fact, I see no clear spatial clustering of the two main new types in the neighborhood. But, unlike before, there is no clear distinction between renters and owners in the new types (i.e. households that rent their dwellings are not mostly concentrated in one particular type anymore).

Table 9 reports how the sorting process differs when the full set of variables is used versus the parsimonious set. "Initial Types" in Table 9 are the original Types A, B, and C using the full set of explanatory variables in the SUR iterative process; "Types After the Second Iterative Process" are the types that emerge after the second run of the iteration using 21 of the original 50 variables. If the second run of the iterative process had given me identical results, I would see all of the numbers in Table 9 on the main diagonal.

Table 9: Initial Types (full independent variable set) vs. New Types (parsimonious set)

	Types After Second Iterative Process		
Initial Types	*1*	2	*Total*
A	22	48	*70*
B	11	173	*184*
C	15	131	*146*
Total	*48*	*352*	*400*

[36] Only eight households sorted into the other six PCA determined groups. Only after allowing these eight households to sort into one of the two major groups do I get the result reported here.

Since the sorting criteria differ (I started with different sets of independent variables used to explain sales price variation), it is not surprising to see observations sort into other types. In Table 9, one can see that a majority of original Type A, B, *and* C households were sorted into the new Type 2. One might conclude that fewer explanatory variables for sales price tend to deemphasize the differences among households at the margin, which leads to fewer statistically aggregable types than if additional variables were included that emphasized more differences between households. One might argue also that, with fewer explanatory variables for sales price (or if PCA-stage variables sorted households into a single group where a single price vector dominates), researchers using this technique may reject the SUR model in favor of OLS if such a high percentage of households is sorted into a single type. But, the significant Breusch-Pagan (1980) test provides strong evidence of three distinct household types that are not aggregable into a single type.

One might conclude that the parsimonious model might not sort households as well as the full set of variables. The analogous situation of variable omission bias in mis-specified regression models suggests that the full set of explanatory variables will provide a more accurate sorting of households into types. This is expected because Types A, B, and C emerge from $\Re^{+}_{50}$ space, whereas in the second run Types 1, 2, and 3 are allowed to emerge from $\Re^{+}_{19}$. In the first run, the iterative process is allowed to update based on a set of approximately 50 explanatory variables, whereas in the second run only 19 variables are used in the update process. So, it is not surprising that I get two different sorts when I use two different sets of explanatory variables, even if it is the parsimonious set of independent variables used to explain sales price variation.

Next, I test whether these alternative type categorizations using a smaller set of independent variables make a difference in the explanation of sales prices. If the t tests for either or both of these discrete variables are significant, then I can conclude that this second run of the SUR iterative process sorts households into groups that explain additional significant variation in sales prices unaccounted for in the previous models. I include a dummy variable for Type 2 in each equation in the model described in Table 5. The results, omitted here, tell me that this second sorting process does give me categorizations of households that significantly explain additional sales price variation for the first two equations in Table 5. There is a significant and negative effect on sales prices for Type A households explained by Type 2 ($t = -3.79$), holding all else constant. Also, for Type B households, there is a significant and positive effect on sales prices explained by households that sorted into Type 2 ($t = 2.48$), all else held constant. This tells me that the re-sorting process has characterized households in a different significant way than the original sorting process. But, the changes in explanatory power of the model are very small. The original RMSEs for Types A, B, and C are 3.88, 4.59, and 4.34; with the new type variable, the RMSEs become 3.85, 4.59, and 4.34, respectively. This, along with an across-equation joint F test of the Type 2 coefficient ($\chi^2 = 18.75$, prob. = .0003), shows a significant improvement to the model. But, the coefficient estimates do not change that much from those reported in Table 5 when the discrete Type 2 variable is included in the model. Therefore, I conclude that Table 5 contains the best coefficient estimates given the limitations of the data and the sorting process.

4.6 Other Data Validity Checks

4.6.1 Validation of Self-Reported Dwelling Market Values

Since the inflation of sales prices is arbitrary, and since house appreciation rates for the City of Atlanta may not best describe the appreciation in Home Park, a logistic growth curve was used to update Multiple Listing Service sales data to their estimated current market values. Figure 5 shows the sales prices by year of last sale for each dwelling in the sample, whether it is currently owner-occupied or rented. One can see that sales prices in Home Park have steadily increased throughout the 1970s through the early 2000s. Using a logistic growth curve[37] in the style of Conrad and Clark (1987), I estimate the following equation:

(8) $$SP_{MLS} = K/[1 + \exp(-a * (x-b))]$$

Using Equation 8, which says that the sales price at the time of the last sale for each homeowner (because I do not have complete data on actual market prices of rental dwellings) is equal to the right-hand side function, the actual sales price graphed in Figure 5 are forecasted up to the year 2002. K is an estimated constant that reflects the highest possible sales price of a dwelling in Home Park for this sample; x is the time between the year of last sale and the year for which the update is calculated (2002) for each dwelling; and a and b are parameters to be estimated. After running this model, K is

[37] A logistic growth curve is used to update sales prices because it allows for nonlinearities as well as different rates of growth over time. For example, Figure 5 shows fluctuations in sales price increases, then a sharp increase in the middle to late 1990s. This, coupled with the current economic slowdown, which means upcoming slower rates of increase in sales prices, warrants a technique that can handle these fluctuations.

estimated by the model to be \$347,900, while a and b are estimated to be 16.2139 and 2.9072, respectively. Then, a 3-4 percent difference (at the mean) was calculated for self-reported market values from owners only and projected market values from MLS data. The mean and standard deviation of owners' self-reported market values (\$226,405 and \$58,419, respectively, for 112 owners) are different than those of the MLS update via logistic growth curve for the same group of owners (\$219,240 and \$88,038, respectively), but not different enough to discourage the use of this variable. Therefore, self-reported sales price is used as the dependent variable.

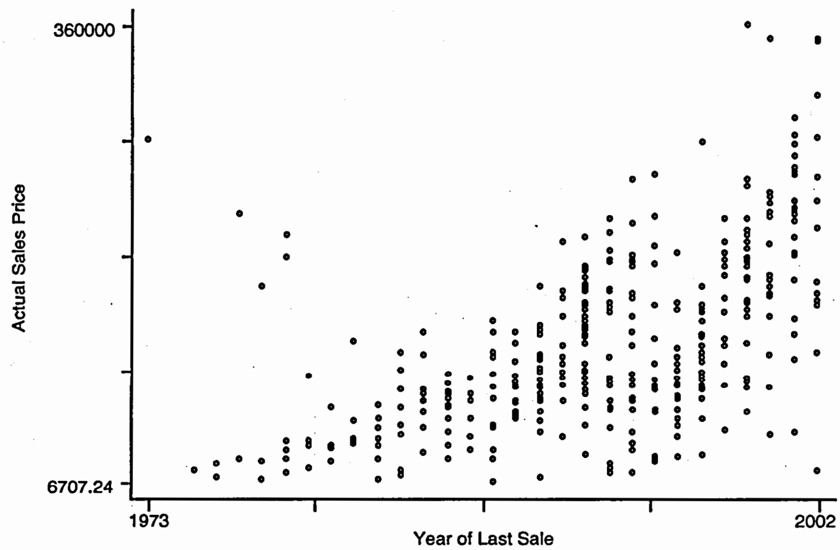


Figure 5: Dwelling Sales Price by Last Year of Sale

4.6.2 Validation of Self-Reported Dwelling Improvements

Building permit summaries (which show information on new startups and upgrades to existing buildings) from Jan. 2000 to Jan. 2003 were requisitioned from the Fulton County Building Inspectors Office. This information was used to verify respondents' answers to the survey question pertaining to residential upgrades in the last three years. I find that 98 different households in Home Park received permits between Jan. 2000 and Jan. 2003. In our survey, 73 homeowners responded that more than $10K worth of upgrades was made. Since renters were not asked about upgrades to their dwellings, the building permits only verify homeowners' responses to this survey question. For the responses that seemed strange, several households had requested permits after the survey was administered; and for several, building houses from scratch did not constitute "upgrades" in their mind.

4.6.3 Out of Sample Comparisons

Twenty surveys were administered to Home Park residents to determine if an out-of-sample bias is present in these results. Given the emphasis here on data disaggregation to learn more about different types of households, the only comparison of the out-of-sample group to the sample households is through descriptive statistics. Table 10 shows the descriptive statistics and difference of means tests for this group. The main concern here is to verify if any statistical differences exist between the full sample with 400 observations and the out-of-sample group of 20 observations. A series of difference of means tests shows that only dwelling sales price and gender are significantly different between the two samples. After checking the self-reported sales prices of the dwellings by type for the main sample, I find that Type A households have the lowest average sales price of $126,582. So, the out-of-sample dwelling sales price average of $111,321 suggests that the additional households sampled either lived in dwellings of relatively lower quality or those sampled lived in dwellings with unusually low rental fees (for example, renters who share part of a dwelling with the owner). Then, for the significant differences in gender between the samples, one must be careful when interpreting this variable. The gender variable represents the gender of the survey respondent, which may not necessarily reflect the gender composition of the household. It is possible for the only male in a household to fill out the survey when actually the household is comprised mostly of women, and vice versa. So, a careful interpretation is required. Nonetheless, this significant gender difference between the samples is not surprising given the significant difference in dwelling sales prices (because of income and status differences across the sexes).

Table 10: Out-of-sample Descriptive Statistics

	Mean	*Std Dev*	*T statistic for difference of means test*	*No. of Observations in Out-of-sample group*
Renter/Owner (Renter = 0, Owner = 1)	.35	.48	0	20
Dwelling Sales Price	111321	46687	**2.84**	14
Age of Respondent	37.82	18.59	-.95	17
Gender (Female = 0, Male = 1)	.35	.49	**2.48**	17
Dwelling Tenure	5.94	9.71	.33	17
Race (Non-white = 0, White = 1)	.53	.44	1.74	17
Full Time Employee?	.35	.48	.72	20
Student?	.41	.47	.37	20

Note: Null hypothesis = no statistical difference between the two samples. All tests are two-tailed using a .95 level of significance, so the critical threshold t statistic = 1.96.

CHAPTER 5

ESTIMATING THE IMPACTS OF LAND USE DESIGNATION CHANGE

Welfare economic theory warns that the improper specification of the first-stage hedonic equation (or equivalently the poor segmentation of markets) can lead to very high R^2 statistics and compromised welfare estimates. Changes in household welfare must be measured with care, so as to prevent value judgments from entering into the interpretations. This chapter briefly reviews the welfare economic implications of any econometric estimates and then estimates the aggregate marginal willingness to pay for two different land use designations. The first, the Home Park green space, is an existing amenity whose value has been fully capitalized into the Home Park housing market. The second, a vacant parcel in North Home Park adjacent to the new Atlantic Station development, is simulated as a "pocket park" that will provide an additional green space opportunity for all Home Park residents, particularly those north of 14th Street.

5.1 Welfare Economic Implications

On a basic level, welfare economics is the branch of economics that measures the impacts on an aggregable group of households of certain institutional changes. These changes may include a policy change mandated by the federal government or the transformation of a particular land parcel from one use to another. Underlying these policy prescriptions is a series of value judgments that are generally accepted by economists. These value judgments, based on the welfare economist's focus on economic efficiency, are social ordering based on individual preferences and the Pareto principle (Boadway and Bruce 1984). First, social ordering is the idea that consumers will rank alternative "states," or bundles of goods at a particular time, according to a set of preference orderings; what makes the ordering of preferences "social" is a series of assumptions that permit individual orderings to be fully aggregated up to the level of a society. Some of the assumptions that have been used to make preference orderings "social" (while they are not necessary for preference ordering aggregation among individuals) include identical preferences and optimal income distributions.

Second, the Pareto principle, named after Italian economist Vilfredo Pareto, in its strictest form says that "if state A is ranked higher than state B for one person, and all other persons rank A at least as high as B, then A should be ranked higher than B in the social ordering" (*Ibid.* 2). However, if a situation occurs where more than one household ranks state A higher than B *and* at least one household ranks state B higher than A, a *Pareto non-comparable* situation arises, meaning that states are rankable only if additional value judgments are made by the researcher.

The point of this discussion is to illustrate the kinds of questions that policy analysts need to think about when deciding how to resolve market imperfections (which in this paper may include cross complementarities between households and spillovers between different land use types). Which households have a negative WTP for additional proximity to the Home Park green space? Which households might move outside of Home Park if a new park is built? These questions focus on the identification of who benefits from particular changes to the neighborhood and who loses.

Recall that when researchers discuss welfare change, they only know an individual household's ranking of states. To measure welfare change, an indefinitely large number of measuring or numbering schemes can be applied that preserve the ordering of the social states. The latter provides economists with the advantage of measuring welfare changes based on the changes in market prices and quantities; this is because the price of a good, even unobserved prices in the case of non-market valuation, measures the marginal benefit to a consumer. But, marginal changes in real estate value alone do not establish a welfare gain! Neither do higher marginal implicit prices for some households indicate *more* welfare gain than other households! So, as I will elaborate in the next section, the objective of this chapter is to simulate the households that remain in or exit from the neighborhood in response to marginal changes in households' distances to the Home Park green space and the proposed park. I use a variation of Freeman's (1995) aggregate marginal WTP calculation to approximate the aggregate changes in sales prices when a new park is constructed. Then, and more importantly for policy analysts, the SUR estimates in Table 5 can be used to estimate the percentage of dwellings that will be occupied by households of different types if a new park is constructed in North Home Park as well as the relative value of the vacant parcel designated as open space, which can be compared to other designations such as commercial and retail uses.

5.2 The Theory of Utility Maximization and its Empirical Applications

What one learns from welfare economics is that changes in household welfare must have normative significance in the sense that if aggregate welfare (estimated by compensating variation [CV] or equivalent variation [EV]) rises, neighborhood residents are better off. Typically, estimates of consumer surplus (the area under the Marshallian demand curve but above the market clearing price) are used to approximate CV and/or EV. Here, I use the marginal implicit prices from Table 5 as measures of the marginal WTP for changes in each independent variable. For example, for Type A and C households in Table 5, living further from the Home Park green space, all else held constant, translates into a marginally lower dwelling sales price. On some level, this *might* indicate that Type A and C households *value* living near the Home Park green space. In hedonic pricing, the assumption of housing market equilibrium suggests that the marginal implicit prices associated with the housing bundle must equal the household's marginal willingness to pay (MWTP) for those characteristics (Freeman 1995, p. 675). One difficulty with marginal implicit prices is that they, along with the quantities of characteristics, are endogenous in the model, suggesting that each individual chooses one point from a price schedule that simultaneously determines that individual's MWTP and quantity of the characteristic. According to Freeman (1995), the most reliable approach for solving the identification problem related to this endogeneity is to find cases where the marginal prices of characteristics vary sufficiently independently of other demand shift variables; or where individuals with the same preferences, income, etc. face different marginal prices. This suggests that market segmentation is one vehicle that can be used to achieve this objective. In essence, this is what the iterative process does – it segments the demand side of the housing market by categorizing households into groups, each of which is treated as a separate line in a SUR model that has its own estimable marginal implicit price vector.

