

## **ABSTRACT**

KLAIBER JR, HENRY ALLEN. Valuing Open Space in a Locational Equilibrium Model of the Twin Cities. (Under the direction of Daniel J. Phaneuf).

This research employs recently developed econometric techniques to extend the literature on open space as well as present one of the first micro-level analyses of housing supply. To estimate horizontal sorting models for housing demand and housing supply, a new database consisting of over 460,000 residential transactions spanning 1990 through 2006 as well as detailed land use data identifying nine types of open space is obtained for the seven county Twin Cities metropolitan area in east-central Minnesota. Estimation of a horizontal sorting model allows the recovery of preferences for local public goods that vary across observable attributes of both households and builders. Several policy implications arising from the estimation results are that: (i) potential development of open space reduces the amenity value open space provides to households, (ii) different types of open space are associated with very different marginal willingness to pay measures indicating the importance of treating open space as a heterogeneous good, (iii) estimation of heterogeneous preferences shows that both households and builders are heterogeneous and accounting for that heterogeneity provides important information to policymakers, and (iv) policies restricting the amount of developable land can influence the location of the supply of new housing by builders.

The use of a horizontal sorting model to estimate both households' demand for housing and builders' supply of new housing allows analysis of not only marginal willingness to pay measures for local public goods, but also willingness to pay measures associated with non-marginal policy counterfactuals. Integrating both the demand and

supply sides of the housing market, I estimate general equilibrium welfare measures where both households and builders adjust their housing decisions in response to policy counterfactuals. To clear the market, I adjust the price of housing in order to remove excess demand and excess supply using an iterative numerical technique. I find that significant differences exist between partial and general equilibrium welfare measures and the incorporation of a supply response plays an important role.

Valuing Open Space in a Locational Equilibrium Model of the Twin Cities

by  
Henry Allen Klaiber Jr.

A dissertation submitted to the Graduate Faculty of  
North Carolina State University  
In partial fulfillment of the  
Requirements for the degree of  
Doctor of Philosophy

Economics

Raleigh, North Carolina

2008

APPROVED BY:

---

Daniel J. Phaneuf  
Committee Chair

---

Raymond B. Palmquist

---

Roger H. von Haefen

---

Christopher Timmins

## **DEDICATION**

To Denise

## BIOGRAPHY

Henry Allen Klaiber Jr. was born in Greensboro, NC on March 18, 1982 to parents Henry Klaiber and Barbara Stevens. While attending Grimsley High School he became interested in computer science and took introductory economics courses at the University of North Carolina at Greensboro. Also at this time, he met his future wife Denise who graduated alongside him in the spring of 2000. Allen pursued a Bachelor of Science degree in Computer Science at North Carolina State University beginning in the fall of 2000. He graduated *summa cum laude* in May of 2003 with a minor in Economics. Continuing to have interests in economics, he enrolled in the Economics Ph.D. program at North Carolina State University in the fall of 2003. The next year, he married his high school sweetheart on June 26, 2004 in Greensboro, NC. After completing his Ph.D., Allen will spend a year as a postdoctoral scholar at Arizona State University working under the direction of V. Kerry Smith and will then join the faculty of The Pennsylvania State University as an assistant professor in the fall of 2009.

## ACKNOWLEDGEMENTS

I am deeply indebted to the many individuals who have made my graduate school experience not only possible, but extremely rewarding on both a personal and intellectual level. I owe a special debt of gratitude to my advisor, Dan Phaneuf, who has taught me to think like an economist and offered constant encouragement throughout my graduate studies. I would also like to thank the other members of my committee: Ray Palmquist, Roger H. von Haefen, and Christopher Timmins. Their constant encouragement and guidance has helped to improve not only my research, but has enriched my graduate school experience and opened doors I never imagined possible.

In addition to my advisors, I have met many other individuals who have shared in the highs and lows of graduate school over the course of the past 5 years. Among these are fellow graduate students Eric Belasco, Mykel Taylor, and Ryan Williams. Each of them has been an important part of my graduate school experience and I look forward to continuing my relationships with them both as friends and colleagues in the future.

Lastly, I owe a great deal of thanks to my family who has supported me throughout this process. My wife Denise has been extremely patient and offers constant encouragement and love. My parents have instilled upon me the foundations of hard work and the importance of education. I will forever remember and appreciate my grandmother for the numerous things she has done to make graduate school both possible and enjoyable while always offering her unconditional love and support.

## TABLE OF CONTENTS

List of Tables .....	ix
List of Figures.....	xi
1 Introduction.....	1
2 Using the Housing Market to Value Local Public Goods.....	7
2.1 Introduction.....	7
2.2 Hedonic Model.....	9
2.2.1 Open space hedonic literature.....	11
2.2.2 Limitations of the hedonic literature.....	13
2.3 Sorting Model .....	15
2.3.1 Vertical sorting model.....	16
2.3.2 Horizontal sorting model .....	18
2.4 Aggregate Supply Models.....	20
2.5 Land Conversion Models.....	22
2.6 Conclusions.....	23
3 Data.....	26
3.1 Introduction.....	26
3.2 Twin Cities Study Area.....	27
3.3 Choice Set of Housing Types .....	29
3.4 Property Transactions .....	31
3.4.1 Single family detached.....	32
3.4.2 Builder transactions .....	33
3.4.3 Land transactions .....	34
3.5 Land Use Data.....	34
3.6 Soil and Supply Characteristics .....	37
3.7 Neighborhood Demographics .....	39
3.8 Household and Builder Characteristics.....	40
3.9 Conclusions.....	42

4	Estimation Strategy and Preliminaries .....	55
4.1	Introduction.....	55
4.2	Theoretical Model of Housing Choice by Agents .....	57
4.3	Empirical Sorting Model.....	60
4.3.1	First stage estimation .....	61
4.3.2	Second stage estimation.....	64
4.3.3	Asymptotic properties of the estimator.....	66
4.4	Price Index Estimation.....	67
4.4.1	Housing price indexes.....	68
4.4.2	Land price indexes .....	71
4.5	Conclusions.....	73
5	Housing Demand.....	80
5.1	Introduction.....	80
5.2	Theoretical Sorting Model of Demand by Households .....	81
5.3	Empirical Estimation .....	83
5.4	Estimation Results .....	85
5.4.1	First stage results.....	86
5.4.2	Second stage results .....	87
5.5	Conclusions.....	90
6	Housing Supply.....	95
6.1	Introduction.....	95
6.2	Theoretical Sorting Model of Builder Location Decisions.....	97
6.3	Choice Set of Housing Types Available to Builders .....	100
6.4	Empirical Estimation .....	101
6.5	Estimation Results .....	104
6.5.1	First stage results.....	104
6.5.2	Second stage results .....	106
6.5.3	Marginal willingness to pay.....	109

6.6	Robustness Checks.....	110
6.6.1	Seconds stage estimation results for a range of quantiles.....	111
6.6.2	Estimation using only chosen new supply alternatives.....	112
6.7	Conclusions.....	114
7	Policy Analysis.....	121
7.1	Introduction.....	121
7.2	Interpretation of Willingness to Pay Measures.....	124
7.3	Expanding Agricultural Preserves .....	125
7.4	Acquiring Additional DNR Land at the Urban Fringe .....	128
7.5	Acquiring Additional DNR Land inside the City .....	129
7.6	Welfare Estimates Excluding Non-Chosen Supply Alternatives.....	130
7.7	Conclusions.....	132
8	Results and Future Research .....	137
	References.....	143
	Appendices.....	147
A	Data Cleaning Procedures.....	148
A.1	Property Transactions .....	148
A.1.1	Data availability summary statistics .....	155
A.1.2	Housing attributes summary statistics .....	157
A.2	Land Transactions.....	159
A.3	Land Use.....	160
A.3.1	Reinvest in Minnesota.....	164
A.3.2	Metropolitan Agricultural Preserve Program.....	166
A.4	Cost Characteristics .....	168
B	Census Tract Spatial Units .....	179
B.1	Introduction.....	179
B.2	Demand Estimation.....	180

B.2.1	First stage estimation results .....	181
B.2.2	Second stage estimation results .....	182
B.3	Supply Estimation.....	183
B.3.1	First stage estimation results.....	184
B.3.2	Second stage estimation results .....	184
B.4	Conclusions.....	185

## LIST OF TABLES

Table 3.1	Housing types with observed transactions .....	44
Table 3.2	Land use by time period in acreage and percentage .....	45
Table 3.3	Spatial divisions for census levels .....	46
Table 3.4	Household summary statistics.....	46
Table 3.5	Builder summary statistics .....	46
Table 4.1	Price index regressions for 2000 census block groups by time period .....	75
Table 4.2	Summary statistics for house price indexes .....	75
Table 4.3	Land price index regressions for 2000 census tracts .....	76
Table 4.4	Summary statistics for land price indexes .....	77
Table 5.1	First stage interaction parameters .....	93
Table 5.2	Second stage results using naïve OLS .....	93
Table 5.3	Second stage estimation results using instrumental variables .....	93
Table 5.4	Marginal willingness to pay heterogeneity .....	94
Table 6.1	First stage interaction parameters .....	116
Table 6.2	Second stage results using naïve quantile regression (0.5).....	116
Table 6.3	Second stage results using instrumental variables quantile regression (0.5) .....	117
Table 6.4	Marginal willingness to pay heterogeneity .....	117
Table 6.5	Instrumental variables quantile regression comparison for supply estimation using all available alternatives .....	118
Table 6.6	First stage supply interaction parameters.....	118
Table 6.7	Second stage chosen alternative results using naïve quantile regression (0.5) .....	118
Table 6.8	Second stage chosen alternative results using instrumental variables quantile regression (0.5).....	119
Table 6.9	Marginal willingness to pay heterogeneity for builders using only chosen alternatives .....	119
Table 7.1	Welfare results for hypothetical policies .....	134

Table 7.2	Welfare results for hypothetical policies using only chosen alternatives for supply .....	134
Table A.1	Transactions by year and county.....	171
Table A.2	Sales and new construction by year .....	171
Table A.3	Repeat sales occurring by year (as observed in the data) .....	172
Table A.4	Summary statistics for housing characteristics by time period and size .....	173
Table A.5	RIM sites located in metro 7 counties.....	173
Table A.6	Breakdown of RIM sites by duration.....	174
Table A.7	Distribution and enrollment of agricultural preserves over time .....	174
Table B.1	Housing types at the census tract .....	187
Table B.2	Price index regressions for 2000 census tracts by time period .....	188
Table B.3	First stage interaction parameters .....	189
Table B.4	Second stage results using naïve OLS .....	189
Table B.5	Second stage estimation results using IV.....	189
Table B.6	Marginal willingness to pay heterogeneity.....	190
Table B.7	First stage supply interaction parameters.....	190
Table B.8	Second stage supply results using naïve quantile regression (0.5).....	191
Table B.9	Second stage supply results using instrumental variables quantile regression (0.5) .....	191
Table B.10	Marginal willingness to pay heterogeneity for builders .....	192

## LIST OF FIGURES

Figure 3.1	Twin Cities metropolitan area.....	47
Figure 3.2	Census 2000 block groups .....	48
Figure 3.3	Division of transactions by square feet.....	48
Figure 3.4	Single family detached transactions.....	49
Figure 3.5	Number of demand transactions comprising each housing type (N=20,444).....	49
Figure 3.6	Histograms of choice set housing characteristics (N=20,444) .....	50
Figure 3.7	Housing supply transactions .....	51
Figure 3.8	Number of supply transactions comprising each housing type (N=4,545).....	51
Figure 3.9	Land transactions .....	52
Figure 3.10	Reinvest in Minnesota conservation easements.....	52
Figure 3.11	Metropolitan agricultural preserves .....	53
Figure 3.12	Developable land by time period .....	53
Figure 3.13	Metropolitan services areas by time period .....	54
Figure 3.14	Number of houses built by builder.....	54
Figure 4.1	Price indexes by time period and house size .....	78
Figure 4.2	Census tracts with missing land transactions.....	79
Figure 4.3	Estimated land price indexes by census tracts .....	79
Figure 6.1	New supply by time period.....	120
Figure 6.2	Estimated alternative specific constants .....	120
Figure 7.1	5% increase in agricultural preserves .....	135
Figure 7.2	2.5% increase in urban fringe open space (non-park).....	135
Figure 7.3	2.5% increase in inner city open space (non-park).....	136
Figure A.1	Histograms of housing characteristics (N=461,695) .....	175
Figure A.2	Price per acre for land transactions (2006 dollars) .....	176
Figure A.3	Areas of poor drainage.....	176
Figure A.4	Areas of very limited agricultural capability .....	177

Figure A.5	Areas of very limited suitability for dwellings .....	177
Figure A.6	Areas of very high slope .....	178

# Chapter 1

## Introduction

Expanding urban areas have raised public awareness of the importance of open space as large tracts of previously undeveloped land are developed. In response to the loss of open space, policymakers have introduced legislation establishing conservation programs, often supported through public funding to preserve existing open space. In a recent survey, Kotchen and Powers (2006) report that between 1998 and 2003, over 1,000 open space referenda have been voted upon in the United States with over 80% passing. Recognizing the importance of open space policy, researchers are increasingly interested in uncovering preferences for open space through the use of housing markets. Recent methodological advancements have allowed researchers to better model, and thus better understand the impacts of open space.

In the current work, I extend recent advancements in the sorting literature to the topic of open space by estimating a flexible set of preferences for heterogeneous open space using a framework capable of valuing non-marginal changes in local public goods. An important

contribution of the current research is the attention paid to the supply side of the housing market, which has received substantially less attention than its demand counterpart in the empirical micro level housing literature. Incorporating an estimated supply of new housing into analysis of local public goods allows a full integration of the supply and demand for housing where price adjusts to changes in demand and supply to clear the market and reach a new equilibrium following a policy shock.

The study area for the current research is the Twin Cities metropolitan area in east-central Minnesota. Rapid expansion over the past 20 years has increased the Twin Cities population from a 1990 census population of 2,538,834 by over 25% to a 2006 estimated census population of 3,175,041. With population growth occurring in the metropolitan area, rapid development has occurred extending the urban rural fringe. From 1990 to 2005, over 250,000 acres of agricultural land were lost to development. To combat the loss of agricultural land for development, policymakers have created numerous conservation programs ranging from permanent conservation easements to a temporary agricultural preserve program.

As a result of the public awareness and concern over the loss of open space, the Twin Cities is an ideal location to explore questions related to open space. Recent examples of the public interest in open space can be found in numerous newspaper articles found in the Star Tribune, the local newspaper for the Twin Cities. An example of these is the May 24, 2006 article outlining the main concerns of residents in which urban sprawl and protection of open space were cited as one of the chief concerns of local residents in regards to their quality of

life. Another example is the June 7, 2006 article outlines home builders' efforts to establish ecofriendly construction projects preserving open space.

In addition to citizens' efforts to preserve open space, policymakers are also actively seeking to preserve open space. A May 18, 2006 article highlights a city council's decision to impose conservation standards on new home building requiring 20 acres of preserved open space set aside for every 20 acres of new development. A March 29, 2006 article outlines proposals for new park construction to provide additional amenities to an existing community. Given the public concern over loss of open space and the policy efforts addressing that concern, additional research on open space serves to better educate policymakers and the public on the potential impacts of proposed open space policy.

Chapter 2 places the current research into context by exploring previous literature focused on open space issues. The existing literature described in chapter 2 consists primarily of equilibrium-based methods including hedonics, sorting models, and aggregate supply and demand models. As virtually no micro-level housing supply literature exists, an overview of the reduced form land conversion literature is presented as many of the same data challenges associated with micro-level data and identification of land conversion variables are related to the data challenges and identification of housing supply variables in the current work.

To analyze open space for the Twin Cities, chapter 3 describes a new dataset containing over 450,000 single family housing transactions for the seven-county Twin Cities metropolitan area spanning the years 1990 through 2006. Combining these transactions with detailed land use data permits analysis of a broad range of open space contained within the

Twin Cities area. Given the large expansion of the metropolitan area over the study period, attention to the supply of new housing by builders seems appropriate and is addressed along with the demand for housing by households.

Expanding the existing open space literature, the current work implements a recently developed horizontal sorting model described in chapter 4. The basis of the horizontal sorting model is a set of decision makers, households and builders, choosing from a finite set of discrete housing types. Aggregation of individual housing transactions to form housing types occurs over the dimensions of location, house size, and time of transaction. Given a discrete choice set, households sort across housing types in order to maximize their utility. Similarly, builders sort over the same set of housing types to maximize utility of profit and supply new housing. It is the interaction of the demand and supply of housing that determine market clearing prices.

The horizontal sorting model not only provides the foundation for a locational equilibrium, but also provides a framework to incorporate a rich set of heterogeneous preferences for open space that vary by observable household and builder characteristics. Estimation results for both housing demand and new housing supply are reported in chapters 5 and 6, respectively. These results highlight the importance of accounting for differences across households arising from observable characteristics of age and household composition as well as across builders based on the size of the home builder.

Insights gained from incorporating heterogeneous preferences for open space are that different forms of open space have very different impacts on households and builders. For example, the value of regional parks to households containing working aged individuals is

over six times greater than the benefit of additional regional parks to households containing only retired individuals. For home builders, the presence of undeveloped agricultural land is worth nearly double to very large builders compared to the value of additional agricultural land to very small builders.

In addition to recovering heterogeneous preferences for a variety of open space, the horizontal sorting model also provides a framework for solving for a new housing equilibrium following policy shocks. Chapter 7 explores the implications of incorporating increasing degrees of model flexibility to evaluate the impacts of several counterfactual open space policies. Using the estimated demand and supply models from previous chapters, three different forms of welfare measurement are explored.

Partial equilibrium welfare measures assume no re-sorting occurs in response to policy changes and house prices do not adjust to clear the housing market. Allowing sorting behavior gives rise to general equilibrium welfare measures where only one side of the market re-sorts and housing prices adjust to clear the market. Lastly, the current work presents a new form of general equilibrium response where both households and home builders are allowed to re-sort following a policy change, and price coupled with the re-sorting behavior adjust to clear the housing market.

Results from these welfare estimation strategies have broad implications for the literature. Firstly, the results corroborate evidence by Walsh (2007) that significant differences in welfare implications arise when households are allowed to re-sort following a non-marginal policy intervention. Secondly, the current research extends Walsh's result to show that incorporating supply responses through builder re-sorting results in substantially

different welfare measures for some policies. Lastly, attention to the spatial location of proposed policy and the decision makers impacted by policy are crucial in determining the welfare implications of proposed changes to open space policy.

The final chapter summarizes key results and discusses how the current work fills several gaps in the existing open space literature. These gaps include defining a set of heterogeneous preferences over a variety of heterogeneous forms of open space, expanding the evidence on the importance of incorporating general equilibrium feedback effects, and introducing an elastic supply of new housing to the existing micro-level housing demand literature. As the current work is a first attempt at addressing these issues in the context of open space, additional work is needed to corroborate and extend the results found in this research. Opportunities to further refine and expand the current research are discussed in the conclusion laying the groundwork for future research building on the innovations and results of the work presented here.

# **Chapter 2**

## **Using the Housing Market to Value Local Public Goods**

### **2.1 Introduction**

This chapter examines the literature using revealed preference methods to value local public goods. In general, determining the value of local public goods is challenging because no direct market exists in which they are exchanged. To overcome the lack of an explicit market, economists have resorted to using observable housing markets in order to recover the value of local public goods. In the current work, the locational equilibrium inherent in a housing market is used to recover preferences for public goods, including open space. This chapter reviews the theoretical and empirical revealed preference methods used in the literature to value open space with an emphasis on market equilibrium facilitating identification of agents' underlying preferences.

The housing market provides two readily observable equilibriums: price and location. Price equilibrium assumes that the observed market prices of housing capture the values agents place on local public goods in the neighborhoods in which a house is located. By observing the spatial variation in prices across the urban landscape, the value of local public goods can in some instances be inferred. The use of price equilibrium gives rise to the hedonic model. In contrast to the hedonic model, locational equilibrium models use information on the location decisions of agents to define a Nash equilibrium in which all agents' location choices are optimal given the location decisions of other agents. By observing the public goods provided in the locations chosen by agents, the underlying preferences for those goods are inferred.

Inherent in the definition of market equilibrium are considerations of both supply and demand. While the majority of the existing literature focuses exclusively on the demand for public goods, a small segment of the literature has begun to recognize the importance of considering the supply side of the housing market in addition to household demand. Until recently, the majority of the research focused on the supply of housing used aggregate time series data. With the availability of more detailed data, recent attempts to treat supply at an individual agent level have been undertaken by Epple et al. (2007) and Murphy (2008). The current research expands on the micro-level agent based treatment of supply by modeling home builders' decisions of where to locate new housing.

The next two sections discuss the open space literature employing the hedonic and locational equilibrium estimation methods, respectively. Those sections are followed by an introduction of the horizontal sorting form of locational equilibrium models and a discussion

of how they differ from the vertical sorting model. Next, supply related literature including attempts at modeling the supply side of the housing market at an aggregate level and the non-equilibrium based land conversion literature are presented showing that they share many similarities with the modeling of supply presented in the current work. Lastly, the chapter concludes by placing the current work in the context of the existing open space literature.

## 2.2 Hedonic Model

Rosen (1974) interpreted the hedonic model in terms of an equilibrium price function established at the tangency between households' bid and suppliers' offer curves. The equilibrium price schedule is a function of housing characteristics, including open space public goods. Denoting the number of houses by  $h$ , and the characteristics of those houses by the vector  $\mathbf{z} = (z_1, z_2, \dots, z_n)$ , where in this treatment  $z_1, \dots, z_n$  is a vector of characteristics for given houses, the equilibrium hedonic price is a function mapping  $\mathbf{z}$  to a particular house expressed by  $P(\mathbf{z}) = P(z_1, z_2, \dots, z_n)$ . In addition to providing an equilibrium price schedule, Rosen also suggested that the underlying demands for housing characteristics,  $\mathbf{z}$ , could be recovered allowing welfare analysis for non-marginal changes to those attributes.

The hedonic price schedule is based on the utility-maximizing behavior of households facing a budget constraint. In particular, household  $i$  is assumed to maximize utility given by

$$u^i = U^i(\mathbf{z}, x, \boldsymbol{\alpha}^i), \quad (2.1)$$

subject to the budget constraint

$$m^i = P(\mathbf{z}) + x. \quad (2.2)$$

In this notation,  $m$  is income,  $x$  is a numeraire good with price normalized to 1, and  $\alpha$  are individual characteristics that vary across agents. At the optimum, the marginal rate of substitution between a housing characteristic and the numeraire good equals the marginal price of the housing characteristic. This relationship forms the basis of the hedonic model and is expressed as

$$\frac{\partial P}{\partial z_k} = \frac{\partial U}{\partial z_k} / \frac{\partial U}{\partial x}, k = 1, \dots, n. \quad (2.3)$$

To determine the entire hedonic price function, multiple households facing a continuous choice set are needed. By recovering the point at which each household equates their marginal willingness to pay for an additional unit of  $z_k$  with the marginal price of  $z_k$ , the price associated with different levels of  $z_k$  are recovered tracing out the hedonic price function.

While the above procedure determines the hedonic price function, the underlying demand function for  $z_k$  is not recovered. To recover this latent demand, Rosen (1974) proposes a two-step approach. The first step estimates a reduced form equation recovering the housing price function,  $P(\mathbf{z})$ . The second step regresses marginal willingness to pay values on housing characteristics and exogenous demand shifters,  $\alpha$ . This strategy for recovering demands for housing characteristics suffers from the fact that multiple demand curves could pass through the single tangency point between a household's bid function and the hedonic price curve. As a result, the original approach by Rosen (1974) does not identify a unique demand curve.

The fundamental limitation arising from the inability to identify the underlying demand curve for housing characteristics is an inability to evaluate responses following large changes in those characteristics. Because only a single point along the demand curve is observed, and the underlying shape of the demand curve is unknown, it is only reasonable to focus on policy changes very near the observed point on the demand curve. From a welfare analysis perspective, this results in an inability to evaluate large changes in housing characteristics that would shift the underlying hedonic price function.

To overcome the identification problem, several strategies have been proposed but are not widespread in the literature and are absent in the existing literature on open space. The two most popular approaches include specifying a utility function along the lines of Bajari and Benkard (2005) and using data from multiple markets as proposed by Brown and Rosen (1982). While each of these methods avoids the identification problem and permits the recovery of the underlying demand curve for housing characteristics, they require either strong preference restrictions or extremely rich data, respectively, which are not widely available. These and other promising approaches to relax the identification problem are appearing more frequently in the literature, but have not yet appeared in the open space literature. For a summary of the identification challenges and methods used to overcome them, see Palmquist (2005).

### **2.2.1 Open space hedonic literature**

The vast majority of the literature focusing on open space falls within the hedonic framework and assumes a fixed supply of housing. These approaches assume that a set of heterogeneous

households are able to costlessly move within the housing market, and select from a set of houses containing a continuum of characteristics in order to maximize their individual utility. The scope of the housing market varies widely from study to study but typically consists of a single metropolitan area and spans several years. The restriction of the choice set to a small number of years is important as it is unlikely the hedonic price function remains constant across long periods of time.

In recent years, the expansion of the urban/rural fringe has brought open space to the forefront of local policy discussions. As mentioned in chapter 1, Kotchen and Powers (2006) report that between 1998 and 2003, over 1,000 open space referenda have been voted upon in the United States with over 80% passing. To aid policymakers in their understanding of open space, recent empirical literature has focused on the amenity value of heterogeneous forms of open space and the importance of location in determining the value of open space.

Heterogeneity of open space is often tied to the expected duration of protected land. Along these lines, authors typically define several broad categories of open space ranging from permanent open space to various types of temporary open space. The difference in duration is often a result of the development potential of the open space. The hypothesis tested by this strain of literature is that the potential for development reduces the amenity value of open space as households do not expect developable land to remain open space forever.

Empirical measurement of the amenity value provided by permanent versus temporary forms of open space has received considerable attention. Smith et al. (2002) estimate hedonic price functions and report that permanent open space provides greater

amenity value than adjustable open space. In similar work, Irwin (2002) performs a hedonic analysis that accounts for endogenous land use spillovers and unobserved spatial correlation using instrumental variables. She finds a price premium is associated with permanently preserved open space compared to developable open space. Geoghegan (2002) also performs a hedonic analysis that finds permanent open space increases land values by over three times the amount of developable open space.

In addition to understanding the importance of open space duration, efforts to account for the importance of the spatial location of open space are increasingly possible as a result of more detailed land use and property transactions data. Anderson and West (2006) account for spatial structure of the housing market by including location specific dummy variables in a hedonic regression. These variables capture spatial varying characteristics common across houses located in the same spatial unit. Using data from the Twin Cities for 1997, Anderson and West find that the value of open space varies with the location of a house and with neighborhood characteristics. This result corroborates results found by Smith et al. (2002) that changes in land use patterns, and in particular development, alter the amenity value provided by open space.

### **2.2.2 Limitations of the hedonic literature**

Hedonic studies of open space have illustrated the importance of accounting for open space heterogeneity resulting from both the location of open space as well as the type of open space. While it is possible to account for spatial differences across the urban landscape using traditional hedonic techniques, the availability of richer spatial data opens the door for

analysis by methods capable of accounting for spatial variation at a much smaller spatial scale. In addition to spatial resolution improvement, data on individual households is becoming more readily accessible allowing estimation of heterogeneous preferences for housing characteristics using recently developed models. These preferences cannot generally be recovered using the traditional hedonic framework.

Despite the advancements in data and modeling, a limitation of the hedonic method is the difficulty in recovering underlying demands for housing characteristics preventing counterfactual analysis of large changes in those characteristics. This limitation arises because large changes in housing characteristics may result in the relocation of households; thereby changing demand and the underlying hedonic price function. As a result of this limitation, the open space hedonic literature is limited to presenting marginal willingness to pay measures and partial equilibrium welfare measures assuming marginal (or very small) changes in housing characteristics.

The marginal nature of the changes analyzed by hedonic methods is also tied to the assumption of a continuum of housing characteristics. In practice, often there are discrete differences in available housing characteristics that cannot to be accounted for using hedonic methods. An alternative approach to the hedonic is the use of discrete choice models where households select from a finite set of housing alternatives in order to maximize utility. A popular form of discrete choice models used in the housing literature is locational equilibrium sorting models. These are the focus of the following section and are the basis of the current research.

### **2.3 Sorting Model**

Sorting models use information on location decisions to infer underlying preferences for local public goods by exploiting the spatial variation in those goods. While a price equilibrium certainly exists in the market, sorting models make use of locational equilibrium concepts in which every agents' location decision is optimal given the location decisions of other agents. Tibout (1956) outlined one of the earliest conceptualizations of a sorting model by assuming mobile households sort across distinct neighborhoods based on preferences for local public goods contained in those neighborhoods. For many years, work on sorting models was restricted to theoretical advancements as the data and empirical methodology needed to implement empirical sorting models was not available.

Recent advances have pushed the use of location equilibrium to infer preferences for local public goods in several directions. Epple and Sieg (1999) and subsequent work by Epple, et al. (2001) introduce a “vertical” sorting model where a locational equilibrium is used to determine preferences for local public goods using moment conditions based on public goods indexes calculated for each discrete community. A different approach to sorting referred to as a “horizontal” sorting model is introduced by Bayer and Timmins (2005), and is based on the random utility model outlined in McFadden (1978). The horizontal model incorporates a rich set of heterogeneous preferences for local goods and incorporates unobserved housing characteristics through a set of choice specific fixed effects. A discussion of each sorting methodology and their differences follows.

### **2.3.1 Vertical sorting model**

The vertical sorting model combines both hedonic and discrete choice techniques. The first stage of estimation involves creating price indexes using hedonic methods as well as indexes for public goods for a set of discrete communities. The model then assumes that households choose a single community, and conditional on that choice, are free to select any continuous quantity of housing within that community. A fundamental assumption of much of the vertical sorting literature is that all agents agree on the ranking of the public goods index across communities giving rise to a vertical structure of preferences. Recent work by Kuminoff (2007) has relaxed this restrictive preference assumption but is not yet widespread in the sorting literature.

Sieg et al. (2004) improve upon the early vertical sorting models by estimating all parameters simultaneously within a one-step GMM framework. In addition to the estimation advancement, they also show that it is possible to calculate general equilibrium welfare effects resulting from non-marginal changes in public goods. To calculate a new equilibrium, the authors assume that large changes in public goods cause some households to choose different communities resulting in the creation of a new set of prices in response to changes in demand for communities. These prices in turn feed back into the models, resulting in the ability to calculate general equilibrium welfare effects for large changes in public goods that are difficult to obtain using hedonic methods.

Walsh (2007) uses the vertical sorting model to evaluate the impacts of open space policies and their effect on urban growth using data from Wake County, North Carolina. In addition to using a model capable of calculating general equilibrium welfare effects, he

incorporates a land conversion decision into the sorting model allowing supply responses in the form of lot conversion. While not accounting for supply of specific houses, this work is the first to include a land conversion supply decision within a sorting framework. Walsh finds that partial and general equilibrium welfare measures vary substantially, offering motivation to modeling frameworks that allow general equilibrium analysis.

In his application, open space is divided into an exogenous component capturing the distance to the nearest open space and an endogenous component capturing the amount of open space within a neighborhood. The inclusion of endogenous open space is made possible by modeling the land conversion process in which open space is converted to development based on prices and lot characteristics. In addition to open space, households also have preferences over communities that vary by land price and amenity level. In equilibrium, households choose optimal neighborhoods, optimal lot sizes, and the land market clears.

While Walsh's work represents a substantial step forward in the modeling of the urban landscape, several extensions are possible. The first is related to the restrictive preferences imposed by the vertical sorting model. It is unlikely that all households have the same preference ranking for communities and more to the point, preferences for open space likely vary by type of open space as well as over households of different socio-economic status. A second extension regards the treatment of supply as a land conversion decision rather than an actual housing supply decision. For the purposes of determining the endogenous amount of open space in a neighborhood, a land conversion decision is sufficient. However, additional insights could be gained by modeling changes in the supply

of specific housing types in response to policy changes. It is likely that different open space policies have different impacts on the types of housing located within a community.

### **2.3.2 Horizontal sorting model**

As an alternative to the vertical sorting model, a class of “horizontal” sorting models discussed by Bayer and Timmins (2007) has been introduced into the literature. The primary advantage of the horizontal model over the vertical model is the incorporation of an extremely flexible set of preferences. While no horizontal sorting models have been estimated and applied to open space questions, the model has appeared in recent empirical work by Bayer et al. (2005), Timmins and Murdock (2007), and Timmins (2007), among others.

Unlike the vertical sorting model, the horizontal model is based on the discrete choice random utility maximization framework (RUM) of McFadden (1978). A distinguishing feature of the horizontal sorting model compared with the traditional RUM framework is the inclusion of a complete set of alternative specific constants (ASCs) capturing components of utility unobservable to the econometrician, but observable to decision making agents. In addition to inclusion of ASCs, the horizontal model incorporates heterogeneous preferences through interactions between individual specific socio-economic variables and alternative specific variables. Inclusion of these heterogeneous interactions permits agents’ preferences to vary over public goods, housing characteristics, and even the demographics of their neighbors.

A second key difference between the vertical and horizontal sorting model surrounds the definition of housing equilibrium. The equilibrium concept present in the vertical model is a type of assignment equilibrium where each household is assigned a specific house and in equilibrium, all households are assigned to different houses such that the total demand for houses and total supply of houses are equal. The equilibrium concept present in the horizontal sorting model does not allocate households to specific houses, but rather assigns a positive probability of each household choosing every house<sup>1</sup> in the market. In equilibrium, the sum of all households' probabilities for locating in a specific housing type is one; indicating that expected demand for each house exactly equals the supply of each house.

Using the locational equilibrium whereby expected demand equals supply, the horizontal sorting model is capable of solving for a new price equilibrium resulting from non-marginal changes to public goods. As with the vertical model, the ability to solve for a new equilibrium allows general equilibrium welfare analysis in response to policy changes. The mechanism by which expected demand and supply are equilibrated following a policy shock is through the price of housing. For any house in which expected demand exceeds supply, the price is raised thereby reducing demand. Likewise, for any house where expected demand is below supply, the price is lowered in order to increase demand. In this process, price serves as the instrument by which supply and demand are equated to clear the housing market.

---

<sup>1</sup> In the current work, houses are represented by housing types defined over the dimensions location, house size, and time of transaction.

## 2.4 Aggregate Supply Models

With the exception of the vertical sorting paper by Walsh (2007), the literature discussed above has not addressed the supply side of the housing market. As the housing market by definition is composed of both a demand for housing and supply of housing, the disparity between the large amount of literature focused on housing demand and the extremely limited amount of literature focused on housing supply is surprising. This discrepancy was highlighted by DiPasquale (1999), who surveyed the existing housing literature finding relatively scant work on housing supply. The empirical housing supply literature that is present typically uses aggregate time series data rather than the micro-level decisions of individual agents. The current work is one of the first to explore housing supply from a micro-level perspective by focusing on the development decisions of builders.

Within the aggregate housing supply literature, the asset market approach developed by Poterba (1984) most closely corresponds to the micro level treatment used in the current work as it incorporates perfectly competitive suppliers of new housing. Poterba assumes that competitive builders supply housing and are principally concerned with the final price of housing. To close the model, Poterba also includes the inverse demand for housing and defines long run equilibrium as the point where the housing stock is constant and prices clear the market. Because his model does not apply to individual decision makers, the author is unable to evaluate the micro-level decisions of optimizing agents. In light of this, it is not surprising that construction costs are found to have little role in determining the aggregate supply of new housing.

The asset approach to modeling the macro housing market was extended by Topel and Rosen (1988) who incorporate intertemporal dynamics into the model of housing supply. The authors specify a model where suppliers are concerned about cost changes across time periods and choose to supply new construction in order to minimize intertemporal costs. To accomplish this, the model introduces expectations of future costs into the decision making process of suppliers. Despite the increased complexity of this model, the authors find that construction costs do not play an important role in the supply of new housing. In addition, neither the work by Porterba (1984) nor Topel and Rosen (1988) model account for land as an input to the production of new housing.

Other attempts at modeling the housing market include the class of urban spatial models pioneered by DiPasquale and Wheaton (1994) which use aggregate time series data to examine shocks to the long-run equilibrium of the housing market. In contrast to the asset approach, these models treat land along, with land price, as important inputs into the supply of new housing. In their model, land price is a function of the existing housing stock and enters into the long run equilibrium supply of housing.

While both the asset market and urban spatial theory approaches address equilibrium in the housing market, neither approach examines the disaggregate decisions of households and builders that form the foundation of the housing market. As such, the aggregate housing market literature provides only limited guidance in terms of the drivers of individual level housing supply decisions. As an example, both aggregate approaches discussed above find that construction costs are not an important driver of new housing supply, a somewhat counterintuitive result. With a greater availability of detailed housing data, it is now possible

to examine housing supply at the micro-level to better understand the drivers of housing supply within an expanding metropolitan area.

## **2.5 Land Conversion Models**

Related to the supply of housing is the idea of land conversion. While not directly modeling housing supply decisions, the existing reduced form land conversion literature does make use of micro-level data to try and understand what motivates the conversion of land from undeveloped to developed land. As a standalone model, the primary criticism of the land conversion literature is the lack of an equilibrium concept connecting supply and demand either through price as shown in the hedonic or through some form of locational equilibrium as used in the sorting literature. As a result, demand responses to changes in land use are not accounted for in the land conversion literature.

An early paper in the land use change literature by Bockstael (1996) addresses land use changes in the rapidly growing Patuxent watershed in Maryland. The paper is motivated by a desire to incorporate economics into ongoing ecological research modeling land use changes that omitted economic considerations. To incorporate economics into the ecological models, a measure of the value of land is estimated using hedonic methods. With this land value in hand, a discrete choice model is estimated that predicts the probability of a parcel of land being converted from an undeveloped to a developed state.

Some notable extensions to the land use change literature are the incorporation of a time dimension of conversion by Irwin and Bockstael (2002) and the introduction of heterogeneous development by Newburn and Berck (2006) using a mixed logit discrete

choice model. A survey of much of the existing land use change literature is performed by Bell and Irwin (2002) and highlights the benefits and challenges present in this line of research; many of which are also present in the horizontal sorting approach employed in the current research.

Among the benefits provided by the land use change literature is the ability to model landscape changes at an extremely small spatial scale, often at the parcel level. The land use literature is similar to the horizontal sorting literature employed in the current work in terms of many of the micro-level variables included in analyses of supply. For example, Newburn and Beck (2006) find that provision of water and sewer services raises the likelihood of high density development. In the current work, I also find that urban services play an important role in supply decisions, and have a greater impact on larger builders than smaller builders.

## **2.6 Conclusions**

This chapter reviewed the empirical housing market literature focusing on advancements in empirical methods used to recover preferences from observed behavior. While much of the existing housing demand literature has focused on households' decisions, empirical housing supply literature is predominately concerned with the aggregate housing market. As a result, the gap between our understanding of households' preferences and those of suppliers of new housing is considerable. The current work uses a newly developed class of horizontal sorting models to recover preferences for both households and builders in the context of a single metropolitan housing market.

The horizontal sorting model overcomes several empirical limitations present in much of the existing housing literature on open space. Due to data limitations, hedonic techniques used to value open space are often unable to recover the underlying demand functions for local public goods. This limitation precludes welfare estimation for policies involving large changes in public goods and has given rise to other methods that are able to solve for a new housing equilibrium following a policy shock. Among the newly developed methods are locational equilibrium models that circumvent the hedonic limitations by exploiting the spatial structure of the housing market.

Locational equilibrium sorting models appearing in the literature fall within two broad categories related to the flexibility of the underlying preference structure imposed by the model. To date, only a single paper by Walsh (2007) has utilized these models to explore preferences for open space. In contrast to the vertical sorting model used by Walsh, recent development of horizontal sorting models capable of recovering preferences that vary by agent and across a diverse set of open space have been introduced. The use of either sorting model permits analysis of general equilibrium welfare effects resulting from large changes in local public goods by recovering new market equilibriums following those changes.

In addition to modeling the demand for housing, recent work has begun to consider the role of supply in empirical housing applications. Incorporating an estimated supply of new housing at a micro-level is novel in that the existing literature has focused almost exclusively on aggregate housing supply or has concentrated on the conversion of land rather than the actual supply of housing at the micro-level. By estimating both housing demand and housing supply the current work is one of the first to incorporate feedback effects between

both sides of the housing market resulting from a policy shock through the joint determination of market clearing prices.

# Chapter 3

## Data

### 3.1 Introduction

Empirical estimation of a horizontal sorting model for both household and builder location decisions requires a rich set of micro-level data. The structure of the assembled data is driven primarily by two aspects of the horizontal sorting model used in this research. The first is the unit of spatial resolution over which housing types are defined. In this research, the spatial unit is defined as the relatively small Census 2000 block group containing between 300 and 500 households. The second driver of the data is the definition of distinct housing types within each spatial block group unit. The definition of housing types defines a common choice set facing both households and builders. Each housing type is defined by a unique combination of location, house size, and time.

Modeling both the demand and supply side of the housing market requires a diverse set of data in order to tease out the drivers of housing decisions made by both households and builders. Of particular importance is determining exclusionary variables for both demand

and supply that are necessary to obtain identification. For demand, exclusionary variables are derived from assumptions regarding households' preferences for amenities and neighborhood composition which enter builders' utility only through the market price of housing. For supply, exclusionary variables are obtained using detailed soil characteristics data which determine the cost of development but are assumed not to directly enter households' utility.

As a primary focus of this research is to analyze open space policies, land use data is especially important. The current chapter examines the data components needed to estimate the empirical sorting model of the Twin Cities housing market. The chapter begins by introducing the study area followed by a description of the choice set facing both households and builders and concluding with description of the various data components assembled to parameterize an empirical sorting model of the housing market.

## **3.2 Twin Cities Study Area**

The Twin Cities serve as the study area for the current research. Located in east-central Minnesota, the Twin Cities metropolitan area consists of 7 counties with the cities of Minneapolis and St. Paul at its center. These cities are surrounded by numerous smaller municipalities providing a diverse set of local amenities. Including the surrounding municipalities, the population of the Twin Cities area is over 3 million people and has grown approximately 20% from a 1990 census population of 2,538,834 to a 2000 census population of 2,968,806. Figure 3.1 provides a map of the Twin Cities highlighting the counties within

the metropolitan area. The counties range in size from the relatively small and heavily developed Ramsey County to the large, primarily agricultural Dakota County.

The location and layout of the study area provides several modeling benefits. The first benefit comes as a result of the Twin Cities metropolitan area being relatively self-contained. This independence from other large neighboring metropolitan areas helps eliminate concerns over the correct scope of the choice set. Having no large metropolitan areas nearby, it is unlikely individuals work outside the Twin Cities area and equally unlikely that other individuals commute into the area for work. In addition to being relatively isolated, the 7 county metropolitan area exhibits a fairly even growth pattern with an extensive urban center surrounded by lower density suburban development and agricultural lands beyond that. The undeveloped land that exists within the study area provides an ideal setting to study the effects of urban development in response to changes in open space features.

The extent and landscape diversity of the Twin Cities area lends itself to the use of a horizontal sorting model capable of capturing households' preference heterogeneity for local amenities as well as builders' preference heterogeneity over exogenous features of the landscape. The remaining sections of this chapter provide an overview of the cleaned data used during estimation. Because the database used in the current research was created from many individual components that had not been used previously, several steps were required to obtain the final data used in estimation. Additional details on the data cleaning procedure are provided in an appendix.

### **3.3 Choice Set of Housing Types**

The horizontal sorting model employed in the current research requires a delineation of the landscape into discrete housing options. To this end, defining housing types over location, house size, and time allows the choice set to better reflect the diverse and dynamic landscape associated with a housing market for a large metropolitan area. The first dimension, location, can be thought of as representing neighborhoods. In order to use the available census data to characterize neighborhood characteristics and social demographics, this dimension is defined by the set of Census 2000 block groups. Block groups are large enough to include a variety of house sizes, yet small enough to be distinguished from one another by unique landscape differences and neighborhood characteristics. Figure 3.2 shows the Census 2000 block groups across the entire seven county area.

The second dimension of a housing type is house size. To capture the size aspect of the choice set, house size is defined by square footage and broken into three ranges based on the 33<sup>rd</sup> and 66<sup>th</sup> percentile of square footage represented in available transactions data. By dividing housing size into small, medium, and large categories, it is possible to approximate differences between subdivisions within a single spatial area. This approximation is possible because most houses within a subdivision fall into a single size category of small, medium, or large. To see an example of this, figure 3.3 shows several block groups with housing transactions falling into bins based on house size identified by different colors and shapes.

The final component of a housing type is the timing of the choice. Dividing the choice set by time raises the issue of how to incorporate agents' expectations about changing neighborhood attributes and future prices as they decide the timing of their housing decision.

To abstract from these difficulties, the time dimension is assumed exogenous. This assumption restricts the choice set facing agents to the same time period in which they are observed. For example, the choice set facing a household observed in the first time period only consists of alternatives where the time dimension of a housing type is one. While clearly a simplification, this assumption is consistent with some behavior observed driving housing purchases such as learning that an employer plans to move locations, having an unexpected lifestyle change, or simply having reached a financial state in which housing becomes affordable. In addition, the assumption of an exogenous time dimension seems plausible when considering that time divisions are defined over multiple years. In particular, time is classified by four periods defined as 1990-1996, 1997-1999, 2000-2004, and 2005-2006. The combination of a unique location, house size, and time define a “choice” available to both households and builders.

The housing types available to households and builders are determined using transactions data on single family detached housing. Table 3.1 provides a breakdown of the number of housing types represented in the transactions data (described below) for different agents for each time period and house size category. There is a considerable discrepancy between the number of housing types in which households are observed to have purchased housing compared to the number of housing types experiencing new construction. This discrepancy is expected as the treatment of supply only concerns new construction with housing supply from existing housing treated as exogenous to the model. The lack of observed new housing supply for many of the housing types will result in additional estimation challenges for supply not faced in demand estimation.

### **3.4 Property Transactions**

Arm's-length property transactions form the basis of estimation as house location and characteristics reveal the preferences for both households and builders. The transactions database assembled for the current research is unique in that it spans 17 years from 1990 through 2006 and covers a large spatial area. The length of time over which housing transactions are observed covers the rapid metropolitan growth that occurred in the late 1990s and early 2000s. Proprietary transactions data was purchased from Plat Research, a local Minnesota company specializing in the collection of real estate data. Their database includes all real estate transactions and transactions prices, not just single family detached transactions, along with a large assortment of attributes associated with each transaction. The primary source of their data is information gathered from local assessors' offices. Because data from multiple assessors' offices is used, there is a wide range of variables and characteristics available across the different counties and municipalities in the study area.

The current research uses single family detached transactions as the basis for estimating the demand and supply of housing, as well as land transactions to estimate the market price of land. Observing transactions for both the input price of land and the output price of housing permits identification of land price separate from the final retail price of housing. Because home builders are concerned with the input and output price of housing, an important contribution of the current work is the ability to simultaneously identify both input and output prices from the perspective of the builder.

### 3.4.1 Single family detached

After cleaning the data, a final set of 461,795 single family detached housing transactions is obtained containing a complete set of housing attributes. Figure 3.4 plots these transactions. The resulting residential transactions form the basis for both demand and supply estimation of horizontal sorting models. The data appendix provides a detailed description of the procedures used to obtain this set of transactions from the raw transactions data. Each house represented in these transactions is assumed to be purchased by a unique household. To generate variables consistent with the definition of the choice set, some aggregation of housing transactions is required.

Transactions are assigned to bins forming alternatives in the choice set where each bin represents a particular housing type. The characteristics of all houses belonging to a particular bin are averaged to obtain the characteristics for a single housing type.<sup>2</sup> The average bin contains approximately 20 transactions. To provide a sense of the distribution of houses within bins, figure 3.5 provides a histogram showing the distribution of the number of transactions comprising each housing type.<sup>3</sup> Due to the relatively small number of houses within each bin and their similar sizes, this aggregation process preserves considerable data variation across housing types. In total, there are 20,444 housing types created from the single family detached housing transactions.

Figure 3.6 provides histograms of key housing characteristics for the set of 20,444 housing types. Looking at the histograms, there is a wide distribution of housing

---

<sup>2</sup> Other approaches for combining individual transactions such as taking the median house or performing some form of weighted average are certainly possible. For ease of construction, the mean is used here.

<sup>3</sup> The aggregation of individual transactions results in some smoothing of the housing characteristics as compared to the raw transactions data described in Appendix A.

characteristics across the various housing types, implying that the aggregation process did not result in the loss of significant data variation. Examining the histogram for square feet, the impact of including house size as a dimension of housing types becomes clear. The creation of three distinct sizes of houses results in a tri-modal histogram for square footage with a peak at approximately the median square footage within each size category. Recall that the square footage divisions used in the creation of housing types are 1248 square feet and 1838 square feet representing the 33<sup>rd</sup> and 66<sup>th</sup> percentile of square foot distribution, respectively. Despite this tri-modal appearance, there is still considerable variation in square feet across housing types.

### **3.4.2 Builder transactions**

To model supply, the set of transactions representing the initial sale of new construction to households is identified from the larger set of transactions data. To identify these transactions, the first sale of any house built between 1990 and 2006 is identified. From this set of transactions, if the first sale and year built coincide, the sale is assumed to be a builder selling a house developed by that builder. The number of sales identified as new home sales consists of 60,588 transactions.

For each transaction, the housing type the transaction falls within is determined based on the choice set definition defined using the complete set of transactions resulting in observable transactions in 4,545 out of the total 20,444 possible housing types. Table 3.1 shows the number of housing types broken down by house size and time period. There appears to be good coverage across time periods and house sizes contained within this data.

To provide a visual reference for the set of transactions identified as new supply, figure 3.7 plots these transactions showing a broad spatial distribution of the transactions.

Figure 3.8 plots a histogram of the number of supply transactions comprising each of the 4,545 housing types. As expected, the number of transactions in each bin is considerably less than those for the full set of housing transactions. Despite the smaller numbers, the average number of houses in a housing bin is 13.3 and the distribution appears smooth with a maximum number of transactions of 456 and many bins containing only a single transaction.

### **3.4.3 Land transactions**

Separate from the housing transactions discussed above are a set of 31,800 vacant land transactions. These transactions are used to identify the price of land across the metropolitan area at the Census Tract level. Figure 3.9 plots these transactions across space showing relatively uniform coverage across the study area, but clearly not as spatially resolved as the full set of single-family residential transactions. Additional details for the land transactions are provided in the data appendix.

## **3.5 Land Use Data**

The availability of detailed land use data at multiple points in time provides important variation across time periods needed to identify changes in the characteristics of the landscape in response to the expansion of the metropolitan area over time. Data provided by the Metropolitan Council, a local governing body for the entire seven-county metropolitan area, forms the basis of the land use data. The data they provide is available for the years

1990, 1997, 2000, and 2005. In addition to providing information on landscape changes, the availability of land use data is the basis for dividing the study period into 4 time periods. Comparing this data to separately obtained parcel maps, the land use data appears extremely accurate allowing identification of changes in land use over very small spatial scales.

Table 3.2 provides a summary of the land use categories by percent of total acreage for each of the four years of land use data for all categories of land use that I am able to identify.<sup>4</sup> Looking at the land use percentages, it is evident that Hennepin and Ramsey counties are the most heavily developed with over 30% of land devoted to residential development, compared to only 15% of land classified as residential across all seven counties. Because of the large amount of developed area in these counties, it is not surprising that they contain the smallest percentage of agricultural and undeveloped land at less than half the 7 county average.

Focusing on the differences in land use across time, one can see that the amount of residential land is increasing as the metropolitan area expands. This increase occurs in all counties and is most noticeable in the more rural counties where there exists the largest amount of undeveloped land available for development. For example, single family residential land use for Anoka County increases from approximately 15% in 1990 to over 21% by 2005. This 6% increase in single family residential land use is accompanied by increases in commercial and industrial land use as well. The category that appears to be providing the majority of this land is the agricultural and undeveloped category that decreases by 12% from over 63% in 1990 to less than 52% by 2005. A similar pattern of

---

<sup>4</sup> The slight discrepancy in total acreage across land use years is due to slight differences in the spatial resolution of the land use boundaries as well as rounding errors when calculating total acreage.

reduction in agricultural lands occurs in all counties resulting in a total loss of approximately 250,000 acres of agricultural and undeveloped land over the 16 year period spanning 1990 to 2005.

Despite the growing metropolitan area, all forms of protected open space show increasing acreage totals over time as efforts are made to preserve open space and plan new development around existing open space. These preservation efforts can be seen by looking at the expansion of regional parks, local parks (which are associated with development), agricultural preserves, and other miscellaneous types of protected open space. The growth of the metropolitan area has also resulted in the adoption of local policies supporting preservation programs to protect farmland and open space. Similar programs to the conservation programs adopted in the Twin Cities are used in other metropolitan areas, making the study of their impacts useful not only to local policymakers interested in open space, but to policymakers in other areas as well.

Two stylized open space programs that exist in the Twin Cities area were created with policy goals of limiting urban sprawl and protecting agricultural land and open space. These programs are the local Metropolitan Agriculture Preserve Program and the statewide Reinvest in Minnesota conservation program. Each program was developed prior to 1990 as policymakers became concerned with the loss of farm land and open space due to the rapidly expanding metropolitan area. These programs serve as natural experiments to evaluate two common conservation vehicles: relatively large, temporary agricultural easements and smaller, permanent conservation easements.

While both of these programs aim to preserve open space, there are two distinct differences between them. First, the Reinvest in Minnesota program consists primarily of small, permanent easements whereas the Metropolitan Agricultural Preserve program consists of much larger, temporary easements. The temporary aspect of the Agricultural Preserve program allows disenrollment of land after an eight year waiting period from the time a request for removal is received. While very little land is removed over the course of the study period, many of the farms enrolled have requested removal and will exit the program in the years to come. The difference in program duration also helps explain the size distinction. Because the purchase of permanent easement rights for large amounts of land is expensive, the Reinvest in Minnesota program primarily enrolls small numbers of acres scattered across the landscape, whereas the Metropolitan Agricultural Preserve program enrolls large tracts of often contiguous farm lands. Figures 3.10 and 3.11 show the enrolled land in each of these programs as of 2006. Institutional details for these conservation programs are contained in the data appendix.

### **3.6 Soil and Supply Characteristics**

Because direct measurements of land development costs are unavailable, the costs of new housing construction are approximated using information on soil characteristics and provision of urban services. Soil characteristics determine how much effort is required before development can begin while the availability of urban services, particularly sewer, determines whether wells or other costly forms of sewer systems are required. The source of soil data is the publicly available NRCS SSURGO data for which both spatial and tabular

data is available for the Twin Cities. Information on the provision of metropolitan services is provided by the Metropolitan Council for the years 1995, 1997, 2000, and 2005 providing good coverage across the time periods used in the current research.

Because home builders are only concerned with potentially developable land when considering land characteristics, the first step in creating cost variables is to isolate the undeveloped land that existed in each time period. The location of undeveloped land is obtained from the land use data for agricultural and undeveloped land coupled with parcel level data indicating the year built for every parcel. All parcels are assumed undeveloped prior to the indicated year built. Figure 3.12 shows the available land in each time period identified using this approach. As expected, the earliest time periods have the most developable land, decreasing rapidly near the urban/rural fringe over time.

The SSURGO data contains information on soil drainage, development suitability, non-irrigated soil capability, and slope. Each of these data layers contains information on the relevant soil characteristics in the form of a series of discrete ratings. The variables used in estimation are created by first extracting from the soil layers only the data corresponding to the set of developable land at each time period. The next step involves combining multiple ratings and determining the percent of land in each block group falling into each rating category.<sup>5</sup> Variables created include poor drainage, limited development suitability, low agricultural capability, and very high slope.

In addition to the SSURGO data, the Metropolitan Council provides data on the availability of urban services across the entire 7 county region at several points in time.

---

<sup>5</sup> Additional details on the aggregation of ratings are described in subsequent sections of this chapter.

Figure 3.13 shows the areas containing urban services at 4 points in time corresponding to the four time periods defined in the choice set. The primary benefit of locating development in an urban services area is the existence of sewer lines provided by local municipalities. As with the SSURGO data, urban services located in undeveloped land are extracted from the urban services shapefile and the percent of each block group containing urban services is calculated.

### **3.7 Neighborhood Demographics**

Census block group data is the smallest publicly available census data for which a full set of individual and housing attributes are available. This data is reported in the STF3A and SF3 census data files for the years 1990 and 2000, respectively. Table 3.3 lists the number of spatial divisions at both the block and block group level for the 1990 and 2000 census. There are 2,027 block groups within the seven county Twin Cities area in the 2000 census forming the basis of the spatial dimension of the choice set. One difficulty in using census data from multiple years is that spatial boundaries change over time as a result of population growth. To use the 1990 block group data, it must be weighted in terms of 2000 block groups to be comparable across census years. This process has been performed by Geolytics and the transformed 1990 block group data is used in place of the original 1990 block group data provided by the Census<sup>6</sup>.

To obtain block group data for the 1997 time period, the 1990 and 2000 block group data is linearly interpolated. Interpolation is possible because each dataset contains the same

---

<sup>6</sup> Tract level data is also obtained from Geolytics using the Neighborhood Change Database.

spatial boundaries. For the 2005 time period, the 2000 block group data is used without interpolation. Using this data, a variety of neighborhood and social characteristics are obtained that vary over both space and time.

### **3.8 Household and Builder Characteristics**

The 1990 and 2000 Census also provide a limited set of characteristics at the block level in the STF1B and SF1 data files, respectively. The block level data is formed from a 100% sample of households and contains information on age and household composition. While not providing individual characteristics per se, census blocks typically contain less than 50 households, which reduce to an average of less than three transactions after dividing the observed housing transactions by time period and house size. As a result of this fine spatial scale, the block level census data provides an approximate measure of individual characteristics allowing all households within a block to be assigned the block level data as individual characteristics.<sup>7</sup>

The linear interpolation of census block data between 1990 and 2000 blocks is not as straightforward due to changes in the spatial boundaries across these years. For this reason, the 1990 block data is used for both the 1990 and 1997 time periods and the 2000 block data is used for the 2000 and 2005 time periods. A set of robustness tests comparing different strategies for interpolating both block group and block data was performed which revealed no significant differences from alternative interpolation approaches.<sup>8</sup>

---

<sup>7</sup> One might be concerned with assigning the 1990 block level data to households observed in the 1997 time period. Linearly interpolating the 1990 and 2000 data for these years did not result in substantial differences.

<sup>8</sup> For additional details, contact author.

Unlike the block group data, because the block level data is at such a small scale, only limited data is publicly available. Data provided at the block level contains the age and number of household members along with the size of the household. Using this data, variables for the average number of children per household, average number of working adults per household, average number of retirees per household, and average household size are created. Table 3.4 shows summary statistics for household characteristics obtained at the block level. Examining these statistics, it is apparent that quite a bit of variation across different census blocks exists. This variation allows identification of heterogeneous preferences during demand estimation.

The ability to obtain builder characteristics is very limited. Virtually the only source of builder characteristics comes from the residential transactions data. Using this data, the initial sales of new construction are isolated from which the name of the homebuilder can be identified. Using this approach, 13,885 builder names are identified and are manually name matched to capture instances where the same builder appears multiple times under slightly different names.<sup>9</sup> After manually matching builder names, a unique set of 10,101 names is obtained and each transaction is tagged with a unique builder identification number.

Using the tagged transactions, the total number of homes built by each home builder is calculated and used to approximate the size of the building company. By calculating the number of homes built in all previous time periods in addition to the current time period, this value changes over time for a particular builder. Figure 3.14 plots a histogram of the number of houses built by each builder where the same builder enters multiple times if observed in

---

<sup>9</sup> Typically, this resulted due to abbreviations in the name and misspelling.

multiple time periods. Table 3.5 shows summary statistics for the number of houses built by builders showing a large variety of builder sizes present in the data. The largest builder is responsible for building over 2000 houses while the average builder is responsible for building only 5.4 houses.

### **3.9 Conclusions**

This chapter introduced the data required for estimating an equilibrium housing model defined over housing types. The data appendix describes in greater detail the steps required to generate the final set of data discussed in the current chapter. A key feature of the applied sorting model used in estimation is the definition of specific housing types defined over the dimensions of location, house size, and time. The definition of a set of discrete housing types form the choice set facing both households and builders. As a result of the choice set definition, data is aggregated to correspond to the dimensions included in the choice set definition. The primary data aggregation involves creating neighborhood variables and variables that vary over house size and time within a neighborhood. For some data types, such as land use, house size does not influence aggregation as land use only varies by neighborhood and time period.

The final data components needed for estimation include residential transactions, land use, social and demographic variables provided by the census, development suitability data in the form of soil characteristics and urban services provision, and agent specific characteristics. An important attribute of the data presented in this chapter is the rich heterogeneity across both space and time improving the ability to identify aspects of

individual preferences that otherwise would not be identifiable. This is particularly helpful in describing preferences for open space, forming the basis of policy counterfactuals discussed in subsequent chapters.

**Table 3.1: Housing Types with Observed Transactions**

Time Period	House Size	Agent		
		Household	Builder (chosen)	Builder (all)
1990 - 1996	Small <sup>a</sup>	1755	577	1663
	Medium <sup>b</sup>	1792	586	1699
	Large <sup>c</sup>	1652	563	1561
	All	5199	1726	4923
1997 - 1999	Small	1640	358	1544
	Medium	1705	368	1597
	Large	1545	328	1440
	All	4890	1054	4581
2000 - 2004	Small	1764	490	1621
	Medium	1859	494	1709
	Large	1721	430	1577
	All	5344	1414	4907
2005 - 2006	Small	1665	118	1415
	Medium	1784	120	1520
	Large	1562	113	1347
	All	5011	351	4282
All	Small	6824	1543	6243
	Medium	7140	1568	6525
	Large	6480	1434	5925
	All	20444	4545	18693

<sup>a</sup>Represents houses  $\leq 1248$  square feet.

<sup>b</sup>Represents houses  $> 1248$  and  $\leq 1838$  square feet.

<sup>c</sup>Represents houses  $> 1838$  square feet.

**Table 3.2: Land Use by Time Period in Acreage and Percentage**

Time Period	Land Use Category	County							Total <sup>a</sup>	
		Anoka	Carver	Dakota	Hennepin	Ramsey	Scott	Washington	Acres	Percent
		Percent	Percent	Percent	Percent	Percent	Percent	Percent		
1990	Agricultural Preserve	0.6%	15.3%	8.7%	2.9%	0.0%	1.8%	2.8%	93,742	4.9%
	Agricultural/Undeveloped	63.3%	68.5%	68.8%	36.4%	18.7%	79.7%	67.8%	1,135,406	59.7%
	Cemetery	0.0%	0.0%	0.1%	0.3%	0.5%	0.0%	0.0%	2,069	0.1%
	Commercial	0.9%	0.2%	0.9%	2.6%	4.0%	0.6%	0.6%	24,012	1.3%
	Golf	0.4%	0.4%	0.5%	1.5%	1.6%	0.2%	0.6%	13,525	0.7%
	Highway (4 lane)	0.9%	0.1%	1.0%	1.9%	3.3%	0.5%	0.8%	20,706	1.1%
	Industrial	1.1%	0.4%	1.7%	4.1%	7.3%	1.0%	1.2%	40,033	2.1%
	Open Space (non-park)	1.3%	1.2%	0.5%	2.0%	1.4%	0.8%	0.9%	21,954	1.2%
	Miscellaneous	2.4%	0.7%	1.4%	4.1%	6.0%	0.7%	1.7%	42,207	2.2%
	Parks (local)	0.6%	0.1%	0.8%	1.6%	1.6%	0.3%	0.3%	14,483	0.8%
	Parks (regional)	7.4%	2.3%	1.6%	4.7%	5.3%	3.9%	2.2%	71,998	3.8%
	Residential (multi-family)	0.5%	0.1%	0.9%	3.0%	3.6%	0.1%	0.4%	22,105	1.2%
	Residential(single-family)	15.0%	5.0%	9.9%	26.1%	37.4%	5.8%	11.2%	277,832	14.6%
	Water	5.4%	5.8%	3.4%	8.9%	9.4%	4.6%	9.4%	123,060	6.5%
Total Number of Acres		285,935	240,366	374,820	387,841	108,655	235,317	270,197	1,903,132	100.0%
1997	Agricultural Preserve	0.6%	20.5%	13.5%	2.9%	0.0%	2.3%	3.0%	126,272	6.6%
	Agricultural/Undeveloped	58.8%	61.4%	58.8%	32.4%	15.1%	75.8%	61.5%	1,022,115	53.7%
	Cemetery	0.0%	0.0%	0.1%	0.3%	0.5%	0.0%	0.0%	2,135	0.1%
	Commercial	1.1%	0.3%	1.2%	3.1%	4.5%	0.6%	1.0%	29,274	1.5%
	Golf	0.5%	0.4%	0.7%	1.7%	1.7%	0.4%	1.1%	17,351	0.9%
	Highway (4 lane)	0.9%	0.1%	1.1%	2.1%	3.2%	0.7%	0.8%	21,936	1.2%
	Industrial	1.4%	0.6%	2.1%	4.1%	8.0%	1.3%	1.4%	44,955	2.4%
	Open Space (non-park)	1.9%	1.4%	0.6%	2.3%	2.0%	0.8%	1.4%	28,022	1.5%
	Miscellaneous	2.1%	0.7%	2.0%	4.1%	6.1%	0.9%	1.7%	44,668	2.3%
	Parks (local)	0.8%	0.1%	1.1%	1.9%	2.0%	0.3%	0.6%	18,989	1.0%
	Parks (regional)	8.2%	2.4%	2.1%	5.0%	5.7%	4.6%	3.0%	81,256	4.3%
	Residential (multi-family)	0.9%	0.2%	1.2%	3.8%	6.2%	0.2%	0.6%	30,905	1.6%
	Residential(single-family)	17.5%	6.0%	12.0%	27.5%	35.7%	7.5%	14.4%	311,399	16.4%
	Water	5.4%	5.8%	3.4%	8.9%	9.4%	4.6%	9.4%	123,060	6.5%
Total Number of Acres		285,136	240,372	374,818	387,834	108,654	235,320	270,203	1,902,337	100.0%
2000	Agricultural Preserve	0.6%	22.3%	14.4%	2.8%	0.0%	2.7%	3.1%	134,728	7.1%
	Agricultural/Undeveloped	55.7%	58.1%	55.5%	29.7%	11.5%	73.9%	58.3%	965,643	50.8%
	Cemetery	0.0%	0.0%	0.1%	0.3%	0.5%	0.0%	0.0%	2,188	0.1%
	Commercial	1.2%	0.3%	1.4%	3.3%	4.9%	0.7%	1.1%	32,264	1.7%
	Golf	0.8%	0.6%	0.9%	1.8%	1.7%	0.5%	1.2%	20,251	1.1%
	Highway (4 lane)	1.0%	0.3%	1.2%	2.4%	3.4%	0.8%	0.9%	25,456	1.3%
	Industrial	1.7%	0.7%	2.6%	4.3%	8.0%	1.5%	1.6%	49,418	2.6%
	Open Space (non-park)	1.8%	1.2%	0.8%	2.6%	3.8%	0.7%	2.2%	33,062	1.7%
	Miscellaneous	2.1%	0.6%	1.6%	3.8%	5.7%	0.7%	1.4%	39,884	2.1%
	Parks (local)	0.8%	0.2%	1.2%	2.1%	2.6%	0.3%	0.8%	21,196	1.1%
	Parks (regional)	8.5%	2.7%	2.7%	5.1%	5.7%	4.9%	3.1%	86,710	4.6%
	Residential (multi-family)	1.0%	0.3%	1.5%	4.0%	6.5%	0.4%	0.9%	34,895	1.8%
	Residential(single-family)	19.3%	7.0%	12.8%	28.8%	36.5%	8.4%	15.9%	333,618	17.5%
	Water	5.4%	5.8%	3.4%	8.9%	9.4%	4.6%	9.4%	123,060	6.5%
Total Number of Acres		285,140	240,378	374,821	387,843	108,657	235,323	270,209	1,902,372	100.0%
2005	Agricultural Preserve	0.6%	34.5%	15.2%	2.8%	0.0%	2.9%	3.2%	168,068	8.8%
	Agricultural/Undeveloped	51.6%	44.0%	52.0%	27.2%	9.4%	69.8%	55.0%	876,565	46.1%
	Cemetery	0.0%	0.0%	0.1%	0.3%	0.5%	0.0%	0.0%	2,205	0.1%
	Commercial	1.5%	0.4%	1.5%	3.4%	5.6%	0.9%	1.2%	35,781	1.9%
	Golf	1.0%	0.7%	0.9%	1.8%	1.8%	0.6%	1.3%	21,832	1.1%
	Highway (4 lane)	1.2%	0.3%	1.4%	2.5%	3.8%	0.9%	1.1%	28,346	1.5%
	Industrial	2.0%	0.7%	3.0%	4.4%	7.4%	1.7%	1.7%	51,993	2.7%
	Open Space (non-park)	3.0%	1.6%	1.3%	3.4%	4.5%	1.2%	2.7%	45,131	2.4%
	Miscellaneous	1.6%	0.7%	1.7%	4.1%	5.9%	0.7%	1.5%	40,682	2.1%
	Parks (local)	0.9%	0.2%	1.2%	2.2%	2.6%	0.4%	0.8%	21,680	1.1%
	Parks (regional)	8.8%	2.9%	2.8%	5.2%	5.7%	5.2%	3.5%	90,651	4.8%
	Residential (multi-family)	1.3%	0.5%	2.0%	4.5%	6.6%	0.6%	1.2%	41,522	2.2%
	Residential(single-family)	21.1%	7.7%	13.6%	29.4%	36.8%	10.4%	17.4%	354,984	18.7%
	Water	5.4%	5.8%	3.4%	8.9%	9.4%	4.6%	9.4%	123,060	6.5%
Total Number of Acres		285,163	240,387	374,838	387,877	108,668	235,334	270,234	1,902,500	100.0%

<sup>a</sup> Slight differences in acreage are due to aggregation rounding and differences in spatial coverage of GIS shapefiles