5.3 Impact Assessment 1: Impact of Home Park on Housing Market

Before I calculate the impact of the existing park on the housing market, I should note that, according to welfare economics, that these impact estimates have no clear interpretation. These estimates, which follow Freeman's (1995) method to signal *at the first stage* whether an amenity is something valuable enough to be policy relevant, are used simply to determine 1) whether open space is a significant contributor to the sales prices of local dwellings and 2) if open space is a viable alternative to other land use designations such as retail and commercial. This is measured through the simulated movements of households into and out of the neighborhood.

The SUR model results in Chapter 4 show that all household types do not have a positive WTP to be closer to the Home Park green space (i.e. view it as an amenity). But, city planners, local officials, and local residents would like to know whether the overall impact of the park on dwelling sales prices is positive or negative, and the magnitude of that impact. Earlier a double log functional form was specified for the first-stage hedonic equation, the results of which can be found in Table 5. To illustrate *one way* of determining the impact of Home Park on neighborhood-wide dwelling sales prices, I rewrite the three-line SUR model as:

(9)

$$\begin{bmatrix} \ln SP_i^A \\ \ln SP_i^B \\ \ln SP_i^C \end{bmatrix} = \begin{bmatrix} \mathbf{X}_i^A, hp_i^A & \cdots & \cdots \\ \vdots & \mathbf{X}_i^B, hp_i^B & \vdots \\ \vdots & \vdots & \mathbf{X}_i^C, hp_i^C \end{bmatrix} \begin{bmatrix} \varphi_A, \beta_A \\ \varphi_B, \beta_B \\ \varphi_C, \beta_C \end{bmatrix} + \begin{bmatrix} \eta_i^A \\ \eta_i^B \\ \eta_i^C \end{bmatrix}$$

Here, ln SP^j_i is the natural log of the sales price, $\mathbf{X}^j_i$ (j = A, B, C) represents all independent variables in the SUR model except the variables that seek to identify the impacts of the distance to the Home Park green space (*hp*) on sales prices, φ_j is a vector of coefficients on $\mathbf{X}^j_i$, β_j is a vector of coefficient estimates for the distance to Home Park variables (hp^j_i), and η^j_i is a 3×1 vector of error terms.

To calculate the marginal price of increasing the size of the Home Park green space by one meter on all sides (or equivalently moving each dwelling one meter closer to the park), I treat each line of the SUR model as independent in the calculation of the marginal prices:

(10)
$$\frac{\partial SP}{\partial DIST\_HP} = \left[\beta_{DIST\_HP}(SP / DIST\_HP)\right]$$

Here, marginal prices are calculated by household type and by distance to Home Park in meters *DIST_HP*. For example, for Type A households, the marginal price of the distance to the Home Park green space depends on each household's distance to the park. For households within 200 meters of the park, their marginal prices are calculated using β = .50; for households more than 200 meters from the park but still within South Home Park (south of 14th Street), $\beta = .11$. These coefficients are multiplied by the ratio of sales price to distance to Home Park for each household and are summed across each type to yield the aggregate marginal WTP for each type, which are reported in Table 11. One can see that Type A and B households have a positive marginal WTP to be closer to the Home Park green space, all else held constant. Cumulatively, the marginal monetary value of the park capitalized into the Home Park housing market, according to the SUR model results, is approximately $10,000. However, as I noted in the sections on welfare economics, these estimates have no strict interpretation beyond the fact that this is a significant impact that shows a welfare gain to the neighborhood if the park's size increases by one meter on all sides or if each dwelling were located one meter closer to the green space. Another caveat that sheds additional light on the welfare impacts of the green space is that this estimate of marginal WTP does not include any observations on the new condominium complex located at 401 Tenth Street, where each condo on the north side of the building has a view overlooking the Home Park green space. If these condominium residents are of the same type as other condo owners in the neighborhood (Type C), then the negative marginal WTP for Type C households might be offset (and potentially become positive) by the location of Type C households closer to the park than the average Type C household currently. Also, the impacts reported in Table 11 are for the sample only; since there are approximately 800 households in the neighborhood, the estimated impact on all households in Home Park would be twice those reported here. Also note that the 95 percent confidence intervals tell me that in repeated sampling, 95 times out of 100 the aggregate MWTP would be between $933.91 and $20,098.16.

Table 11: Estimated Aggregate MWTP for Home Park: Summary

	Aggregate MWTP	95% c.i.
Type A	*$1629.41*	*$75.97 / $3,562.94*
Type B	*$8696.58*	*$3,913.46 / $13,479.70*
Type C	*-$339.50*	*$-3,055.52 / $3,055.52*
GRAND TOTAL	*$9986.49*	*$933.91 / $20,098.16*

5.4 Impact Assessment 2: The Effect of a Vacant Parcel Transformation on Types

Next, I use the coefficients estimated for the distance to the Home Park green space variable to simulate the change in the types of households that occupy dwellings if a currently vacant parcel is transformed into a public green space. Using the distances to this proposed park and the reasonable (i.e. consistent with the hedonics literature) coefficient magnitudes in Table 5, I believe the impact of a new park on household behavior (to stay or leave the neighborhood) can be estimated robustly and applied to the entire neighborhood.

Since the impact of this proposed park is simulated, several assumptions must be made. First, I assume that the organization/entity that purchases the vacant parcel (the City of Atlanta or the Home Park Community Improvement Association) does not affect the price at which the vacant land is sold on the existing real estate market that zones the property as fee simple residential use. Second, I assume that the directional impacts of the proposed park on housing prices for each household type are the same as those for the Home Park green space, meaning that the same positive and negative coefficients on the logged distance to Home Park variable are directly translatable to the proposed park.

Third, it is important also to treat the coefficients from the distance to the Home Park green space variable (and the resulting estimates of aggregate MWTP) as upper bounds; the average North Home Park household's distance to the proposed park is much less (572 meters) than its average distance to Home Park (818 meters). If one uses the same coefficients from the distance to Home Park variable to simulate the impact of the proposed park, one must realize that in essence another independent variable is being added to the right-hand side of the model without adding anything to the left-hand side dependent variable, which can exaggerate the average impact of the proposed park on sales price given that no other coefficients are altered in the simulation. Also, as a substitute for the Home Park green space, this new park is likely to decrease the marginal prices of the other park covariates, which may have a substantial effect on which types of households locate to the neighborhood.

Fourth, the double-log functional form of the first-stage hedonic equation is one of many that can be used to estimate the impact of a new park on the composition of a neighborhood and on the sales prices of Home Park dwellings. Other functional forms, such as Box-Cox regression, semi-log, and quadratic forms, are plausible for the first-stage hedonic equation. In this paper, the double-log form is used here to scale down the differences in the numerical values of variables between households and to ease the interpretation of regression coefficients (as elasticity or percent change estimates).

To estimate the dwellings that will be occupied by different household types, I save the predicted sales prices from the SUR model (Table 5). Then, assuming that households of particular types will have the same marginal prices for the new park as for the Home Park green space (at the various distances of less than 200 meters, 200 meters or more, etc.), holding all other estimated coefficients constant, I add the product of the estimated coefficient and the actual logged distance to the new park to the predicted sales prices from the SUR model: $Y_i^{NEW} = Y_i^{SUR} + b * LNDIST\_NEWPARK_i$. So, for Type A households that are 200 meters or less to the new park, $b = .51$; for Type A households that are more than 200 meters away from the new park, $b = .11$[38]; for Type B households, $b = .41$ for those 200 meters or less to the new park and $b = .50$ for those more than 500 meters away from the park; for Type C households, $b = -.02$ for households located more than 200 meters from the new park. Then, a similar statistic to that used earlier to determine statistically aggregable types, $Min\left|\left(\ln SP_i - Y_i^{NEW}\right)/\sigma_i\right|$, is used to determine the types that will occupy dwellings after the new park is constructed.

Adding another independent variable (and SUR-estimated coefficient) to the model without changing the dependent variable presents a problem, that the price vectors for Types A and B are increased and for Type C decreased. This suggests that Type A and B households might be "knocked" into a "higher" price vector that better explains sales prices; the empirical results that would support this idea include a drop in the number of Type A and B households, which might occur as Type B and Type C price

[38] I assume here that all Type A households that are more than 200 meters from the new park have a .11 coefficient, whether they live in North or South Home Park.

vectors better explain sales prices. But, since an accurate estimate of the coefficients on the distance to the new park is not possible (since the park does not exist and its value in the local real estate market has not been capitalized), this exercise serves only as a "proof of the concept" of estimating the dwellings that might be occupied by different household types. So, I only present the changes in types that occupy dwellings in Table 12.

Table 12: Estimated Impact of Proposed Park: Summary

	% Change Types	*Sorting Based on New Park (A,B,C)*
Full value of coefficients	32%	5,283,112
Half value of coefficients	29%	6,256,138

Note: The original iterative process sorted 70, 184, and 146 households into Types A, B, and C, respectively

As one can see here, I estimate that between 29 percent – 32 percent of dwellings will be occupied by different types of households if the new park is constructed. Given that approximately 92 percent of survey respondents report that they wish to move from their current dwelling in the next two years, this estimate suggests strongly that a high percentage of dwellings will be occupied by the same types of households as reside their currently.[39] So, Table 13 shows the distribution of household Types under different coefficient values. As I predicted, the majority of dwellings currently occupied by Type A households become occupied by other household types, mostly Type B households. But, contrary to the prediction, 53 and 37 of the 146 dwellings originally occupied by Type C households (the "highest" price vector because of the large number of homeowners) become occupied by Type B households (a "lower" price vector relative to Type C) under different coefficient structures. In one sense, this result suggests that the price vector associated with Type B "attracts" a large share of households away from the price vectors of Types A and C, as households preference orderings do not change in this analysis. Also, one can see that as the coefficient values get lower (i.e. approach the initial SUR estimates where this new park variable and hypothesized coefficient did not exist), households tend to distribute more closely to the original Types (70 in A, 184 in B, and 146 in C).

In summary, re-sorting occurs because households are allowed to align with other households that have similar implicit price vectors. Non-economists should care about

[39] Interestingly, from the housing survey, one learns that every household in the sample that lives south of 14th Street wishes to move from its current dwelling in the next three years, whereas 69 percent of residents in North Home Park wish to move in the same period.

whether households will move from or stay in the neighborhood in response to a new park; identifying households that value certain land use designations positively is important for planners and policymakers to know if they attempt to attract a certain "type" of resident to an area. Matching the incentives that accompany urban redevelopment efforts with the types of households that respond well to those incentives will make for better policy and planning endeavors that avoid the problems with policy implementation well documented in that literature.

Table 13: Distribution of Household Types with Different Park Coefficients

(a) Full value of coefficients

	Types after Construction of New Park			
Initial Types	*Type A*	*Type B*	*Type C*	*Total*
Type A	4	55	11	70
Type B	1	175	8	184
Type C	0	53	93	146
Total	5	283	112	400

(b) Half value of coefficients

	Types after Construction of New Park			
Initial Types	*Type A*	*Type B*	*Type C*	*Total*
Type A	5	49	16	70
Type B	1	170	13	184
Type C	0	37	109	146
Total	6	256	138	400

5.5 Marginal WTP Estimates

The key feature that makes marginal changes in WTP analogous to marginal welfare change is that the coefficients on the natural log of the distance to Home Park variable are a reasonable estimate of the value of the coefficients on the new park if it were constructed. It is likely that the *actual* estimated coefficients on distance to the new park are *lower* as 1) the proposed park is approximately one-half the size of the Home Park green space, 2) the average distance for all Home Park residents to the proposed park is higher than the average distance to the Home Park green space, and 3) the addition of a new park adds interesting substitution effects as residents have two green spaces from which to choose. An earlier sensitivity analysis showed that the percentage change in dwellings that are occupied by a different type of household is estimated to be between 29 percent and 32 percent. The significance of this change is in what one can say about the neighborhood and its residents. Holding all else constant, if I assume that each household of a particular Type has a marginal WTP for new park proximity (one meter change) that is identical to the marginal WTP for a one meter change in the distance to the Home Park green space (presented in Table 11 as MWTP per household of a particular Type), then the aggregate marginal WTPs for the neighborhood Type compositions (under different coefficient structures) after the new park is constructed are \$2,753,720 and \$2,494,282, which provide a rough estimate of the variability in aggregate marginal WTP when the coefficients change. In this case, both coefficient structures reflect a welfare gain to the households in Home Park if the new park is constructed. This is an extremely important point because bare land zoned for commercial or industrial uses usually commands the highest price in real estate markets.

With this information, this parcel zoned as green space definitely trumps the "value added" by residential and may be a viable economic alternative to retail use. In other words, changing this parcel into a green space could potentially increase overall welfare more than other uses for the same land. In future research, I hope to conduct a similar analysis of the commercial and industrial sectors to estimate the bare land value of real estate in the same geographical area.

To conclude this chapter, one can look at the geographical distribution of household types in Home Park in different scenarios. Figure 6, which shows the distribution of the original household typologies, can be compared to Figure 7, which shows the simulated distribution of household types if a new park is constructed in North Home Park (using the full coefficient values). Here I see no clear clustering of household types with the single exception of the condominium owners at 287 14th Street, who all remain classified as Type C households in the original typology and the typology simulated after the new park is constructed. It seems from these figures that no obvious spatial clustering patterns exists in Home Park currently or after the new park is built.