**Table 3.3: Spatial Divisions for Census Levels**

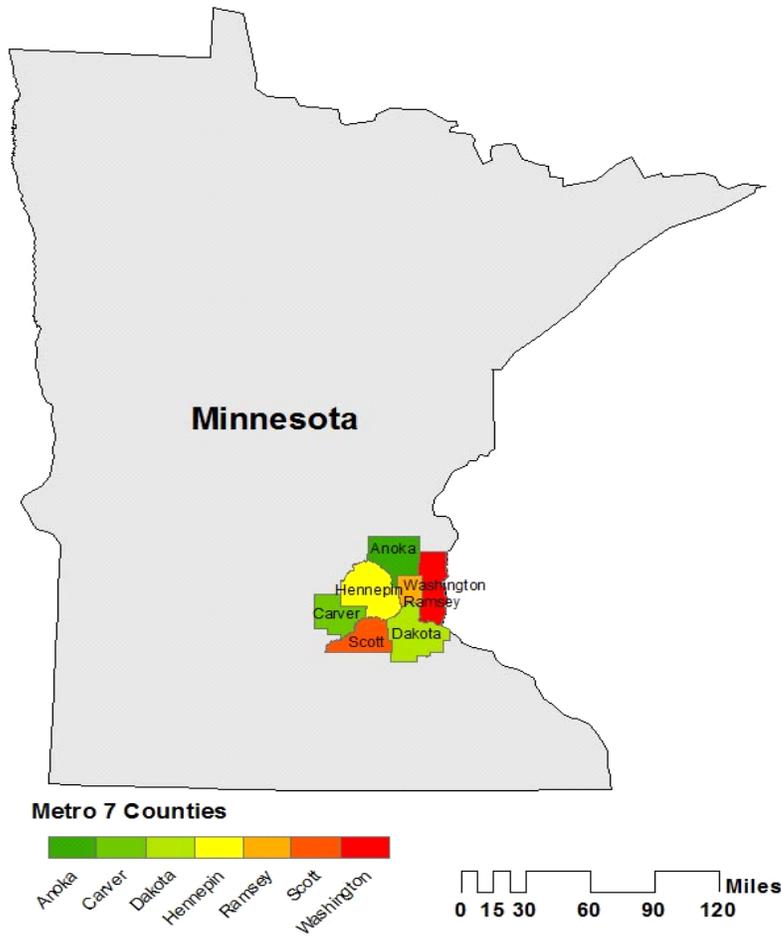
Year	Division	County							All
		Anoka	Carver	Dakota	Hennepin	Ramsey	Scott	Washington	
1990	Block	3,970	1,537	3,805	15,916	7,278	1,540	2,654	36,700
	Block Group	172	35	177	1,058	482	49	141	2,114
	Tract	60	12	55	298	137	15	35	612
2000	Block	4,883	1,785	5,120	17,301	7,443	2,117	3,389	42,038
	Block Group	230	36	194	995	401	54	117	2,027
	Tract	80	18	87	300	136	21	50	692

**Table 3.4: Household Summary Statistics**

Characteristic	Statistic			
	Mean	Std	Min	Max
# of Children	0.83	0.41	0.00	5.00
# of Working Adults	1.78	0.33	0.00	5.90
# of Retirees	0.22	0.20	0.00	3.00
Household Size	2.83	0.54	1.00	7.33

**Table 3.5: Builder Summary Statistics**

Time Period	Number of Bldrs	Houses Built			
		Mean	Std	Min	Max
1990 - 1996	19,202	3.81	29.39	1	2031
1997 - 1999	4,280	8.50	56.84	1	1644
2000 - 2004	6,746	7.53	49.77	1	1744
2005 - 2006	349	13.98	116.72	1	2153



**Figure 3.1: Twin Cities Metropolitan Area**

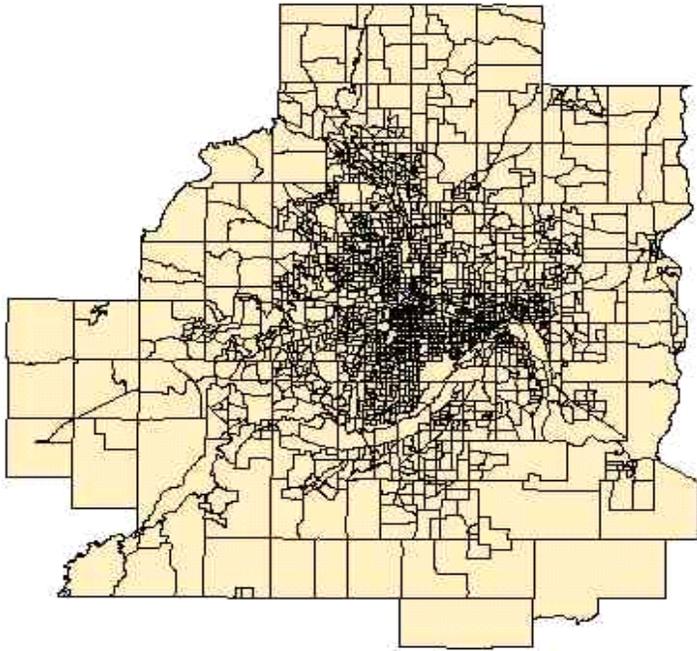


Figure 3.2: Census 2000 Block Groups

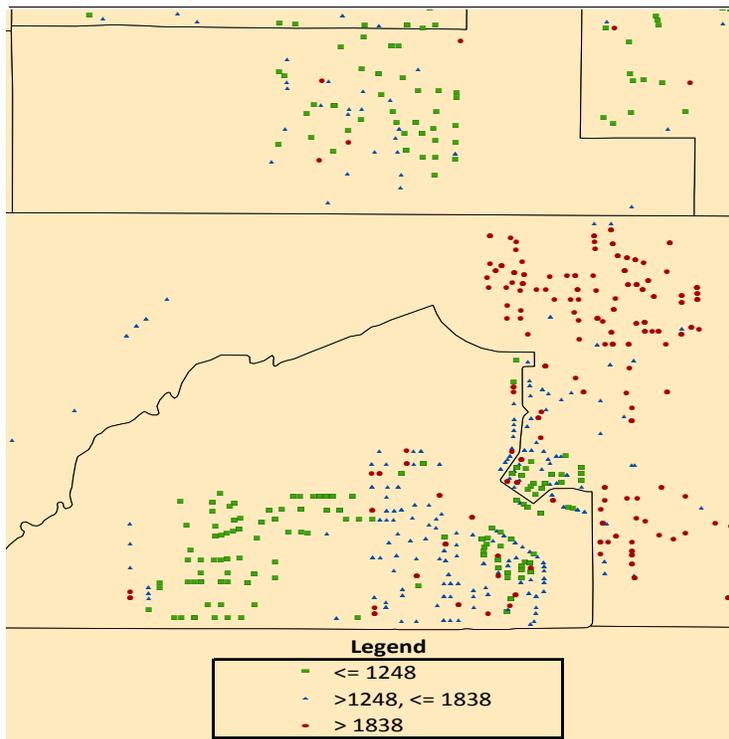
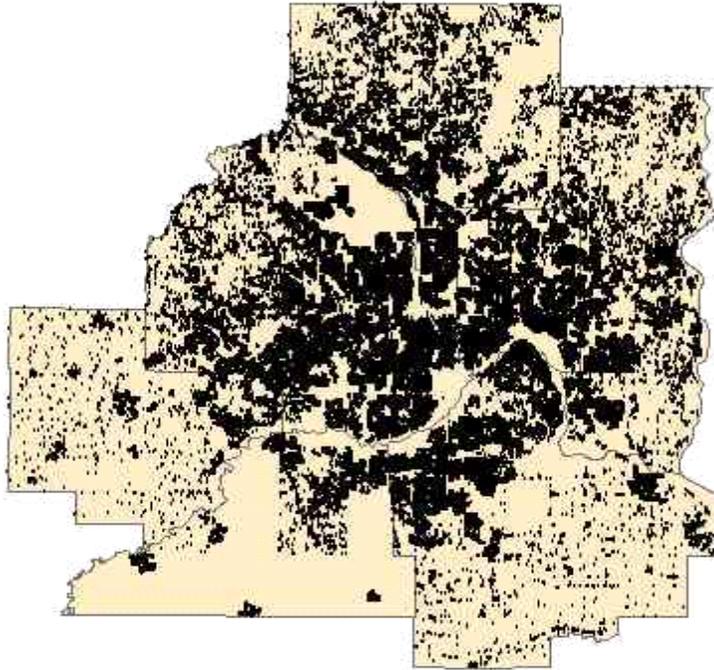
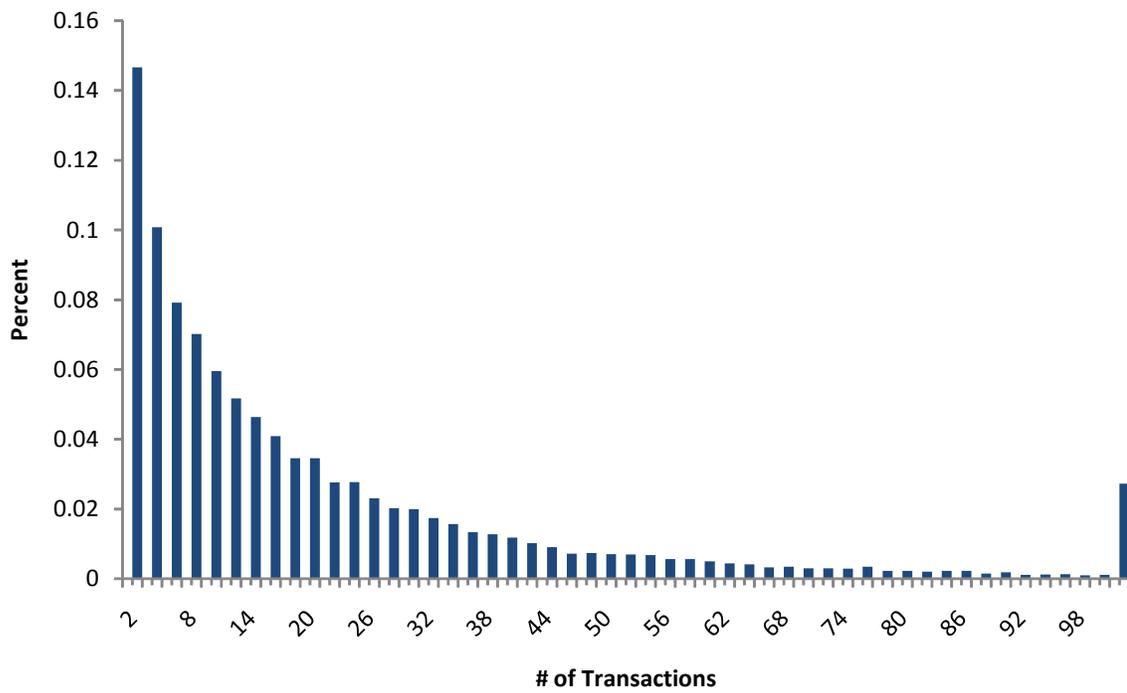


Figure 3.3: Division of Transactions by Square Feet



**Figure 3.4: Single Family Detached Transactions**



**Figure 3.5: Number of Demand Transactions Comprising Each Housing Type (N=20,444)**

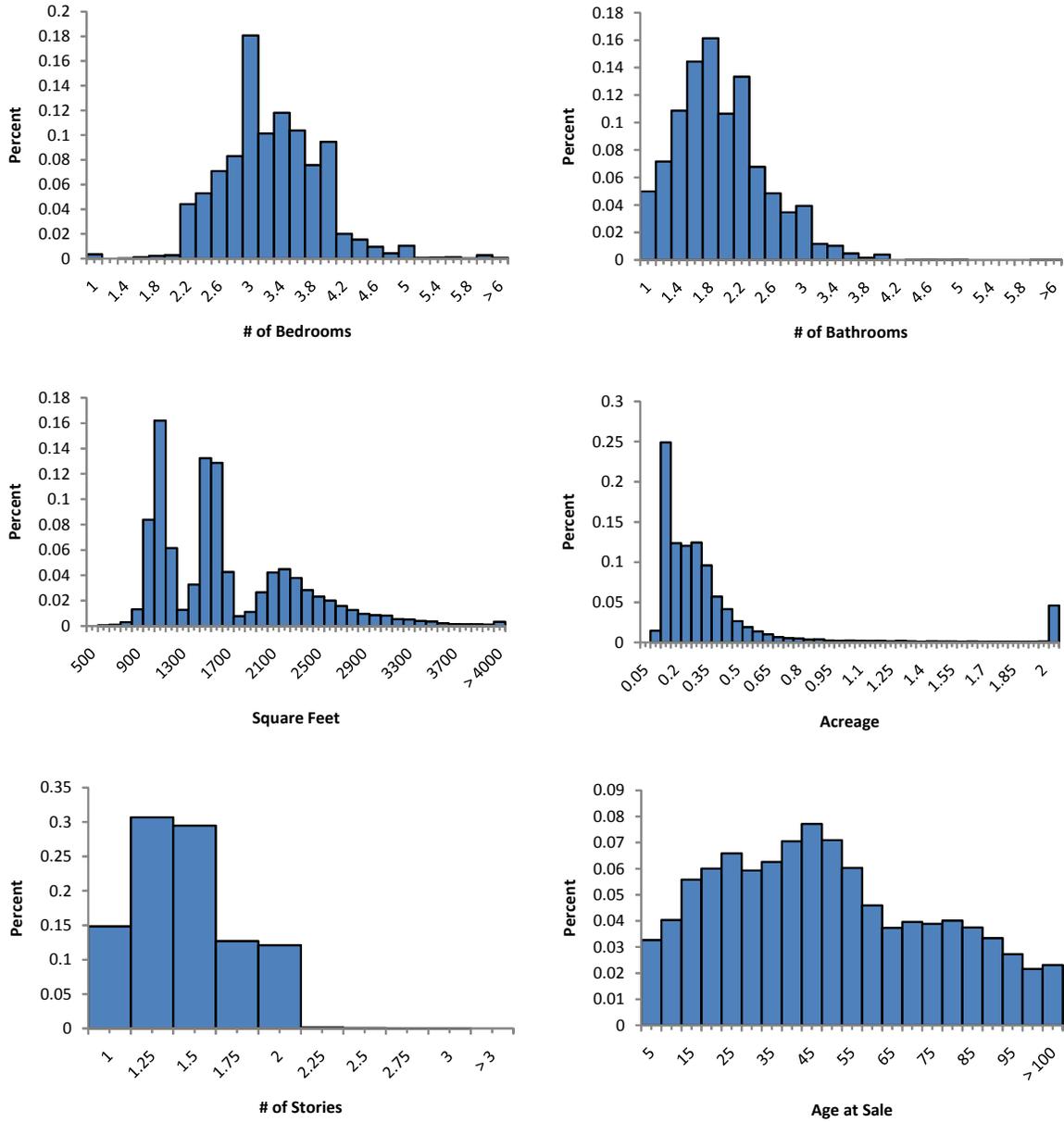
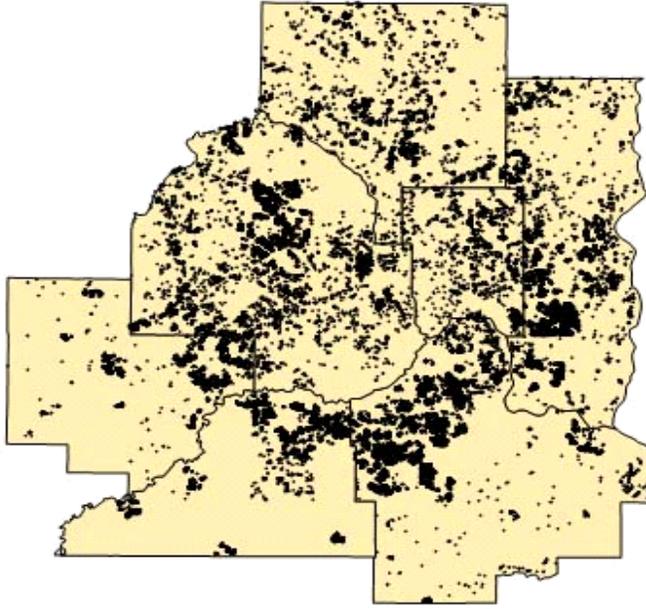
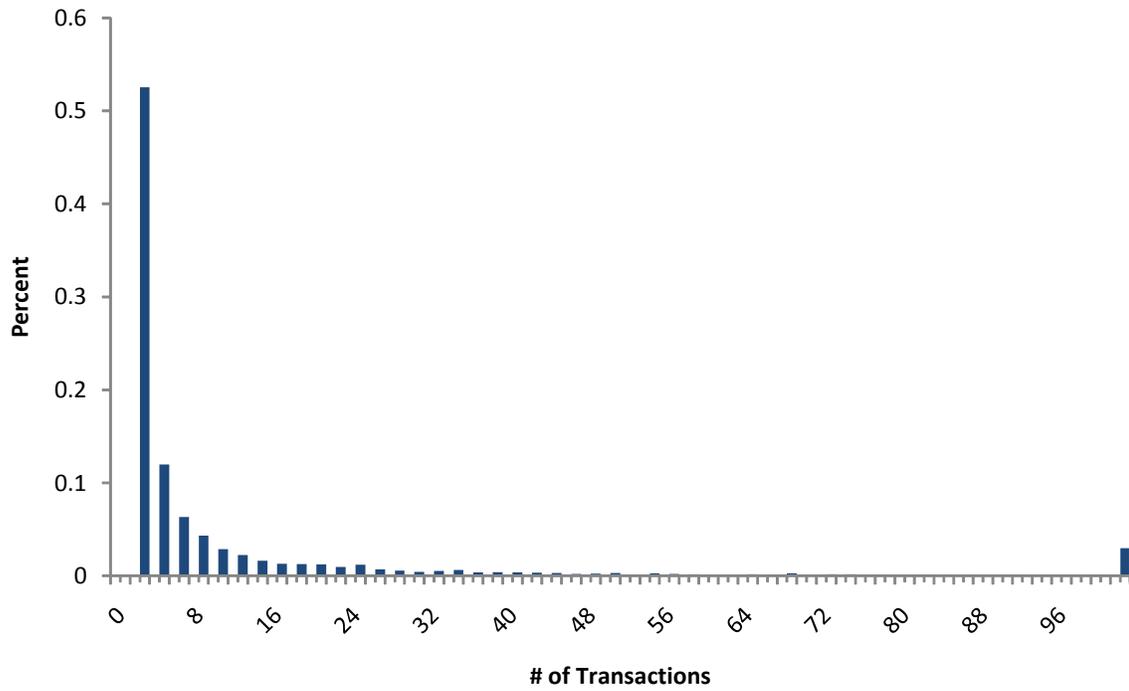
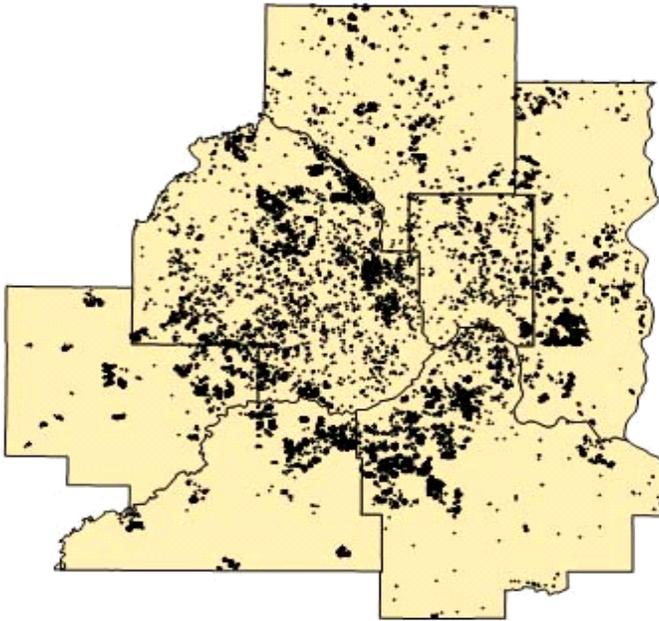


Figure 3.6: Histograms of Choice Set Housing Characteristics (N=20,444)

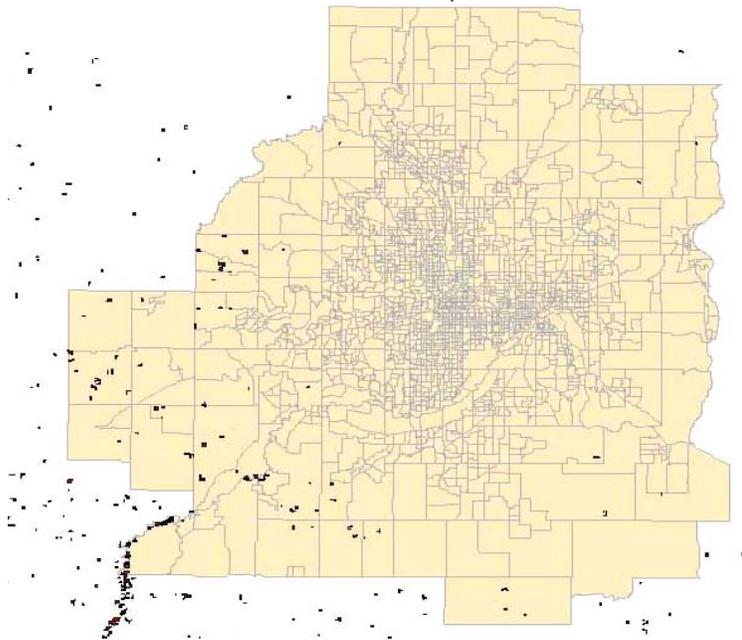


**Figure 3.7: Housing Supply Transactions**

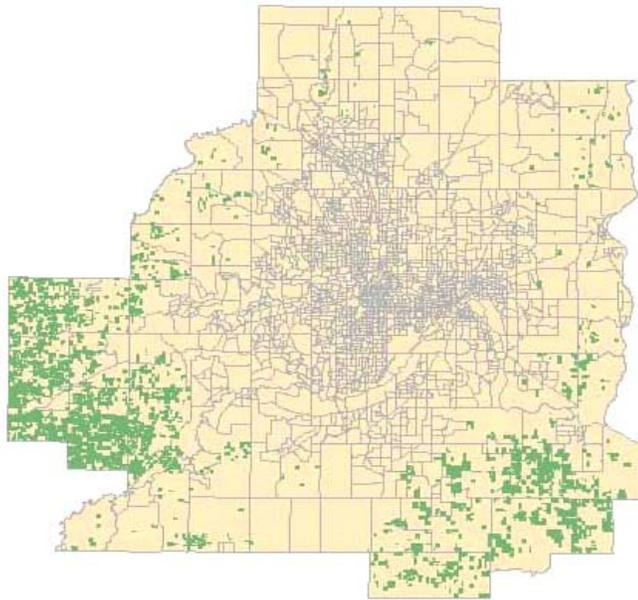




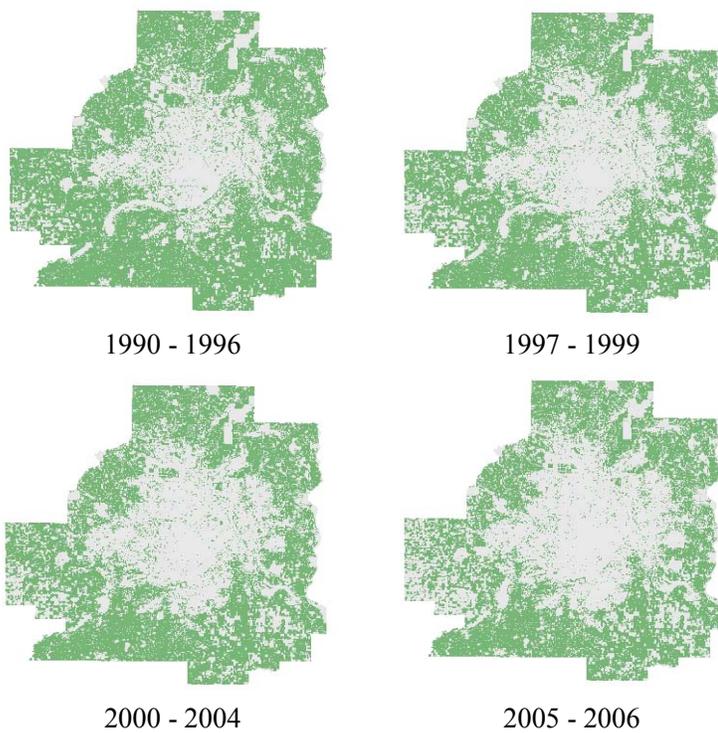
**Figure 3.9: Land Transactions**



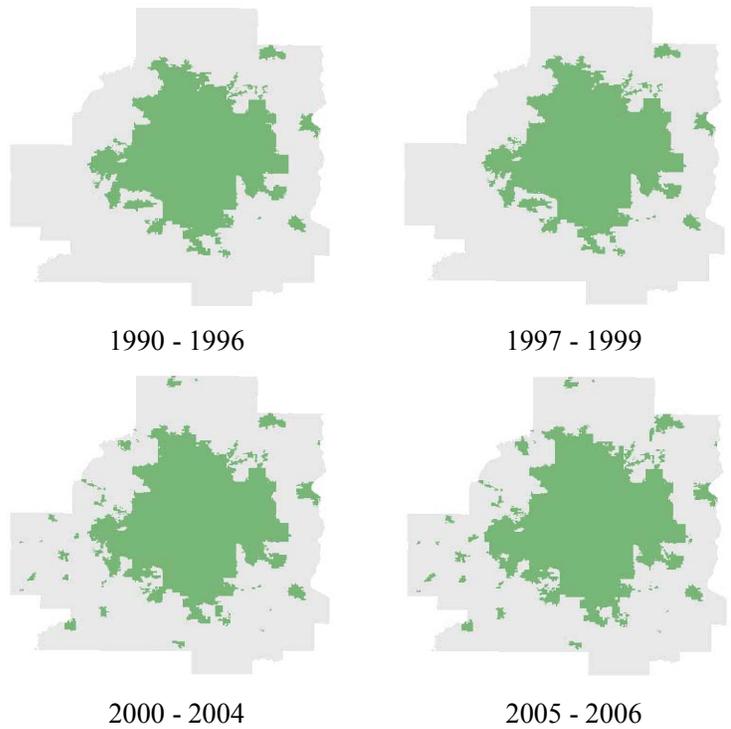
**Figure 3.10: Reinvest in Minnesota Conservation Easements**



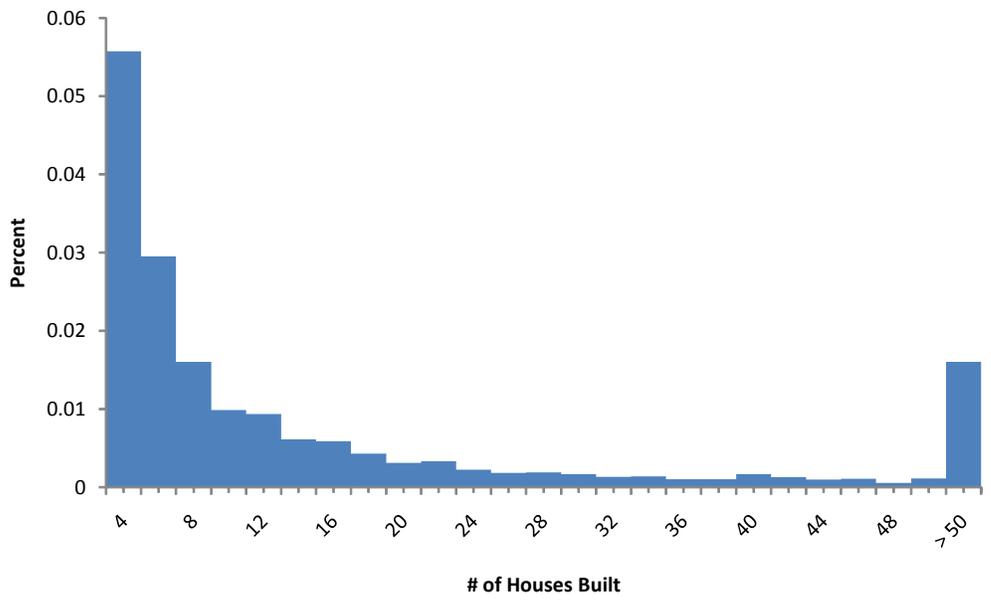
**Figure 3.11: Metropolitan Agricultural Preserves**



**Figure 3.12: Developable Land by Time Period**



**Figure 3.13: Metropolitan Services Areas by Time Period**



<sup>a</sup>Approximately 82% of builders are observed building less than 4 houses and are omitted.

**Figure 3.14: Number of Houses Built by Builder<sup>a</sup>**

# Chapter 4

## Estimation Strategy and Preliminaries

### 4.1 Introduction

The purpose of this chapter is to outline the theoretical model of agents choosing over distinct housing types and present the empirical model used to estimate agents' behavior.

Both the demand and supply side of the housing market are estimated using the methodology presented in this chapter and are ultimately combined for policy analysis. As discussed in the literature review, the idea of agents choosing over a finite set of alternatives is not new, nor is the idea of using revealed preference information about agents' choices to determine their underlying preferences for attributes. What is novel about the recently developed sorting approach presented in this chapter is the ability to use the sorting process itself to solve for a new equilibrium in response to changes in attributes entering the utility function of agents. This innovation unlocks the potential to examine large changes in local public goods as opposed to marginal changes examined in much of the existing literature.

In addition to allowing greater flexibility in the type of counterfactuals examined, the sorting framework is designed to capture locational aspects of the alternatives over which agents make choices. Attention to the locational aspects of the choices made by agents allows the sorting model to capture changes in spatially varying attributes that are difficult to capture using other methodologies. In addition, the horizontal sorting model uses information about the agents themselves to recover preferences that vary by the observable heterogeneity of the agents. When analyzing housing decisions made by agents, it is very likely that different agents have different preferences over local public goods and the information obtained by exploiting the observable heterogeneity among agents results in a rich specification that gives insights into the micro-level decisions that comprise the larger housing market.

While both the demand and supply sides of the housing market are modeled using the same horizontal sorting framework, the actual variables entering the agents utility differ and therefore, the models differ. In addition, each side of the market presents unique empirical challenges that are explored in greater detail in their respective chapters to come. The remainder of this chapter describes the utility maximizing agent choosing from a set of alternatives followed by a discussion of how an agent's behavior is transformed into an empirical sorting model. At the end of this chapter, some estimation preliminaries are presented as these results are used in the empirical sorting models presented in subsequent chapters.

## 4.2 Theoretical Model of Housing Choice by Agents

The foundation of the horizontal sorting model is the familiar random utility framework where each agent chooses a particular housing type,  $h$ , in order to maximize her utility. It is typical in the sorting literature to define housing types based on their location in space where space is thought of as a neighborhood. The size of those neighborhoods can range from literal neighborhoods to metropolitan areas or countries depending on the application.

Because the sorting model exploits differences across space, the decision of how large an area constitutes a neighborhood determines, in large part, the type of data required and the underlying economic questions that are answerable. In the current work, the definition of a housing type is partitioned not only by spatial location, but also by the components of house size and transaction time. Taken together, a unique housing type is a combination of all three components of location, size, and time which are denoted by  $j$ ,  $k$ , and  $t$ , respectively.

With this definition in place, an agent,  $a$ , receives utility from choosing a particular housing type,  $h$ , defined by the following utility function

$$U_{jkt}^a = U_h^a = V(X_h, A^a, P_h, \omega_h) + \epsilon_h^a, \quad (4.1)$$

where indirect utility,  $U_h^a$ , consists of attributes common to a housing type,  $X_h$ , observable characteristics specific to an agent,  $A^a$ , prices,  $P_h$ , an unobserved measure of attributes associated with a housing type,  $\omega_h$ , and an idiosyncratic error component,  $\epsilon_h^a$ , that varies across agents and housing types. The housing type variables,  $X_h$ , do not vary at an individual agent level, but rather vary over any number of dimensions of a housing type. For example,

some variables, such as lot slope, may be constant across time and house size in which case they only vary by location,  $j$ , and report the average slope for the spatial unit.

What differentiates the model in equation (4.1) from a traditional discrete choice model is the inclusion of alternative specific constants,  $\omega_h$ . These fixed effects capture any attributes of a particular housing type that are observable to agents but are unobservable to the econometrician. In a housing application, unobservable attributes could range from school quality to the presence of an old gas station that looks unattractive.

A second component of the utility function that plays a critical role in the horizontal sorting model is the inclusion of characteristics specific to particular agents. These agent specific characteristics are interacted with alternative varying characteristics to define preferences varying across agents. It is these interactions, and the subsequent estimation of heterogeneous preferences, that give rise to the horizontal nature of the sorting model.

The final key component of the utility specification in equation (4.1) is the additive error term,  $\epsilon_h^a$ . The linear in parameters utility function with an additive error term resembles a traditional discrete choice model and, depending on the specification of the error term, results in a familiar, well understood estimation framework. The most common discrete choice model used in the horizontal sorting literature is a multinomial logit model obtained by specifying a type I extreme value distribution for the error component of utility. The logit specification is employed in the current research.

To make explicit the heterogeneous preferences defined by the utility specification in equation (4.1), utility is rewritten as

$$U_h^a = \alpha_X^a X_h + \alpha_P P_h + \omega_h + \epsilon_h^a, \quad (4.2)$$

where  $\alpha_X^a$  is defined as a composite of a mean parameter common across agents and an agent specific component that varies across agents such that

$$\alpha_X^a = \alpha_{0X} + \sum_{q=1}^{\lambda} \alpha_{qX} A_q^a, \quad (4.3)$$

where  $q$  indexes the agent specific characteristics of which there are  $\lambda$ . In this notation,  $\alpha_{0X}$  is the mean parameter and  $\alpha_{qX}$  are parameters that vary across the observed heterogeneity in agents. Using this parameter specification, it is possible to introduce a flexible set of heterogeneous preferences by specifying interactions between characteristics common to a housing type with agent characteristics that vary across each decision-making agent.

Given the specification of utility in equation (4.2), an agent is assumed to choose a particular housing type if the utility from choosing that housing type is at least as large as the utility from choosing any other housing type. More formally, an agent will choose housing type  $h^*$  if

$$U_{h^*}^a \geq U_h^a, \quad \forall h^* \neq h. \quad (4.4)$$

To determine the market equilibrium quantities for a particular housing type, note that each agent's choice of housing type depends on the characteristics of other housing types not chosen by the agent but available in her choice set. The dependence on non-chosen alternatives allows the probability of an agent choosing a particular housing type to be expressed as

$$\Pr_h^a = f(A^a, \mathbf{A}, \mathbf{X}, \mathbf{P}, \boldsymbol{\omega}), \quad (4.5)$$

where  $A^a$  is a particular agent's characteristic,  $\mathbf{A}$  is the full set of all agents characteristics,  $\mathbf{X}$  are alternative attributes,  $\mathbf{P}$  is price, and  $\boldsymbol{\omega}$  is a set of alternative specific constants. Using these probabilities, the equilibrium amount of housing type,  $h$ , is obtained by aggregating all agents' probabilities for selecting that particular housing type to obtain

$$Amount_h = \sum_a Pr_h^a, \quad (4.6)$$

where  $Amount_h$  refers to either demand or supply of new housing for a particular housing type,  $h$ . To clear the housing market, the total demand for each housing type must equal the total supply of each housing type where supply is a composite of the exogenous amount of existing supply and the estimated amount of new supply. This market clearing condition is expressed by

$$Demand_h = Supply_h^{exog} + Supply_h, \quad \forall h, \quad (4.7)$$

where  $Supply_h^{exog}$  is the pre-existing exogenously determined supply of housing type  $h$  and  $Supply_h$  is the amount of new housing type  $h$  predicted by the sorting model for builders' decisions to supply new housing. In this sense, market clearing implies that the aggregate choice probabilities for every housing type equal the empirical share of each housing type observed in the data.

### 4.3 Empirical Sorting Model

It is not possible to jointly identify all the components of indirect utility specified by equation (4.2) in a single estimation step due to the inability to jointly identify both the vector of unobserved components of utility along with the set of mean preference parameters. As a

result, estimation proceeds in two steps following Berry (1994). To see the mechanics of this two-stage procedure, the indirect utility defined by equation (4.2) is rewritten as

$$V_h^a = \Theta_h + \Gamma_h^a + \epsilon_h^a, \quad (4.8)$$

where  $\Theta_h$  defines variables common to all agents and  $\Gamma_h^i$  defines variables unique to agents created by interacting agent specific characteristics,  $A$ , with housing type specific characteristics,  $X$ . Explicitly,  $\Theta_h$  is defined as

$$\Theta_h = \alpha_{0X}X_h + \alpha_{0P}P_h + \varphi_h, \quad (4.9)$$

and  $\Gamma_h^a$  is defined as

$$\Gamma_h^a = \left( \sum_{q=1}^{\lambda} \alpha_{qX} A_q^a \right) X_h. \quad (4.10)$$

By dividing utility in this way, each component is recoverable in a single estimation step.

The first stage of estimation recovers all of the heterogeneous parameters,  $\alpha_{qX}$ , along with the mean indirect utility,  $\Theta_h$ , using maximum likelihood techniques. The second stage decomposes the mean indirect utility,  $\Theta_h$ , into observable components,  $\alpha_{0X}$ , and unobservable components,  $\omega_h$ , according to equation (4.9) .

### 4.3.1 First stage estimation

To estimate the first stage,  $\epsilon_h^a$  is assumed to take on an i.i.d. extreme value distribution resulting in the traditional multinomial logit probability of an agent choosing a particular housing type,  $h$ , given by

$$Pr_h^a = \frac{e^{\Theta_h + \Gamma_h^a}}{\sum_j e^{\Theta_j + \Gamma_j^a}}. \quad (4.11)$$

The log-likelihood function is defined as

$$\ell\ell = \sum_a \sum_h Y_h^a \ln (Pr_h^a), \quad (4.12)$$

where  $Y_h^a$  equals 1 if an agent chooses housing type  $h$  and equals zero otherwise. Behavior consistent with this stage of estimation is characterized by a Nash sorting equilibrium where every agent's housing decision is optimal given the housing decisions of all other agents and the set of observed market clearing prices.

First stage estimation is relatively straightforward with the exception of two specific issues. The first is related to the large number of agents and the large number of housing types over which optimization is required. Given current computing power, it is extremely difficult to estimate the entire model at once due to physical memory constraints. Fortunately, because the logit specification maintains the independence of irrelevant alternatives (IIA) assumption, it is possible to follow McFadden (1978) and estimate the model using a randomly selected subset of non-chosen alternatives for each individual along with their chosen alternative. This estimation strategy yields unbiased estimates, yet sacrifices some efficiency of first stage estimates.

The second issue concerns how to obtain parameter estimates for literally thousands of different mean taste parameters, of which there is one for each housing type. As with the previous issue, the specification of the logit model can once again be exploited to solve this problem. Because of the structure of the multinomial logit model, it is possible to estimate

all of the mean parameters,  $\Theta_h$ , through a contraction mapping routine introduced by Berry (1994). To see the rational for the use of a contraction mapping in this application, it is useful to examine the first order conditions of the log-likelihood function with respect to each  $\Theta_h$ :

$$\frac{\partial \ell \ell}{\partial \Theta_h} = \sum_{a \in h} \frac{\partial \ln (Pr_h^a)}{\partial \Theta_h} + \sum_{a \notin h} \frac{\partial \ln (Pr_h^a)}{\partial \Theta_h} = W_h - \sum_a Pr_h^a = 0, \quad (4.13)$$

where  $W_h$  is the observed share of agents choosing housing type  $h$ . These first order conditions imply that in order to maximize the log-likelihood, the observed share of agents choosing housing type  $h$  must perfectly match the sum of the agent probabilities for choosing the same housing type. Using this result, it is possible to specify a contraction mapping whereby the mean parameter,  $\Theta_h$ , can be recovered indirectly within the maximum likelihood routine using the following contraction mapping

$$\Theta_h^{c+1} = \Theta_h^c - \ln \left( \sum_a \frac{\widehat{Pr}_h^a}{W_h} \right), \quad (4.14)$$

where  $c$  indexes the iteration of the contraction mapping. Using the contraction mapping in equation (4.14), it is possible to quickly solve for a new vector of mean indirect utilities during every iteration of the maximum likelihood optimization process. This process provides consistent, maximum likelihood, estimates of the mean indirect utility parameters without the need to estimate them numerically within the maximum likelihood routine.

### 4.3.2 Second stage estimation

The second stage of estimation decomposes the mean indirect utilities recovered in the first stage into observable and unobservable components as shown in equation (4.9) . Were there no endogeneity concerns, the second stage could be estimated by ordinary least squares returning the mean taste parameters for observable attributes. However, the sorting of agents into housing types is undoubtedly related to factors which are unobserved to the econometrician. These unobservable factors are likely to be correlated with observed prices also included in the second stage. This correlation occurs because better locations, often reflecting factors unobservable to the econometrician, are more likely to command a higher price and thus correlation between unobservable factors and prices arises. Because of the endogeneity problem created by the correlation between price and the unobservable components of a particular housing type, it is necessary to find a suitable instrument for housing prices and estimate the second stage via instrumental variables techniques.

Bayer et al. (2005) recognized that because observed prices arise as a result of market equilibrium, it is possible to create an instrument for house prices using the intuition that neighborhoods in distant areas of the housing market are likely to influence demand for local housing, but are almost certainly uncorrelated with local unobservable characteristics of a particular housing type. By accounting for all observable characteristics contributing to housing prices within a suitable range of a particular neighborhood, it is possible to then use the exogenous attributes of distant neighborhoods as instruments for prices. This instrumentation strategy works as a result of the implicit price equilibrium that exists within a particular housing market. In other words, it is reasonable to assume that prices in distant

neighborhoods influence prices in local neighborhoods but the characteristics of those distant neighborhoods are unlikely to be correlated with local unobservable components of utility.

The first step in generating instruments for price is to rearrange equation (4.9) as

$$\Theta_h - \alpha_{0P}P_h = \alpha_{0X}X_h + \varphi_h, \quad (4.15)$$

where the price term has been moved to the left hand side. The next step involves guessing a plausible value for the price coefficient,  $\alpha_{0P}$ , and including additional variables to account for neighborhoods that border the current neighborhood on the right hand side. These additional right hand side variables are constructed from neighborhood characteristics,  $N_{jt}$ , as well as neighborhood social demographics,  $Z_{jt}$ , for all census block group centroids within 1,2,3,4, and 5 mile rings from the current census block group centroid. With these variables in place, equation (4.15) is estimated via OLS. By setting the OLS residual,  $\varphi_h$ , to zero and calculating a new set of prices to equate the left hand side and right hand side of equation (4.15), an instrument for prices is obtained as follows

$$P_h^{iv} = \frac{\Theta_h - (\widehat{\alpha}_{0X}X_h + \widehat{\alpha}_{0N}N_h + \widehat{\alpha}_{0Z}Z_h)}{\alpha_{0P}}. \quad (4.16)$$

The power of this instrument is derived from exogenously varying features of the housing market outside the five mile rings included in the OLS regression.

A potential concern raised by this approach is that the instrument for price is dependent on an initial conjecture for the price coefficient. To eliminate this dependence, after determining the initial price instrument and running IV, the estimated price coefficient is obtained and the entire process of determining the price instrument is re-run using the new price coefficient as the initial guess. By repeating this process several times, the price

coefficient eventually stabilizes and the initial dependence on the conjecture for the price coefficient is removed. In practice, after approximately five iterations the estimated price coefficient does not change.

### 4.3.3 Asymptotic properties of the estimator

The use of a two-stage estimator where the second stage decomposes an estimated vector of alternative specific constants into mean components of utility raises the question of the asymptotic properties of this approach. Asymptotic concerns arise because the true vector of alternative specific constants is not known, but rather only an estimate of those alternative specific constants is known. Fortunately, the asymptotic properties of such an estimator have been developed by Berry, et al. (2004) and are summarized here.

Were the true vector of alternative specific constants known, consistency of second stage estimates and asymptotic normality are achieved as long as the number of housing types approaches infinity. The reliance on estimated alternative specific constants creates additional requirements in order to achieve consistency and asymptotic normality involving the relationship between the number of agents and number of housing types. These additional asymptotic requirements ensure that the estimation uncertainty from the first stage of estimation has no effect on the estimation error from second stage estimation. In other words, the only effective error in second stage estimation is  $\varphi_h$ .

To ensure consistency, the number of agents must grow large relative to the number of housing types. This is achieved as long as  $\frac{h \ln h}{n} \rightarrow 0$ , where  $h$  is the number of housing types and  $n$  is the number of agents, or choice occasions. By defining distinct housing types

over which agents choose, it seems reasonable that the condition for consistency of second stage estimates is achieved. The requirement for asymptotic normality of second stage estimates is that the ratio  $\frac{h^2}{n}$  is bounded. As with the requirement for consistency, it seems plausible that this condition is met given the construction of a finite set of housing types and the relatively large sample size of agents.

#### **4.4 Price Index Estimation**

Dividing housing transactions into specific housing types as opposed to treating each transaction as a single housing type necessitates the creation of price indexes to convert multiple housing sales with different prices and characteristics into a single housing price. The approach used in this section for both house sales and land sales accounts for differences in characteristics across the particular transactions to be combined. Because of this, the resulting set of price indexes represents the price for a unit of housing services and thus is comparable across space and time regardless of changes in the housing stock.

In addition to controlling for changes in characteristics across space, year dummy variables are also included which soak up any generic price dynamics over time. The general price changes captured by these variables include inflation but would also include supply and demand changes that alter the general equilibrium price level in the metropolitan area. The final set of price indexes estimated in this section replace the price variable,  $P$ , used in previous sections of this chapter. The remainder of this section describes the price index

creation process for housing prices as well as land prices and the differences between the approaches used to estimate each price index.

#### 4.4.1 Housing price indexes

The price of a unit of housing services for a particular housing type  $h$  reflects characteristics based on the spatial location, the size, and the time period in which the housing transaction occurred, as well as any omitted amenities unique to a particular housing type. The approach described here is similar to that employed by Bayer et al. (2006) but is extended to include a size component of housing in addition to location and time.<sup>10</sup> To capture the value of a particular housing type, the following hedonic regression is estimated

$$P_{jkt}^i = \rho_{jkt} e^{S_{kt}^i \phi_t + Y_{\tau}^i \theta_{\tau} + \epsilon_{jkt}^i}, \quad (4.17)$$

where  $P_{jkt}^i$  is an observed sale price<sup>11</sup> in the time period  $t$ <sup>12</sup>,  $S_{kt}^i$  includes observable characteristics unique to an individual house, and  $\rho_{jkt}$  is a constant term that captures components of price that are excluded from  $S_{kt}^i$ . The inclusion of  $Y_{\tau}^i \theta_{\tau}$  adds a dummy variable for the year a house sold,  $\tau$ , within a time period,  $t$ , where the first year in each time period is omitted. Variables included in the vector of housing characteristics include acreage, square footage, number of bedrooms, number of bathrooms, stories, and any other observable housing features as well as information on views such as view of a lake or view of a park that vary across households within a housing type. The set of year dummy

---

<sup>10</sup> The approach used in this paper controls for changes in the stock of housing across space, time, and size. An alternative approach by Chay and Greenstone (2005) uses median house price.

<sup>11</sup> All sales prices were converted to January 2006 dollars using seasonally adjusted CPI excluding food and energy.

<sup>12</sup> Recall that there are four time periods spanning the years 1990-1996, 1997-1999, 2000-2004, and 2005-2006.

variables,  $Y_{\tau}^i$ , are included to account for any changes that occur within a time period, and more specifically, to account for appreciation or depreciation of the general price of housing within a time period.

Note in equation (4.17) that the parameters for housing characteristics,  $\phi_t$ , are allowed to vary over time period but not over the size dimension. Allowing these parameters to vary over time periods serves two purposes. First, this allows the hedonic price function to shift over time. Given that the observed transactions span 17 years from 1990 through 2006, it is likely that some changes in the hedonic equilibrium occurred over this period. The second, and more practical, reason is that by allowing the function to change over time, it is possible to estimate price indexes for each time period independently. Because the set of housing transactions is extremely large, the ability to estimate each time period independently greatly aids in the ability to calculate price indexes while allowing interactions across housing sizes within a time period to remain.

To convert equation (4.17) into an estimable form, the natural log of both sides is taken and  $h$  is substituted for the subscript  $jkt$  resulting in the following

$$\ln P_h^i = \ln \rho_h + s_{kt}^i \phi_t + Y_{\tau}^i \theta_{\tau} + \epsilon_h^i . \quad (4.18)$$

This equation defines the price of a unit of housing services of type  $h$  as  $\rho_h$  and the amount of services provided by house  $i$  of size  $k$  at time period  $t$  as  $S_{kt}^i = e^{s_{kt}^i \phi_t + Y_{\tau}^i \theta_{\tau}}$ . The coefficients  $\phi_t$  are obtained by estimating equation (4.18) using Census 2000 block groups as the dimension for  $j$  and a division of house size based on square footage for dimension  $k$ .

The estimation results for the parameters  $\phi_t$  and  $\theta_\tau$  from equation (4.18) are provided in table 4.1. All coefficients for housing characteristics as well as views have the expected signs and appear to have similar magnitudes across time periods. In addition, all of the coefficient estimates except the view of a cemetery are significant at the 5% level. Looking at the coefficient estimates for the year dummy variables, one can see that with the exception of the early 1990s and 2006, they take on positive and significant values. The positive coefficients capture the general rise in property values experienced over much of the 1990s and into the first half of the 2000s.

To provide a sense of the spatial coverage of transactions as well as the general price distribution for the various sizes of houses, table 4.2 provides summary statistics for the price indexes estimated using equation (4.18). From these results, it is clear that despite adjusting all prices by CPI, prices still rise over time due to appreciation between time periods. By omitting the first year's dummy variable for each time period, the estimated prices are effectively the prices in 1990, 1997, 2000, and 2005 for each time period in year 2006 dollars.

When looking at these estimates, it should be noted that a housing transaction is not observed in every block group for every house size in each time period and thus the dimension of the resulting vectors of price indexes is not constant across time periods. Despite the varying number of transactions and years across time periods, the number of spatial and house size divisions is fairly consistent. This is seen by comparing the value for  $N$  from table 4.1 across the various time period and house size divisions.

Figure 4.1 provides a graphical depiction of the estimated price indexes for each time period and house size category. The relative magnitudes of prices indexes across the landscape appear consistent across house sizes within each time period as would be expected if the driving force behind price differences is from the amenities provided by a specific location. Comparing prices across time periods, there is evidence of increased housing prices farther away from the central business district. This increase reflects the expansion of the urban/rural fringe and the creation of additional amenities further from the city center.

#### **4.4.2 Land price indexes**

The price of a unit of land services is calculated using land transactions data separate from the residential transactions data used in estimation of housing price indexes. Because the land transactions data has considerably fewer transactions, the spatial and temporal coverage of this data is not as refined as that used in the creation of housing price indexes. As a result, several assumptions are required to recover land price indexes along with limited interpolation of missing land price indexes.

To recover land price indexes, a similar approach to that employed to recover house price indexes is used that accounts for changes in the types of land across space while calculating the value of a homogeneous unit of land in all census tracts containing land transactions. The use of census tracts as opposed to block groups is required because there is not enough data to estimate a sufficiently large number of land price indexes at the block group level. To account for land use changes across space, the following hedonic is specified

$$P_u^i = \zeta_u e^{N_{jt}\phi + Y_{\tau}^i \theta_{\tau} + \epsilon_u^i}, \quad (4.19)$$

where  $i$  indexes a transaction,  $u$  indexes census tract,  $j$  indexes block group,  $t$  indexes time period, and  $\tau$  indexes year of sale. The parameters  $\phi$  are constant across all time periods and capture the impact of neighborhood land use on the sale price of land. The parameters  $\theta_{\tau}$  are estimates of year dummy variables capturing general changes in land price across time.

To convert equation (4.19) into an estimable form, the natural log of both sides is taken resulting in the following

$$\ln P_u^i = \ln \zeta_u + N_{jt}\phi + Y_{\tau}^i \theta_{\tau} + \epsilon_u^i. \quad (4.20)$$

The value of a unit of land in census tract  $u$  is  $\zeta_u$  and the services provided by a unit of land in census tract  $u$  are  $L_u = e^{N_{jt}\phi}$ . The estimated price of a unit of land services in each census tract is assigned to all block groups within that tract resulting in a set of land prices included in estimation of builder location decision.

Table 4.3 provides estimation results for equation (4.20). Because the set of census tracts covered by the land price transactions is not complete, interpolation is necessary to determine land prices in the missing census tracts. Figure 4.2 highlights the census tracts for which no land price is estimated as a result of missing land transactions data. To interpolate the land prices in the highlighted areas, inverse distance weighting is performed using the spatial analyst toolbox in ArcGIS by setting the number of points to 12 and the smoothing power to 2.

The final set of land price indexes after interpolation is shown in figure 4.3. The interpolation process produced what appear to be realistic estimates of the land price in

missing census tracts. Summary statistics for the final set of land price indexes are shown in table 4.4. Examining these summary statistics along with the graphical depiction of land prices shows considerable land price variation across the metropolitan area. In particular, the highest land prices appear located just outside the metropolitan center as well as in isolated more distant areas corresponding to smaller municipalities located well away from the metropolitan center.

## **4.5 Conclusions**

The current chapter presented the theoretical motivation for the horizontal sorting model and explained the empirical procedure needed to estimate such a model. After presenting the empirical model, some preliminary estimation steps to recover price indexes for housing and land were presented whose results are used in the estimation of sorting models for demand and supply discussed in subsequent chapters.

By presenting a generic sorting model of agents choosing over discrete housing alternatives, the similarities between the forthcoming demand and supply empirical models can be seen. Both models consist of agents choosing from a finite set of housing types in order to maximize utility. The horizontal component of preferences is obtained through interactions between agent specific characteristics and characteristics of the housing types over which agents choose. In addition to these interactions, fixed effects are also estimated for each housing type capturing not only the mean taste parameters, but also components of utility unobservable to the econometrician but observable to the decision making agent.

The inclusion of fixed effects for each alternative necessitates estimation in two stages in order to obtain identification. To facilitate estimation, several empirical techniques are employed both to speed up estimation and to reduce the number of parameters estimated using gradient based methods within maximum likelihood. In addition, the rationale for instrumenting for endogenous prices in the second stage of estimation was presented showing how to use the spatial structure of the urban landscape to form an appropriate instrument.

The following two chapters present demand and supply estimation, respectively, using the methodology presented in this chapter. While both the demand and supply sorting models use the same methodology, the variables entering the utility functions of households and builders vary considerably. For demand, households are treated as agents choosing from a set of housing types. The agents in supply estimation are perfectly competitive builders choosing housing types to supply. Builders face the same choice set as households, except some housing types for which no developable land exists are excluded.

**Table 4.1: Price Index Regressions for 2000 Census Block Groups by Time Period**

Variables	Time Period							
	1990-1996		1997-1999		2000-2004		2005-2006	
	Estimate	t-statistic	Estimate	t-statistic	Estimate	t-statistic	Estimate	t-statistic
<b>Housing Characteristics</b>								
acreage	0.023	21.160	0.026	20.550	0.030	35.430	0.043	32.680
bedrooms	0.024	22.190	0.026	19.450	0.022	23.500	0.020	15.950
square feet	0.000	80.510	0.000	68.720	0.000	93.870	0.000	73.110
stories	0.019	7.180	0.018	5.860	0.012	5.630	0.008	2.730
garage	0.097	27.240	0.132	31.500	0.110	42.840	0.064	17.310
bathrooms	0.060	40.350	0.056	30.980	0.056	44.670	0.057	32.730
age	-0.003	-53.620	-0.003	-44.590	-0.003	-61.810	-0.003	-42.870
<b>Views</b>								
park	0.026	5.730	0.027	5.340	0.018	5.150	0.011	2.310
golf	0.083	7.720	0.132	12.570	0.105	15.400	0.085	8.560
cemetery	-0.054	-2.870	-0.036	-1.560	-0.022	-1.460	-0.075	-3.230
open space (noi)	0.025	5.640	0.021	4.710	0.013	4.300	0.008	2.310
water	0.203	44.920	0.225	40.040	0.212	52.110	0.209	34.670
<b>Year Dummy Variables</b>								
Year 1990	n/a	n/a						
Year 1991	-0.033	-10.450						
Year 1992	-0.036	-12.570						
Year 1993	-0.024	-8.350						
Year 1994	0.000	-0.150						
Year 1995	0.014	4.820						
Year 1996	0.050	17.200						
Year 1997			n/a	n/a				
Year 1998			0.052	22.490				
Year 1999			0.149	63.280				
Year 2000					n/a	n/a		
Year 2001					0.117	50.810		
Year 2002					0.198	89.850		
Year 2003					0.273	124.350		
Year 2004					0.345	156.380		
Year 2005							n/a	n/a
Year 2006							-0.006	-3.270
<b>Model Information</b>								
N	160344		89445		155871		59567	
R <sup>2</sup>	0.9994		0.9995		0.9996		0.9998	

**Table 4.2: Summary Statistics for House Price Indexes**

Statistic	Time Period											
	1990 - 1996			1997 - 1999			2000 - 2004			2005 - 2006		
	Size (square feet)											
	Small <sup>a</sup>	Medium <sup>b</sup>	Large <sup>c</sup>	Small	Medium	Large	Small	Medium	Large	Small	Medium	Large
N	1789	1861	1679	1697	1776	1566	1801	1897	1744	1706	1828	1537
Min	\$26,512	\$18,455	\$14,997	\$23,688	\$23,506	\$12,066	\$29,224	\$14,193	\$8,912	\$20,107	\$36,834	\$37,035
Max	\$267,745	\$219,947	\$240,181	\$261,092	\$420,525	\$232,848	\$263,105	\$294,116	\$405,688	\$384,366	\$469,442	\$560,811
Mean	\$79,292	\$83,813	\$86,970	\$81,668	\$87,195	\$92,129	\$110,015	\$117,483	\$123,225	\$163,691	\$174,146	\$185,322
Std	\$21,828	\$23,123	\$29,703	\$22,785	\$26,470	\$32,371	\$23,465	\$30,348	\$41,221	\$32,738	\$46,017	\$63,106

<sup>a</sup>Represents houses <= 1248 square feet.

<sup>b</sup>Represents houses > 1248 and <= 1838 square feet.

<sup>c</sup>Represents houses > 1838 square feet.

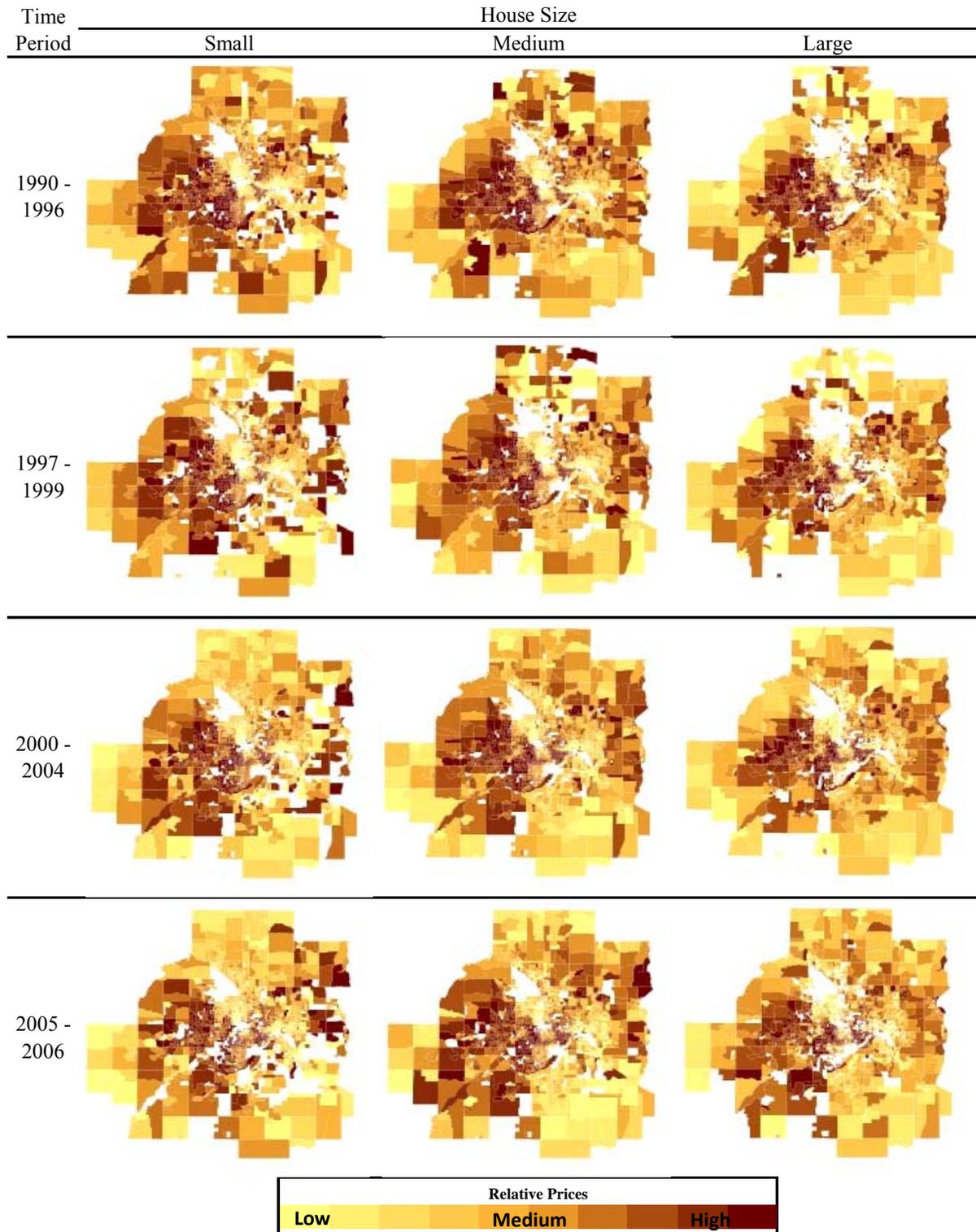
**Table 4.3: Land Price Index Regressions for 2000 Census Tracts**

Variable	Estimate	Std Err	t-statistic
acre	<b>0.1955</b>	0.0028	70.08
acre <sup>2</sup>	<b>-0.0013</b>	0.0000	-44.83
% Agricultural / Undeveloped	<b>-0.0559</b>	0.0238	-2.34
% Agricultural Preserve	<b>-0.3381</b>	0.0858	-3.94
% Cemetary	0.2908	0.2923	0.99
% Regional Park	-0.0718	0.0488	-1.47
% Local Park	<b>-0.2300</b>	0.1150	-2.00
% Open Space - Non Park	-0.0224	0.0874	-0.26
% Golf	<b>0.9209</b>	0.1000	9.21
% Water	<b>0.5534</b>	0.0447	12.39
% Commerical	-0.1736	0.0976	-1.78
% Industrial	<b>0.6060</b>	0.0702	8.63
% Highway	-0.2147	0.1603	-1.34
<b>Year Dummy Variables<sup>a</sup></b>			
Year 1990	<b>-1.0206</b>	0.0542	-18.84
Year 1991	<b>-1.0657</b>	0.0541	-19.69
Year 1992	<b>-1.0822</b>	0.0539	-20.08
Year 1993	<b>-1.0744</b>	0.0539	-19.95
Year 1994	<b>-1.0002</b>	0.0540	-18.51
Year 1995	<b>-0.9928</b>	0.0540	-18.38
Year 1996	<b>-0.9682</b>	0.0539	-17.98
Year 1997	<b>-0.9261</b>	0.0537	-17.23
Year 1998	<b>-0.8894</b>	0.0535	-16.62
Year 1999	<b>-0.8476</b>	0.0535	-15.85
Year 2000	<b>-0.7317</b>	0.0539	-13.56
Year 2001	<b>-0.6015</b>	0.0538	-11.19
Year 2002	<b>-0.5367</b>	0.0539	-9.96
Year 2003	<b>-0.3846</b>	0.0540	-7.13
Year 2004	<b>-0.2344</b>	0.0543	-4.32
Year 2005	<b>-0.1109</b>	0.0548	-2.02
Year 2006	-0.0386	0.0555	-0.70
<b>Model Information</b>			
N		31800	
R <sup>2</sup>		0.9988	

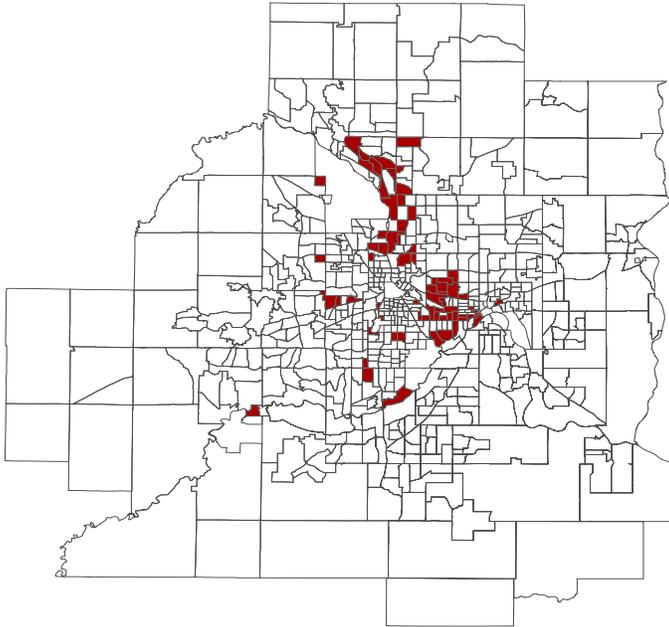
<sup>a</sup>The 2007 year dummy variable was omitted.

**Table 4.4: Summary Statistics for Land Price Indexes**

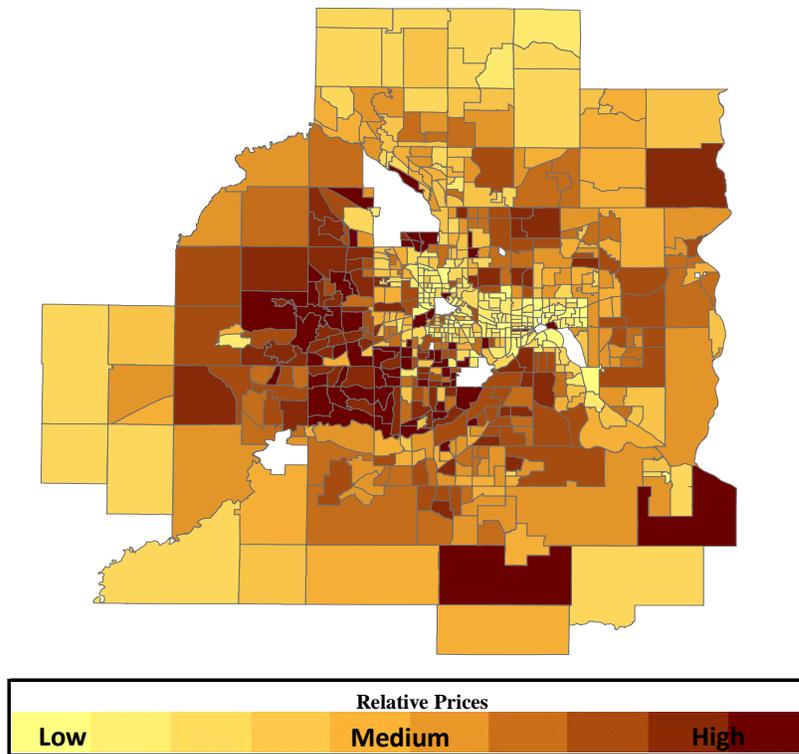
County	Statistic				
	N	Min	Max	Mean	Std
Anoka	79	\$49,978	\$233,475	\$91,645	\$27,687
Carver	18	\$67,739	\$201,965	\$123,560	\$37,343
Dakota	84	\$37,319	\$277,568	\$121,318	\$39,342
Hennepin	71	\$1,928	\$1,954,783	\$141,992	\$146,373
Ramsey	31	\$10,547	\$277,679	\$82,739	\$50,498
Scott	20	\$64,721	\$148,234	\$108,833	\$19,048
Washington	48	\$34,073	\$156,081	\$109,503	\$23,741
All	51	\$1,928	\$1,954,783	\$117,367	\$101,842



**Figure 4.1: Price Indexes by Time Period and House Size**



**Figure 4.2: Census Tracts with Missing Land Transactions**



**Figure 4.3: Estimated Land Price Indexes by Census Tract**

# Chapter 5

## Housing Demand

### 5.1 Introduction

Analysis of local land use amenities has typically been restricted to demonstrating amenity impacts on property values and obtaining partial equilibrium responses resulting from policy counterfactuals. In the context of open space and land use, much of the existing research applies hedonic methods to value open space amenities and reports marginal willingness to pay measures associated with changes to those amenities. Recent examples of literature in this area include Irwin and Bockstael (2001) and Smith et al. (2002). These authors recognize the need to treat open space as a heterogeneous good and therefore account for different types of open space in their hedonic analysis. In this chapter, I incorporate notions of heterogeneous open space amenities and build on a new class of sorting models that utilize the concept of locational equilibrium to model individuals' choices across space. Modeling innovations provided by these sorting methods include the ability to capture the rich spatial structure of the urban landscape, to account for unobserved components of locations, as well

as the ability to model general equilibrium feedback effects in response to changes in open space amenities.

The work presented in this chapter is most closely related to the approach employed by Bayer et al. (2005) who use the horizontal sorting structure to exploit micro-level data covering both residential transactions and individual characteristics. The individual level data allows the authors to specify heterogeneous preferences for local amenities, giving rise to preference orderings that vary across agents. The analysis in the current chapter proceeds using a similar approach to Bayer, et al. (2005), with an emphasis on open space and a choice set definition that consists of distinct housing types.

I extend the existing open space literature by applying the recently developed horizontal sorting framework to a new, and uniquely rich, dataset covering the seven county Twin Cities metropolitan area in Minnesota. The dataset spans 17 years from 1990 through 2006 and includes over 450,000 real-estate transactions, detailed spatially explicit land use data, as well as neighborhood demographics and individual characteristics provided by Census 1990 and Census 2000. An important feature of this dataset is the ability to identify nine types of open space including two conservation programs unique to the Twin Cities.