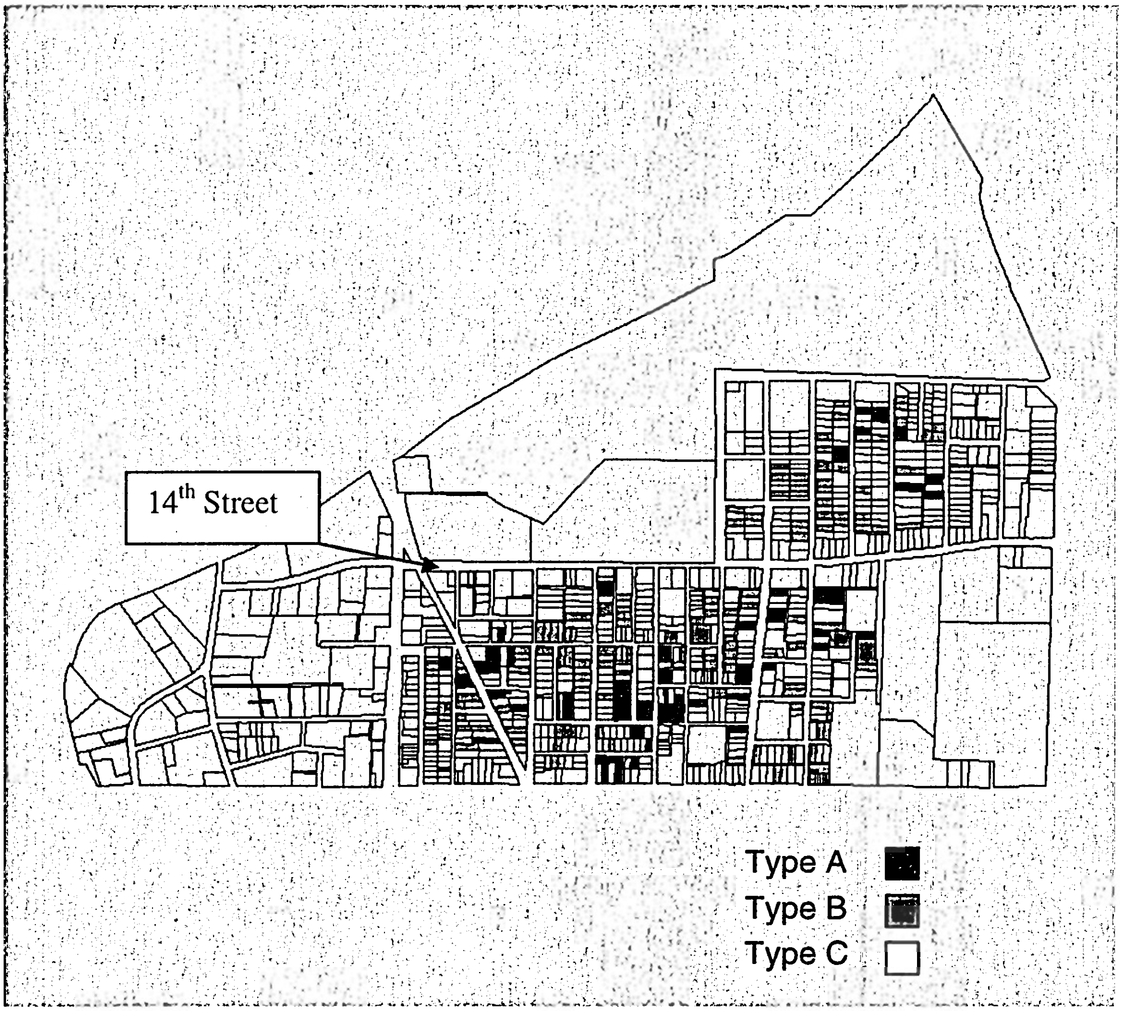


Figure 6: Original Household Types

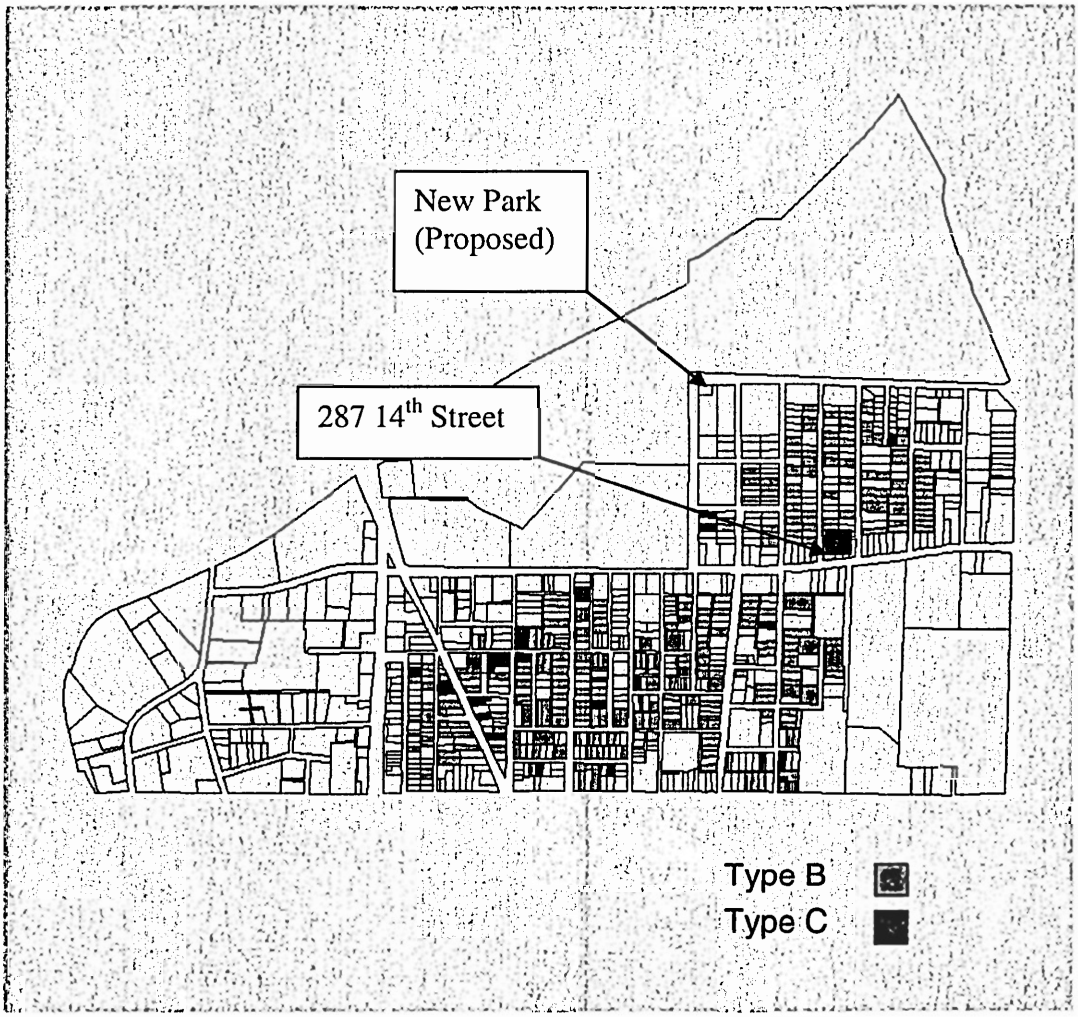


Figure 7: Simulated Change in Household Types (full coefficient values)

CHAPTER 6

POLICY IMPLICATIONS AND CONCLUSIONS

In this final chapter, I discuss some policy implications of the results from Chapters 4 and 5. The extent to which implications reflect the actual estimates in Chapters 4 and 5 are briefly discussed in the section on welfare estimation. Next, after a description of how planning decisions occur in the real world, I discuss the ways in which my method of neighborhood economic redevelopment is better than current planning methods. Then, I discuss the generalizability of my results and methods in terms of benefits transfer and model replication. Next, I discuss several land-use development scenarios in metropolitan Atlanta and the data requirements necessary to produce results similar to those presented in this paper. Then I close with some implications for planners and regional scientists.

6.1 A Brief Summary of the Implications for Welfare Estimation

Tiebout argued that households "vote with their feet" and sort themselves into neighborhoods that most satisfied their preferences for public goods. This self-organization process, in its original rendition at the limit, does not allow for multiple household types to coexist in the same neighborhood space. My idea, that neighborhoods, however defined, simultaneously can accommodate households with different statistically aggregable sets of preferences, is perfectly analogous to Tiebout's argument at the *inter-neighborhood level* that certain citizen-types tend to "cluster" in specific neighborhoods within a metropolitan area. The difference is that I argue for a

Tiebout-style treatment of neighborhoods at a different scale than the original work, at the *within-neighborhood level* in which I address housing supply diversity in dwelling features and local amenities. I acknowledge that in one sense my work contradicts Tiebout's conclusions. But, my paper is written in the "Tiebout spirit" because of my application of Tiebout-style principles to an urban model of a single neighborhood, specifically the incorporation of socioeconomic characteristics and preferences for certain neighborhood amenities into the household sorting process. Modeling households in this way reveals, more accurately I hope, the commonalities in preferences of certain types of households within a single neighborhood. In doing so, the aggregation problem is mitigated enough to permit the household sorting exercise described in earlier chapters; as a consequence, I side-step some of the theoretical and empirical challenges of so-called second stage hedonic estimation.

The importance of identifying household types lies in the welfare estimates one can make. Traditionally, researchers make simplifying assumptions about error structures in first-stage hedonic studies to compensate for preference heterogeneity. In this way, a proper generalized least squares (GLS) estimate can be made to infer efficient welfare change measurements from changes in local amenity access. Common welfare measurements such as consumers' and producers' surplus require 1) estimates of demand and supply at the first-stage (Hanley and Spash 1993) and 2) the assumption that the hedonic price schedule (the intersection of all bid and offer curves) is estimated consistently (Chay and Greenstone 2000) if a second-stage estimate is necessary. Then, a change to the physical structure of the neighborhood, such as the construction of a new park, alters the market equilibrium price and equilibrium quantity such that overall welfare to the demanders and suppliers of the particular good changes. The implicit assumption is that all demanders are similar enough in some respect to be perfectly aggregable, that all demanders have identical preferences - or equivalently that differences are tractable enough to manage in a GLS-corrected first-stage hedonic or orderly enough to allow for a consistent second-stage estimate.

This paper shows that a policy change like the new park in one neighborhood will indeed have different impacts on different classes of households, impacts that would not be noticeable if I had not allowed households to sort and re-sort into statistically aggregable types within the same neighborhood. So, only in the case of perfect or near perfect household aggregability can a *single market* inverse demand curve be estimated to extract implicit marginal prices for a dwelling characteristic or a neighborhood amenity/disamenity. This single market demand curve, to reiterate, masks important information revealed in this research: accounting for preference heterogeneity in a structurally coherent way permits the full power of welfare economic estimates to be used to interpret the changes to dwelling values in the neighborhood.

6.2 Current Planning and Economic Development Decisions

From the planning literature on how decisions are made and who makes them, it is clear that even the most seasoned planners do not have a clearly defined role in the planning *design* process. According to Babcock (1966), planning is comprehensive "fiscal, economic, and physical development" (p. 60) within which zoning is the planner's most important tool. But, the planner's impact on the actual physical development of areas remains limited, as shown by Babcock's comment that "despite the rules, the private citizen is still the catalyst of most changes to a landscape" (p. 41). Peiser (1990) echoes this sentiment when he says that "Planners are the police of contemporary urban America. They enforce the rules, they regulate the process, they stop the offenders. But their influence in effecting change is in recession" (p. 496). Even though planners "no longer create the grand design" (Ibid. 496), their power in the planning process is not completely impotent. Generally, planners serve the interests of the public whereas developers serve private interests; and each tends to view the other as an obstacle to the completion of their desired tasks. Despite these tensions, Peiser (1990) defends that developers will need to rely on planners more than ever given the changing state of cities, particularly the fiscal crises in many areas that limit upgrades to infrastructures necessary for developers' new construction projects [which is a timely concern given the fiscal crises in state and local governments in the late 1990s and early 2000s - see Jenny (2002)], the amount of impact fees charged to developers, the increase in power and influence in homeowner and community associations, traffic and environmental concerns, and the increasing complexity in the development process. A related point is that while the local planning process tends to be "costly, conservative, non-innovative, and highly subjective," it "should not become the basis for regulation under the state's police power" (DiMento 1980, p. 262). Yet, comprehensive plans do provide, in terms of the law, "the Court with a yardstick by which to measure the reasonableness and fairness of the particular action taken by the local authority" (*Ibid.* p. 262).

So, how is planning actually done? The latest guide available to planners with "smart growth" in mind is the *Growing Smart Legislative Guidebook*. Since my concern is the local level, this analysis will focus only on local neighborhood planning concerns. According to the *Guidebook*, the planner first should analyze the existing institutional framework, then design a local comprehensive plan, and finally implement it. At some point in this process a land market monitoring system can be created to track the supply and demand for "buildable" land as well as to identify parcels that may be important to acquire for future transportation projects. Common ingredients of local plans at any scale include Census Bureau demographic data (with the most disaggregate scale being the block group), transportation related data (the street network, measures of vehicle occupancy and congestion, etc.), and economic data (Multiple Listing Service data on dwelling sales prices, dwelling structural characteristics, etc.). The problem with this kind of data in light of the results in this paper is that it is not specific enough to each household (one household per parcel) to permit a coherent impact analysis whenever the *individual household* is the unit of analysis (who wins and who loses in terms of economic welfare). The planner's demographic data are not specific to each household, only to some group of households in the same area.

To illustrate the differences between the outcomes predicted by my model versus that of the planner/economic developer, one only needs to look at Home Park. In creating the Home Park Master Plan, the economic development firm EDAW, Inc. wanted to provide "community members an organizational framework that can guide land use and development decisions within a dynamic urban context (Greater Home Park Master Plan 2001, p. 3). To do this, EDAW collected previously published reports, conducted site visits, and collected data from secondary sources. Of particular interest are these secondary data sources, which include the 2000 Census and CACI Marketing Systems Group. Data gleaned from the 2000 Census include demographic characteristics (based on four block groups that cover approximately the same area as Home Park), age composition, and existing and future kinds of land uses; data from CACI include income trends and distribution across households. In this Master Plan one can see that land use recommendations are based on trends over time (trends that are usually based on data collected at large intervals such as the 1990 Census and 2000 Census) and gross averages over a two to three year period prior to the time of the study. *At no point in this Master Plan are demographic characteristics and attitudes of current Home Park residents analyzed or combined with information on the dwelling features that attracted each household to its particular dwelling.* This, combined with the fact that EDAW, Inc. used previous neighborhood studies and the demographic characteristics from the *1990 Census block groups* to create a Master Plan for Home Park, shows that the planner cannot ascribe a certain set of characteristics to each household (this is an ecological fallacy). Therefore, the conclusions that EDAW, Inc. can make about Home Park are limited to groups of households/parcels, not individual households/parcels; even these conclusions are suspect.