## 5.2 Theoretical Sorting Model of Demand by Households

A household's utility from choosing a particular housing type is defined as

$$U_{jkt}^i = U_h^i = V(S_h, N_{jt}, Z_{jt}, I^i, \rho_h, \xi_h) + \epsilon_h^i \quad (5.1)$$

where indirect utility,  $U_h^i$ , is composed of housing attributes,  $S_h$ , neighborhood attributes,  $N_{jt}$ , neighborhood social characteristics,  $Z_{jt}$ , individual characteristics,  $I^i$ , prices,  $\rho_h$ , an unobserved measure of attributes associated with a housing type,  $\xi_h$ , and an idiosyncratic error component,  $\epsilon_h^i$ .

To make explicit the heterogeneous preferences defined by the utility specification in equation (5.1), utility is rewritten as

$$U_h^i = \alpha_S^i S_h + \alpha_N^i N_{jt} + \alpha_Z Z_{jt} + \alpha_\rho \rho_h + \xi_h + \epsilon_h^i, \quad (5.2)$$

where  $\alpha_X^i$ ,  $X = \{S, N\}$  is defined as a composite of a mean parameter common across individuals and an individual component that varies across individuals such that

$$\alpha_X^i = \alpha_{0X} + \sum_{q=1}^{\lambda} \alpha_{qX} I_q^i. \quad (5.3)$$

With this parameter specification, it is possible to introduce a flexible set of heterogeneous preferences by specifying interactions between characteristics associated with a housing type with individual characteristics that vary across households.

To define market equilibrium, note that a household's choice of housing type depends on the characteristics of other housing types within their choice set. Because of this dependence, it is possible to assign a probability of choosing any housing type,  $h$ , to every household as

$$\Pr_h^i = f(I^i, \mathbf{I}, \mathbf{S}, \mathbf{N}, \mathbf{Z}, \boldsymbol{\rho}, \boldsymbol{\xi}), \quad (5.4)$$

where  $\mathbf{I}$  refers to individual characteristics,  $\mathbf{S}$  are characteristics of housing,  $\mathbf{N}$  are neighborhood characteristics such as land use,  $\mathbf{Z}$  are social demographics obtained from the

census,  $\boldsymbol{\rho}$  is a vector of housing prices, and  $\boldsymbol{\xi}$  is a vector of characteristics unobservable to the econometrician, but observable to the household. Using these probabilities, the market demand for a specific housing type,  $h$ , is obtained by aggregating all of the individual probabilities for selecting that particular housing type to obtain

$$Demand_h = \sum_i Pr_h^i. \quad (5.5)$$

Assuming an exogenous supply of housing, the market clears when demand for every house type exactly equals the exogenous supply of every house type. This relationship is expressed as

$$Demand_h = Total Supply_h, \quad \forall h, \quad (5.6)$$

where  $Total Supply_h$  is the total supply of housing type  $h$ . In this sense, market clearing implies that the aggregate choice probabilities for every housing type equal the empirical share of each housing type observed in the data.

### 5.3 Empirical Estimation

Empirical estimation follows the procedure described in chapter 3. This section briefly reviews the estimation strategy using the variables included in the theoretical model of demand. It is not possible to jointly identify all the components of indirect utility specified by equation (5.2) in a single step due to the inability to simultaneously identify both a vector of unobserved components of utility along with a set of mean preference parameters. As a result, estimation proceeds in two steps following Berry et al. (1995). To see the mechanics of this two-stage procedure, the indirect utility defined by equation (5.2) is rewritten as

$$V_h^i = \Theta_h + \Gamma_h^i + \epsilon_h^i, \quad (5.7)$$

where  $\Theta_h$  defines variables common to all households and  $\Gamma_h^i$  defines variables unique to households through interactions with individual characteristics. Explicitly,  $\Theta_h$  is defined as

$$\Theta_h = \alpha_{0S}S_h + \alpha_{0N}N_{jt} + \alpha_{0Z}Z_{jt} + \alpha_{0\rho}\rho_h + \xi_h, \quad (5.8)$$

and  $\Gamma_h^i$  is defined as

$$\Gamma_h^i = \left( \sum_{q=1}^{\lambda} \alpha_{qS} I_q^i \right) S_h + \left( \sum_{q=1}^{\lambda} \alpha_{qN} I_q^i \right) N_h. \quad (5.9)$$

With this division, the first stage of estimation recovers all of the heterogeneous parameters,  $\alpha_{qX}$ , along with the mean indirect utility,  $\Theta_h$ , via maximum likelihood. The second stage decomposes the mean indirect utility,  $\Theta_h$ , into observable components,  $\alpha_{0X}$ , and unobservable components,  $\xi_h$  according to equation (5.8).

Estimation of the first stage was performed using Matlab version 7.1 with the `fminunc` optimization package to perform maximum likelihood estimation. For purposes of estimation, analytical gradients were provided allowing use of the ‘Large Scale’ option provided by the optimizer. All convergence criteria were set at 1e-6, including the criteria for contraction mapping convergence. To estimate the model using naïve starting values took approximately 2 days. Starting with informed initial values, estimation can be performed in several hours.

Because the dimension of the choice set is so large, IIA was exploited to improve the estimation speed as described in chapter 3. Each individual was subject to a random draw of 1000 alternatives, including the one chosen by the individual. All alternatives for an

individual were drawn from the available alternatives in the time period the household was observed. The reason for drawing alternatives only from the observed time period is to maintain the assumption of exogenous time; thereby forcing the probability of choosing any housing type in a different time period to zero.

Second stage estimation was also performed in Matlab where instrumental variables techniques were used to account for the endogeneity of price as described in chapter 4. Second stage estimation is considerably faster than first stage estimation and can be completed in under a minute as no numerical optimization is required. Because IV is a mean fitting regression, the results reported for the second stage decomposition of the alternative specific constants correspond to the value associated with the mean alternative specific constant.

## **5.4 Estimation Results**

This section presents estimation results from the sorting model of households location decisions presented in the previous section. When interpreting the results from second stage estimation, it is important to recall that the overall impact of a particular variable consists of both the first stage result as well as the second stage result. In addition, land use variables have no spillover effects into neighboring spatial units. This means that households receive amenity value only from the open space located within their spatial unit. Combining both stages of estimation permits analysis of heterogeneous preferences where households' preferences vary by household composition. To drive home this point, marginal willingness to pay is presented for several sample households of different compositions.

### **5.4.1 First stage results**

First stage estimation recovers interaction parameters as well as a vector of mean indirect utilities for the 20,444 housing types in the sample. The first stage estimation results for the interaction parameters between household characteristics and housing and neighborhood characteristics are reported in table 5.1. These results show intuitive signs. For example, the interaction between number of retired people per household and number of stories takes on a negative sign reflecting the preferences of older individuals for single story housing. Similarly, interaction terms between number of children and number of bedrooms are positive and significant as is the interaction between household size and number of bathrooms.

To capture preferences for open space, interactions between household characteristics and golf, local parks, regional parks, water, and acreage are included. Appendix A contains additional information on these variables. The interactions between number of retired individuals and golf as well as the interaction between number of retired individuals and water are positive and significant implying households with more retired individuals prefer living near golf courses and water bodies. Interactions between local parks and children is positive and significant reflecting families preferences for neighborhoods containing local parks and the interaction between number of working age adults and regional parks is also positive and significant. The positive and significant sign for the interaction with regional parks and working adults possibly accounts for the physical demands needed to utilize the full variety of amenities provided by regional parks such as hiking, biking, paddling, etc., for

which households with more working aged adults may be better able to take advantage of as compared to other households.

An area of interest to policymakers is whether larger lot sizes provide similar open space value to households as provided by other, more public, forms of open space. Interacting number of children with acreage returns a positive coefficient that is statistically significant at the 5% level. The interaction between number of children and local parks is also positive and significant indicating that households with more children prefer both local parks and larger lot sizes compared to households with fewer children. To fully examine the value of larger lot sizes and local parks, it is important to incorporate the mean effect of each type of land use which is the focus of second stage estimation.

#### **5.4.2 Second stage results**

Second stage estimation results for the naïve ordinary least squares (OLS) regression of equation (4.9) are reported in table 5.2. These results have unexpected signs for several variables including a positive price coefficient as well as a negative coefficient on golf courses and water, both of which are considered desirable community characteristics. To address price endogeneity inherent in this regression, an instrument is formed following equation (4.16) by estimating equation (4.15) with a variety of neighborhood variables included to account for observable determinants of housing prices. These variables include all second stage parameters as well as 13 land use characteristics<sup>13</sup> for the 1,2,3,4, and 5 mile

---

<sup>13</sup> The omitted land use category is the other/miscellaneous category.

ring around each block group centroid as well as census characteristics<sup>14</sup> for 1,2,3,4, and 5 mile rings around each block group centroid.

After creating an instrument for price, an instrumental variable (IV) regression of equation (5.8) is performed with the results reported in table 5.3. After instrumentation, the coefficient on price is negative and significant and the coefficients on golf courses and water are both positive and significant, as expected. Of particular interest to policymakers are the variables regarding open space which include agricultural preserves, golf courses, local parks, regional parks, cemeteries, water, the number of Reinvest in Minnesota sites, acreage, and miscellaneous open space. The results for these variables strongly support the notion that open space is a heterogeneous good as the coefficients associated with open space take on a range of signs and significance.

The first class of open space focuses on public open space that is present across much of the study area. Included in this class of open space are cemeteries, local parks, regional parks, water, and other protected open space not designated as parks. Within this group, both cemeteries and regional parks have a negative sign, although neither is significant at the 5% level. While the coefficient on cemeteries may not be surprising, the negative effect of regional parks is counter to what one might expect. When analyzing these second stage results, it is important to consider the interaction terms from the first stage of estimation in which the interaction between regional parks and working aged individuals is positive and significant. Assuming the average household has at least one working aged individual, the marginal utility of regional parks is positive. Other forms of open space within this category

---

<sup>14</sup> Census characteristics include % black, % working, % born US, % born MN, % child, % vacant, and % owner-occupied.

are local parks, water, and non-park protected open space. All of these categories have positive and significant coefficients, as expected.

The second category of open space consists of semi-private and private forms of open space including golf courses and private lot size. While the positive and significant coefficient of golf courses is expected, it is worth noting that the coefficient on acreage is negative and significant. The negative coefficient on acreage likely reflects the increased costs and time required to upkeep larger lots which outweigh the open space benefits provided by larger lots. Recall, however, that the heterogeneous preference parameter interacting number of children and acreage from the first stage of estimation is positive, indicating that the magnitude of the negative coefficient on acreage likely varies over household composition. Despite the positive first stage interaction effect, these results imply that private open space in the form of larger lot size is not a substitute for publicly provided open space in the form of local parks.

The final category of open space includes policy based attempts to preserve land and limit urban sprawl. Open space included in this category includes temporary agricultural preserves, permanent conservation easements, and unprotected agricultural/undeveloped land. The coefficient on agricultural preserves is positive, but not significant, indicating that agricultural preserves provide little if any amenity value to households. In contrast, the coefficient on the permanent Reinvest in Minnesota (RIM) easements is positive and significant indicating that households receive positive marginal utility from these sites. An explanation for the difference in sign and significance for these two programs is that households are concerned with the potential development of open space, resulting in a higher

marginal utility from permanent types of open space than open space which has the potential for development in the future. Comparing these results with the negative and significant coefficient on agricultural/undeveloped land, a story where development potential reduces the value of open space is even stronger. The relationship between the various forms of open space in terms of development potential is examined further when evaluating marginal willingness to pay associated with a variety of types of open space.

## **5.5 Conclusions**

To provide a consistent basis for evaluating the various types of open space, table 5.4 provides marginal willingness to pay measures for all second stage variables at the data averages as well as for three specific household compositions. Providing willingness to pay measures that vary by household composition is possible as a result of the heterogeneous parameters obtained from first stage estimation of interaction terms. The reported willingness to pay highlights the distributional impacts of open space across households and provides important insight into the effects of proposed open space policy.

Focusing on the marginal willingness to pay measures, several policy implications arise. The first implication arises as a result of households having a negative willingness to pay for increased lot size. The negative willingness to pay for an increase in lot size potentially reflects the increased maintenance costs as well as time costs associated with upkeep of larger lot sizes. This result suggests that policies which impose minimum lot size restrictions provide little amenity value to households. Contrasting the negative willingness to pay for increased lot size with the positive willingness to pay associated with local parks;

it appears the two types of open space are not substitutes. For policymakers, this suggests that if providing open space amenities at a neighborhood level is the desired goal, zoning requiring local parks within new development is preferred to creating minimum lot size restrictions.

A second policy implication is that permanent open space in the form of small conservation easements command a higher marginal willingness to pay as compared to semi-permanent open space in the form of agricultural preserves, which in turn have a higher marginal willingness to pay than unprotected open space in the form of agricultural and undeveloped land. In fact, both forms of conservation easements are associated with a positive willingness to pay while the unprotected land has a negative willingness to pay. The result that permanent open space is more valuable than potentially developable open space is consistent with previous literature finding that land perceived as developable provides less amenity value than protected open space.<sup>15</sup>

The third policy implication concerns the treatment of open space as a heterogeneous good over which different households have different preferences. Focusing on the marginal willingness to pay for local parks and water, it is clear that the degree to which households are willing to pay for additional amounts of these types of open space depend critically on the composition of the households under consideration. For example, households with no children and two working aged adults have a 20% lower marginal willingness to pay for local parks than a household with two children and two working adults. For water, a household with two retirees is willing to pay more than double the sample mean for an additional

---

<sup>15</sup> Of course, this result does not take into account other policy goals such as limiting urban sprawl.

amount of open space in the form of water. The ability to observe these differences in willingness to pay arises as a result of the heterogeneous preferences defined in the model and should form an important part of any policy discussion.

In addition to parks, other forms of protected open space also provide significant amenity value to households. Examples of protected open space not included in the local or regional parks categories include greenways, wetlands, and natural areas. These forms of open space are captured by the non-park open space category and are associated with a positive marginal willingness to pay. The marginal willingness to pay for water and golf courses is also positive. These results suggest that a variety of open space provides benefits to households, and policies should not be limited in scope to a single form of open space, but should instead evaluate a variety of open space when deciding how to allocate conservation funds.

**Table 5.1 First Stage Interaction Parameters**

Variables (Neighborhood-X-Household)	Estimate	Std Err <sup>a</sup>	t-stat
% golf-X-# of retirees	<b>4.5537</b>	0.1393	32.6997
% industrial-X-# of children	<b>-0.3942</b>	0.0496	-7.9519
% local park-X-# of children	<b>0.2467</b>	0.0609	4.0525
% regional park-X-# of working aged adults	<b>1.2167</b>	0.0572	21.2711
% water-X-# of retirees	<b>2.4963</b>	0.0765	32.6156
% born US-X-# of retirees	<b>7.0736</b>	0.1692	41.8088
% owner occupied-X-# of children	<b>1.6661</b>	0.0234	71.0521
acreage-X-# of children	<b>0.0318</b>	0.0054	5.8546
# of baths-X-household size	<b>0.2879</b>	0.0057	50.1591
# of stories-X-# of retirees	<b>-5.9223</b>	0.0352	-168.1252
# of bedrooms-X-# of children	<b>0.4985</b>	0.0098	50.6855
distance to CBD-X-# of working aged adults	<b>0.1406</b>	0.0005	270.6814

<sup>a</sup>Standard errors calculated using the inverse of numerically approximated hessian

**Table 5.2: Second Stage Results Using Naïve OLS**

Variables	Estimate	Std Err	t-statistic	Variables	Estimate	Std Err	t-statistic
Constant	<b>12.50</b>	0.34	36.44	% born in US	<b>-14.26</b>	0.33	-43.00
% Agricultural Preserves	<b>2.67</b>	0.67	3.97	% Owner-occupied	<b>1.00</b>	0.11	8.94
% Agricultural/Undeveloped	<b>-0.65</b>	0.13	-4.94	% Vacant houses	<b>-11.60</b>	0.65	-17.92
% Cemetary	<b>2.46</b>	0.58	4.24	Acreage	<b>-0.30</b>	0.03	-11.62
% Commercial	<b>1.24</b>	0.30	4.19	Age of house	<b>0.01</b>	0.00	5.25
% Golf	<b>-2.89</b>	0.39	-7.38	# of baths	<b>-1.04</b>	0.05	-21.11
% 4-lane highways	<b>2.34</b>	0.41	5.69	Garage	<b>1.02</b>	0.08	12.31
% Industrial	<b>1.23</b>	0.19	6.32	# of stories	<b>0.69</b>	0.08	8.21
% Open Space (non-park)	<b>1.32</b>	0.43	3.07	# of bedrooms	<b>-0.44</b>	0.04	-11.14
% Local Parks	<b>0.90</b>	0.32	2.80	Square feet (in 100s)	-0.01	0.00	-1.68
% Regional Parks	<b>-2.85</b>	0.25	-11.27	Distance to CBD	<b>-0.14</b>	0.00	-29.80
% Water	<b>-1.77</b>	0.19	-9.20	Price (in \$1000s)	<b>0.04</b>	0.00	97.81
# of RIM Sites	0.04	0.05	0.97				

**Table 5.3: Second Stage Estimation Results Using Instrumental Variables**

Variables	Estimate	Std Err	t-statistic	Variables	Estimate	Std Err	t-statistic
Constant	<b>20.4656</b>	0.7670	26.6827	% born in US	<b>-17.1743</b>	0.6422	-26.7435
% Agricultural Preserves	0.7880	1.3423	0.5870	% Owner-occupied	<b>3.4094</b>	0.2449	13.9234
% Agricultural/Undeveloped	<b>-2.9411</b>	0.2665	-11.0344	% Vacant houses	<b>-38.8343</b>	1.7829	-21.7812
% Cemetary	-0.7289	1.1073	-0.6582	Acreage	<b>-0.2750</b>	0.0550	-5.0017
% Commercial	<b>1.9902</b>	0.5741	3.4667	Age of house	<b>0.0531</b>	0.0028	19.2929
% Golf	<b>3.2637</b>	0.8865	3.6815	# of baths	<b>0.9136</b>	0.1337	6.8319
% 4-lane highways	<b>4.9225</b>	0.8154	6.0371	Garage	-0.2040	0.1911	-1.0673
% Industrial	-0.6057	0.3805	-1.5918	# of stories	<b>1.0909</b>	0.1827	5.9699
% Open Space (non-park)	<b>13.1443</b>	1.0781	12.1920	# of bedrooms	<b>-0.9369</b>	0.0898	-10.4335
% Local Parks	<b>1.9064</b>	0.6165	3.0925	Square feet (in 100s)	<b>0.0429</b>	0.0100	4.2908
% Regional Parks	-0.2772	0.5122	-0.5412	Distance to CBD	<b>-0.0890</b>	0.0096	-9.2384
% Water	<b>2.3736</b>	0.4466	5.3147	Price (in \$1000s)	<b>-0.0594</b>	0.0040	-14.7085
# of RIM Sites	<b>0.2276</b>	0.0971	2.3447				

**Table 5.4: Marginal Willingness to Pay Heterogeneity**

Household Attribute	Household Structure				
	Mean	A	B	C	D
# children	0.83	0	2	0	0
# working age	1.78	2	2	0	1
# retirees	0.22	0	0	2	1
household size	2.83	2	4	2	2
Variables (Change)	Marginal WTP (\$)				
% Agricultural Preserves (1%)	133	133	133	133	133
% Agricultural/Undeveloped (1%)	-495	-495	-495	-495	-495
% Cemetary (1%)	-123	-123	-123	-123	-123
% Commercial (1%)	335	335	335	335	335
% Golf (1%)	721	549	549	2,083	1,316
% 4-lane highways (1%)	829	829	829	829	829
% Industrial (1%)	-157	-102	-235	-102	-102
% Open Space (non-park) (1%)	2,213	2,213	2,213	2,213	2,213
% Local Parks (1%)	355	321	404	321	321
% Regional Parks (1%)	317	363	363	-47	158
% Water (1%)	494	400	400	1,240	820
# of RIM Sites (1)	3,832	3,832	3,832	3,832	3,832
% born in US (1%)	-2,625	-2,891	-2,891	-510	-1,700
% Owner-occupied (1%)	807	574	1,135	574	574
% Vacant houses (1%)	-6,538	-6,538	-6,538	-6,538	-6,538
Acreage (0.1)	-419	-463	-356	-463	-463
Age of house (1)	894	894	894	894	894
# of baths (.5)	14,550	12,537	17,384	12,537	12,537
Garage (0.1)	-343	-343	-343	-343	-343
# of stories (0.5)	-1,963	9,183	9,183	-90,519	-40,668
# of bedrooms (1)	-8,804	-15,773	1,012	-15,773	-15,773
Square feet (in 100s) (1)	722	722	722	722	722
Distance to CBD (1)	2,707	3,236	3,236	-1,498	869
Price (in \$1000s) (1)	-1,000	-1,000	-1,000	-1,000	-1,000

# Chapter 6

## Housing Supply

### 6.1 Introduction

Housing markets are inevitably comprised of both a demand and supply component. While the supply component of the housing market has been examined at an aggregate level, there is relatively little micro-level analysis of housing supply. In this chapter, I explore the supply side of the housing market by focusing on the decision making process of home builders who must decide where to locate new housing. As compared to the land conversion literature which focuses on the probability of a particular parcel of land being developed, the current treatment of supply focuses on the actual placement of houses across neighborhoods defined by Census 2000 block groups.

To determine which neighborhoods to build housing, along with what type of housing to build, home builders are concerned with the costs of construction, the sale price of the finished house, and the availability of land dictating where new supply of housing can occur. Because home builders in my model do not own undeveloped land, they face a market for

land where equilibrium land price is determined by a competitive market. An important contribution of the current research is the ability to disentangle the input price of land from the output price of housing through exclusionary restrictions.

To model supply, builders are treated as agents who seek to maximize the utility of profit by supplying new housing from a set of housing types defined over dimensions of location, size, and time. As is assumed on the demand side, time is treated as exogenous thereby preventing builders observed providing housing services in one time period from having any impact on the aggregate housing services provided in a different time period. In addition to the assumption that time is exogenous, the total amount of new supply is also assumed exogenous. In other words, the model does not determine the total amount of housing services provided by new construction, but rather seeks to determine the location and type of new housing services.

The remainder of this chapter outlines the theoretical sorting model of builder location decisions followed by a look at the choice set of housing types facing builders. Next, the empirical model is presented followed by estimation results. Following this, I present a series of robustness checks to address potential concerns arising from the definition of the choice set of housing types used during supply estimation. Lastly, this chapter concludes by describing the contributions of the current research in the context of the existing literature as well as policy implications.

## 6.2 Theoretical Sorting Model of Builder Location Decisions

Treating home builders as utility maximizing agents choosing to provide housing services by selecting a particular housing type to supply allows estimation using a horizontal sorting model based on the random utility framework. From the perspective of the supply side of the market, each builder chooses to provide a particular new housing type,  $h$ , in order to maximize their utility of profit. The utility of profit received by builder,  $b$ , from providing a particular housing type,  $h$ , is defined as

$$U_{jkt}^b = U_h^b = V(C_{jt}, L_{jt}, B^b, \zeta_j, \rho_h, \psi_h) + \epsilon_h^b, \quad (6.1)$$

where  $C$  represents cost variables,  $L$  represents land availability,  $B$  represents builder characteristics,  $\zeta$  represents land price,  $\rho$  represents the sale price of a unit of housing services,  $\psi$  is an alternative specific constant, and  $\epsilon$  is an idiosyncratic error term.

Converting equation (6.1) into an estimable equation yields the following indirect profit function

$$U_h^b = \beta_C^b C_{jt} + \beta_L L_{jt} + \beta_\zeta \zeta_j + \beta_\rho \rho_h + \psi_h + \epsilon_h^b, \quad (6.2)$$

where heterogeneous preferences are introduced through the interaction of builder characteristics,  $B$ , and cost variables,  $C$ , according to

$$\beta_C^b = \beta_{0C} + \sum_{q=1}^{\lambda} \beta_{qC} B_q^b, \quad (6.3)$$

with  $q$  indexing builder characteristics. This definition results in preferences over cost variables,  $C$ , that consist of a component common to all builders,  $\beta_{0C}$ , as well as a set of parameters varying based on a specific builder's characteristics,  $\sum_{q=1}^{\lambda} \beta_{qC} B_q^b$ .

The profit specification for builders differs in several ways from the utility specification of households. The most important difference is that the land use categories, specifically many of the amenity providing open space categories, contained in a household's utility are omitted. From a household's perspective, these land use characteristics provide amenity value associated with living in a particular neighborhood. Builders are not concerned with amenity value directly, but rather are interested in the ultimate sale price of the home,  $\rho$ , as it directly impacts revenue. From the builder's perspective, the observed variation in housing prices captures the different amenity levels provided by land use varying across the spatial dimension of housing types.

The second key component of a builder's profit function is the inclusion of cost parameters in the form of soil characteristics,  $C$ , which proxy for the cost of preparing the land for construction as well as the price of the land itself,  $\zeta$ . While these variables enter directly into the builder's profit, they enter only indirectly in the household's utility as the final sale price of the home,  $\rho$ , should partially reflect the costs associated with providing housing services. It is important to note that although the price of a unit of housing services reflects these costs, it is not the sole driver of the final sale price of a house nor are these costs highly correlated with the final sales price.

A component of a builder's profit function closely related to the land use characteristics used in demand estimation is that of available land,  $L$ . This component includes the total amount of available land in the form of agricultural and undeveloped land, as well as policy variables consisting of non-park protected open space, agricultural preserves, and Reinvest in Minnesota conservation easements. The inclusion of policy

variables allows analysis of policies that restrict the amount of potentially developable land within a neighborhood while allowing the price of a unit of housing services to soak up the amenity value provided by other forms of land use.

The final component of a builder's profit function is a set of alternative specific constants allowing the model to capture any components of a builder's profit that are known to the builder but unobservable to the econometrician. As discussed in the following sections, the inclusion of these alternative specific constants raises several empirical challenges not faced in demand estimation. These challenges arise because not all housing types are observed to have new supply. Fortunately, numerical approaches are available that overcome these challenges, as discussed in subsequent sections of this chapter.

To determine the aggregate supply of new housing, the probabilities for each builder choosing each housing type are summed with the probability of builder  $b$  choosing to supply housing type  $h$  given by

$$\Pr_h^b = f(\mathbf{B}^b, \mathbf{B}, \mathbf{C}, \mathbf{L}, \boldsymbol{\zeta}, \boldsymbol{\rho}, \boldsymbol{\Psi}), \quad (6.4)$$

where  $\mathbf{B}$  refers to builder characteristics,  $\mathbf{C}$  are cost characteristics such as soil suitability,  $\mathbf{L}$  represents neighborhood land use characteristics such as available land,  $\boldsymbol{\zeta}$  are land prices,  $\boldsymbol{\rho}$  is a vector of housing prices, and  $\boldsymbol{\Psi}$  is a vector of characteristics unobservable to the econometrician, but observable to the builder. Given this definition, the aggregate supply of new housing is expressed as

$$Supply_h = \sum_b \Pr_h^b. \quad (6.5)$$

An important distinction between the aggregate demand in equation (5.5) and aggregate supply in equation (6.5) is that the aggregate supply does not fully characterize the total market supply of housing. The total market supply of each housing type is the sum of the exogenous supply of existing housing and the builder determined supply of new housing. For purposes of policy counterfactuals in the following chapter, total supply is equated with demand to clear the housing market.

### **6.3 Choice Set of Housing Types Available to Builders**

All builders face a subset of the housing types available to households. The reason builders do not face the full set of housing types available to households is that for some housing types, there is no undeveloped land available on which to build. Over time, the number of block groups having no developable land has expanded so that the available set of housing types facing builders decreases in the more recent time periods. Table 3.1 provides a breakdown of block groups containing developable land for both households and builders at each time period for each house size.

In addition to the restriction on the number of available housing types resulting from the lack of developable land, the housing types available to builders are also restricted to the set of housing types observed in the full transactions data. For example, if a particular block group does not have all three house sizes in the first time period, only the housing types with observed transactions within the block group during the first time period are included in the set of builder housing types. In essence, this excludes housing types not contained on the demand side of the market. Excluding these housing types serves the practical purpose of

preventing missing data problems where no houses were available from which to create price indexes or housing characteristics for particular housing types.

Figure 6.1 shows new home construction included in the supply side of the housing market broken out by time period in which the new supply occurs. The pattern of new development appears as expected in that significant growth of the housing market is occurring along the expanding urban/rural fringe. Potentially more important from a modeling perspective is that not all housing types available to builders contain new housing construction. As a result, identification of the alternative specific constant,  $\psi_h$ , associated with a housing type for which no new housing construction occurred is problematic. One would expect that because no builder chose the particular housing type, the resulting value of the alternative specific constant should be sufficiently negative to ensure the expected supply of the non-chosen alternative is negligible. The empirical implications resulting from the inclusion of non-chosen alternatives are addressed in the empirical section that follows.

## 6.4 Empirical Estimation

Empirical estimation proceeds in two stages as described in chapter 3 where utility of profit is rewritten as

$$V_h^b = \Theta_h + \Gamma_h^b + \epsilon_h^b, \quad (6.6)$$

to emphasize the two stage estimation strategy. With this partitioning of the profit function,  $\Theta_h$  defines variables common to all builders and  $\Gamma_h^b$  defines variables unique to builders through interactions with builder characteristics. For supply estimation,  $\Theta_h$  is defined as

$$\Theta_h = \beta_{0S}C_h + \beta_{0L}L_{jt} + \beta_{0\zeta}\zeta_{jt} + \beta_{0\rho}\rho_h + \psi_h, \quad (6.7)$$

and  $\Gamma_h^b$  is defined as

$$\Gamma_h^b = \left( \sum_{q=1}^{\lambda} \beta_{qS} B_q^b \right) C_h. \quad (6.8)$$

A key difference between this model and that of demand is reflected in the expected sign on the coefficient for price,  $\beta_{0\rho}$ . From a builder's perspective, a positive coefficient reflects builders desire to have higher housing values, all else equal, whereas households prefer lower prices. In addition, a second price coefficient,  $\beta_{0\zeta}$ , is included capturing the price of land which is assumed undesirable from the builder's perspective as higher land prices are viewed as a cost by builders.

First stage estimation recovers heterogeneous parameters that arise through interactions between builder specific characteristics and cost characteristics. In addition, alternative specific constants associated with each housing type are also estimated. The builder specific variable used to form interactions is the size of the builder calculated as the total number of houses built by a particular builder in the current time period and all previous time periods combined. The hypothesis tested by these interactions is that larger builders have a stronger preference to avoid building in more costly areas as compared to smaller builders.

The second stage of estimation decomposes the alternative specific constants recovering preferences for cost variables, land price, house price, as well as available land and land use restrictions. The expectation is that builders do not prefer more costly

development over less costly development but do prefer higher housing prices. In addition, builders are assumed to prefer areas with larger amounts of developable land and are expected to be averse to neighborhoods with conservation limiting the amount of developable land.

An empirical challenge present in supply estimation concerns the definition of the choice set. For builders, the choice set includes not only those housing alternatives in which new supply is observed but also alternatives for which no new supply is observed. In order to estimate the model with non-chosen alternatives included in the choice set, a numerical patch is required that initializes all empirical shares to a positive, but small, value epsilon. For purposes of estimation, the value of epsilon chosen was  $1e-8$ . This patch is required because the contraction mapping routine used to recover alternative specific constants requires all non-zero empirical shares. The non-zero share requirement arises because the probability of choosing each housing type is strictly positive for every builder.

The implication of including non-chosen alternatives in the choice set is that the alternative specific constants associated with them take on much smaller values than the values of alternative specific constants for alternatives actually chosen. As a result, mean estimation techniques are subject to influence by outliers. To partially compensate for this effect, second stage estimation proceeds using median regression techniques rather than the mean techniques used in demand estimation. The technique employed for second stage estimation is instrumental variables quantile regression as described in Chernozhoukov and Hansen (2008) and employed by Timmins and Murdock (2007). The same process to instrument for price is used in this approach as was used for demand estimation. In addition,

the variables included in the price instrument creation regressions are the same as the variables included in the corresponding regressions during demand estimation.

## **6.5 Estimation Results**

The estimation results for builders show expected signs and provide intuitive results for the heterogeneous parameters capturing different preferences for cost variables depending on the size of the builder. An important contribution of the results presented in this chapter is the ability to gain identification of both land price and retail house price simultaneously. As expected, retail price takes on a positive coefficient while a negative coefficient is estimated for land price. The ability to identify price variables, cost variables, and policy variables involving availability of land contribute to the understanding of the supply decisions made by builders. The remainder of this section examines the estimation results in detail.

### **6.5.1 First stage results**

First stage estimation recovers heterogeneous cost interaction parameters as well as an alternative specific constant for each of the 18,693 housing alternatives. Because only 4,545 of the housing alternatives are observed to contain new housing supply, the majority of the alternative specific constants take on significantly smaller values than the ASCs for the chosen alternatives. Graphically, this is shown by the bimodal distribution of the estimated alternative specific constants in figure 6.2. The reason for the bimodal appearance is that ASCs for alternatives not chosen by any builder had to take on a small enough value relative to the alternative specific constants associated with chosen alternatives to ensure that the

probability of any builder choosing a non-chosen alternative approaches zero. The implications of this are examined in discussion of the second stage results.

For the parameter estimates associated with the interactions between the size of builder and soil characteristics approximating costs of development, the hypothesis that larger builders have a stronger aversion to more costly areas than do smaller builders is confirmed. Table 6.1 shows the first stage interaction results where the size of builder is interacted with both poor drainage and limited development suitability. The interaction with poor drainage has the expected negative coefficient indicating larger builders have stronger preferences for areas with better drainage than smaller builders as these areas are likely less time consuming to develop and potentially less costly. The interaction with limited dwelling suitability has a negative and significant coefficient also reflecting the stronger preference of larger builders to supply housing in areas rated more suitable for dwelling than smaller builders. A possible explanation for this result is that larger builders often build many houses in close proximity whereas smaller builders typically build single houses and, therefore, are less concerned with the amount of preparation required for development compared with larger builders. The interaction between size of builder and urban services returns a positive coefficient indicating that larger builders have stronger preferences for neighborhoods where sewer lines exist than do smaller builders. Once again, this result may indicate that the cost of supplying water and sewer to a large subdivision is significantly higher for a single house.

The interaction between the amount of developable land and the size of builder returns a positive and significant coefficient indicating that larger builders have a stronger preference for areas that have potential for future development, measured by available land,

than do smaller builders. This result could also indicate that larger builders often build multiple houses in an area and need more developable land than do smaller builders who may be more likely to build only a single house in an area and are therefore less concerned with the amount of developable land.

The statistically significant first-stage estimates support that notion that builders are not homogeneous in their preferences, but rather that builders preferences vary depending on the size of the builder. From a policy perspective, the variation in preferences across builders is important for understanding the distributional impacts of policies attempting to either curtail or encourage development in a particular area.