In contrast, the data collected for this paper *at the household level* can be used to determine the best land-use designation for *individual* parcels. From more detailed data on households as well as the local real estate market, I learn in this research that:

- different types of households reveal different willingnesses-to-pay for the same dwelling structure characteristics
- Type A and C households have a positive willingness-to-pay to be closer to

Piedmont Park, a green space outside of the neighborhood but seemingly influential on residential sales prices; it is possible that the effort to triangulate sales prices is responsible for the significance of this variable

- Different "submarkets" with respect to Home Park green space proximity within each type are able to be identified through more detailed household data
- Type B households located above State Street level reside in dwellings with significantly higher sales prices than State Street households on street level – this might be a decent first approximation of the value of a view of the City of Atlanta skyline
- Occupants of dwellings want to move from the neighborhood (more than 90 percent in the survey), a claim supported by the simulation in Chapter 5 that suggests that approximately 30 percent of current dwellings will be occupied by a household of a different type; of course, maybe the addition of a green space that complements the existing green space will entice current residents to remain in their current dwellings, thus the lower percentage in the simulation
- The Home Park green space, with some evidence from Chapter 5, is likely the "highest and best use" of that particular land parcel; as for the parcel that may accommodate a new park, green space may be outbid by retail or commercial uses.

Through discussions with the Home Park Community Improvement Association (HPCIA), key issues and concerns were identified through focus groups and used to inform the construction of the housing survey administered in August 2002. This led to the collection of more specific data relevant to *this* neighborhood, which in turn resulted in more specific conclusions about the neighborhood generally to assess which land use designations are likely to fit well within the overall design plan of Home Park specifically. The focus is not only the physical boundary of Home Park. Indeed the motivations for this paper were external neighborhood changes, a growing Georgia Tech campus, a redeveloped West End area, and of course a multiple-use redevelopment of a 140-acre former steel mill.

The insights offered by my method (neighborhood-specific data and type-specific regression models) over those of the traditional planner are many, but the policy goals are the same. Compared to the data collected by planners who must confront neighborhood planning, my data describe more specific differences between households than would be captured in more macro-level data collected by planners. More household-specific data translates into better decision-making on specific conflicts and tensions within the Home Park neighborhood. For example, in the case of which land-use designation is most appropriate for the vacant parcel discussed in Chapter 5, traditional planning methods most likely would result in that parcel being zoned for a commercial or retail use that would either complement or compete against similar establishments in Atlantic Station. The model developed in this paper suggests that open space should be considered as a viable land-use alternative for this vacant parcel. Also, this method offers a way to empiricize the movement of certain "types" of households into and out of the neighborhood and to rationalize community input as an important factor in the determination of "types." Therefore, this economic optimization approach to land-use planning complements the zoning-driven methods of the planner by including greater involvement of and consultation and dialogue with local residents.

To reiterate, changes in real estate value alone do not establish a welfare gain. So from a policy perspective, the importance of collecting household level data for this study is not to assess or plan aggregate land-use zoning changes at the city level, but to reasonably predict the types of households (and eventually businesses) that are most likely to move into and out of the neighborhood. Zoning officials may care about aggregate land-use zoning changes for taxation purposes (Fischel 1995), but I care about economic welfare changes. Unfortunately, simple real estate value changes nor simple OLS estimation answers a primary planning question – "is the Home Park green space, in the greater context of the neighborhood, the best use of that particular parcel of land?" The range of marginal WTP estimates (approximately $933.91 to $20,098.16) for the green space in Chapter 5 is quite wide. But, in this policy context, it is enough to suggest that the Home Park green space is very likely the appropriate use of that parcel.

6.3 Generalizability of Methods and Results

To assure that one's methods are not suspect, one can check for threats to internal and external validity. Internal validity is present when no other causes can be used to explain why something changes. In this case, internal validity exists if the SUR model is specified correctly – that no other independent variable that influences sales price has been excluded from the model. The SUR model specification in Chapter 3 is consistent with the hedonics literature and adds a few more spatial variables that are less common in the hedonics literature. The primary purpose for the spatial variables that are within the neighborhood (distance to Home Park green space, distance to the Muslim school, etc.) is to identify "submarkets"; the primary purpose of the spatial variables that measure distances to landmarks outside the neighborhood boundaries (particularly the distance to Piedmont Park) is to triangulate sales prices – to find any other directional changes in the hedonic price surfaces. For this reason, the coefficients on these variables are not interpreted strictly as the coefficients on structure characteristics such as the number of acres or number of bedrooms. So, for the spatial variables outside the neighborhood, one might say that I have internal validity but not construct validity.

In terms of generalizability (or external validity) of the methods and results *to other neighborhoods*, some caveats apply. First, the generalizability of the methods used in this paper (the use of focus groups to identify key issues in the neighborhood, the housing survey design, the theory and econometric method developed) to other neighborhoods is very high. It is my experience that residents generally care about the livability of their neighborhood and wish to know their neighbors. Unlike existing planning models often saddled with the same data from neighborhood to neighborhood, the approach here would clearly tailor concerns of interest to the neighborhood directly into the survey instrument. Also, the theory and empirical translation of that theory developed in this paper are generalizable to other neighborhoods seeking to determine the "highest and best use" of particular land parcels.

The generalizability of the particular implicit price results to other neighborhoods is more difficult. For neighborhoods in the City of Atlanta that have similar infrastructure features (sidewalks, green space, good access to public transportation, and nearby shopping opportunities) and similar demographic and economic compositions, some of the coefficients can be used to estimate welfare gains if certain kinds of development activity enter the neighborhood in the spirit of "benefits transfer"; and it remains to be determined how far analysts can go. However, as other neighborhoods become more different from Home Park in multiple dimensions, the generalizability of coefficients becomes more suspect for all the reasons that motivate this study. For example, the use of the coefficients on dwelling structure features in the Bankhead neighborhood (in West Atlanta) may be appropriate across household "types." However, the use of the coefficients on distance to the nearest brown industry (Meineke establishment in Home Park) would be highly suspect due to the prevalence of brown industries in the Bankhead neighborhood. Again, with respect to the distance variables, the point of triangulation is to discern a statistically significant spatial pattern in sales prices. Therefore, these coefficients are unlikely to translate well to other neighborhoods.

Next, *at the city scale*, the generalizability of the methods and results are even more suspect. The method proposed in this paper is geared toward neighborhood planning, not city or regional planning. Planning on a more aggregate, city scale requires different data, such as decennial Census data aggregated by ZIP code or Census tract, that answer more large-scale questions about land-use mixes. Large-scale planning captures the "macro-market" effects of new employment clusters, new road infrastructure, etc. on neighborhoods, tracts, and ZIP codes generally. But, this brings up the original question of planning appropriateness, estimation feasibility, and implementation underscored by Babcock (1966), Babcock and Siemon (1985), and Peiser (1990), which seem almost bereft of references to the literature and warnings of policy implementation in public management and policy highlighted by Pressman and Wildavsky (1984), Majone (1989), Sabatier and Jenkins-Smith (1993), Grin and Van De Graaf (1996), and Farmer, Lipscomb, and McCarthy (2003). In contrast, my approach better captures the "micro-market" effects of new parks, new public transportation access points, new commercial strips, among others. To analyze an entire city using my approach, the data requirements alone would take several years and millions of dollars to complete. Then, the calibration of an econometric model that follows the one developed in this paper would likely exhaust the total degrees of freedom if one were to include all of the theoretically relevant independent variables. An even more parsimonious model with fewer independent variables than the one presented in this paper would not capture the differences that might classify a household as a different type. This kind of model misspecification repeated throughout a city would produce suspect results that might lead to horrific outcomes. For these reasons, my model is strictly geared toward analyses of neighborhoods; but the ideas from the model can be used to complement city planning models through the incorporation of real estate market information (i.e. implicit prices for various dwelling structure characteristics and local amenities) otherwise not included in planning models.

To conclude this section, I briefly discuss the concept of generalizability to theory. Yin (1994) says that while studies do not necessarily need a minimum number of cases, or randomly selected cases, researchers should work with the situation that presents itself. Then, the researcher should structure the best possible study given the data. In this paper, data on a random sample of 400 Home Park households were collected and analyzed using various quantitative methods (basic descriptive statistics to regression analysis). Variables that are theoretically consistent with the literature were used to sort households into multiple types, each of which was analyzed as a separate line in a seemingly unrelated regression model. The variables used to explain sales price in the SUR model were consistent with the hedonics literature, including the spatial variables. Then, these results were used to simulate changes in dwelling occupancy if a new park were constructed in the neighborhood, the results of which were consistent with descriptive statistics.

6.4 Assessment of Urban Planning Alternatives

In this section, I assess three urban planning alternatives in the Atlanta metropolitan area: the *a priori* zoning rules for Georgia 400 (a toll road); the proposed Northern Arc that will connect Interstate 75 to Interstate 85 between Cartersville and Gwinnett County; and the widening of Northside Drive to connect Atlantic Station with the Georgia Aquarium and Centennial Olympic Park. The idea here is to showcase the usefulness of my model in light of these road projects. The use of road projects, I believe, best shows how the model presented in this paper can be used to enhance current planning models and lead to more desirable outcomes for transportation officials and planners.

6.4.1 Georgia 400 Corridor

Georgia 400, a north-south corridor that connects downtown Atlanta with the northern Fulton County suburbs, was designed to alleviate traffic from the Interstate 285 perimeter and provide an easier commute for residents in the northern Fulton County suburbs. However, Georgia 400 is now recognized as a planning failure in that the land nearest to the corridor was incorrectly zoned as commercial, which causes much traffic congestion during daylight hours, but particularly the "rush hours" during the week.

Using the exact model developed in this paper would not be appropriate for a land-use analysis of the Georgia 400 corridor. A slightly different form of the model that would be appropriate for an analysis of the Georgia 400 corridor would include the following data: general demographic characteristics of households within two or three miles of the proposed corridor (no information on individual households is required); general economic data on households within two or three miles of the proposed corridor (again, no information of individual households is required); data on the price per square foot of space zoned retail, commercial, and residential (with the price per square foot in space zoned residential determined from a regression model similar to that developed in this paper that uses mostly publicly available Multiple Listing Service data); and data on the prices of square feet of space for each land-use designation (retail, commercial, and residential) if they are adjacent to one of the other land-use designations (to account for cross market complementarities in the real estate market). Then, this data can be used to determine the economically efficient land use patterns for parcels designated to be within the Georgia 400 corridor.

The use of my model developed in this paper to analyze the Georgia 400 corridor would suggest a different land-use pattern than exists currently. Similar to the land-use patterns in Home Park, my model would suggest that mostly residential housing (interspersed with retail and commercial establishments) locate closest to Georgia 400 as a buffer that separates the highway from the retail and commercial districts. In this situation, traffic whose destination is a particular interchange off of Georgia 400 is allowed to drive through a residential district before it encounters typically traffic intensive retail and commercial parcels. This situation avoids the bottleneck conditions emblematic of planning and zoning decisions that place retail and commercial parcels very close to highway interchanges.

The counter-argument here is that placing residential housing closest to the highway and major arterial road only will shift the congestion burden from retail and commercial establishments to the residential housing sector. While this is a valid argument, road congestion arises when too many traffic intensive destinations are placed very close to one another. The residential housing "buffer" is appropriate as it limits the number of destinations for households that seek retail and commercial destinations. Also, my model would suggest that public transportation is a viable alternative that maintains traffic dispersion instead of traffic attraction. However, the availability of public transportation was fairly uniform throughout Home Park and therefore was not considered to vary enough from household to household to enter the model directly.

The key to urban redevelopment, as I stated in Chapter 1, is to reinvest funds in the city – make the city as attractive in terms of amenities as suburban areas. Only then will traffic congestion problems and the development of blighted areas (i.e. brownfields are included here) truly be able to be solved. Then, once these general land-use designations for an area have been established, one can use a model (and the more detailed household-level data) very similar to the one presented in this paper to determine the "highest and best use" of each individual parcel. In the case of Georgia 400 interchanges, it is possible that some of the results from the econometric model presented in this paper, such as the coefficient estimates for being near green space, are transferable to other neighborhoods. Others, however, like the coefficient estimates for distance to the Georgia Institute of Technology or the distance to the nearest brown industry, may not have the transferability capabilities as other coefficients for estimation purposes. The degree of coefficient transferability depends on the physical design of the neighborhood, to some extent the racial composition, and the distribution of income across households.

6.4.2 The Northern Arc

The Northern Arc is a proposed intercounty highway that will have one interchange per county. Its primary objective is to reduce commuting times for those who travel between Interstate 75 in Bartow County, Cherokee County, Georgia 400, and Interstate 85 in Gwinnett County. In addition to the interesting eminent domain issues that have taken place over the past six years (the State of Georgia has taken possession of many acres within the proposed corridor), proponents of the Northern Arc argue that the road is a necessary step to alleviate traffic congestion on nearby interstate highways. Critics of the Northern Arc disagree on grounds that alleviation in traffic congestion will be short-lived compared to the urban sprawl that will encroach on the areas near the Northern Arc a few years after its development. Both sides have made very good points on the subject. However, as of this writing (November 2003), the Northern Arc is not under consideration by Governor Sonny Perdue.

So, what data are required to determine the "winners" and "losers" in terms of welfare gains if the Northern Arc is eventually constructed? As with each scenario described here, the key is that *simple aggregate increases in sales prices do not imply that all households are "winners" or experience welfare gains!* To measure who achieves welfare gains/losses, a model similar to the one described in this paper must be constructed. Earlier in this chapter I stated that my model was less than appropriate beyond the neighborhood scale. For an analysis of the Northern Arc, I am suggesting that the same kind of econometric methods (PCA and SUR) can be used, but the data requirements will be less intensive. In fact, I see the data requirements for a Northern Arc analysis to be very similar to those for an analysis of the Georgia 400 corridor (please see the previous section). Georgia 400 is a road with many interchanges that permit commuters and other travelers relatively easy access to retail, commercial, and residential parcels. The Northern Arc is different in that it is intended as a limited access road with a single exit per county, which suggests that only landowners in the proposed path of the road and the parcels immediately near the proposed interchanges should be included in a determination of "winners" and "losers."

The estimation of welfare changes in this scenario can be accomplished in at least two ways. First, recipients of welfare gains can be determined through a change in sales price approach where current market values are estimated before the construction of the Northern Arc [which also accounts for anticipatory changes in sales prices - see Kiel and McClain (1995)] and compared to the estimated market values of the same dwellings after the Northern Arc is built. In this case, an increase in sales price holding all else constant would suggest that the presence of the Northern Arc increases sales prices for certain classes of residents. The aggregation problem in this situation might be mitigated by identifying two household "types:" those who will be forced to move due to eminent domain reasons and those who do not move. Then, overall welfare gains to all households affected by the Northern Arc's construction can be determined by the sales price increases to households who are not forced to move; if the sum of these sales price increases is greater than the costs incurred by those households forced to move, then a potential Pareto improvement situation arises where all households can be made off better. This approach, however, requires a lot of data on individual households.

A less data-intensive way to estimate the impact of the Arc on households is to randomly survey landowners and other residents near the path of the Northern Arc development for their willingness to pay (WTP) for the construction of the road or willingness to accept (WTA) compensation for the road that "injures" them in some way. The data requirements for this option are less stringent than the first option, but the estimates are of WTP or WTA, not true changes in welfare. In any case, these WTP/WTA numbers are directly revealed (as opposed to indirectly revealed calculations of hedonic marginal prices) through survey responses, but do not rely on actual market data (market sales price) to estimate the impact of the Northern Arc.

6.4.3 Northside Drive Corridor

A project designed to widen Northside Drive between Atlantic Station and the Georgia Aquarium would require household-level data (demographic, attitudinal, and economic) very similar to that collected in this paper for each neighborhood along the Northside Drive corridor. Since Northside Drive is a major State highway that tangentially services several neighborhoods (Underwood Hills, Berkeley Park, Home Park, Marietta Street, Vine City, etc.), and since each of these neighborhoods is different in terms of residential composition and local amenities, a more detailed analysis is required to determine those households likely to prefer a widened Northside Drive. In addition to data on households, an analysis of Northside Drive would not be complete without additional data on local retail and commercial establishments. In fact, Northside Drive itself is not adjacent to many residential parcels. So, compared to a study of Home Park in isolation, an analysis of Northside Drive must contain information on the cost per square foot of retail and commercial space, data on other establishments that are complementary (e.g. the location of an ink supplier close to a printing establishment), and data on the reasons for particular establishments to locate to their particular parcels. In summary, results on a neighborhood-by-neighborhood basis are required to discern the "winners" and "losers" in terms of welfare gain.

The planning implication of this "winner/loser" designation is that "winning" households might be made by law to compensate the "losing" households in a (potential) Pareto improvement scenario. One option, although it seems fraught with difficulty, might be to charge higher property taxes to the "winners" and lower property taxes to the "losers" in amounts such that the "losers" are not made worse off by the widening of Northside Drive. A second option might be the assessment of "impact fees" on users of Northside Drive. To avoid congestion issues associated with the collection of fees at the point of "impact," these fees might be paid on a yearly basis at the same time as property taxes. This is definitely the most economically efficient way to pay for the widening of Northside Drive, as well as a good way to identify those who benefit from the use of the road. This method makes it difficult for a household to "mask" its true values of Northside Drive. So, while the details of assessing these impact fees is beyond the scope of this paper, the idea of households "paying for what they use" in terms of road usage is appealing and, if prices are adjusted high enough to discourage certain classes of travellers (single occupancy vehicles, for example), then public transportation access on

Northside Drive may become a viable alternative to travellers between Interstate 75 and the Marietta Street and Centennial Olympic Park corridors.

To conclude this section, this paper suggests several approaches beyond traditional planning techniques to inform local neighborhood economic development. First, whereas planners concern themselves with devising comprehensive plans that preserve a certain set of social values based on one scale (such as the city level), households in this paper are allowed to express their social values or preferences for certain local community features through their answers to a neighborhood housing survey. Again, more specific data about households permit the drawing of finer distinctions between households that locate to a particular neighborhood; and the econometric model accounts for the fact that small changes in demographic characteristics or attitudes toward certain neighborhood features can cause a household to be classified as a different "type." This paper suggests that there is more diversity within a single neighborhood than one might think originally, and certainly more than Tiebout (1956) homogeneity predicts. This in itself is a significant finding, especially when one considers that all current research known to me indicates that no one is looking at demographic characteristics and attitudes of individual households to motivate the market segmentation argument.

In addition, the fact that households have different preferences suggests that one might learn in this research *how* and *why* households vary in their preferences. One suggestion for *why* preferences differ can be seen in the coefficients on dwelling structure variables in Table 5. The suggestion that Type A households prefer more bedrooms (liberal interpretation of a positive coefficient) to additional square feet of living space or the number of baths (liberal interpretation of a negative coefficient) is an important discovery about these households. That Type A households might be dissuaded from the rent or purchase of a dwelling that has fewer bedrooms but more living space is very important to real estate agents and absentee landlords. This kind of information throughout Table 5 can be used to identify those households that are more likely to prefer living in a given dwelling.

Second, the form of land-use planning (based on economic optimization) described in this paper parallels the effort of the real estate market to ascribe certain land-use designations to particular parcels. Whereas traditional planning can be viewed as a set of institutional constraints (e.g. 20 percent of all land within the city must be designated as open space) that may or may not mitigate the externalities associated with the real estate market (which in the extreme case may allocate all land to the use that commands the highest price, like industrial uses), my method does not account for these institutional constraints. Instead, my method takes the existing land-use configuration and determines the value of a particular vacant parcel if instead it were designated as open space. By extension, if I collected data on the local commercial, retail, and industrial real estate markets, I could determine (with some confidence) the kind of land use that is most likely to consume a vacant parcel given the current land-use scheme in the neighborhood. As a result, local neighborhood planning is enhanced as externalities are accounted for directly in the econometric model instead of arbitrarily imposed by local planners.

6.5 Implications for the HPCIA and Planners

6.5.1 HPCIA

The results suggest some implications for the neighborhood. First, HPCIA can target its fundraising efforts better once it knows which households are likely to value proximity to green space more than others. The results tell me that Type A and B households have a positive marginal WTP for proximity to the Home Park green space. Assuming that the direction (and maybe the magnitude) of the influence remains the same for the new park, one can say that these households are more likely to contribute funds to purchase the land designated for the new park than Type C households, who with such a small negative aggregate marginal WTP for proximity to the Home Park green space, may merely have an existence value for both parks; Type C households may not be willing to pay to be closer to the green space or the new park. But, having the parks "available" may entice some Type C households to contribute funds to the purchase of the land designated for the new park. In practical terms, there is always a problem with taking a result for a group of people and attributing it to all of the members of that group. The targeting of certain households for fundraising contributions is a general scheme that can be followed by those seeking funds if one chooses to follow the coefficient signs derived from this model. Otherwise, one may blindly solicit funds from the entire neighborhood without knowing which households are more likely to support an additional green space, for example. So, if the HPCIA can raise enough funds to purchase the vacant parcel for an amount less than the total increases in sales prices of the Home Park dwellings, the total consumer surplus of the neighborhood increases with this parcel transformation. An added benefit of this arrangement is that HPCIA, not the City of Atlanta, controls the destiny of this green space, which provides more local control over this particular parcel.

Second, regarding the political climate in Home Park, two neighborhood associations prevail. The Home Park Community Improvement Association (HPCIA), comprised of mostly homeowners, is the City of Atlanta's officially designated metropolitan planning organization (MPO) for the Home Park neighborhood. The other group, the North Home Park Coalition (NHPC), is another policy voice that desires to maintain the section of Home Park that is north of 14th Street as a purely "residential" community. NHPC believes that Atlantic Station (a large mixed-use project scheduled to be completed in the summer of 2004) poses a great threat to the relatively quiet residential atmosphere of Home Park. In particular, this group believes that a south entrance (exit) to (from) Atlantic Station through the Home Park community (via State Street, Atlantic Drive, Holly Street, Barnes Street, and/or Francis Street) will cause an unacceptable amount of noise pollution and greater escape access for criminals. NHPC favors cul-de-sacs being placed at the proposed entrances to the Atlantic Station site, so that areas in Home Park north of 14th Street retain their residential flavor. HPCIA, in contrast, opposes these cul-de-sacs because they limit access to Atlantic Station from the south. Easy access to Atlantic Station, argues HPCIA, will offset potential increases in crime with higher home values and higher rental fees that encourage higher-income households to locate in Home Park. In the language of economists, HPCIA desires to maximize rents by eliminating cul-de-sacs in hopes of driving up the home values in Home Park, which happen to be owned by a significant number of HPCIA members. However, the problem with higher rental fees is that the current population of students that resides in Home Park is likely to be displaced. As a result, other neighborhoods near Home Park (i.e. Berkeley Park, Collier Hills, Underwood Hills, and Loring Heights) are likely to accommodate more students if rental fees increase after Atlantic Station opens.

Third, the SUR estimates in Chapter 4 suggest that different household types will perceive different sales price impacts of a physical change to Home Park parcels if that change alters the numerical value of one of its independent variables (e.g. if the construction of the new park decreases the distance a particular household would have to travel in order to access a green space). In the case of the cul-de-sac issue in North Home Park, I expect the approval or disapproval of cul-de-sacs to have a significant impact on sales prices and the types of households that will move into Home Park. The construction of cul-de-sacs will simultaneously 1) provide one less outlet for criminals looking for escape access from the neighborhood, 2) increase the driving distances from these North Home Park residences to Atlantic Station and all other points north, and vice versa, 3) impose physical and mental "barriers" that separate North Home Park from Atlantic Station, among others. Therefore, once the cul-de-sacs have been constructed and capitalized into the housing market, the change in sales prices at the margin will depend on the degree to which these three factors and others are viewed as benefits or liabilities to current neighborhood residents. Unfortunately, welfare economics says that net benefits of changes to the neighborhood can be estimated only if one assumes that the neighborhood is static, that no one moves from or moves into the neighborhood.

Fourth, what should HPCIA do to vacant lots? As mentioned in the literature review, a particular type of land use for a particular vacant lot would maximize the total dwelling values of surrounding homes/apartments in their current state; one goal could be to find that particular land use that maximizes total value across the neighborhood. The aggregate marginal WTP in Chapter 5 shows one way in which an econometric model can be used to determine the "highest and best use" of a particular parcel under the assumption of neighborhood stasis. Then, in essence, one can do incrementalist city planning in the Lindblom style according to a set of rules established by planners, real estate developers, local government officials, and neighborhood residents. These rules might look like the maximization of an objective function (i.e. maximize the total sales prices of all parcels in Home Park) subject to some constraints, such as the current land use designations, the amount of green space in the neighborhood, planning rules, local laws, etc. Then, one can model the changes to the existing land use pattern if a change to a particular parcel is made. While these estimates limit the amount of analysis and implementation that can be done without knowing more about the cross-market relationships between different land use designations (i.e. residential versus commercial, commercial versus industrial, residential versus industrial), a good start has been made. These estimates are valuable data that can be used by local decision-makers in their policy decisions.[40]

6.5.2 Implications for Planners

Planners generally operate at scales larger than a single neighborhood. Their

[40] The problems here are not new and echo concerns for benefits transfer in other non-market valuation (NMV) studies, such as Bingham *et al.* (1992), Lampietti (1995), and others. The most comprehensive study I found is by Rosenberger and Loomis (2001), who present guidelines for the use of particular estimates from past NMV studies to analyze current management and policy actions. In their report, benefits estimates for 21 recreational activities are summarized for studies published between 1967 and 1998.

planning attempts seek to identify an appropriate mix of land uses across some physical space. Usually some kind of stakeholder analysis (focus groups, statistics about the neighborhood, etc.) is used to gauge the kinds of land uses that might fit well in the overall planning scheme. Typically, though, the data used to formulate those plans do not permit a *systematic assessment of households' preferences* for certain kinds of developments, nor do they identify any value of willingness to pay for particular local amenities. This is a problem because planners cannot devise a plan that achieves the basic objective of planning – creating the optimal zoning mix - without having some knowledge about the right mix of amenities that will attract the desired households to a neighborhood. One result from this paper, that different household types perceive different impacts on sales price of the same dwelling structure characteristics, etc., suggests that current planning is not as effective as it could be if planners could identify those who benefit and those who lose (in terms of welfare) when certain policy changes occur; policy changes include zoning changes that re-designate particular parcels from one use to another. So, until an econometric analysis like this one is conducted to link changes in sales prices at the margin to proximity to certain kinds of land use, planners will continue to guess about what mixes of land uses are most appropriate in an area and the dynamics of what types of households are likely to be enticed to the neighborhood.

The next logical question is what happens if I estimate welfare changes across multiple urban neighborhoods as in the case of the Northside Drive corridor described earlier? Similar to "benefits transfer," households of one type can be compared only to households of the same or similar type. To compare one type of household to a different type is to violate the aggregability conditions defined throughout the sorting process.

True welfare economic comparisons can be made only on the same group of households whether one is studying a one or more neighborhoods. If some households of a particular type move out of a neighborhood, then estimates of welfare *change* will not be based upon the same set of households; in other words, *ex ante* and *ex post* dwelling sales prices are not comparable unless the neighborhood is assumed to be in a period of stasis (all households remain in the neighborhood); yet the violation of this assumption is almost an object of urban redevelopment itself where households are classified into different "types" or the mix of residences changes. This makes strict comparisons of welfare changes related to changes in amenities/disamenities invalid. So, for the proposed park in North Home Park, one can only estimate the changes in welfare for the neighborhood as it is currently inhabited and then report demographic shifts. The establishment of the new park will certainly entice new residents and force out others, meaning that the reach of welfare economics as a tool for measuring neighborhood economic value change is limited but not impotent. This is because the welfare estimates for one set of households are unique to that set of households, and that a change to that set causes changes in neighborhood composition and the "way" that welfare estimates are calculated.

While the reach of welfare economics does not extend beyond the neighborhood's ability to remain static, the power of welfare economic estimates should not be dismissed. The estimates obtained from the first-stage hedonic equation illustrate households' different marginal WTPs to have more or less of the local amenities and dwelling characteristics *and* the potential difficulties associated with bringing together the appropriate stakeholders. The fact that households are systematically different in multiple dimensions indicates that some common policy ground for future neighborhood redevelopment may be a challenge. But, with open discourse and a commitment to compromise, the path chosen for neighborhood redevelopment will be palatable to all parties involved.

The results presented above also suggest that making large, capital-intensive neighborhood redevelopment plans based on estimates at a different scale (a different level of aggregation than that on which the redevelopment is to occur) is very risky. Often one will hear of attempts to solve a problem at one scale, say the household level, using estimates based on another scale, say, the city level. In the case of single neighborhood redevelopment, the appropriate scale or unit of analysis is the household. Therefore, data to be collected and analyzed should reflect household preferences at the household level. Since the smallest scale of publicly available data is the block group, one must collect data at the household level when attempting this level of detailed neighborhood analysis. This suggests that any conclusions drawn from this research are specific to the Home Park neighborhood, are mildly generalizable to other nearby residential neighborhoods, and are very suspect to generalization to the city level. Yet, the approach itself is replicable to other neighborhoods and cities.