### **6.5.2 Second stage results**

Second stage estimation decomposes the alternative specific constants recovered in the first stage into mean components common to all builders according to equation (6.7). The major difference between second stage estimation recovering builder preferences and that for household preferences is the inclusion of many non-chosen alternatives in the choice set of builders. Because these alternatives result in a bimodal distribution of alternative specific constants, quantile regression is used to reduce the impact of outliers. This estimation technique allows estimates associated with any quantile, as opposed to the mean regression methods of OLS and IV, to be recovered. The decision to report the median quantile as opposed to a mean based estimate ensures that the estimates reported fall within the actual data and are not influenced by outliers.

Were there no concerns of price endogeneity, estimation of the second stage decomposition of alternative specific constants is possible using quantile regression without instrumentation. Table 6.2 shows estimates for the median quantile recovered using this approach. All coefficients have the expected sign, but this does not imply that endogeneity concerns are misplaced. It is quite probable that some components of a builders' utility are unobserved to the econometrician and correlated with the retail price of housing. For example, builders may care about zoning restrictions as they dictate what type of neighborhood potential houses are located in and thereby have an influence on house price.

If price is endogenous, one would expect the naïve estimate of the price coefficient to be depressed. Using the instrumentation approach described in chapter 4 and performing instrumental variables quantile regression, the results in table 6.3 are generated.<sup>16</sup> Consistent with an endogeneity story, the coefficient on price increases in magnitude suggesting naïve estimation suffered from price endogeneity problems. While the coefficient for price increases in magnitude, the other coefficients continue to exhibit intuitive signs.

Restrictions to the amount of developable land in the form of agricultural preserves, conservation easements, and non-park open space all have negative coefficients as expected. Of these three land restriction variables, only non-park open space has a negative and significant coefficient indicating that builders prefer not to supply new housing in neighborhoods with large amounts of non-park open space, but are relatively indifferent to the other forms of land restricting open space. An explanation for this result is that

---

<sup>16</sup> A more traditional instrumentation strategy using exclusionary variables from demand estimation as instruments in supply and exclusionary variables in supply estimation for instruments in demand is possible. This instrumentation strategy is difficult due to missing supply variables for some housing types contained in demand for which no developable land is available.

conservation easements and agricultural preserves are viewed as more of an amenity to households and less a constraint on developable land than is non-park open space. In addition, the two conservation programs primarily involve private land and not public land; possibly indicating less of a role for policy to have an impact on the amount of developable land available in the future.

As expected, the amount of agricultural and undeveloped land takes on a positive and significant coefficient indicating that the availability of developable land is an important consideration for builders choosing where to supply new housing. Combined with the first stage results, this positive effect is even stronger for large builders who may be concerned with the ability to expand development in the future or are developing larger areas such as subdivisions.

In terms of cost variables, all soil characteristics associated with costly development take on negative coefficients. The two coefficients with negative and significant signs are poor drainage and limited development potential while the coefficients for very poor agricultural land and very high slope are negative and insignificant. The difference in significance is likely a result of the spatial scale at which these variables vary. The poor drainage and limited development variables change gradually over space whereas the poor agricultural land and degree of slope change quickly. Although the unit of measure is the block group, this spatial unit may be too large to identify variables changing rapidly across space.

The coefficient for metropolitan services is positive and significant indicating that the provision of sewer lines is attractive to builders. The positive coefficient is expected as the

existence of sewer lines results in less work for builders; ultimately saving costs compared with development in which private wells, septic tanks, or other forms of sewer are required. The positive and significant coefficient implies that policy decisions of where to place sewer lines have an important impact on development patterns and are an important policy instrument.

The final category of supply variables includes the price of housing and the price of land. These can be thought of as analogous to an input price in the form of land and an output price for housing after value added through the construction process. As expected, the coefficient on the price of land takes on a negative and significant sign while the coefficient on the price of housing takes on a positive and significant sign. The positive coefficient on house price indicates that builders are influenced by the interaction between supply and demand which determines equilibrium prices, as are households. As a result of this, general equilibrium analysis in the next chapter uses house price as the mechanism to equilibrate supply and demand in response to policy changes altering the demand and supply of housing across the landscape until a new equilibrium is achieved.

### **6.5.3 Marginal willingness to pay**

Combining first and second stage estimates is possible by examining the marginal willingness to pay for variables included in estimation. These values are reported in table 6.4 for a variety of builder sizes in order to emphasize the importance of builder heterogeneity. While no observable heterogeneity is present for variables not interacted in the first stage, the

marginal willingness to pay values for those variables provide a consistent way to evaluate the differences between willingness to pay for various cost characteristics.

Examining the variables with observable heterogeneity, the marginal willingness to pay for more agricultural/undeveloped land and for metro services differ by large magnitudes comparing small builders of single digit numbers of homes and larger builders supplying upwards of 1000 homes. While not differing as substantially, the marginal willingness to pay values for poor drainage and very limited development suitability also differ across builders of different size.

The marginal willingness to pay results once again confirm the hypothesis that larger builders have stronger preferences to avoid more costly development than do smaller builders. In addition, the large differences in willingness to pay for increased urban services and additional agricultural land provide policymakers with important information on the degree to which proposed policies affect builders differently. Specifically, the reduction of agricultural land through conservation programs would result in a greater impact on larger builders than smaller builders while the expansion of urban services would benefit larger builders to a greater extent than smaller builders.

## **6.6 Robustness Checks**

A principal concern with the estimation results presented in the previous sections is the existence of a large number of non-chosen alternatives. The presence of non-chosen alternatives necessitates using a numerical patch initiating all shares to some small, yet positive value in order to allow estimation of alternative specific constants for those

alternatives. As a result, alternative specific constants recovered using this approach are not well identified.

To partially overcome this issue, the use of a median fitting quantile regression estimation technique is employed rather than the mean fitting OLS or IV estimation methods. Despite the use of this technique, the median ASC still falls outside of the range of chosen alternatives prompting robustness checks presented in the current section. These include presenting estimates from a range of quantiles spanning both the chosen and non-chosen alternatives as well as estimation using only the chosen alternatives as the basis for the choice set. Each of these approaches is discussed in turn.

#### **6.6.1 Second stage estimation results for a range of quantiles**

The use of instrumental variables quantile regression allows estimation of parameters at any quantile of the dependent variable. Table 6.5 reports quantile results for a range of quantiles including estimates consistent with the ASCs for chosen alternatives. As approximately only 25% of the supply alternatives are chosen by builders, the 80<sup>th</sup> and 90<sup>th</sup> quantiles fall within that range while the remaining quantiles correspond to ASCs for non-chosen alternatives. Comparing the parameter estimates across different quantiles, it is clear that the magnitudes of the estimates increase further away from the range of chosen alternatives.

In terms of coefficient signs and significance, all statistically significant coefficient estimates take on the same sign across all quantiles with the exception of the coefficient on agricultural preserves and distance to central business district. The coefficients on agricultural and undeveloped land, urban services, house price, and land price are significant

at all quantiles and maintain the same sign across quantiles. For the remaining variables, the estimated coefficients are generally consistent with the exception of quantiles very far from the median.

The similarity of coefficient estimates for various quantiles gives confidence that while the inclusion of a large number of non-chosen alternatives may change the magnitudes of the estimated coefficients, the qualitative results are robust. The one notable exception to this is the coefficient associated with agricultural preserves. As most of the agricultural preserves are located in areas receiving little or no supply, they are likely underrepresented in the set of chosen supply alternatives and thus, it is not surprising the coefficient estimates vary substantially across quantiles for this variable.

### **6.6.2 Estimation using only chosen new supply alternatives**

A second approach to assess the impact of including large numbers of non-chosen alternatives during supply estimation is to estimate the supply model using only chosen alternatives. Excluding non-chosen alternatives reduces the number of alternatives available to suppliers to 4,545 from 18,693 or approximately a 75% reduction. Compared with estimation using the entire set of available supply alternatives, estimation using only the chosen alternatives is about 5 times faster requiring less than an hour to estimate using relatively good starting values.

First stage estimation results using only chosen alternatives are shown in table 6.6 and are virtually identical to estimates using all alternatives, as expected. The reason these two sets of estimates are identical is that the first stage only recovers observable heterogeneity in

preferences along with ASCs. By definition, the observed heterogeneity is only observed if an alternative is chosen. In other words, the only data used to identify first stage interaction parameters comes from the set of observed housing supply transactions.

Excluding non-chosen alternatives results in estimation of a smaller vector of alternative specific constants during first stage estimation compared with estimation using all alternatives. As the estimated vector of ASCs forms the dependent variable for second stage regressions, it is this difference that drives potential changes in second stage estimates. Table 6.7 shows quantile regression results where no instrument for price is used and table 6.8 shows instrumental variables quantile regression results using an instrument for house price.

Comparing the second stage results in table 6.8 to those using all available supply alternatives in table 6.3, several differences arise. The first difference is that the magnitudes of the coefficient estimates are much smaller using only chosen alternatives which corresponds to the results from quantile comparison in table 6.5 where quantiles corresponding to chosen alternatives resulted in smaller magnitudes. The second difference is the loss of significance for several of the variables. The loss of statistical significance is not surprising given that the number of alternatives, left hand side variables, is reduced by 75%.

In terms of coefficient estimates, only the coefficient estimates for agricultural preserves and very high slope change sign with only the agricultural preserve coefficient significant. The sign difference for agricultural preserves may reflect the underrepresentation of agricultural preserves in the set of chosen alternatives as many areas with agricultural preserves lie outside the urban/rural fringe and have experienced limited development.

Despite this, the qualitative results for the majority of the variables remain the same using only the chosen supply alternatives compared with the full set of available alternatives.

Examining the marginal willingness to pay values in table 6.9 once again highlight the decrease in magnitude for the coefficient estimates as the marginal willingness to pay values for many of the variables are slightly smaller in magnitude than for the results using the full set of alternatives. In addition to the decreased magnitudes, the reversal of sign for the coefficients on agricultural preserves and very high slope result in the opposite sign for the respective marginal willingness to pay values.

## **6.7 Conclusions**

This chapter presented one of the first empirical analyses of supply by builders using micro-level data. The horizontal sorting model allowed this analysis to capture important heterogeneity across builders based on observed builder size. In addition to accounting for observed heterogeneity, the sorting model treats the landscape as heterogeneous, allowing identification of builders' preferences for variables that vary across both time and the spatial landscape. By defining distinct housing types consistent with those used in demand estimation, it is possible to equate supply and demand for purposes of policy counterfactuals presented in the next chapter.

The insights gained from the supply model are largely intuitive suggesting the model does a good job of capturing the determinants of the actual behavior of builders. In particular, builders are concerned with the availability of land and prefer to supply new housing in areas with more available land, but are also conscious of the costs associated with

preparing land for development. Another important contribution is the ability to identify both input land prices and output housing prices. The result that builders prefer less expensive land and prefer higher priced housing, all else equal, is intuitive.

Policy implications arising from this chapter are that the provision of urban services has a significant impact of the development decisions of builders. In addition, policymakers can influence the amount of developable land by acquiring privately owned open space. The ability to use policies that restrict developable land provides a means by which policymakers can implement smart growth strategies or achieve other political objectives. The next chapter explores the use of this type of policy tool in the context of both partial and general equilibrium welfare analysis.

**Table 6.1 First Stage Interaction Parameters**

Variables (Characteristic-X-Builder)	Estimate	Std Err <sup>a</sup>	t-stat
% agricultural/undeveloped-X-# of houses <sup>b</sup>	<b>0.3198</b>	0.0046	68.9739
% poor drainage-X-# of houses	<b>-0.0472</b>	0.0058	-8.0703
% very limited dwelling-X-# of houses	<b>-0.0607</b>	0.0079	-7.65
% urban services-X-# of houses	<b>0.1888</b>	0.0029	64.1834

<sup>a</sup>Standard errors calculated using analytical score

<sup>b</sup>Number of houses is in 100s

**Table 6.2: Second Stage Results Using Naïve Quantile Regression (0.5)**

Variables	Estimate	Std Err	t-statistic
constant	<b>6.9478</b>	2.3150	3.0012
% Agricultural Preserves	-2.3148	12.1034	-0.1912
% Agricultural/Undeveloped	<b>22.6219</b>	2.6721	8.4659
% Open Space (non-park)	-1.1757	0.6986	-1.6828
# of RIM Sites	0.3453	0.2988	1.1555
Distance to CBD	<b>0.5272</b>	0.0905	5.8223
% Poor Drainage	<b>-0.9639</b>	0.3680	-2.6193
% Very Limited Development	<b>-0.9314</b>	0.3638	-2.5604
% Urban Services	<b>-7.8697</b>	2.2752	-3.4589
% Very Poor Ag Potential	-0.3367	0.3673	-0.9166
% Very High Slope	-0.0900	0.3379	-0.2662
Price (1000s)	<b>0.0241</b>	0.0007	32.3260
Land Price (1000s)	-0.0030	0.0006	-5.2033

**Table 6.3: Second Stage Results Using Instrumental Variables Quantile Regression (0.5)**

Variables	Estimate	Std Err	t-statistic
constant	<b>-33.4069</b>	2.3514	-14.2074
% Agricultural Preserves	-0.2428	10.6285	-0.0228
% Agricultural/Undeveloped	<b>42.574</b>	1.4036	30.3331
% Open Space (non-park)	<b>-33.7212</b>	3.9568	-8.5223
# of RIM Sites	-0.358	0.6988	-0.5123
Distance to CBD	<b>1.3139</b>	0.3232	4.0654
% Poor Drainage	<b>-5.6304</b>	1.3133	-4.2873
% Very Limited Development	<b>-6.5646</b>	1.3423	-4.8905
% Urban Services	<b>9.1775</b>	1.3072	7.0208
% Very Poor Ag Potential	-0.9557	1.3483	-0.7088
% Very High Slope	-0.5773	1.3865	-0.4164
Price (1000s)	<b>0.2759</b>	0.0147	18.7534
Land Price (1000s)	<b>-0.0351</b>	0.0039	-9.0833

**Table 6.4: Marginal Willingness to Pay Heterogeneity**

Builder Attribute	Builder Size Structure				
	Mean	A	B	C	D
# of houses built	5.4059	1	50	100	1000
Variables (Change)	Marginal WTP (\$)				
% Agricultural Preserves (1%)	-9	-9	-9	-9	-9
% Agricultural/Undeveloped (1%)	1,544	1,543	1,549	1,555	1,659
% Open Space (non-park) (1%)	-1,222	-1,222	-1,222	-1,222	-1,222
# of RIM Sites (1)	-1,298	-1,298	-1,298	-1,298	-1,298
Distance to CBD (1)	4,762	4,762	4,762	4,762	4,762
% Poor Drainage (1%)	-204	-204	-205	-206	-221
% Very Limited Development (1%)	-238	-238	-239	-240	-260
% Urban Services (1%)	333	333	336	339	401
% Very Poor Ag Potential (1%)	-35	-35	-35	-35	-35
% Very High Slope (1%)	-21	-21	-21	-21	-21
Price (1000s) (1)	1,000	1,000	1,000	1,000	1,000
Land Price (1000s) (1)	-127	-127	-127	-127	-127

**Table 6.5: Instrumental Variables Quantile Regression Comparison for Supply Estimation using All Available Alternatives**

Variables	Quantile								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
constant	<b>-55.3463</b>	<b>58.3408</b>	<b>-54.8518</b>	<b>-44.027</b>	<b>-33.407</b>	<b>-23.6579</b>	<b>-9.1301</b>	<b>-5.9431</b>	<b>-4.5324</b>
% Agricultural Preserves	<b>20.1864</b>	<b>28.0469</b>	<b>30.1015</b>	18.6361	-0.2428	<b>-11.4252</b>	-1.7251	0.8268	0.7385
% Agricultural/Undeveloped	<b>48.856</b>	<b>53.7467</b>	<b>53.5747</b>	<b>48.1776</b>	<b>42.574</b>	<b>29.4714</b>	<b>6.5792</b>	<b>2.3177</b>	<b>0.8418</b>
% Open Space (non-park)	<b>-56.6008</b>	<b>-56.1642</b>	<b>-53.9466</b>	<b>-43.1583</b>	<b>-33.7212</b>	<b>-20.8958</b>	<b>-5.5481</b>	-1.3113	0.6299
# of RIM Sites	-0.6424	-0.4441	-1.1168	-0.4857	-0.358	0.0923	0.0565	0.0191	0.0706
Distance to CBD	<b>-1.703</b>	<b>-0.9883</b>	0.1073	<b>0.9548</b>	<b>1.3139</b>	<b>1.2623</b>	<b>0.7336</b>	<b>0.7271</b>	<b>0.8106</b>
% Poor Drainage	<b>-10.0284</b>	<b>-10.3536</b>	<b>-8.6487</b>	<b>-7.5073</b>	<b>-5.6304</b>	<b>-3.4295</b>	-0.874	-0.2754	0.1785
% Very Limited Development	<b>-9.6149</b>	<b>-10.4628</b>	<b>-10.8149</b>	<b>-8.7297</b>	<b>-6.5646</b>	<b>-4.767</b>	-1.1248	-0.5245	-0.5753
% Urban Services	<b>10.8436</b>	<b>14.4572</b>	<b>15.0587</b>	<b>12.1221</b>	<b>9.1775</b>	<b>7.1816</b>	<b>2.5855</b>	<b>1.6989</b>	<b>1.0986</b>
% Very Poor Ag Potential	<b>-6.5581</b>	<b>-5.3751</b>	-2.9872	-1.1175	-0.9557	0.1696	-0.2235	-0.2939	<b>-0.4511</b>
% Very High Slope	-2.4637	-1.1298	-0.0195	0.0388	-0.5773	-0.2439	0.2053	0.1955	0.1136
Price (1000s)	<b>0.4901</b>	<b>0.4726</b>	<b>0.4384</b>	<b>0.3793</b>	<b>0.2759</b>	<b>0.1844</b>	<b>0.1161</b>	<b>0.0931</b>	<b>0.0647</b>
Land Price (1000s)	<b>-0.0418</b>	<b>-0.0564</b>	<b>-0.0615</b>	<b>-0.051</b>	<b>-0.0351</b>	<b>-0.0244</b>	<b>-0.0143</b>	<b>-0.0135</b>	<b>-0.0138</b>

**Table 6.6: First Stage Supply Interaction Parameters**

Variables (Characteristic-X-Builder)	Estimate	Std Err <sup>a</sup>	t-stat
% agricultural/undeveloped-X-# of houses <sup>b</sup>	<b>0.3198</b>	0.0046	68.9738
% poor drainage-X-# of houses	<b>-0.0472</b>	0.0058	-8.0703
% very limited dwelling-X-# of houses	<b>-0.0607</b>	0.0079	-7.6499
% urban services-X-# of houses	<b>0.1888</b>	0.0029	64.1834

<sup>a</sup>Standard errors calculated using analytical score

<sup>b</sup>Number of houses is in 100s

**Table 6.7: Second Stage Chosen Alternative Results Using Naïve Quantile Regression (0.5)**

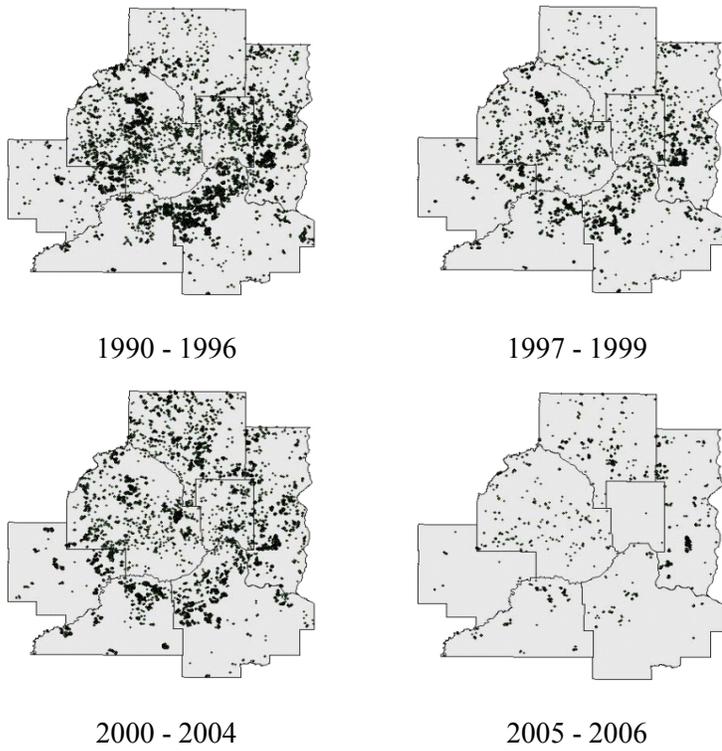
Variables	Estimate	Std Err	t-statistic
constant	0.2711	0.1911	1.4185
% Agricultural Preserves	1.9279	1.0682	1.8048
% Agricultural/Undeveloped	<b>1.2609</b>	0.2151	5.8616
% Open Space (non-park)	<b>2.4906</b>	0.7264	3.4286
# of RIM Sites	0.0699	0.0410	1.7059
Distance to CBD	0.1017	0.0540	1.8837
% Poor Drainage	-0.4039	0.2167	-1.8642
% Very Limited Development	0.2298	0.2274	1.0104
% Urban Services	-0.1840	0.1477	-1.2455
% Very Poor Ag Potential	<b>-0.8130</b>	0.2822	-2.8814
% Very High Slope	-0.2598	0.2540	-1.0227
Price (1000s)	<b>0.0253</b>	0.0009	28.3563
Land Price (1000s)	<b>-0.0061</b>	0.0005	-12.0487

**Table 6.8: Second Stage Chosen Alternative Results Using Instrumental Variables Quantile Regression (0.5)**

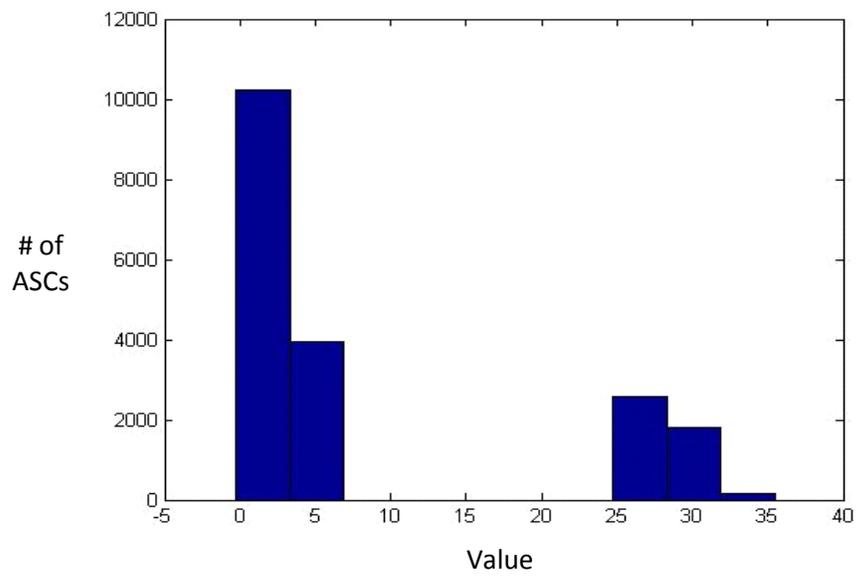
Variables	Estimate	Std Err	t-statistic
constant	<b>-4.0406</b>	0.3133	-12.8959
% Agricultural Preserves	<b>4.7060</b>	1.0089	4.6644
% Agricultural/Undeveloped	<b>2.8937</b>	0.2752	10.5133
% Open Space (non-park)	-1.0111	1.0567	-0.9568
# of RIM Sites	-0.1139	0.0900	-1.2659
Distance to CBD	0.0875	0.0741	1.1812
% Poor Drainage	<b>-1.1438</b>	0.3408	-3.3565
% Very Limited Development	-0.6948	0.3596	-1.9322
% Urban Services	<b>0.5567</b>	0.1913	2.9097
% Very Poor Ag Potential	<b>-1.4526</b>	0.5325	-2.7280
% Very High Slope	0.0885	0.3845	0.2301
Price (1000s)	<b>0.0704</b>	0.0025	28.2511
Land Price (1000s)	<b>-0.0166</b>	0.0012	-14.3730

**Table 6.9: Marginal Willingness to Pay Heterogeneity for Builders using only Chosen Alternatives**

Builder Attribute	Builder Size Structure				
	Mean	A	B	C	D
# of houses built	5.4059	1	50	100	1000
Variables (Change)	Marginal WTP (\$)				
% Agricultural Preserves (1%)	668	668	668	668	668
% Agricultural/Undeveloped (1%)	413	411	434	456	865
% Open Space (non-park) (1%)	-144	-144	-144	-144	-144
# of RIM Sites (1)	-1,618	-1,618	-1,618	-1,618	-1,618
Distance to CBD (1)	1,243	1,243	1,243	1,243	1,243
% Poor Drainage (1%)	-163	-163	-166	-169	-230
% Very Limited Development (1%)	-99	-99	-103	-107	-185
% Urban Services (1%)	81	79	92	106	347
% Very Poor Ag Potential (1%)	-206	-206	-206	-206	-206
% Very High Slope (1%)	13	13	13	13	13
Price (1000s) (1)	1,000	1,000	1,000	1,000	1,000
Land Price (1000s) (1)	-236	-236	-236	-236	-236



**Figure 6.1: New Supply by Time Period**



**Figure 6.2: Estimated Alternative Specific Constants**

# Chapter 7

## Policy Analysis

### 7.1 Introduction

In addition to the marginal willingness to pay measures obtained directly from parameter estimates, the horizontal sorting model is capable of providing several additional welfare measures. These welfare measures are divided into partial and general equilibrium depending on whether households and/or builders are permitted to change their housing decisions in response to policy interventions. To compute willingness to pay measures for a policy change, the familiar log-sum rule is applied

$$\Delta E(CS^i) = \frac{1}{\alpha_{0\rho}} \left[ \ln \left( \sum_h e^{v_h^{i1}} \right) - \ln \left( \sum_h e^{v_h^{i0}} \right) \right], \quad (7.1)$$

where the 1 superscript refers to after the policy change and the 0 superscript refers to before the policy change.

Modeling both the supply and demand of housing facilitates calculating additional welfare measures resulting from the interaction between both the supply and demand sides of

the housing market. While no cross equation restrictions are imposed, supply and demand jointly determine the market clearing prices which in turn enter into the optimizing decisions of both households and builders. As a result, the interaction between the demand and supply sides of the market is required to solve for a new equilibrium following a policy shock.

Before discussing how to integrate the two sides of the market, it is important to understand the procedure by which a new equilibrium is obtained on the demand side of the market as this serves as the basis for integrating supply and demand later on. To provide some intuition for the logic behind general equilibrium welfare measures, recall that the structure of the horizontal sorting model provides a Nash equilibrium framework where each household's decision of which housing type to consume is optimal given every other household's decision. The general equilibrium response arises when households re-sort following a policy change resulting in a new market demand. Assuming a fixed supply, any new equilibrium must equate market demand to the exogenous supply of housing. The mechanism by which the market clears is straightforward and involves increasing the price for housing types with excess demand and decreasing the price associated with housing types having excess supply. The adjustment of prices to clear the market is performed using an iterative approach allowing households to re-sort after each price change until a new equilibrium is obtained with demand equal to exogenous supply.

Relaxing the exogenous supply restriction only slightly modifies the approach needed to solve for a new equilibrium. In the simplest case, the policy change only affects supply through price. For this case, the modified numerical approach uses price to remove excess demand by increasing prices, and removes excess supply by decreasing prices in the same

way as with an assumed exogenous supply. The difference is that after each change in prices, supply is allowed to adjust in response to the changing prices. Because the effect of price is opposite for builders compared to households, an increase in price results in additional supply and a decrease in price reduces supply, all else equal. For a housing type initially having excess demand, the increase in price not only drives away some demanders, but it also results in increased supply. In this way, the effect of a price change is exaggerated by including an elastic supply response.

A more intensive numerical procedure is needed when the policy shock affects demand and supply variables as well as prices. In this scenario, supply changes both as a direct result of the policy and as a result of market clearing changes in price. While the same approach is used to solve for a new equilibrium as used when only price changes impact supply, the initial disparity between supply and demand is greater as a result of the direct supply shift in response to the policy. As a result, the numerical approach used to solve for a new equilibrium is significantly slower.

In order to demonstrate the importance of general equilibrium with a fixed supply or demand, and also general equilibrium including supply and demand feedback effects, three policy scenarios are examined. The policies involve increasing enrollment in the temporary agricultural preserve program, the purchase of open space by the department of natural resources at the urban fringe, and the purchase of undeveloped land by the department of natural resources inside the city. Each policy is described in the sections that follow.

## **7.2 Interpretation of Willingness to Pay Measures**

Given the calculation of willingness to pay measures shown in equation (7.1), it is clear that the estimated price coefficient plays a critical role in determining the willingness to pay measures associated with the three welfare policies discussed below. As such, it is important to recall the method used for price index creation as those price indexes are the basis for the estimated price coefficient. The estimated prices are nothing more than a fixed effect for each housing type recovered using a hedonic regression accounting for observable housing characteristics. The inclusion of observable housing characteristics in the hedonic regression accounts for changes in the housing stock across time and space.

Adjusting transactions prices by CPI, the estimated prices are roughly equivalent to the median home price for all transactions of a particular housing type after adjusting for inflation. As such, the prices included in estimation are a one-time price to acquire housing services. Housing services vary across housing types and are based on observable housing characteristics included in the price index regression. As no scaling of the estimated prices is performed, willingness to pay values are neither annualized nor recurring measures, but rather a one-time willingness to pay. This onetime welfare measure is similar to the mortgage fee used to fund agricultural preserves imposed on all sales of new houses. For builders, the willingness to pay measures are equivalent to building permit fees levied a single time at the onset of building a house.

One final aspect of the housing price approach used during estimation is that absentee landlords are assumed to receive all capital gains resulting from land use changes over time or general price appreciation. As a result, the agents modeled in the current work are very

much like renters who pay a single fee for lifetime rental rights. Because land price is assumed not to adjust in response to supply and demand changes, the entire price effect resulting from welfare policies is captured through changes in the housing price.

### **7.3 Expanding Agricultural Preserves**

The primary objectives of the agricultural preserve program are to encourage investment in agriculture and reduce the loss of agricultural lands resulting from increased development pressure. To examine the impact of an expansion of the agricultural preserve program, a hypothetical policy is designed that enrolls 5% of unprotected agricultural land into the agricultural preserve program for 20 block groups located greater than 10 miles from the center of either Minneapolis or St. Paul. The increased agricultural preserve enrollment and decreased amount of unprotected agricultural and undeveloped land is assumed to exist in all time periods. Figure 7.1 highlights the block groups impacted by this policy. In total, 212 elements of the choice set are affected by this policy.

To evaluate the welfare impacts resulting from an expansion in agricultural preserves, a variety of welfare measures are employed focusing not only on the effects to households but also the effects of the policy on home builders. These welfare measures include partial equilibrium measures where no price changes occur in response to the policy, general equilibrium measures where only one side of the market adjusts assuming a fixed amount of either demand or supply but allowing resorting of agents and using price to clear the market, and general equilibrium measures where both demand and supply re-sorting is allowed and prices adjust to clear the market.

Households' marginal willingness to pay for agricultural/undeveloped land is negative while their marginal willingness to pay for agricultural preserves is positive. This suggests that the policy should raise households' welfare keeping everything else the same in a partial equilibrium sense. While the expected partial equilibrium effect is intuitive, the degree to which the various measures of general equilibrium welfare effects differ from the partial equilibrium effect are not straightforward. To bring the market into equilibrium assuming a fixed supply, prices adjust to decrease demand in the neighborhoods experiencing excess demand, which are likely the neighborhoods receiving additional agricultural preserves, and increase demand in neighborhoods with excess supply. Because the total demand is exogenous, in order for some housing types to have excess demand, others must have excess supply. To bring the market back into equilibrium, prices must rise in neighborhoods with excess demand and fall in neighborhoods with excess supply.

For general equilibrium where supply is not fixed, the marginal willingness to pay for a loss of agricultural and undeveloped land is negative from the perspective of the builders and the marginal willingness to pay for additional agricultural preserves is near zero, but negative. Therefore the expected partial equilibrium effects to builders are the opposite of households resulting in an expected welfare loss from the policy. For reasons outlined above, the expected change in welfare moving to a general equilibrium framework where demand is assumed exogenous is unclear as prices adjust across the urban landscape to bring the market into equilibrium.

The most realistic welfare measures is achieved when both supply and demand adjust to the policy and price equates the two sides of the market, but at a different equilibrium than

prior to the policy. In this scenario, the initial shock to households and builders is opposite, resulting in a greater disparity in supply and demand initially following the policy shock than for welfare measures where only one side of the market adjusts to the policy shock. Because the effect of price is opposite for households and builders, the same price change used in general equilibrium with a fixed side of the market results in a faster rate of convergence when both sides of the market are allowed to adjust. The larger adjustment of the market necessitates making smaller price adjustments in order to solve for the new equilibrium price vector.

Welfare results are reported in table 7.1 and as expected show a welfare increase for households and a welfare decrease for builders resulting from the increase in agricultural preserves. There is a large difference in welfare measures between partial equilibrium measures, where no price adjustment has occurred in order to equate supply and demand, and general equilibrium measures equating supply and demand through price changes. Focusing on the difference between partial equilibrium and general equilibrium assuming only one side of the market adjusts; there is a 10% difference between welfare results for households and over a 200% difference in welfare measures for builders.

Allowing both sides of the market to adjust to the policy, the difference between partial equilibrium and full general equilibrium is smaller with only a 2% difference for households and approximately a 70% difference for builders. It is important to remember that the direction of the welfare changes resulting from a full general equilibrium model compared to a general equilibrium model assuming the policy only affects one side of the market is not a general result, and is dependent on the particular under analysis.

## **7.4 Acquiring Additional DNR Land at the Urban Fringe**

Unlike agricultural preserves, land owned by the department of natural resources (DNR) is considered permanently protected open space. The vast majority of this land consists of sensitive wildlife habitats and non-productive agricultural lands. However, a small portion of DNR lands consist of maintained trails and greenways and provide a different form of amenities than the preserved wildlife habitat that constitutes the majority of DNR protected open space. Regardless, the vast majority of DNR land provides very different amenities than those provided by parks.

The growth of the metropolitan area has placed development pressure on undeveloped land at the urban/rural fringe resulting in the loss of wildlife habitat along with development of fallow fields. The policy evaluated in this section explores the welfare impacts to both households and builders from preserving 2.5% of agricultural and undeveloped land in 20 block groups selected randomly at the urban/rural fringe. For purposes of this policy, the urban/rural fringe is defined as areas between 8 and 14 miles from either the center of Minneapolis or St. Paul. The randomly selected block groups affected by this policy are shown in figure 7.2. In total, 188 housing types are impacted by the policy across the 4 time periods.

The marginal willingness to pay for non-park protected open space (DNR land) is positive for households and negative for builders as was the case for the agricultural preserve policy in section 7.3. The welfare results reported in table 7.1 are as expected in that households experience a positive welfare change and builders experience a negative welfare

change as a result of the increased non-park open space and decreased amount of agricultural/undeveloped land. Also, there are significant differences between partial and general equilibrium welfare measures for both households and builders, but those differences are smaller when both supply and demand are impacted by policy as compared to when only one side of the market is impacted. As with the agricultural preserve policy, this result is not universal and depends on the particular policy under evaluation.