Welfare estimates for Home Park are also affected by other land-use designations (commercial, industrial, and retail) within the neighborhood and in neighboring areas. The externalities between households and other land use designations within the same neighborhood have been captured in the hedonic pricing model presented above, where the distances to commercial/retail enterprises and the local brown industry explain sales price variation. Positive and negative coefficients on these explanatory variables indicate whether households perceive these local features as negatively or positively affecting sales price, all else held constant. While *intra-neighborhood* externalities between parcels have been captured in the model in various ways, most *inter-neighborhood* externalities, particularly between the residential housing sectors of multiple neighborhoods, remain unmodeled. Given Home Park's relative isolation from other residential neighborhoods, an immediate incorporation of the externalities *between neighborhoods* is beyond the scope of this research largely because retail and commercial units serve consumers in several adjacent neighborhoods. In the future, however, data on another local neighborhood will be collected, which permits a similar analysis across two neighborhoods in Atlanta along the Northside Drive corridor. To have such micro-level data for two neighborhoods makes for interesting analyses over the next several years. Even though this current research focuses on one neighborhood, the econometric analysis is easily extended to multiple neighborhoods in the same metropolitan area where an SUR model of households would remain an appropriate econometric analysis tool.

Ideally, if I could get price data for certain kinds of land uses (commercial, retail, and industrial), I could account for the cross-market effects of being adjacent to or near these other land use designations in a systematic way to model the land use patterns of non-residential sectors as well. In this case, a full system of simultaneous equations may be possible rather than the unmodeled causes of correlation between equations in the SUR model. In addition, using Home Park data as well as data on the neighborhood to the north (Berkeley Park), I can test the urban economic assumption of monocentricity (that prices decrease with increased distance to the central business district [CBD]), test the difference between two neighborhoods with similar percentages of renters and homeowners, and estimate true welfare gains in each neighborhood for activities such as the construction of a major road that increases each neighborhood's accessibility (however defined) to certain kinds of activities and physical places. The main caveat with across-neighborhood analyses, as alluded to earlier, is that households must be of a similar enough "type" for one to draw any static comparisons.

One final implication for planners is that they generally do not seem as concerned about issues related to policy *implementation* as they do their traditional planning methods. By implementation I mean "to ensure actual fulfillment of concrete measures" (Merriam Webster Dictionary <http://www.m-w.com/cgi-bin/dictionary>). The policy implementation literature is replete with examples of good planning but bad implementation, barriers to successful implementation in different contexts, among others. This literature also in part suggests that planning, implementation, and evaluation are virtually co-temporal, non-separable activities. Planners seem to assume that the problems associated with the implementation process are either not there or will not hinder their plan. The evidence of this includes planners' citations of very few (if any) key articles and books on policy implementation. If anything, planners should be reading this literature and be aware of the implementation concerns of others working on similar projects. The potential hurdles (mainly political, bureaucratic, and/or financial) that may hinder the implementation of a plan should be considered early in the planning process. This new approach to planning (i.e. the consideration of barriers to implementation) has a positive aspect in that planners will be more knowledgeable of what works and what types of planning designs raise suspicion to politicians, local government officials, developers, and other parties. The negative aspect of this is that planning schemes formerly designed in relative "isolation" is lost, as planners who are aware of the policy implementation literature will plan "around" these known implementation barriers.

6.6 Implications for Policy Analysts and Regional Scientists

In addition to the same policy implementation issues that face planners, other implications face policy analysts and regional scientists. First, the results indicate for policy analysts and regional scientists that the theoretical full aggregation of households (all households assumed to be similar enough to be analyzed using a single-line regression method like OLS) is not appropriate *unless* the empirical sorting process results in a single household type. While the two-stage do-loop process is not the only way to sort households, it is a reasonable first approach that assures that households with similarities in several dimensions, including similarities in marginal prices for certain dwelling and/or neighborhood features, are grouped together for analysis purposes. In addition, the Breusch-Pagan (1980) test (null hypothesis is that the variance-covariance error matrix is diagonal) described earlier provides a decision rule regarding the hypothesis of full aggregation.

Second, regional scientists and geographers need to question the assumptions that are made when using the spatial weights matrix approach. At certain levels of aggregation, the spatial weights matrix approach makes sense. However, when one has data on individual households, spatial weights matrices do not necessarily need to be used to model spatial dependency. Once the extent of one's neighbors (who are my neighbors) has been determined, either by assumption of through the direct questioning of residents, variables that describe across-parcel externalities or some kind of clustering can be constructed and used directly in the analysis as independent variables to explain sales price variation. The common dichotomous variable approach used to determine a household's neighbors can be foregone when one has data at the household level. To my knowledge, no other housing hedonic study can offer the level of detailed estimates of dwelling-specific, spatial, and adjacency variables on the sales prices of subgroups of a single sample as those found here.

To conclude, this paper theoretically and empirically supports the idea of multiple hedonic price lines in a single urban neighborhood. Some conceptual issues remain to be addressed in future work. In the meantime, I reasonably can say that the spatial theory developed here, complemented with the incorporation of detailed household-level dwelling and spatial data, provides a powerful tool for economic developers and planners who wish to perform neighborhood-level planning and economic redevelopment. The power of the tool lies in its distinction of multiple household types, each of which has different marginal WTPs for dwelling structure characteristics, for being "near" certain neighborhood landmarks, and for being "near" households with certain characteristics.

APPENDIX A: HOME PARK HOUSING SURVEY

Dear Home Park Resident:

You have been randomly selected to complete a survey that will be used to implement the Home Park Master Plan. If you are a renter, please complete the **"I AM A RENTER"** survey **only**. If you own your home, please complete the **"I AM A HOMEOWNER"** survey **only**.

Timothy State, President of the Home Park Community Improvement Association, Inc., approved this survey. You will not be paid and there is no cost to you to participate. As our way of saying thank you, please find an enclosed stick of gum that will last longer than this five-minute survey. There is no direct benefit or foreseeable risk to you to participate in this survey. If you have questions about your rights as a research subject please contact the Office of Research Compliance at 404-894-6944.

After completing the survey, please return it to me using the self-addressed stamped envelope enclosed.

We need your help. All information that identifies you and your dwelling is confidential and will be purged upon completion of this study. Your answers will <u>not</u> be provided to any companies for any reason. If you have questions about the survey please contact Dr. Michael Farmer at 404-894-6458 or Cliff Lipscomb at 770-387-2518.

The Home Park Community Improvement Association will receive a summary report from this research. Your contribution is very useful to us, but <u>no</u> information on individual households will be shared with outside sources.

Your participation is completely voluntary. There are no right or wrong answers; we only want your opinion if you want to provide it to us.

Thank you in advance for your time in completing this survey- the Home Park Community Improvement Association and I very much appreciate your help.

Cliff Lipscomb, Ph.D. Candidate
Georgia Tech School of Public Policy

I AM A HOMEOWNER

1.How many adults (18 years old or older) live in your dwelling? _____
2.How many children (under 18 years of age) live in your dwelling? _____
3.How many adults in your dwelling are employed (including self-employed adults)? _____
4.How old are you? _________ years
5.What is your gender? ___ Female ___ Male
6.Which best describes your living situation:
___ I own the house, live in it, and do not rent any of it.
___ I own the house, live in it, and rent part of it.
7.Approximately how long have you lived in your current dwelling?
________ years
8.Which best describes your race?
___ African-American ___ White ___ Other
___ Hispanic ___ Asian
9.Check the ONE statement that best describes YOU. I am:
___ unemployed. ___ a student.
___ a homemaker. ___ retired.
___ employed full-time. ___ employed part-time (non-student).
10.If you are currently employed (self-employment is included here), which statement that best describes YOU:
___ I am the highest wage earner in the dwelling.
___ Another household resident is the highest wage earner.
___ All residents who work make approximately the same wages.
11.Which of these best describes YOU (*please check only one*):
___ I did not complete high school.
___ I am a high school graduate.
___ I am currently an undergraduate student.
___ I have an undergraduate degree.
___ I am currently a graduate student.
___ I have a graduate degree.
12. Which category best describes *your* total annual income before taxes (include all sources of income, including fellowships, alimony, retirement income, etc.)?
___ Less than $15,000 ___ $50,000 to $69,999
___ $15,000 to $24,999 ___ $70,000 to $89,999
___ $25,000 to $34,999 ___ $90,000 or more
___ $35,000 to $49,999
13. Which *one* issue besides national security is more important to you?
___ Crime ___ Education
___ Environment ___ Social Security
14. Politically speaking, do you generally consider yourself (*check one only*):
___ Conservative ___ Liberal ___ Moderate

(over)

15. What is the approximate market value of your house today?
(*Check only one.*)

___ $90,000 to $149,999	___ $225,000 to $249,999
___ $150,000 to $174,999	___ $250,000 to $274,999
___ $175,000 to $199,999	___ $275,000 to $299,999
___ $200,000 to $224,999	___ More than $300,000

16. From the list below, please check the *two* most important factors in your decision to live in the Home Park *neighborhood.*

___ Close to religious facilities	___ Tree cover
___ Close to Georgia Tech	___ Crime rate
___ A feeling of community	___ Noise levels
___ Diversity of residents	___ Close to parks
___ Investment potential	___ Close to your workplace
___ Close to entertainment	___ Close to public transportation

17. From the list below, please check the *two* most important factors in your decision to live in *your particular house.*

___ Kitchen	___ Lots of windows
___ Number of bathrooms	___ Hardwood floors
___ Views of the neighborhood	___ Internet connection
___ Recent renovations to the house	___ Other ____________
___ Number of bedrooms	

18. To your knowledge, have total improvements to your home over the last three years exceeded $10,000? ___ Yes ___ No

19. Which statement best describes you: "In the next two years…"
___ I do not wish to move.
___ I want to move inside Home Park.
___ I want to move outside Home Park.

20. (optional) Feel free to add any of your own comments concerning what you like or do not like about the Home Park neighborhood and/or your house.
____
____

Thank you for your help!

I AM A RENTER

1.How many adults (18 years old or older) live in your dwelling? ______
2.How many children (under 18 years of age) live in your dwelling? _____
3.How many adults in your dwelling are employed (including self-employed adults)? _____
4.How old are you? ____________ years
5.What is your gender? ___ Female ___ Male
6.Which best describes your living situation:
___ I rent an entire house or apartment.
___ I rent a unit in an apartment complex.
___ I rent a room in a house owned by someone else.
7.Approximately how long have you lived in your current dwelling?
________ years
8.Which best describes your race?
___ African-American ___ White ___ Other
___ Hispanic ___ Asian
9.Check the ONE statement that best describes YOU. I am:
___ unemployed. ___ a student.
___ a homemaker. ___ retired.
___ employed full-time. ___ employed part-time (non-student).
10.If you are currently employed (self-employment is included here), which statement that best describes YOU:
___ I am the highest wage earner in the dwelling.
___ Another roommate is the highest wage earner.
___ All roommates who work make approximately the same wages.
11.Which of these best describes YOU (*please check only one*):
___ I did not complete high school.
___ I am a high school graduate.
___ I am currently an undergraduate student.
___ I have an undergraduate degree.
___ I am currently a graduate student.
___ I have a graduate degree.
12.Which category best describes *your* total annual income before taxes (include all sources of income such as fellowships, retirement income, etc.)?
___ Less than $15,000 ___ $50,000 to $69,999
___ $15,000 to $24,999 ___ $70,000 to $89,999
___ $25,000 to $34,999 ___ $90,000 or more
___ $35,000 to $49,999
13.Which *one* issue besides national security is more important to you?
___ Crime ___ Education
___ Environment ___ Social Security
14.Politically speaking, do you generally consider yourself (*check one only*):
___ Conservative ___ Liberal ___ Moderate

(over)

15. How much do you and your roommates pay *total* for your rented dwelling each month?

___ \$400 to \$499 ___ \$800 to \$899
___ \$500 to \$599 ___ \$900 to \$999
___ \$600 to \$699 ___ More than \$1,000
___ \$700 to \$799

16. From the list below, please check the *two* most important factors in your decision to live in the Home Park *neighborhood.*

___ Close to religious facilities ___ Tree cover
___ Close to Georgia Tech ___ Crime rate
___ A feeling of community ___ Noise levels
___ Diversity of residents ___ Close to parks
___ Investment potential ___ Close to your workplace
___ Close to entertainment ___ Close to public transportation

17. From the list below, please check the *two* most important factors in your decision to live in *your particular rental unit.*

___ Kitchen ___ Lots of windows
___ Number of bathrooms ___ Hardwood floors
___ Views of the neighborhood ___ Internet connection
___ Recent renovations to the house ___ Other ______________
___ Number of bedrooms

18. Which ONE statement best describes you: "In the next two years...

___ I do not wish to move.
___ I want to move inside Home Park.
___ I want to move outside Home Park.

19. (optional) Feel free to add any of your own comments concerning what you like or do not like about the Home Park neighborhood and/or your rental unit.

____
____

Thank you for your help!

APPENDIX B: DERIVATION OF THE SUR MODEL

Under the assumptions of the classical linear regression model, the least squares estimators of the regression coefficients are unbiased and efficient. This assumes, of course, that ALL of the information known about the relationships between the independent and dependent variables is *known and accounted for* in the model. However, if some other piece of information about these relationships is known *but not included in the model*, then the least squares estimators are no longer established theoretically. If this other piece of information is knowledge concerning the correlation of the error term from one equation with the error term from a different equation, then we have violated the classical assumptions of the linear regression model, namely the expected value of $\boldsymbol{\varepsilon} = 0$ and $E(\boldsymbol{\varepsilon}_m \boldsymbol{\varepsilon}_m') = \sigma_{mm}\mathbf{I}_T = \boldsymbol{\Omega}$. Instead, when the error terms of different equations are correlated, we get $E(\boldsymbol{\varepsilon}_m \boldsymbol{\varepsilon}_p') = \sigma_{mp}\mathbf{I}_T$ (m, p = 1, 2, ..., M). This is the important subtle link that makes a system of M equations *seemingly unrelated* (Kmenta 1986).