### **7.5 Acquiring Additional DNR Land inside the City**

The purpose of the third policy is to examine the impacts of the location of open space on welfare. To do this, the identical non-park open space policy is employed but to a different set of spatial locations. For the current policy, 20 block groups within 10 miles of the cities of either Minneapolis or St. Paul were randomly selected for an increase of 2.5% of total land classified as non-park open space and a loss of 2.5% of total land classified as agricultural/undeveloped. Figure 7.3 highlights the block groups affected by this policy. In total, 201 housing types are impacted.

Comparing the willingness to pay measures for this policy with the same policy targeting the urban/rural fringe, substantial welfare differences are evident. For households, welfare change from this policy is approximately 40% smaller than that for the policy impacting the urban/rural fringe. For builders, the welfare changes are even more dramatic as fewer suppliers of new housing are impacted due to the lower level of new development in the affected areas. Another difference between the inner city increase in open space and the urban fringe increase is that the difference in general equilibrium welfare for households

between the policies where supply is assumed fixed and supply is allowed to adjust is smaller for the inner city policy. A likely cause of this is that less development is occurring in the more urban areas than at the urban/rural fringe resulting in a smaller supply adjustment response.

A final result is worth noting regarding the change in general equilibrium measures for builders between the policy assuming a fixed demand and the policy allowing demand to adjust in response to the policy shock. Unlike all of the other policies evaluated, the welfare effect for the later policy is of a greater magnitude than the welfare measures assuming a fixed demand. This result confirms that the difference between a policy affecting only a single side of the housing market and one impacting both supply and demand is highly dependent on the particular policy being evaluated.

## **7.6 Welfare Estimates Excluding Non-Chosen Supply Alternatives**

Chapter 6 presented supply estimation results using only chosen alternatives, reducing the number of available supply alternatives by approximately 75%. Using those results, the identical 3 welfare policies discussed in the current chapter were run and the results are shown in table 7.2. The first obvious difference from the results using the full set of supply alternatives is the number of alternatives impacted by the policy on the supply side is substantially reduced. Not only does the welfare policy not impact non-chosen supply alternatives, but the full general equilibrium welfare measures integrating the supply and demand side of the market are restricted to only allow new supply to adjust within the set of chosen supply alternatives.

The expected implication of reducing the number of supply alternatives is to lessen the impact of integrating supply and demand in general equilibrium welfare calculations. Comparing the differences between the two general equilibrium measures for households, this difference is evident. For all three policies, the direction of welfare change is the same when allowing a supply response, but the magnitudes of those changes are much smaller. For the policy increasing non-park open space at the urban fringe, the increase in welfare by allowing a general equilibrium supply response is less than \$3 compared with over \$13 dollars using the full set of supply alternatives.

The welfare effects on the builder side of the market are also quite different when the choice set facing builders is reduced compared with the full choice set including initially non-chosen alternatives. Unlike demand in which only the fully integrated supply and demand general equilibrium welfare measure changes, all three welfare measures change on the supply side as a result of different parameter estimates and different numbers of affected alternatives. As with households' willingness to pay, reducing the number of supply alternatives dampens the effect of integrating both the supply and demand sides of the market.

Recall that estimation using only the chosen alternatives resulted in a positive coefficient on agricultural preserves as opposed to the negative coefficient using the full set of available supply alternatives. The impacts of this sign difference are seen in the positive partial equilibrium welfare measures associated with increasing agricultural preserves of \$21.28 compared with the negative partial equilibrium welfare measure of -83.08 using the

full set of supply alternatives. This difference highlights the important role the estimation results play in determining the welfare measures from the underlying econometric models.

Despite the differences in welfare measures for both households and builders, the importance of incorporating both sides of the housing market into general equilibrium welfare measures is still evident. While reducing the number of supply alternatives dampens the effect of integrating the market due to the reduced ability for supply responses to policy shocks, there remains an important supply response. Clearly, the degree to which supply responses are relevant depend on the particular policies under analysis and the spatial impact of those policies within the metropolitan area.

## **7.7 Conclusions**

The policy scenarios evaluated in this chapter highlight the importance of incorporating general equilibrium welfare measures into potential policy discussions as opposed to simply including partial equilibrium measures or marginal willingness to pay values. In addition to the substantial differences between partial and general equilibrium welfare measures, the inclusion of both sides of the housing market results in substantial differences in welfare measures as opposed to welfare measures holding supply constant. From a modeling perspective, this result highlights the importance of incorporating an estimated supply of housing into policy analysis evaluating open space proposals.

Several other implications from the policy work presented here are that the magnitude and direction of welfare changes between partial and general equilibrium analysis are highly dependent on the particular policy in question and that the spatial nature of open space policy

plays a critical role in determining the welfare implications from policy as well as the degree to which supply and demand interact. The importance of space corroborates work by Walsh (2007) finding that the location of proposed open space plays an important role in determining the welfare impacts to households.

In addition to these results, the results presented in the current chapter highlight that proposed open space policies affect both households and suppliers and that the effects of these policies on each group are often opposite. Given that the empirical literature on micro-level supply decisions is virtually nonexistent; this result should highlight the importance of filling in this gap in the literature first outlined by DiPasquale (1999). It is worth noting that the signs and magnitudes of the policies evaluated in the current chapter are highly dependent on the estimation results and are quite sensitive to the parameter estimates for policy variables.

**Table 7.1: Welfare Results for Hypothetical Policies**

Policy	Impacted Alternatives		WTP (\$)					
	Demand	Supply	Households			Builders		
			Partial <sup>a</sup>	General (fix) <sup>b</sup>	General <sup>c</sup>	Partial	General (fix)	General
Increase agricultural preserves by 5%	212	212	48.12	43.19	47.24	-83.08	-284.28	-145.54
Increase urban fringe non-park protected open space 2.5%	188	188	75.77	59.84	72.88	-59.76	-177.79	-80.44
Increase inner city non-park protected open space 2.5%	201	201	44.59	40.23	41.14	-5.06	-16.86	-41.61

<sup>a</sup>No price adjustments occur

<sup>b</sup>Only price adjustments occur, fixing the other side of the market

<sup>c</sup>Price adjustments occur and both sides of the market adjust

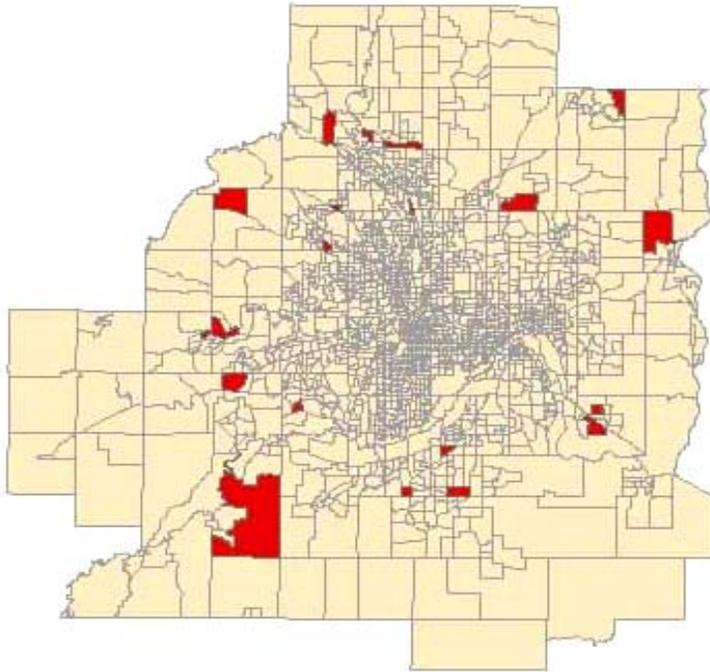
**Table 7.2: Welfare Results for Hypothetical Policies using only Chosen Alternatives for Supply**

Policy	Impacted Alternatives		WTP (\$)					
	Demand	Supply	Households			Builders		
			Partial <sup>a</sup>	General (fix) <sup>b</sup>	General <sup>c</sup>	Partial	General (fix)	General
Increase agricultural preserves by 5%	212	106	48.12	43.19	43.91	21.28	21.27	62.81
Increase urban fringe non-park protected open space 2.5%	188	58	75.77	59.84	62.73	-30.00	-31.56	41.98
Increase inner city non-park protected open space 2.5%	201	37	44.59	40.23	40.36	-2.42	-2.84	-27.96

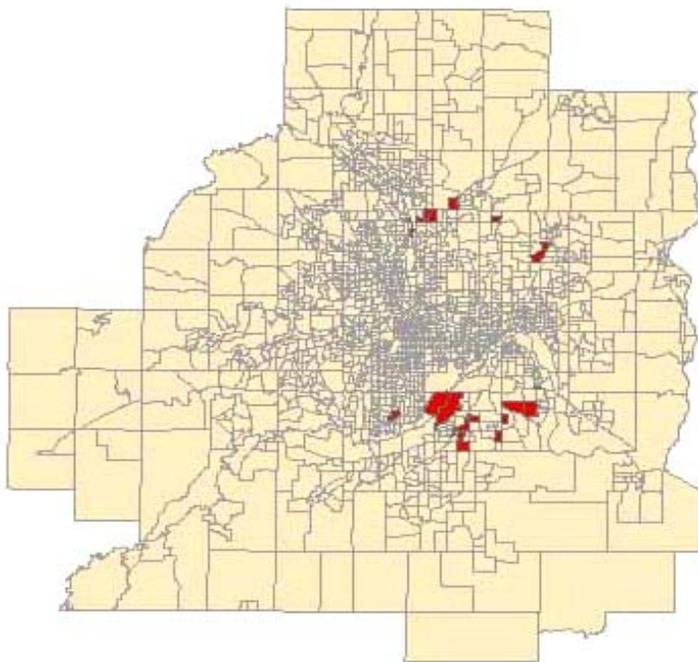
<sup>a</sup>No price adjustments occur

<sup>b</sup>Only price adjustments occur, fixing the other side of the market

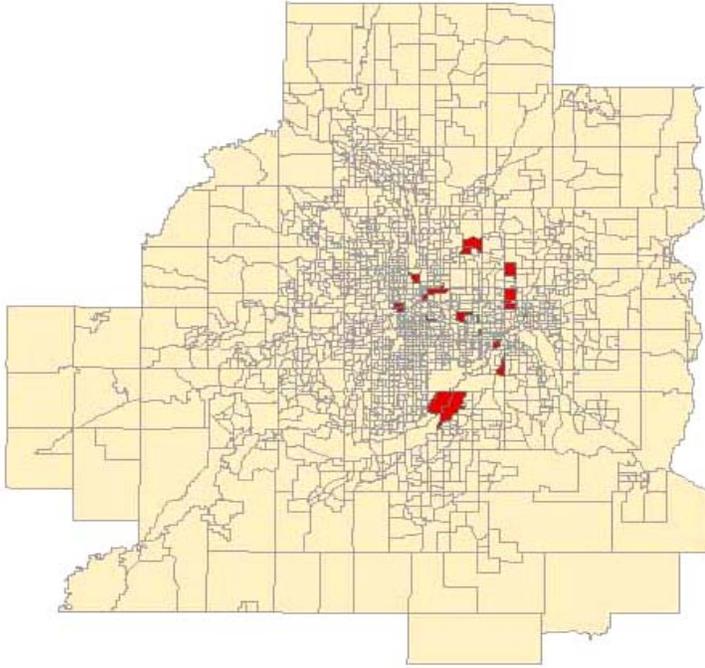
<sup>c</sup>Price adjustments occur and both sides of the market adjust



**Figure 7.1: 5% Increase in Agricultural Preserves**



**Figure 7.2: 2.5% Increase in Urban Fringe Open Space (non-park)**



**Figure 7.3: 2.5% Increase in Inner City Open Space (non-park)**

# Chapter 8

## Results and Future Research

Housing markets have long provided a window through which researchers can examine households' preferences for local public goods. Housing markets not only allow researchers to view market clearing prices, but also the locations of the houses involved in market transactions. The research presented here uses observable housing transactions as the basis to estimate both household and builder preferences for the first horizontal sorting model focusing on open space. In addition to advances provided by the horizontal sorting model for demand estimation, the estimation of micro level supply of housing provides one of the first empirical insights into the preferences of home builders. This chapter reviews the key empirical results of the current research and discusses areas of future research expanding the work presented here.

As virtually all of the existing open space literature has focused on households, and not suppliers of new housing, the results obtained from demand estimation are most directly comparable to existing literature. Defining a variety of heterogeneous forms of open space

reveled that households have a larger willingness to pay for more permanent forms of open space than for temporary or unprotected open space. This result corroborates much of the existing open space literature finding that more permanent forms of open space provide greater benefits to households than potentially developable open space. In addition to confirming that duration of open space is important, this result also highlights the importance of differentiating among the many forms of open space present in a large metropolitan area.

By estimating not only the preferences for a diverse set of open space, but also allowing those preferences to vary over household composition, I have shown that incorporating observable heterogeneity is crucial in order to fully understand the impacts of proposed open space policy. The importance of incorporating heterogeneous preferences into analysis has motivated the development of horizontal sorting models discussed by Bayer and Timmins (2007) as well as recent work extending the vertical sorting model by Kuminoff (2007). The current work confirms the notion that much can be learned from incorporating more flexible preferences into the analysis of local public goods and should serve to encourage efforts to continue investigating methods that incorporate richer heterogeneity and exploit increasingly detailed data on socio-economic characteristics of agents.

A key contribution of the current work is the estimation of micro-level builder decisions of new housing supply. As outlined by DiPasquale (1999), a sizeable hole in the empirical literature is the lack of micro level housing supply analysis. Drawing on the land conversion literature, the current paper is one of the first to estimate actual new housing supply in an equilibrium framework provided by the horizontal sorting model. The emphasis on the equilibrium nature of supply estimation should not be understated as it is that

equilibrium that allows integration of supply and demand for policy analysis and permits feedback effects from demand to influence the supply of housing through market prices.

As there is little existing micro-level supply literature to appeal to, results from supply estimation are evaluated based on the degree to which they match our intuition on what drives supply decisions of builders. The finding that larger builders have stronger preferences to avoid more costly development than smaller builders seems sensible. Likewise, the finding that higher costs, as proxied by land characteristics, along with higher land prices have a negative marginal willingness to pay confirms our intuition on what drives builders' decisions. Recovering a positive house price effect is also quite intuitive as that directly influences a builder's revenue.

In terms of the impact of potential open space policy on builders, the current research provides evidence that some policies appear effective at influencing urban expansion and supply decisions. First, builders strongly prefer areas with undeveloped land and un-protected agricultural land, which provide the land for new supply. Finding that protected open space has negative marginal willingness to pay values opens the door to consider policies that restrict the amount of developable land in an effort to discourage development. Second, provision of urban services is found to encourage development as the existence of water and sewer lines likely reduce costs of development. For policymakers, decisions to expand urban services are likely to result in increased new supply as the areas receiving urban services become more attractive to home builders.

Using both housing demand and housing supply estimates, welfare policies are evaluated showing that the re-sorting of agents in response to policy changes can result in

large differences in welfare measurements. In addition, incorporating supply feedbacks in response to demand changes, and vice versa, are shown to result in welfare measures that differ considerably from those obtained when only one side of the housing market is allowed to adjust following a policy shock.

I find that increasing open space along the urban/rural fringe results in greater welfare gains to households than a similar policy targeting inner city neighborhoods. This finding highlights the importance of accounting for space when describing policy and should serve to encourage the use of econometric techniques designed to account for space at smaller spatial resolutions. As improved spatial data is rapidly becoming available, the integration of better data with spatially explicit modeling techniques is likely to provide a springboard for future advancements in the literature on local public goods.

While the current work has improved our understanding of preferences for open space and laid the groundwork for incorporating micro-level supply into policy analysis, many areas for future research remain. These range from expanding the current modeling framework using additional data to investing more realistic characterizations of the supply of housing by builders. Through acquisition of additional data, development of new theory, and posing slightly different research questions, many of these extensions can be implemented given only minor adjustments to the data and framework presented in the current research.

The limited availability of micro-level socio-economic data in the current work precludes inclusion of non-linear income effects using the block group spatial characterization of the choice set of housing types. To incorporate income effects through a non-linear budget constraint, either additional data is required or aggregation of housing

types to a larger spatial scale is needed. The results in appendix B present comparable estimation results to those reported here using the larger census tracts as the spatial unit of a housing type. Estimation of that model shows that tradeoffs certainly occur by moving to larger spatial levels. In particular, the ability to tease out statistical significance for the large number of heterogeneous forms of open space is diminished as averaging over larger neighborhoods removes important identifying variation in some forms of open space.

Despite the drawbacks to moving to a larger spatial scale, a slightly different research question such as comparing more broadly defined open space categories would allow integration of a non-linear budget constraint into the RUM framework used to estimate the horizontal sorting model. The inclusion of a budget constraint would highlight affordability and preclude low income households from choosing high price neighborhoods that are likely unaffordable.

Another important advancement to the current work involves the expansion of the treatment of new housing supply. As the current research provides an encouraging initial attempt at estimating builder behavior, future work extending the model of supply to include non-perfectly competitive home builders seems appropriate. Given data on the size of home builders, it seems reasonable to conclude that larger home builders hold some form of market power. Including non-competitive home builders would likely reduce supply of new housing while raising prices.

Potentially reducing supply of new housing raises an important, unresolved empirical challenge that exists in almost all of the micro-level housing literature. In the current work, and in many other works, the total supply of housing is assumed exogenous. While the

current work relaxed the exogenous assumption of where new supply is located, it did not attempt to determine the total amount of new supply in the housing market. Rather, I simply allowed the location of an exogenous amount of new supply to be determined endogenously. Future work to endogenously determine the total amount of new supply would provide important improvements to the current analysis.

Lastly, it is worth remembering that the current research is specific to a particular metropolitan area and a particular time frame. As most of the existing housing literature, including the current work, focuses on large metropolitan areas, our understanding of the impacts of open space policies on smaller municipalities is sparse. Future work using different data would provide important corroboration for many of the empirical findings from the current research.

## References

- Anderson, S. T., and S. E. West. "Open Space, Residential Property Values, and Spatial Context." *Regional Science and Urban Economics* 36(2006): 773-789.
- Bajari, P., and C. L. Benkard. "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach." *Journal of Political Economy* 113(2005): 1239-1276.
- Bajari, P., and M. E. Khan. "Estimating Housing Demand with an Application to Explaining Racial Segregation in Cities." *Journal of Business and Economic Statistics* 23(2005): 20-33.
- Bayer, P., N. Keohane, and C. Timmins (2006) Migration and Hedonic Valuation: The Case of Air Quality.
- Bayer, P., and C. Timmins. "Estimating Equilibrium Models of Sorting Across Locations." *The Economic Journal* 117(2007): 353-374.
- Bayer, P., and C. Timmins. "On the Equilibrium Properties of Locational Sorting Models." *Journal of Urban Economics* 57(2005): 462-477.
- Bayer, P. J., R. McMillan, and K. S. Rueben. "An Equilibrium Model of Sorting in an Urban Housing Market." *NBER working paper series*, no 10865(2005).
- Bell, K. P., and E. G. Irwin. "Spatially explicit micro-level modelling of land use change at the rural-urban interface." *Agricultural Economics* 27(2002): 217-232.

- Berry, S., J. Levinsohn, and A. Pakes. "Automobile Prices in Market Equilibrium." *Econometrica* 63(1995): 841-890.
- Berry, S., O. Linton, and A. Pakes. "Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems." *Review of Economic Studies* 71, no. 3(2004): 613-654.
- Berry, S. T. "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics* 25, no. 2(1994): 242-262.
- Bockstael, N. E. "Modeling Economics and Ecology: The Importance of a Spatial Perspective." *American Journal of Agricultural Economics* 78, no. 5(1996): 1168-1180.
- Brown, J. N., and H. S. Rosen. "On the Estimation of Structural Hedonic Price Models." *Econometrica* 50(1982): 765-768.
- Chay, K., and M. Greenstone. "Does Air Quality Matter? Evidence from the Housing Market." *Journal of Political Economy* (2005): 376-424.
- Chernozhukov, V., and C. Hansen. "Instrumental Variable Quantile Regression: A Robust Inference Approach." *Journal of Econometrics* 142, no. 1(2008): 379-398.
- DiPasquale, D. "Why Don't We Know More About Housing Supply?" *Journal of Real Estate Finance and Economics* 18, no. 1(1999): 9-23.
- DiPasquale, D., and W. C. Wheaton. "Housing Market Dynamics and the Future of Housing Prices." *Journal of Urban Economics* 31, no. 3(1994): 337-359.
- Epple, D., B. Gordon, and H. Sieg. "A New Approach to Estimating the Production Function for Housing." Working Paper (2007).
- Epple, D., T. Romer, and H. Sieg. "Interjurisdictional Sorting and Majority Rule: An Empirical Analysis." *Econometrica* 69(2001): 1437-1466.
- Epple, D., and H. Sieg. "Estimating Equilibrium Models of Local Jurisdictions." *Journal of Political Economy* 107, no. 4(1999): 645-681.

- Geoghegan, J. "The Value of Open Space in Residential Land Use." *Land Use Policy* 19(2002): 91-98.
- Irwin, E., and N. Bockstael. "The Problem of Identifying Land Use Spillovers: Measuring the Effects of Open Space on Residential Property Values." *American Journal of Agricultural Economics* 83, no. 3(2001): 698-704.
- Irwin, E. G. "The Effects of Open Space on Residential Property Values." *Land Economics* 78, no. 4(2002): 465-480.
- Irwin, E. G., and N. E. Bockstael. "Interacting Agents, Spatial Externalities, and the Evolution of Residential Land Use Patterns." *Journal of Economic Geography* 2, no. 1(2002): 31-54.
- Kotchen, M. J., and S. Powers, M. "Explaining the Appearance and Success of Voter Referenda for Open-Space Conservation." *Journal of Environmental Economics and Management* 52(2006): 373-390.
- Kuminoff, N. "Recovering Preferences from a Dual-Market Locational Equilibrium." *Working Paper* (2007).
- McFadden, D. *Modeling the Choice of Residential Location*. Spatial Interaction Theory and Planning Models. Edited by A. Krlqvist, L. Lundqvist, and F. Snickars. Amsterdam: North Holland, 1978.
- Murphy, A. "A Dynamic Model of Housing Supply." Working Paper (2008).
- Newburn, D. A., and P. Berck. "Modeling suburban and rural-residential development beyond the urban fringe." *Land Economics* 82, no. 4(2006): 481-499.
- Palmquist, R. B. (2005) *Property Value Models*, ed. K. Gören-Mäler, and J. Vincent, vol. 2, North Holland Press.
- Poterba, J. M. "Tax Subsidies to Owner Occupied Housing: An Asset Market Approach." *Quarterly Journal of Economics* 99, no. 4(1984): 729-752.

- Rosen, S. "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition." *Journal of Political Economy* 82(1974): 34-55.
- Sieg, H., et al. "Estimating the General Equilibrium Benefits of Large Changes in Spatially Delineated Public Goods." *International Economic Review* 45(2004): 1047-1077.
- Smith, V. K., C. Poulos, and H. Kim. "Treating Open Space as an Urban Amenity." *Resource and Energy Economics* 24(2002): 107-129.
- Tiebout, C. M. "A Pure Theory of Local Expenditures." *Journal of Political Economy* 64(1956): 416-424.
- Timmins, C. "If You Cannot Take the Heat, Get out of the Cerrado Recovering the Equilibrium Amenity Cost of Nonmarginal Climate Change in Brazil\*." *Journal of Regional Science* 47, no. 1(2007).
- Timmins, C., and J. Murdock. "A Revealed Preference Approach to the Measure of Congestion in Travel Cost Models." *Journal of Environmental Economics and Management* 53(2007): 230-249.
- Topel, R., and S. Rosen. "Housing Investment in the United States." *Journal of Political Economy* 96, no. 4(1988): 718-740.
- Walsh, R. "Endogeneous Open Space Amenities in a Locational Equilibrium." *Journal of Urban Economics* 61, no. 2(2007): 319-344.

# Appendices

# Appendix A

## Data Cleaning Procedures

### A.1 Property Transactions

Housing transactions form the basis of sorting estimation as they capture the decisions made by both households and builders. The attributes associated with housing provide guidance for dividing the housing stock into sizes,  $k$ , and also provide the basis for defining a rich set of heterogeneous preferences through interactions between housing attributes and individual characteristics. The assembled transactions database is unique as it spans 17 years from 1990 through 2006 and covers a large spatial area. The length of time over which housing transactions are observed captures the rapid metropolitan growth that occurred in the late 1990s and early 2000s. Because the transactions data covers the entire 7 county metropolitan area, the growth of the Twin Cities area and the location choices made by individuals and builders are fully captured.

After cleaning the data, there are 461,795 transactions with a full set of housing attributes that are able to be geocoded using ArcGIS. These transactions are spread across

the entire study area as shown in figure 4.2 and provide ample spatial coverage over which small neighborhoods are defined. The following discussion looks at the cleaning process applied to the raw housing data followed by a discussion of the attributes contained within the data. Discussion of the data cleaning steps is broken into three components; the first removes erroneous or duplicate data, the second addresses data quality, and the third addresses data availability.

The housing transactions data was downloaded through a web based interface from Plat Research, a local Minnesota company specializing in the collection of real estate data. Their database includes all real estate transactions, not just single family transactions which are the focus of this research, as well as a large assortment of housing attribute variables obtained from assessors' offices. Because data from multiple assessors' offices is used, there is a wide range of variables and housing characteristics available across the different counties and municipalities in the study area. This results in the need for substantial data cleaning prior to using the downloaded data.

Several common problems with the original data are a result of reliance on multiple assessors' data stored in different formats with different variables. This results in many variables that are specific to a particular county and not reported or blank for other counties. In addition, some of the variables that are reported for every county, such as acreage, are reported under different variable names and measurement units for different counties. One last data issue is that duplicates are reported for transactions containing more than one grantee or grantor. Despite these shortcomings, the data appears quite accurate and contains a wide variety of housing attributes available across the entire study region.

The original downloaded data contains property transactions from all 7 metro counties spanning 1990 through 2006. The first cleaning step was to limit attention to single family residential transactions. The downloaded data contains a property type indicator that is consistent across the study area allowing division of the transactions into smaller groups based on the type of the transaction. The property types contained in the data are: residential, vacant land (non-residential), vacant land (residential), commercial, apartment, triplex, double bungalow, townhouse, farm, utility, industrial, seasonal, condominium, miscellaneous, and exempt. Restricting attention to the residential category results in a dataset containing 787,776 rows of single family residential data.

Because this data contains not only the most recent sale of a property, but also contains historical transactions for a property, some of the data observations are for sales occurring before 1990. Because these are sparse, and often appear incomplete, the second data step involves limiting the data to transactions occurring from 1990 through 2006. This further reduces the number of data rows to 719,346. One weakness in the repeat sales aspect of the data is that the housing attributes are only available for the most recent transaction. The lack of historical characteristics makes it difficult to determine the exact attributes of houses at the time of a transaction, requiring additional cleaning steps to remove the majority of remodeled homes from the sample.

Having reduced the sample to single family residential transactions spanning 1990 through 2006, the next cleaning step removes duplicate transactions from the data. Duplicates occur when there is more than one grantee or grantor and the same transaction is recorded for each grantee or grantor on a separate line. To remove these, a unique identifier

composed of the property id, sale date, and sale price is created for each transaction. By maintaining only the first occurrence of each unique identifier, a modest reduction of the sample size from 719,346 to 718,961 transactions occurs. Inspection of the remaining data shows that some houses changed ownership multiple times on the same day, but for dramatically different prices. It is possible that this represents multiple parties financing the same house, or splitting the transaction into two components representing land and buildings. Rather than attempt to guess at what the correct transaction price should be, all transactions where the same house sold on the same day for different prices are removed. This lowers the number of transactions from 718,961 to 683,531.

The above cleaning steps result in a fairly clean dataset containing only single family residential transactions, spanning the years 1990 through 2006, and free of duplicates. The remainder of the data cleaning steps focus on data quality issues as opposed to data validity issues as discussed above. These quality issues involve removing houses likely to have undergone significant structural changes, land transactions to builders that are included as a repeat sale of an existing house, and removal of transactions that do not contain enough structural characteristics to be used in estimation.

The first quality issue involves transactions where the sale price fell by more than 20% between sales. Because housing prices generally rose during the entire study period, it is unlikely that an arm's length transaction could occur at such a discounted price if the housing quality remained comparable between sales. Potential reasons for this magnitude of a price drop include foreclosure, condemnation, and other significant quality degradation of the property. Rather than subjectively remove the single sale in question, all sales from a

property experiencing such a dramatic price drop are removed from the data set. The decision to remove all sales may seem excessive, but typically houses experiencing this form of downward price fluctuation seem erratic across all sales and change ownership much more frequently than other houses in the sample. The removal of these properties reduces the number of transactions from 683,531 to 663,001; a reduction of 20,530 transactions.

The next and more challenging cleaning step involves identifying an initial land purchase by a builder and separating that purchase from the subsequent transactions that include both the land and constructed house. This cleaning step is accomplished in several steps based on grantee name as well as price appreciation in the years immediately following construction. The first step in identifying these sales involves matching grantee names with names likely to be construction companies for transactions occurring in the same year as the house was built as indicated by the year built variable. In addition to restricting attention to the year built, only properties selling for less than \$150,000 are removed as most land purchases are below this price. The matching process searches grantee names for any of the following keywords: “homes”, “construction”, “builder”, “development”, “llc”, “inc”, “corp”, “const”, “properties”, “develop”, and “investm.” The matching criterion identifies 27,800 transactions which are subsequently removed, reducing the sample size to 635,201.

Because of missing data, not all transactions contain information on grantee name. For this reason, a second attempt at identifying and removing builder purchases of land is necessary. This step makes use of the repeat sales data identifying any transactions involving houses built after 1988 in which a 40% annually compounded appreciation rate is observed. The majority of these transactions occur in the first year after a property was built as

indicated by the year built variable. Removing these transactions eliminates an additional 3,815 transactions bringing the number of remaining transactions down to 631,376.

The final data quality issue involves identifying houses whose structural characteristics are likely to have changed significantly over time. These changes are likely the result of remodeling either by individuals or private companies investing in real estate. Given the later, the same matching criteria used to identify builders based on grantee names is applied to houses older than 1 year. Also, similar to the builder sales, any houses built prior to 1989 that appreciate at a compounded annual rate of greater than 40% between sales are identified and removed. The last set of transactions identified as remodeled homes makes use of a year remodeled variable that exists for a small sample of the transactions. Because the data only contains the current characteristics of a property, it is important to remove all sales occurring prior to the remodeling date. Sales occurring after the remodeling date should contain accurate housing characteristics. Removal of sales occurring prior to remodeling removes an additional 30,401 transactions leaving a total of 601,023 usable transactions.

In theory, all 601,203 transactions remaining are suitable for use during estimation; however, no attention to the availability of data attributes or geocoding has been addressed. Addressing availability of attributes for a given transaction is the first step in further refining the dataset. This process removes any transactions not containing core housing attributes defined as square footage, acreage, number of bathrooms, number of bedrooms, number of stories, an indicator for the presence of a garage, year built, sale date, and sale price. In

addition to attribute availability, outliers as identified by approximately the 1<sup>st</sup> and 99<sup>th</sup> percentile are removed leaving only transactions satisfying the following criteria:

- $\$25,000 \leq \text{sale price} \leq \$1,250,000$
- $0.05 \leq \text{acreage} \leq 20$
- $1 \leq \text{number of baths} \leq 8$
- $1 \leq \text{number of bedrooms} \leq 8$
- $500 < \text{square footage} \leq 8000$
- $1 \leq \text{number of stories} \leq 6$
- $\text{age of house} \leq 120$
- $\text{sale price} / \text{square footage} \leq 500$

These steps further reduce the number of transactions to 467,909.

Closely related to the housing transactions data is the ability to locate the individual houses precisely in space. Through the acquisition of parcel maps for the entire 7 county metro area, a one to one correspondence between housing transactions and their location is possible. In comparison to an approximate mapping using ArcGIS geocoding via street address, this precise location facilitates the creation of “view” or “on” variables that are discussed in the price index creation section of the estimation chapter. A second important benefit of the precise spatial geocoding is the ability to accurately overlay census data on the individual transactions.

Given the importance of geocoding, the penultimate reduction in the number of property transactions results from an inability to match a unique identifier composed of county code and property id to the same code in the parcel level data. The parcel level data is

current as of April 2007, so the most likely cause of the inability to match properties is due to keying error either in the original property data or in the parcel data. After removing the relatively small number of transactions whose properties are not matched through the geocoding process, the resulting data set contains 465,227 transactions. This number is further reduced to 461,795 after removing transactions for which there are problems matching the spatial location of transactions with other sources of data. The final set of 461,795 single-family detached housing transactions serve as the basis for the empirical work undertaken in the current research.

#### **A.1.1 Data availability summary statistics**

To provide a sense of spatial coverage, table A.1 shows transactions broken down by year and county. The counties of Hennepin and Ramsey contain the cities of Minneapolis and St. Paul, respectively. It is not surprising that over half of the total transactions occur in these counties as they are the oldest portion of the Twin Cities area and are the most heavily developed. Looking at the remaining 5 counties, an increasing number of transactions are observed over time. This trend reflects the growth of the metropolitan area and underscores the rapid expansion of the metropolitan area. The growth and expansion of the area over time is even more evident when looking at land use in the next section.

Focusing on the number of new houses constructed provides additional insight into the growth of the metropolitan area. Table A.2 uses the year built for each transaction as an indicator for houses constructed. Looking at this data raises several questions about the nature of the housing data. First, there are more houses constructed than sold in particular

counties, which could occur for several reasons. The first is that although repeat sales data is included in the transactions database, it is possible some repeat sales are not included which would reduce the number of sales in earlier years of the data but have no effect on the year built variable. A second reason this occurs is due to the way assessors coded the year built variable. Just because a house is built in a particular year does not imply it sold in the same year. Thus, there are some properties whose year built is in an earlier year than the first recorded sale. A third reason there are more sales than homes built is due in part to a modeling assumption that all transactions are treated as independent and consideration of repeat sales is not explicitly accounted for. The treatment of each transaction as a separate house results in multiple entries for year built from a single house if multiple sales of the house occur.<sup>17</sup>

On the other end of the time scale, one should note that there are fewer new homes built in 2006 than in previous years. This is potentially the result of slow recording of new homes by assessors' offices or the data vendor. Another explanation is that sales observed in 2006, of which there are many, are primarily of existing homes due to the lag between construction and sale. The year 2006 is kept in the dataset as there are a large number of sales for the year as shown in table A.2, and there is no reason to believe a systematic exclusion of new home sales exists in the data that would introduce bias during estimation. This is particularly true in light of the assumption that time is treated as exogenous to the decision makers.

---

<sup>17</sup> For supply estimation, only the first sale of a property is included resulting in a smaller dataset than indicated by this table.