For our purposes here, a SUR system of M equations (when compressed) looks like the following:

(B1)
$$\begin{bmatrix} SP_i^{\,j} \\ \vdots \\ SP_i^{\,M} \end{bmatrix} = \begin{bmatrix} \beta_0^1 + \alpha_i S + \delta_i D + \psi_i A \\ \vdots \\ \beta_0^M + \alpha_M S + \delta_M D + \psi_M A \end{bmatrix} + \begin{bmatrix} \varepsilon_j \\ \vdots \\ \varepsilon_M \end{bmatrix}$$

On page 247 in Breusch and Pagan (1980), the B-P test for independence, a form of the

Lagrangian multiplier test, was first derived. It tests the diagonality of the error covariance matrix in a seemingly unrelated equation (SUE) model. Under the null hypothesis, the Breusch-Pagan test says the variance-covariance matrix is diagonal, meaning that the error terms are not correlated between equations, evidenced by zeros in all of the off-diagonal elements and $\mathbf{N}\sigma_{kk}^{-1}\sigma_{ll}^{-1}$ on the principal diagonal. Here, **N** is the number of observations and $\sigma_{kk}^{-1}\sigma_{ll}^{-1}$ is the inverse of the terms that comprise the variance-covariance matrix. Then, the Lagrangian multiplier test that determines the diagonality of the error covariance matrix is:

(B2) $$\mathbf{N}\sum_{i=1}^{m}\sum_{i=1}^{i-1} r_{ij}^2$$

In Equation B2, r_{ij} is $\mathbf{N}^{-1}(\sigma_{kk}\sigma_{ll})^{-1/2}\mathbf{u}_i'\mathbf{u}_i$, where $\mathbf{u}_i$ are the error terms from the system of equations. The statistic has a χ^2 distribution with (m/2 x (m-1)) degrees of freedom. Therefore, if the null hypothesis is confirmed, a system of equations will reduce to a single-line OLS regression equation. Otherwise, a rejection of the null hypothesis means the system of equations is dependent, meaning that the error terms between equations are correlated.

APPENDIX C: HUMAN CAPITAL CROSS-MARKET EFFECTS

Problems exist with models that purport to compare the results of one household type to another household type. As articulated earlier, only households of the same type can be compared. The aggregability conditions here do not permit household estimates for one type to be compared to estimates of another type. Aggregation implies a certain relationship between a particular type of household and the inter-neighborhood externalities that impact the estimates of that particular type. In fact, if one had identical data on an adjacent neighborhood, estimates could be compared only if the same type appeared in the other neighborhood, because they would be of the same statistically aggregable type. This sorting process, which mitigates the specification and aggregation problems, does not simply map the area based on household level data. Distinguishing between household types also identifies the directional impact of cross-neighborhood market effects. Immediately one can see this in some of the independent variables (with significant coefficients of different signs) that reflect the impact of certain outside-the-neighborhood land uses on Home Park sales prices.

These externalities are not limited to factors outside the neighborhood; other non-parcel-specific factors *within* Home Park influence sales prices as well. These include the college education levels of one's neighbors, the home improvements made by one's neighbors, the racial composition of one's neighbors, and others. In particular, I focus on the neighbors' education levels as an indication for the impact of "human capital externalities" on sales prices.

A relatively small literature examines the idea that human capital externalities (measured here by the education level of my neighbors) play a role in determining whether or not households are attracted to certain neighborhoods. The theory, from Rauch (1991) and Zhang (2002), is that new households want to be near other households that are "like" them in some respect. Rauch (1991) used rent gradients to explain how living near educated people can enhance one's productivity. For example, a person with a Master's degree or Ph.D. may choose to locate to an area based on the general level of knowledge exhibited by his/her future neighbors. This person may prefer educated neighbors because he/she cares about his/her intellectual health; interacting with similarly educated and/or intelligent persons keeps that person sharp intellectually, exposes that person to the ideas of others, and provides that person immediate assimilation into a type of social network.

In the labor market, Rauch concludes that wages in high human capital cities should be relatively higher. Assuming a fixed housing supply, as a more highly educated and technical workforce, which commands higher wages, moves into an area, one will see an increase in dwelling sales prices. Several effects (direct and indirect) cause this increase in sales prices: higher wages (households with higher education levels earn more income); higher house premiums (households will pay more to live near other households with higher levels of education); and firms wanting to be near other high-tech firms. To separate the effects of households that command higher wages from the effects of cross-market human capital externalities on sales prices, the analyst needs information on 1) the local real estate market (including residential, retail, and commercial), on 2) the relationships between particular real estate sectors (i.e. the value of residential land located adjacent to commercial, etc.), 3) firm-level wage data, 4) whether a firm is "low-technology" or "high-technology," and others. In the real estate market only, a positive effect of human capital will translate into higher sales prices of houses, particularly if one assumes that a household would be willing to pay a higher sales price for being "near" educated neighbors. The "type" concept can be used to differentiate the impact of educated neighbors on different subsets of the Home Park sample. At the establishment level, increases in wages that correlate with increases in formal employee education can be used as an indicator of the "high-technology capacity" of firms.

Part of this theory, that human capital externalities influence household location decisions, can be tested empirically by looking at the marginal implicit prices of being adjacent to college educated households. My theory suggests that households will be willing to pay a higher premium to live near households that have higher education attainment levels. This suggests that clusters of human capital exist in the neighborhood. Data from the housing survey follow this theory of clusters in that 33.5 percent of households in Home Park located to Home Park because it was near their workplace(s) and 57.7 percent located there because their chosen dwelling was near Georgia Tech. Since theory suggests that neighbors' characteristics influence the household location decision, then an examination of neighbor characteristics in the housing hedonic model will provide a decent first approximation of the willingness to pay for a particular class of residents to live near households with certain characteristics, holding all else constant. Given the specification of the first-stage hedonic model in Table 5, I argue that certain adjacency variables measure the impact of neighbors' attributes/characteristics on the sales price of a particular dwelling. Those estimates are located in Table 14.

Table 14: Iterated SUR Estimates from Table 5 (abbreviated)

Adjacency Variables	Type A	Type B	Type C
Are you Adjacent to a Renter?	1.74*** (.67)	-2.29*** (.79)	.24 (.74)
Are you Adjacent to Undergrad Student?	1.25** (.56)	1.39** (.67)	-2.65*** (.63)
Are you Adjacent to a College Educated person?	-.19 (.63)	-.82 (.75)	1.17* (.71)
Are you Adjacent to a Person of a Different Race Than You?	-1.92*** (.47)	-2.69*** (.55)	4.58*** (.52)
Are you Adjacent to a Dwelling that made Home Improvements?	.69 (.48)	-3.98*** (.57)	3.61*** (.53)
Are you Adjacent to Someone who Lives in Home Park Because it is Near Georgia Tech?	1.47*** (.57)	2.61*** (.67)	-4.13*** (.63)

As an example of the impact of a "social" externality on sales prices, the estimates in Table 14 show that Type A households at the margin have a 1.92 percent decrease in sales price if their neighbors are of a different race than them, all else held constant. Similarly for Type B and C households, the marginal percentage changes in sales prices decrease by 2.69 percent and increase by 4.58 percent when households of other races are neighbors, respectively. So, we see here that Type C households have a positive marginal valuation for neighbors of different racial backgrounds than their own. This provides a rough estimate for the value of neighborhood diversity.

Another interesting and counterintuitive result is that Type A households (mostly student renters) are not willing to pay to near other college educated persons (who already have a degree), but are willing to pay to be near other renters. Strangely, Type B households (mostly renters) have a significant and negative WTP for being near renters and Type C households an insignificant and positive WTP. Given that Type A households are comprised of 35 percent undergraduate and 11 percent graduate students, the positive coefficients on "adjacent to renters" and "adjacent to undergraduate students" are not surprising. Then, Type B households are comprised of 28 percent undergraduate and 22 percent graduate students. The negative and significant coefficient on renter adjacency but positive and significant coefficient for undergraduate adjacency for Type B households is a likely signal of the value households place on being near folks who are pursuing their formal education. Then, for Type C households, of which 55 percent is homeowners, a positive and insignificant marginal sales price for adjacent renters and a negative and significant marginal WTP for adjacent undergraduate students reflects the general perception that older renters are mindful of how their actions impact others and that less-educated students (underclassmen) make lots of noise on weekend nights and do not keep up their yards (as several survey respondents noted), adding to the stigma placed on students by Home Park owners.

REFERENCES

Alonso, William. 1964. *Location and Land Use.* Cambridge: Harvard University Press.

Anas, Alex, Arnott, Richard J., and Small, Kenneth A. 2000. "The Panexponential Monocentric Model." *Journal of Urban Economics* 47: 165-179.

Anas, Alex, Arnott, Richard J., and Small, Kenneth A. 1998. "Urban Spatial Structure." *Journal of Economic Literature* XXXVI(3): 1426-1464.

Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models.* Kluwer Academic Publishers: Dordrecht.

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

Anselin, Luc. 2003. "Spatial Externalities, Spatial Multipliers, and Spatial Econometrics." Forthcoming in *International Regional Science Review.*

Arguea, Nestor M. and Cheng Hsiao. 2000. "Market Values of Environmental Amenities: A Latent Variable Approach." *Journal of Housing Economics* 9, 104-126.

Babcock, Richard F. 1966. *The Zoning Game: Municipal Practices and Policies.* University of Wisconsin Press: Madison, WI.

Babcock, Richard F. and Siemon, Charles L. 1985. *The Zoning Game Revisited.* Oelgeschlager, Gunn, and Hain in association with the Lincoln Institute of Land Policy: Boston, MA.

Becker, Gary. 1965. "A Theory of Allocation of Time." *Economic Journal* 75, 493-517.

Benson, Earl D., Julia L. Hansen, Arthur L. Schwartz, Jr., and Greg T. Smersh. 1998. "Pricing Residential Amenities: The Value of a View." *Journal of Real Estate Economics and Finance* 16(Jan.): 55-73.

Bell, Kathleen P. and Nancy E. Bockstael. 2000. "Applying the Generalized-Moments Estimation Approach to Spatial Problems Involving Microlevel Data." *Review of Economics and Statistics* 82(1): 72-82.

Bingham, Tayler H., Kealy, Mary Jo, David, Elizabeth, LeBlanc, Michael, Graham-Tomassi, Theodore, and Leeworthy, Robert. 1992. *Benefits Transfer: Procedures, Problems, and Research Needs.* Association of Environmental and Resource Economists Workshop, Snowbird, Utah.

Black, D., G. Gates, S. Sanders, and L. Taylor. 2002. "Why do gay men live in San Francisco?" *Journal of Urban Economics* 51(1): 1-23.

Bolitzer, B. and N.R. Netusil. 2000. "The impact of open spaces on property values in Portland, Oregon." *Journal of Environmental Management* 59: 185-193.

Bourassa, Steven C., Foort Hamelink, Martin Hoesli, and Bryan D. MacGregor. 1999. "Defining Housing Submarkets." *Journal of Housing Economics* 8: 160-183.

Box, G.E.P. and Cox, D.R. 1964. "An analysis of transformations." *Journal of the Royal Statistical Society, Series B.* 211-252.

Boyce, David, Peter Nijkamp, and Daniel Shefer (Eds.). 1991. *Regional Science: Retrospect and Prospect.* Springer-Verlag: Berlin. p. 412-413.

Breusch, T.S. and A.R. Pagan. 1980. "The Lagrangian Multiplier Test and its Applications to Model Specification in Econometrics." *Review of Economic Studies* 47(1): 239-253.

Brock, W. and S. Durlauf. 2001. "Interactions-Based Models." *The Handbook of Econometrics 5,* James J. Heckman and Edward E. Leamer, eds. Elsevier Science.

Chattopadhyay, Sudip. 2000. "The effectiveness of McFadden's nested logit model in valuing amenity improvement." *Regional Science and Urban Economics* 30: 23-43.

Chay, Kenneth Y. and Michael Greenstone. 2000. "Does Air Quality Matter? Evidence from the Housing Market." Working Paper: University of California at Berkeley.

Clark, Terry N., Richard Lloyd, Kenneth K. Wong, and Pushpam Jain. 2002. "Amenities Drive Urban Growth." *Journal of Urban Affairs* 24(5): 493-515.

Cohen, Michael. 1990. "A note on the cost of crime to victims." *Urban Studies* 27(1): 139-46.

Colwell, Peter F. and Gene Dilmore. 1999. "Who was first? An examination of an early hedonic study." *Land Economics* 75(4): 620-626.

Conrad, J. and Clark, C. 1987. *Natural Resource Economics: Notes and Problems.* Cambridge University Press. New York, NY.

Coulson, N. Edward. 1991. "Really Useful Tests of the Monocentric Model." *Land Economics* 67(3): 299-307.

Coy, Peter. 2003. "A Home Buyer's Froth Detector: Checking out rental rates can help you determine whether housing prices are inflated." *Business Week* 16 June 2003.

Cramer, J.S. 1986. *Econometric applications of Maximum Likelihood methods.* Cambridge University Press: Cambridge, England, UK. p. 4

Cropper, Maureen L., Deck, Leland B., and Kenneth E. McConnell. 1988. "On The Choice of Functional Form for Hedonic Price Functions." *Review of Economics and Statistics* 70: 668-675.

De Leon, Peter. 1992. "Policy Formulation: Where Ignorant Armies Clash by Night." *Policy Studies Review* 11(3/4): 389-405.

DiMento, Joseph. 1980. *The Consistency Doctrine and the Limits of Planning.* Oelgeschlager, Gunn, & Hain: Cambridge, MA. Quoted in Babcock, Richard F. and Siemon, Charles L. *The Zoning Game Revisited.* Oelgeschlager, Gunn, & Hain in association with the Lincoln Institute of Land Policy: Boston, MA.

Dixit, Avinash K. and Joseph E. Stiglitz. 1977. "Monopolistic competition and optimum product diversity." *American Economic Review* 67(3): 297-308.

Ekeland, Ivar, James J. Heckman, and Lars Nesheim. 2002. "Identifying Hedonic Models." *American Economic Review* 92(2): 304-309.

Epple, Dennis. 1987. "Hedonic Prices in Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products." *Journal of Political Economy* 95 (1): 59-80.

Farmer, Michael C. 2003. "Re-Investing in the Built Environment: A Spatial Network Approach." *European Planning Studies* 11(1): 57-76.

Farmer, Michael C., Joshua Reiss, and William Ditto. 2000. "Re-Investing in the Built Environment." *European Economic Review*, forthcoming.

Farmer, Michael C., Lipscomb, Clifford A., and McCarthy, Patrick S. 2003. "How Alcohol-Related Crashes of Different Severity Interrelate and Respond to Local Spatial Characteristics: An Evaluation of a Common Site Sales Ban on Alcohol and Gasoline. Under review.

Fischel, William. 1995. "Zoning and the Urban Environment." In Bromley, Daniel W., Ed. *The Handbook of Environmental Economics.* Basil Blackwell: Oxford. 61-88

Florida, Richard. 2002. *The rise of the creative class.* New York: Basic Books/Perseus.

Freeman, A. Myrick. 1993. *The Measurement of Environmental and Resource Values: Theory and Methods.* Resources for the Future: Washington, D.C., 104-105.

Freeman, A. Myrick. 1995. "Hedonic Pricing Methods." In Bromley, Daniel W., Ed. *The Handbook of Environmental Economics.* Blackwell: Oxford.

Glaeser, Edward. 1998. "Are Cities Dying?" *Journal of Economic Perspectives* 12(2): 139-160.

Glaeser, E., Kahn, M., and Rappaport, J. 1997. "Why do poor people live in cities." Working Paper. Harvard University Department of Economics.

Goodman, Allen C. 1978. "Hedonic prices, price indices, and housing markets." *Journal of Urban Economics* 5: 471-484.

Goodman, Allen C. 1989. "Topics in empirical urban housing research" in *The Economics of Housing Markets.* R.F. Muth and A.C. Goodman, eds. Chur, Switzerland: Harwood Academic, 49-143.

Goodman, Allen C. 1998. "Andrew Court and the Invention of Hedonic Price Analysis." *Journal of Urban Economics* 44(2): 291-298.

Greater Home Park Master Plan. 2001. EDAW, Inc., Atlanta, GA.

Greene, William H. 1993. *Econometric Analysis, 2nd edition.* New York, NY: Macmillan Publishing Co.

Greene, William H. 2003. *Econometric Analysis, 5th edition.* Upper Saddle River, NJ: Prentice-Hall.

Grin, John and Henk Van De Graaf. 1996. "Implementation as communicative action: An interpretive understanding of interactions between policy actors and target groups." *Policy Sciences* 29: 291-319.

Growing Smart Legislative Guidebook. http://www.planning.org/guidebook/Guidebook

Guilkey, David K. and Peter Schmidt. 1973. "Estimation of Seemingly Unrelated Regressions with Vector Autoregressive Errors." *Journal of the American Statistical Association* 68(343): 642-647.

Haab, Timothy C. and Robert L. Hicks. 1997. "Accounting for Choice Set Endogeneity in Random Utility Models of Recreation Demand." *Journal of Environmental Economics and Management* 34: 127-147.

Halstead, John M., Rachel A. Bouvier, and Bruce E. Hansen. 1997. "On the Issue of Functional Form Choice in Hedonic Price Functions: Further Evidence." *Environmental Management* 21(5): 759-765.

Hanemann, W. Michael. 1984. "Discrete/Continuous Models of Consumer Demand." *Econometrica* 52(3): 541-562.

Hanley, Nick and Clive L. Spash. 1993. *Cost-Benefit Analysis and the Environment*. Edward Elgar: Northampton, MA, USA, 74-76.

Harvey, Andrew. 1990. *The Econometric Analysis of Time Series*, 2d ed. The MIT Press: Cambridge, MA, 328.

Heckman, James J. and James M. Snyder, Jr. 1997. "Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators." *RAND Journal of Economics* 28(0): S142-S189.

Helmuth, Obata, & Kassabaum, Inc., and Raytheon Infrastructure Services, Inc. 1997. Sea-Tac International Airport Impact Mitigation Study. Washington Department of Community, Trade and Economic Development. [Cited March 13, 2000]. http://www.rcaanews.org/rcaa/hoksec9.html.

Hendry, D.F. 1971. "Maximum Likelihood Estimation of Systems of Regression Equations with Errors Generated by a Vector Autoregressive Process." *International Economic Review* 12: 257-272.

Hoyt, William H. and Stuart S. Rosenthal. 1997. "Household Location and Tiebout: Do Families Sort According to Preferences for Locational Amenities?" *Journal of Urban Economics* 42, 159-178.

Huettman, F. and A.W. Diamond. 2001. "Using PCA Scores to Classify Species Communities: An Example for Pelagic Seabird Distribution." *Journal of Applied Statistics* 28(7): 843-853.

Hughes, William T., Jr. and Geoffrey K. Turnbull. 1996. "Uncertain Neighborhood Effects and Restrictive Covenants." *Journal of Urban Economics* 39: 160-172.

http://cepa.newschool.edu/het/essays/sequence/sequence.htm. History of Economic Thought website, New School University, New York, NY, USA.

Immergluck, Daniel. 1999. "Neighborhoods, race, and capital: The effects of residential change on commercial investment patterns." *Urban Affairs Review* 34(3): 397-411.

Irwin, Elena G. and Nancy E. Bockstael. 2001. "The Problem of Identifying Land Use Spillovers: Measuring the Effects of Open Space on Residential Property Values." *American Journal of Agricultural Economics* 83 (Aug.): 698-704.

Irwin, Elena G. and Nancy E. Bockstael. 2002. "Interacting Agents, Spatial Externalities, and the Evolution of Residential Land Use Patterns." *Journal of Economic Geography* 2: 31-54.

Jenny, Nicholas W. 2002. "Continued Weakness in State Tax Revenue." *State Revenue Report* No. 50. Fiscal Studies Program, Rockefeller Institute of Government.

Kahn, Matthew. 1996. "The Silver lining in Rust Belt decline: Killing off pollution Externalities." Working Paper.

Kakwani, N. 1967. "The Unbiasedness of Zellner's Seemingly Unrelated Regression Equation Estimators." *Journal of the American Statistical Association* 62: 141-142.

Kelejian, Harry and Ingmar Prucha. 1999. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model." *International Economic Review* 40: 509-533.

Kiel, Katherine A. and McClain, K.T. 1995. "House Prices During Siting Decision Stages: The Case of an Incinerator From Rumor Through Operation." *Journal of Environmental Economics and Management* 28(2): 241-255.

Kim, Chong Won, Phipps, Tim T., and Anselin, Luc. 2003. "Measuring the Benefits of Air Quality Improvement: A Spatial Hedonic Approach." Forthcoming in *Journal of Environmental Economics and Management.*

Kmenta, Jan. 1986. *Elements of Econometrics, Second Edition.* Macmillan Publishing Company: New York. p. 635-649.

Krugman, Paul. 1980. "Scale Economies, product differentiation and the pattern of trade." *American Economic Review* 70: 950-959.

Krugman, Paul. 1991. *Geography and Trade.* MIT Press: Cambridge, MA.

Krugman, Paul. 1996. "Urban concentration: the role of increasing returns and transportation costs." *International Regional Science Review* 19: 5-30.

Krugman, Paul. 1998. "Space: the final frontier." *Journal of Economic Perspectives* 12(2): 161-174.

Kulshreshtha, Suren N. and Jon A. Gillies. 1993. "Economic Evaluation of Aesthetic Amenities: A Case Study of River View." *Water Resources Bulletin* 29 (April): 257-266.

Lampietti, Julian. 1995. "Case Study: Benefits Transfer in Indonesia." World Bank Institute.

Lancaster, Kelvin. 1966. "A New Approach to Consumer Theory." *The Journal of Political Economy* 74(2): 132-157.

Lindblom, Charles. 1959. "The Science of Muddling Through." *Public Administration Review* 19. In Theodoulou, Stella and Can, Matthew, Eds. 1995. *Public Policy: The Essential Readings.* Englewood Cliffs, NJ: Prentice Hall. 113-127.

Lipscomb, Clifford A. 2001. "The Capitalization of Amenities and Disamenities into Local Housing Markets: The Case of College Park, Georgia." *Colloqui: Cornell Journal of Urban Planning*. Ithaca, NY: Cornell University, 2001.

Luttik, Joke. 2000. "The value of trees, water and open space as reflected by house prices in the Netherlands." *Landscape and Urban Planning* 48: 161-167.

Majone, Giandomenico. 1989. *Evidence, Argument, and Persuasion in the Policy Process.* New Haven: Yale University Press.

McFadden, Daniel. 1978. "Modelling the Choice of Residential Location." In Karlqvist, A., L. Lundqvist, F. Snickars, and J. Weibull (eds.) *Spatial Interaction Theory and Planning Models.* North Holland: Amsterdam, 75-96. Reprinted in J. Quigley (ed.), *The Economics of Housing, volume 1.* 1997. Edward Elgar: London.

McLeod, Paul B. 1984. "The Demand for Local Amenity: An Hedonic Price Analysis." *Environment and Planning A* 16 (Mar.): 389-400.

Merriam-Webster Dictionary. <www.m-w.com>

Minor, Elliott. "Preservation key to improving quality of life." The Daily Tribune News March 1, 2002, pg. 5A.

Morrell, Peter and Cherie H.-Y. Lu. 2000. "Aircraft noise social cost and charge mechanisms- a case study of Amsterdam Airport Schiphol." *Transportation Research Part D* 5: 305-320.

Murty, M.N., Gulati, S.C., and Banerjee, A. 2003. "Hedonic Property Prices and Valuation of Benefits from Reducing Urban Air Pollution in India." Delhi University enclave, Institute of Economic Growth Working Paper.

Muth, Richard F. 1960. "The Demand for Non-Farm Housing," in *The Demand for Durable Goods*. A.C. Harberger, Ed. Chicago: University of Chicago Press, 29-96.

--------------------. 1969. *Cities and Housing*. Chicago: University of Chicago Press.

Neyman, J. and Scott, E.L. 1948. "Consistent Estimates Based on Partially Consistent Observations." *Econometrica* 16: 1-32.

Palmquist, Raymond B. 1991. "Chapter IV: Hedonic Methods" in *Measuring the Demand for Environmental Quality*, Braden, John B. and Charles D. Kolstad, Eds. Elsevier Science Publishers B.V. (North-Holland).

Papageorgiou, Y.Y. and T.R. Smith. 1983. "Agglomeration as Local Instability of Spatially Uniform Steady-States." *Econometrica* 51(4): 1109-1120.

Parsons, George R. 1986. "An Almost Ideal Demand System for Housing Attributes." *Southern Economic Journal* 53(2): 347-363.

Paterson, Robert W. and Kevin J. Boyle. 2002. "Out of Sight, Out of Mind? Using GIS to Incorporate Visibility in Hedonic Property Value Models." *Land Economics* 78(3): 417-425.

Peiser, Richard. 1990. "Who Plans America? Planners or developers?" *Journal of the American Planning Association* 56(4): 496-503.

Pressman, Jeffrey and Wildavsky, Aaron. 1984. *Implementation: How great expectations in Washington are dashed in Oakland; or, Why it's amazing that Federal programs work at all, this being the saga of the Economic Development Administration as told by two sympathetic observers who seek to build morals on a foundation of ruined hopes.* 3rd edition. Berkeley, CA: University of California Press.

Rapaport, Carol. 1997. "Housing Demand and Community Choice: An Empirical Analysis." *Journal of Urban Economics* 42(2): 243-260.

Roback, Jennifer. 1982. "Wages, Rents and the Quality of Life." *Journal of Political Economy* 90(6): 1257-1278.

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

Rosenberger, Randall S. and Loomis, John B. 2001. *Benefit transfer of outdoor recreation use values: A technical document supporting the Forest Service Strategic Plan (2000 revision).* Gen. Tech. Rep. RMRS-GTR-72. Fort Collins: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 59 pp.

Sabatier, Paul A. and Hank C. Jenkins-Smith (eds.) 1993. *Policy Change and Learning: An Advocacy Coalition Approach.* Westview Press: Boulder, CO.

Samuelson, Paul A. 1954. "The pure theory of public expenditures." *Review of Economics and Statistics* 36: 387-389.

Sieg, Holger, V. Kerry Smith, H. Spencer Banzhaf, and Randy Walsh. 2000. "Estimating the general equilibrium benefits of large policy changes: the Clean Air Act revisited." National Bureau of Economic Research, Working Paper 7744.

Sivitanidou, Rena. 1995. "Urban Spatial Variations in Office-Commercial Rents: The Role of Spatial Amenities and Commercial Zoning." *Journal of Urban Economics* 38: 23-49.

Smith, Barton A. 1976. "The Supply of Urban Housing." *The Quarterly Journal of Economics* 90(3): 389-405.

Smith, V.K. and Huang, Ju-Chin. 1995. "Can Markets Value Air Quality? A Meta-analysis of Hedonic Property Value Models." *Journal of Political Economy* 103: 209-227.

Stata Base Reference Manual. 2003. Volume 4, S-Z, Release 8. 157

Texas Transportation Institute. 2001. "2001 Urban Mobility Report." College Station, Texas: Texas A&M University.

Thorsnes, Paul. 2000. "Internalizing Neighborhood Externalities: The Effect of Subdivision Size and Zoning on Residential Lot Prices." *Journal of Urban Economics* 48: 397-418.

Thünen, Johann Heinrich von. 1826. *Der Isolierte Staat in Bezeihung auf Landwirtschaft und Nationalökomie* (Rostock, Leopold), partially translated by Peter Hall in 1966 as *von Thünen's Isolated State*, (Oxford, U.K, Pergammon).

Tiebout, Charles. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64: 416-424.

Tolley, George. 1974. "The Welfare Economics of City Bigness." *Journal of Urban Economics* 1(3): 325-345.

Tyrvainen, Liisa, and Antti Miettinen. 2000. "Property Prices and Urban Forest Amenities." *Journal of Environmental Economics and Management* 39: 205-223.

Varian, Hal. 1999. *Microeconomic Analysis, Third Edition.* Norton Publishers.

Wheaton, William C. 1977. "Income and Urban Residence: An Analysis of Consumer Demand for Location." *The American Economic Review* 67(4): 620-631.

Wissen, Leo Van and Annemarie Rima. 1988. *Modelling Urban Housing Market Dynamics: Evolutionary Patterns of Households and Housing in Amsterdam*, in *Studies in Regional Science and Urban Economics*, Ake E. Anderson, Walter Isard, and Peter Nijkamp, eds. Volume 18, North Holland Publishing Co.: Amsterdam.

Yin, Robert. 1994. *Case Study Research: Design and method (2nd edition).* Beverly Hills, CA: Sage.

Zellner, Arnold. 1962. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias." *Journal of the American Statistical Association* 57: 348-368.

VITA

Clifford A. Lipscomb is a native of Cartersville, Georgia. He received his Bachelor's degrees in Economics and Sociology from Berry College, Mount Berry, Georgia, in 1998. His interests include environmental and urban economics, statistics, and public finance. As of this writing, Dr. Lipscomb is an Economist with the Regional Economic Analysis Division, Bureau of Economic Analysis, U.S. Department of Commerce, stationed in Washington, DC.