One final data aspect to examine involves the assumption that all transactions are separate houses and the decision not to treat repeat sales differently from other sales. This assumption does not preclude analysis based on changing land use and in fact prevents problems arising by not knowing if a particular sale is the first sale for a property after construction or if the sale is simply the first occurrence in the data. Table A.3 provides a breakdown of sales and the corresponding properties history of sales for each year. Approximately 33% of transactions in the data are for properties that have appeared as sales in earlier years. Despite this, the majority of sales are either the first or second occurrence of the property in the data; but there are several properties that have appeared up to 7 times over the course of the 17 years of data. Given the long time period of data, these numbers seem reasonable.

### **A.1.2 Housing attribute summary statistics**

Discussion of housing transactions has so far focused on data cleaning and data availability. Perhaps the most important feature of the housing transactions data is the rich heterogeneity contained within the housing attributes. This heterogeneity provides the basis for defining a rich set of household and builder preferences and forms the basis for the size dimension of the choice set. To provide a sense of the housing stock, table A.4 provides summary statistics for the main housing attributes. The average house in the sample is approximately 1750 square feet with 3 bedrooms, 2 baths and is located on a little less than half an acre. One additional thing to note from this table is that age is calculated as the year of existence a

house is in, meaning that a newly built house is in its first year so the age of that house is one, not zero.

Providing a more detailed look at selected housing characteristics, figure A.1 provides histograms for the housing attributes square feet, number of bathrooms, number of bedrooms, acreage, year built, and price. The histograms associated with these attributes appear fairly smooth and positively skewed. The histogram for year built shows that a large percentage of houses were built from 1980 onward reflecting the rapid growth of the metropolitan area over the past 25 years. As might be expected, the distribution of bedrooms and bathrooms appears normally distributed while square footage takes on a log-normal appearance. The difference in distribution implies that larger houses do not necessarily have significantly more rooms than smaller houses, but rather that the rooms in larger houses are often bigger than in smaller houses.

The acreage histogram appears bimodal reflecting a tradeoff between house size and land. To save costs, smaller houses are often associated with very small lots. In addition, the acreage histogram has a very positively skewed tail indicating that properties with large lots are available in the choice set. The properties in the tail primarily occur in rural areas outside the central metropolitan area and are often associated with homesteads being sold from farmlands. The diversity of land types from dense inner city development to open farm homesteads provides a large degree of heterogeneity across housing types over which the choice set facing both households and builders is defined.

## **A.2 Land Transactions**

Land transactions are obtained from Plat Research through several methods. The most direct source of land transactions are those flagged as vacant land sales containing both residential and non-residential sales. The original vacant land sales data contained 6,749 transactions. The second source of land transactions is the residential transactions data described in the previous section. The identification of the initial purchase of land by builders adds 31,615 transactions to the initial set of land transactions.

To prepare these transactions for estimation, several cleaning steps were required. The first cleaning step involved removing duplicates and any transactions not containing positive values for price or acreage. After removing these transactions, there were 37,293 remaining transactions. The next cleaning step removed outliers identified as any transaction in the top 5% or bottom 5% of price per acre. The cutoffs used for price per acre were \$20,000 and \$700,000. This step removed an additional 3,629 transactions resulting in a set of 33,664 land transactions.

The remaining transactions were geocoded at the parcel level by the “pin” attribute. It was possible to match 31,800 of the transactions to individual parcels. These transactions serve as the basis for land price index estimation. Figure 4.7 shows the distribution of the land transactions across the 7 county metro area. Figure A.2 plots a histogram of price per acre showing considerable variation in prices across the Twin Cities area.

### **A.3 Land Use**

Detailed land use data facilitates analysis of open space by providing snapshots of the local amenities provided by open space and the expansion of the urban/rural fringe at several points in time. In particular, the ability to observe the state of land use at multiple points in time directly, without having to work backwards and construct the landscape from current data, is a part of what makes this study area and data set unique. To capture a large number of types of open space, multiple land use sources are combined with the primary source being a detailed set of land use data covering the entire 7 county metropolitan area developed by the Metropolitan Council, a local committee made up of elected representatives from across the 7 counties. In addition to this extremely spatially resolute data, a series of more specialized datasets are used to further tease out land use categories that are not explicitly separated in the Metropolitan Council's land use files.

Obtaining detailed land use data is important for both the demand and supply sides of the housing market. On the demand side, the desirability of a particular community is driven in large part by the surrounding land use and the amenities provided by various types of land use. Moreover, identifying open space such as agricultural land plays an important role in determining the location of new housing supply as agricultural land provides the majority of new land for development. This section discusses the various data sources and how they are integrated to identify a detailed set of land use data for a wide variety of land use classifications at multiple points in time.

The Metropolitan Council's detailed land use data forms the basis for all further land use analysis as it is the most detailed, and likely the most accurate, data set available for this

area. To generate the land use data included in this dataset, local governments manually interpreted aerial photographs and combined those interpretations with property data from assessors' offices as well as political boundary data. As a result of the local involvement with the data, this data is quite accurate and appears to match the parcel level residential transactions data closely. Over the 17 year study period, four versions of land use data are available corresponding to the years 1990, 1997, 2000, and 2005. With each update, the categories of land use have become finer, thus requiring aggregation of categories in more recent years in order to generate comparable categories to earlier years. A weakness of this data is that although land use categories have become more refined over time, the categories involving open space are not as refined as one would hope. For example, large nature preserves and local parks are grouped within the same land use category. This deficiency is overcome by integrating external sources of land use data provided by a variety of sources with the Metropolitan Council's land use data.

The first step to render the land use data usable is to develop comparable categories across all 4 years of land use data. Combining categories from recent years to match categories in previous years identifies eight broad land use categories including: single-family residential, multi-family residential, commercial, industrial, agriculture and undeveloped, parks and recreation, water, four lane highways, and a catch all miscellaneous category. A common adjustment made to all land use measures regards the water category. In 2000, the minimum acreage classified as water was lowered from 5 acres to 3 acres. As a result, there is a substantial difference in the amount of water reported for 2000 and 2005 as compared to earlier years. To eliminate the impacts of this classification change, the 2005

water land use category is assumed constant across time. This assumption permits removal of any land contained within the 2005 water category from all previous years land use categories. To do this, the areas designated water in 2005 are overlaid on previous years and classified as water regardless of their actual designation.

The parks and recreation category contains the majority of the open space land that this research project aims to evaluate. Unfortunately, it is not possible given the land use data alone to determine the types of open space contained within this category. The remainder of this section discusses the additional data and methods involved in extracting more detailed information on the open space contained within this category. In addition to the data in the parks and recreation category, additional open space data is obtained from the agricultural and undeveloped category identified in the land use data.

The easiest category to separate from the parks and recreation grouping is the golf course category. For both the 2000 and 2005 land use data, the parks and recreation category already separates golf courses from other parks and recreation whereas the 1990 and 1997 land use data does not. Extracting the 2000 and 2005 golf course data adds an additional category for those years. To generate a comparable golf category for the earlier years, the 2000 golf course layer is overlaid on the 1997 parks and recreation category, of which golf is included. Assuming any land with the potential to be a golf course in 1997 (as indicated by inclusion in the parks and recreation category) and identified as a golf course in 2000, must have been a golf course in 1997 allows identification of a golf course category for 1997 separate from parks and recreation. The same procedure is used to obtain a golf course layer for 1990 using the 1997 golf course layer as the base.

To obtain a more refined breakdown of the land remaining in the parks and recreation category, several additional data sets are combined with the land use data. The additional data include the ESRI Parks 2006 data layer, the ESRI Cemetery 2006 data layer, and the Regional Recreation Open Space dataset developed by the Metropolitan Council. Using these datasets, categories for local parks, regional parks, cemeteries, and other protected open space are obtained. In addition to these types of open space, land enrolled in an agricultural preserve program is also separated from the agricultural and undeveloped land use category.

Overlaying the ESRI Cemetery data with the area remaining in the parks and recreation category identifies cemeteries for each time period. After removing this land from the parks and recreation layer, overlaying the ESRI Parks and regional parks datasets with the remaining parks and recreation layer identifies all protected open space identified as parks and preserves. This layer is further divided into local and regional parks by classifying any land contained in the Regional Parks dataset as “regional parks” with the remainder of the land classified as “local and county parks.” The “regional parks” category contains the following types of land use that are unable to be identified separately: natural history areas, nature centers, park reserves (regional), regional parks, scientific and natural areas, special recreation features (regional), special use (zoo), state parks, state recreation areas, wildlife management areas, and wildlife refuges. The land remaining in the original parks and recreation layer that was not identified as belonging in any additional category is considered “non-park protected open space” and contains greenways, protected wetlands, and department of natural resource owned land.

Two open space programs unique to the Twin Cities area are the Metropolitan Agricultural Preserve program and the Reinvest in Minnesota conservation program. Each of these programs was developed prior to 1990 with the goal of limiting urban sprawl and preserving open space either to protect agriculture or protect environmentally sensitive areas. These programs serve as natural experiments to evaluate two popular types of conservation: relatively large, temporary agricultural easements and smaller, permanent conservation easements. The institutional details of these programs are discussed below.

### **A.3.1 Reinvest in Minnesota**

The Reinvest in Minnesota conservation reserve program allows private individuals to voluntarily enroll sensitive environmental land in a conservation easement. When started in 1986, this program helped pioneer the idea of encouraging private conservation of sensitive environmental areas. To compensate landowners who enroll in this program, they are paid a portion of the assessed value of the enrolled land. Funding for this program is obtained through a variety of public and private methods. The primary source of funds is obtained through bonding and private donation, with vanity license plate sales also contributing to the availability of funds. Over the past 20 years, the state of Minnesota has appropriated over \$23 million dollars, the sale of vanity license plates has generated over \$3 million dollars, and private contributions have totaled over \$26 million dollars.

The primary recipients of these monies are private individuals agreeing to enroll their land in a permanent easement for purposes of conservation. Types of land eligible for enrollment include wetlands, riparian lands, marginal cropland, pastured hillsides, and

important groundwater areas. While the state controls the easement, the landowner is allowed to determine access to the land making this a popular program among hunters seeking to protect game lands.

One feature that stands out when looking at the coverage of this program is that it consists of relatively small and often non-contiguous lands. The scattered nature of enrolled land is due to the prohibitive costs associated with purchasing easements for large tracts of land. Table A.5 provides summary statistics illustrating the distribution of RIM sites as well as the sizes and costs associated with enrolling them. One thing that is immediately apparent is that these sites are only present in 5 of the 7 counties making up the metropolitan area. In 2005, the average acreage of an enrolled site is just over 17 acres and cost slightly over \$900 per acre to acquire.

The high costs associated with acquiring these lands prohibit acquisition of large contiguous tracts of land. When the program first started in the late 1980s and early 1990s several 20 year easements were purchased in addition to permanent easements. Table A.6 provides a breakdown of permanent vs. temporary easements. No temporary easements were purchased after 1997 and less than 10% of the easements in existence are temporary as of 2005. In addition to the switch to permanent easements, new lands are being acquired by the RIM program as additional monies are acquired, highlighting the continued expansion of permanent easements in the study area despite increasing land prices.

### **A.3.2 Metropolitan Agricultural Preserve Program**

Enrolling significantly more land than the Reinvest in Minnesota program, the Metropolitan Agricultural Preserve program allows landowners owning more than 40 acres of land to temporarily enroll their land in a preservation program in return for reduced tax loads. Started in 1980, the goal of this program is to limit development and sprawl into agricultural lands while providing incentives for agricultural investment. By enrolling land, the landowner agrees to limit development to a maximum of one dwelling per 40 acres. In contrast to the RIM program, landowners may opt out of this program at any time but must wait 8 years before developing enrolled land. Because enrollment is temporary, the landowner does not give up the total option value of the land, resulting in a program capable of preserving large tracts of land relatively inexpensively.

The primary source of enrollment data for this program is a GIS layer depicting enrolled parcels. To more accurately characterize the preserved land, this layer is overlaid on the land use data and any land not categorized as agriculture is removed from the agricultural preserve data. The process of removing non-agricultural land from the agricultural preserve layer primarily removes homesteads and other dwellings located within enrolled parcels. In addition to providing a list of enrolled parcels, a date signifying enrollment is present for all but the counties of Anoka and Hennepin. The enrollment date allows division of the agricultural preserve data into distinct time periods corresponding to the land use periods of 1990, 1997, 2000, and 2005. For Anoka and Hennepin counties it is assumed that all enrolled parcels corresponding to agricultural land, as determined from the land use data, are enrolled in the agricultural preserve program. This assumption implies that all parcels

presently enrolled for Hennepin and Anoka counties were enrolled prior to 1990. This assumption likely overstates the actual level of enrollment in the early time periods.

As of 2005, 189,200 acres of land were enrolled according to a report on the status of the agricultural preserve program. Of these reported acres, 168,068 acres are identified by the spatial GIS layer after removing homesteads and other non-agricultural land. The discrepancy in acres enrolled likely indicates that some lands enrolled have not been geocoded and are thus left out of the data used in the current analysis. Enrollment in this program results in over \$290,000 in reduced annual taxes spread over the enrolled farms. The source of the tax savings, called conservation credits, results from taxing agricultural lands as if they were located away from the rapidly developing metropolitan area. Changing the tax rate for enrolled land results in tax savings because demand for land near the metropolitan area is larger than that for more rural land in other parts of the state resulting in higher assessed values and taxes for urban areas near the metropolitan area. To more appropriately value agricultural land for agricultural use, the prior year's statewide tax rate for non-urban townships is multiplied by 105% and used as the basis for the tax rate as opposed to the local tax rate. Landowners are subject to the lower of the average tax rate or their local tax rate with a minimum rate of \$1.50 per acre. Up until the late 1990's, this resulted in large reductions in tax rates for enrolled land but has gradually diminished as tax rates in non-urban areas have increased.

To compensate local communities for the loss of tax revenue resulting from lower tax rates for enrolled land, all mortgage registrations and deed transfers include a \$5.00 fee of which half is given to counties directly to compensate them and the other half is placed in a

general conservation fund to cover any shortages a particular county may have. If a county does not use all of the monies allotted to it, it is free to spend the excess money on other conservation planning or programs. For all but Carver County, the \$2.50 share received by the county is more than enough to cover the tax revenue loss. Despite the tax savings, over the past 5 years many landowners have signaled they wish to exit the program and have started the 8 year waiting period. However, new enrollment in areas further from the metropolitan center is occurring, resulting in a fairly steady number of enrolled acres.

Table A.7 uses the spatial GIS layer and the associated enrollment year to calculate the number of parcels and acres enrolled for the four land use years of 1990, 1997, 2000, and 2005. Ramsey County is the only county with no enrolled acres as that county was heavily developed prior to the adoption of the agricultural preserve program. In comparison with the enrollment of slightly over 2,100 acres in the RIM program, the agricultural preserve program has enrolled over 168,000 acres for approximately the same cost when adding up tax savings over the years 1990 to 2005. Thus, one of the main goals of enrolling large tracts of agricultural land at a relatively low cost has been achieved but at the cost of allowing future development as landowners opt out of the program.

#### **A.4 Cost Characteristics**

Cost characteristics are obtained from the NRCS's SSURGO soil database and from the Metropolitan Council's urban services data. The primary use of these characteristics is to provide approximations to building costs influencing the location of new housing supply across the metropolitan area. Because these variables are assumed to affect builders

exclusively, the relevant scope of land for which these variables are calculated consists of undeveloped land not contained in conservation easements or otherwise preserved. Figure 4.10 shows the available undeveloped land for each of the 4 time periods used in estimation.

To determine neighborhood cost characteristics, I overlay each of the available land layers on the SSURGO spatial layers and then calculate acreages for different characteristics within each block group. This is the same process used for land use data except the data is clipped by the developable land layers prior to calculating percentages of land. The following data are obtained as cost characteristics: drainage class, non-irrigated capability, suitability for dwelling without basement, slope, and metropolitan urban services area. Using this data, I calculate the percentage of land in each block group falling within a particular category as described below.

The drainage data contains 7 categories ranging from “very poorly drained” to “excessively drained” indicating the natural drainage of the soil. I have constructed three variables based on this data which are poorly drained, moderately drained, and well drained. During estimation, I include the poorly drained variable which is comprised of the lowest three drainage classes. Figure A.3 identifies areas categorized as poor drainage.

The non-irrigated agricultural data represents the potential agricultural capability of the land and is reported in eight levels numbered 1 through 8 where lower numbers represent better production capability. Using this data, I have constructed 4 variables and include the poor production variable, categories 7 and 8, in estimation. Figure A.4 shows the areas considered poor production.

Suitability for dwelling without basement determines whether the soil is suitable for building. This measure reflects the ability of the soil to support a foundation, whether the soil is corrosive to concrete, as well as other variables impacting development. The data is reported in three classes: “not limited,” “somewhat limited,” and “very limited.” I include the “very limited” category in estimation. Figure A.5 highlights the very limited category.

The final SSURGO data used during estimation is slope data. Slopes are reported in ½ percent increments. From this data, I create variables for low slope ( $\leq 5\%$ ), medium slope ( $>5\%$ ,  $\leq 10\%$ ), high slope ( $<10\%$ ,  $\leq 15\%$ ), and very high slope ( $> 15\%$ ). I include very high slope in estimation and these areas are shown in figure A.6.

Data on the location of metropolitan services is used to identify the location of sewer service provided by local municipalities. This data is provided by the Metropolitan Council and provides data on the extent of urban services for the 4 years 1995, 1997, 2000, and 2005. There is very little difference between the 1995 and 1997 layer and therefore I use the 1995 layer as the basis for the 1990-1996 time period. For all other time periods, I use the corresponding year. Using this data, I calculate the percent of undeveloped land in each block group that falls within the metropolitan services area.

**Table A.1: Transactions by Year and County**

Year	County							Total
	Anoka	Carver	Dakota	Hennepin	Ramsey	Scott	Washington	
1990	10	440	3417	10344	1752	674	2272	18909
1991	2	491	4108	10739	2115	545	205	18205
1992	3	791	4537	12479	4077	755	1936	24578
1993	501	859	5578	12927	4490	927	2195	27477
1994	217	775	4585	10290	4274	932	1667	22740
1995	545	834	3970	10444	3919	888	2092	22692
1996	183	953	4517	11870	3711	1246	3263	25743
1997	200	998	4830	11999	4398	1152	2991	26568
1998	197	1383	5768	14294	5304	1541	4090	32577
1999	182	1341	5631	13153	4841	1654	3498	30300
2000	830	620	2510	12274	1931	964	2398	21527
2001	3761	894	3771	10293	4772	1680	2944	28115
2002	3963	1315	5555	12925	5439	1909	3486	34592
2003	4131	1465	5911	13195	5586	2143	3743	36174
2004	4255	1353	5760	13231	5089	2108	3667	35463
2005	3749	1182	4745	12580	5550	1848	3838	33492
2006	2689	872	3519	10513	4650	1118	2714	26075
All	25418	16566	78712	203550	71898	22084	46999	465227

**Table A.2: Sales and New Construction by Year**

Year	# of Sales	# Built in Year Sold
1990	18909	8941
1991	18205	7857
1992	24578	9891
1993	27477	9867
1994	22740	8630
1995	22692	7284
1996	25743	8018
1997	26568	7075
1998	32577	7995
1999	30300	8211
2000	21527	5940
2001	28115	6525
2002	34592	5968
2003	36174	5868
2004	35463	4242
2005	33492	2110
2006	26075	63
All	465227	114485

**Table A.3: Repeat Sales Occurring by Year (as observed in the data)**

Sale Year	Sale Number							Total
	1	2	3	5	5	6	7	
1990	18756	153	.	.	.	.	.	18909
1991	17724	476	5	.	.	.	.	18205
1992	23104	1432	41	1	.	.	.	24578
1993	24400	2976	99	2	.	.	.	27477
1994	18969	3570	188	13	.	.	.	22740
1995	17792	4527	358	14	1	.	.	22692
1996	18675	6367	666	34	1	.	.	25743
1997	18028	7405	1042	85	8	.	.	26568
1998	20648	9932	1814	170	13	.	.	32577
1999	18245	9568	2216	252	17	2	.	30300
2000	13173	6094	1926	303	27	4	.	21527
2001	17141	7859	2649	421	42	2	1	28115
2002	19001	10792	3894	788	104	13	.	34592
2003	19082	11460	4445	1026	145	16	.	36174
2004	17746	11511	4712	1251	201	36	6	35463
2005	14780	11543	5273	1546	319	28	3	33492
2006	10109	9530	4589	1452	326	63	6	26075
All	307373	115195	33917	7358	1204	164	16	465227

**Table A.4: Summary Statistics for Housing Characteristics by Time Period and Size\***

Time Period	Size	Square Feet		# of Bedrooms		# of Baths		# of Stories		Acreage		Age at Sale		Presence of Garage		Price (adjusted by	
		Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean	Std	Mean (\$)	Std (\$)
1990-1996	Small	1024	150	2.73	0.77	1.44	0.50	1.17	0.29	0.26	0.53	42	26	0.91	0.28	111,637.70	36,386.35
	Medium	1520	169	3.16	0.74	1.77	0.58	1.32	0.36	0.36	0.83	36	27	0.91	0.29	139,419.20	49,276.58
	Large	2661	801	3.79	0.85	2.44	0.81	1.65	0.40	0.64	1.32	21	26	0.89	0.32	226,294.20	128,261.90
	All	1765	848	3.25	0.90	1.90	0.77	1.39	0.41	0.43	0.98	33	28	0.90	0.30	161,221.60	97,174.46
1997-1999	Small	1023	151	2.72	0.78	1.45	0.54	1.17	0.30	0.25	0.47	47	28	0.91	0.28	125,157.00	42,164.21
	Medium	1522	168	3.13	0.75	1.80	0.60	1.32	0.36	0.35	0.84	39	29	0.91	0.29	157,165.40	57,208.78
	Large	2703	819	3.79	0.86	2.49	0.81	1.66	0.41	0.64	1.31	22	26	0.89	0.31	260,487.20	136,443.60
	All	1795	874	3.24	0.91	1.94	0.80	1.40	0.42	0.43	0.98	35	29	0.91	0.29	184,616.20	108,452.90
2000-2004	Small	1021	147	2.75	0.78	1.46	0.54	1.19	0.29	0.30	0.63	48	28	0.90	0.30	182,005.60	50,106.26
	Medium	1514	169	3.13	0.76	1.85	0.65	1.33	0.38	0.38	0.88	40	30	0.91	0.29	227,240.80	78,512.70
	Large	2673	812	3.74	0.84	2.53	0.81	1.67	0.41	0.65	1.35	22	26	0.88	0.32	353,500.90	165,848.50
	All	1721	846	3.20	0.89	1.94	0.80	1.39	0.42	0.44	1.00	37	30	0.90	0.30	252,796.10	131,063.90
2005-2006	Small	1021	147	2.74	0.77	1.46	0.53	1.18	0.29	0.29	0.61	53	27	0.91	0.29	210,579.10	50,577.57
	Medium	1513	168	3.14	0.76	1.85	0.64	1.33	0.37	0.36	0.81	45	30	0.92	0.28	262,400.80	84,725.92
	Large	2617	759	3.73	0.85	2.53	0.82	1.66	0.41	0.60	1.25	27	28	0.89	0.31	405,195.70	177,613.60
	All	1666	792	3.17	0.89	1.91	0.79	1.37	0.41	0.41	0.92	43	30	0.91	0.29	286,567.60	139,499.50
All	Small	1022	149	2.74	0.78	1.45	0.53	1.18	0.29	0.28	0.57	47	27	0.91	0.29	152,944.40	58,864.72
	Medium	1518	169	3.14	0.75	1.81	0.62	1.33	0.37	0.36	0.85	39	29	0.91	0.29	187,437.20	81,534.53
	Large	2669	804	3.77	0.85	2.49	0.81	1.66	0.41	0.64	1.32	22	26	0.89	0.32	294,937.90	163,181.70
	All	1743	847	3.22	0.90	1.92	0.79	1.39	0.41	0.43	0.98	36	29	0.90	0.30	212,450.00	126,747.70

\* Housing characteristics are reported as the mean characteristics of all houses of a given housing type

**Table A.5: RIM Sites Located in Metro7 Counties**

Time Period	County															All		
	Carver			Dakota			Hennepin			Scott			Washington					
	Sites	Acres	Cost	Sites	Acres	Cost	Sites	Acres	Cost	Sites	Acres	Cost	Sites	Acres	Cost	Sites	Acres	Cost
1990	13	229	\$194,123	2	40	\$14,946	2	22	\$9,015	19	394	\$291,233	.	.	.	36	685	\$509,316
1997	28	383	\$328,869	7	94	\$59,754	12	193	\$268,912	35	779	\$612,404	1	8	\$4,770	83	1,457	\$1,274,708
2000	35	576	\$459,418	7	94	\$59,754	24	325	\$506,672	37	822	\$661,863	2	15	\$16,979	105	1,833	\$1,704,687
2005	38	699	\$524,191	7	94	\$59,754	29	364	\$575,512	43	939	\$729,212	2	15	\$16,979	119	2,111	\$1,905,647

**Table A.6: Breakdown of RIM Sites by Duration**

Time Period	County											
	Carver		Dakota		Hennepin		Scott		Washington		All	
	20 Years	Permanent	20 Years	Permanent	20 Years	Permanent						
1990	2	11	2	.	2	.	1	18	.	.	7	29
1997	2	26	3	4	2	10	1	34	.	1	8	75
2000	2	33	3	4	2	22	1	36	.	2	8	97
2005	2	36	3	4	2	27	1	42	.	2	8	111

**Table A.7: Distribution and Enrollment of Agricultural Preserves Over Time**

Time Period	County													
	Anoka		Carver		Dakota		Hennepin		Scott		Washington		All	
	Farms	Acres	Farms	Acres	Farms	Acres	Farms	Acres	Farms	Acres	Farms	Acres	Farms	Acres
1990	29	1,672	272	36,687	295	32,596	132	11,064	42	4,165	101	7,557	871	93,742
1997	29	1,678	353	49,391	442	50,757	132	11,054	55	5,365	105	8,027	1116	126,272
2000	29	1,761	379	54,970	472	55,079	132	11,478	62	6,460	109	8,625	1183	138,374
2005	29	1,690	536	82,979	499	56,895	132	11,045	68	6,768	114	8,692	1378	168,068

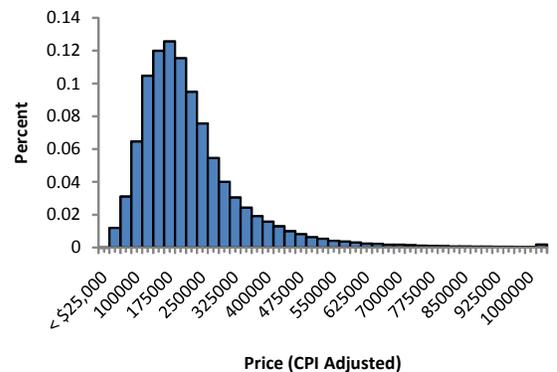
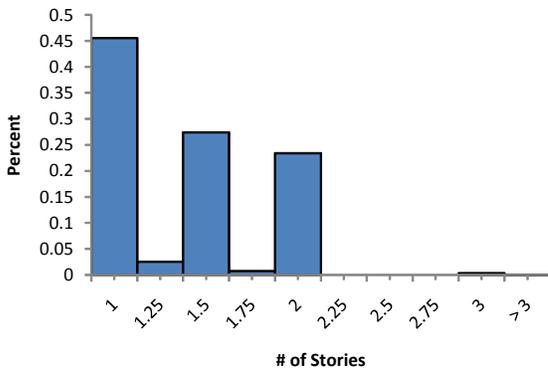
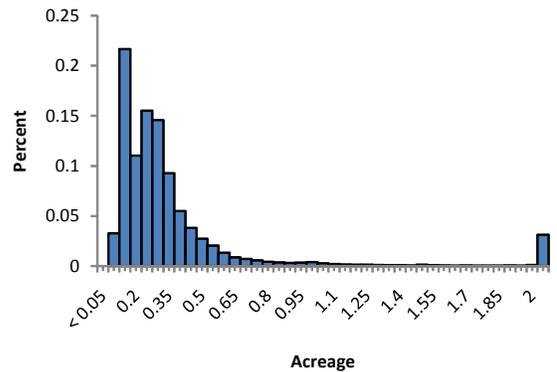
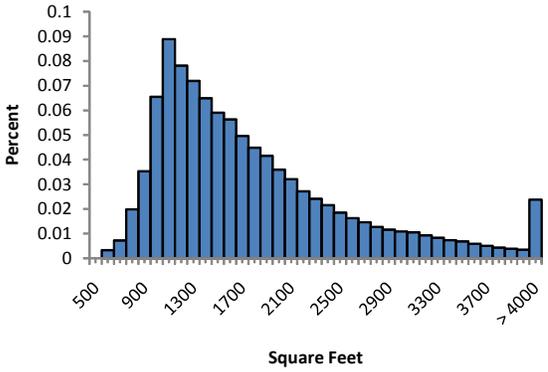
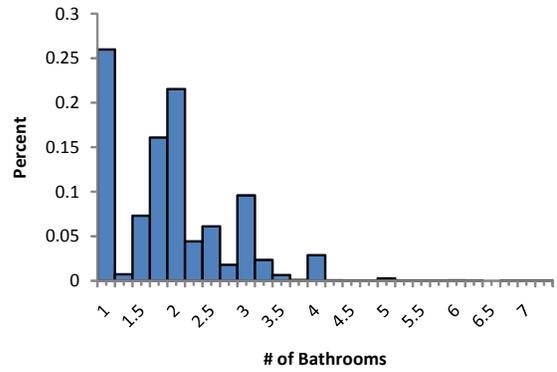
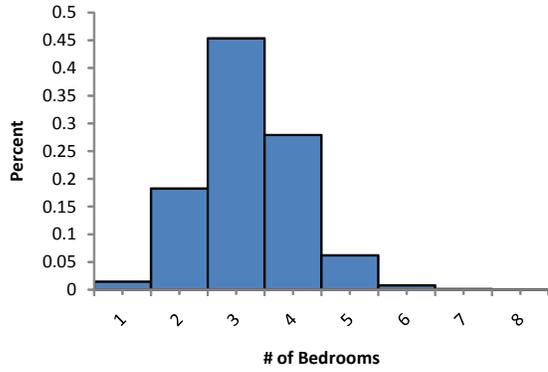
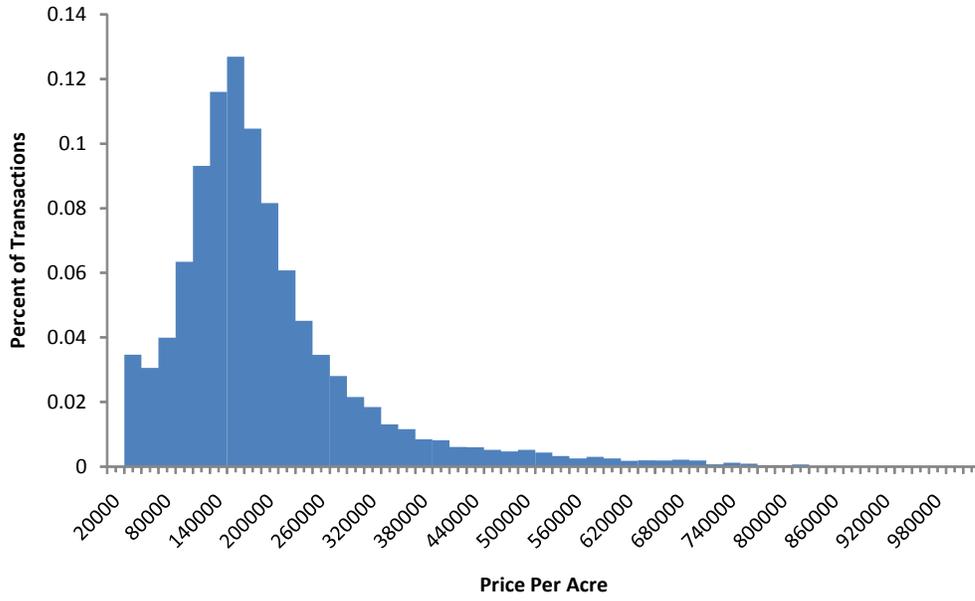
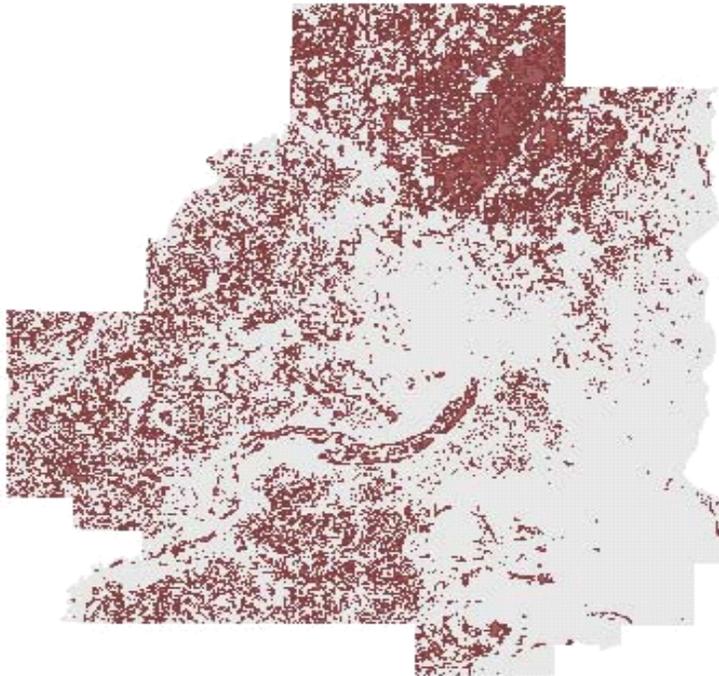


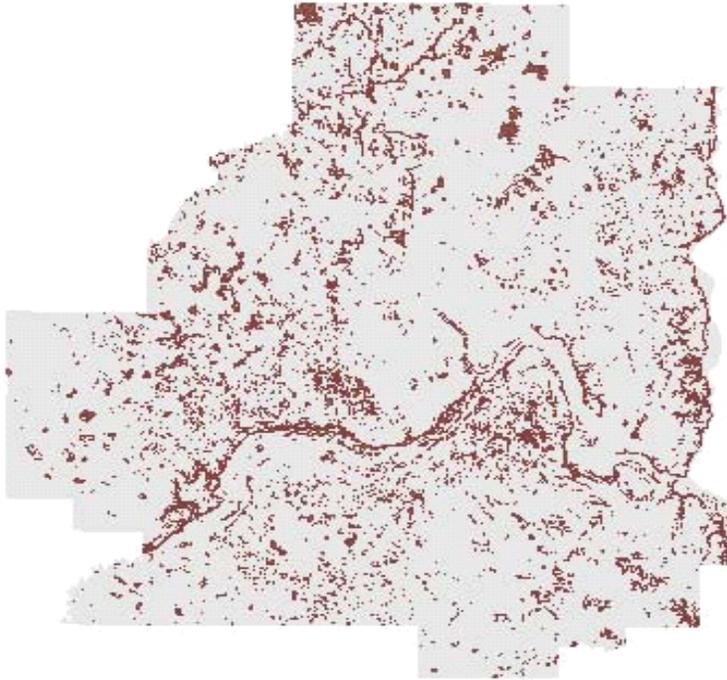
Figure A.1: Histograms of Housing Characteristics (N=461,695)



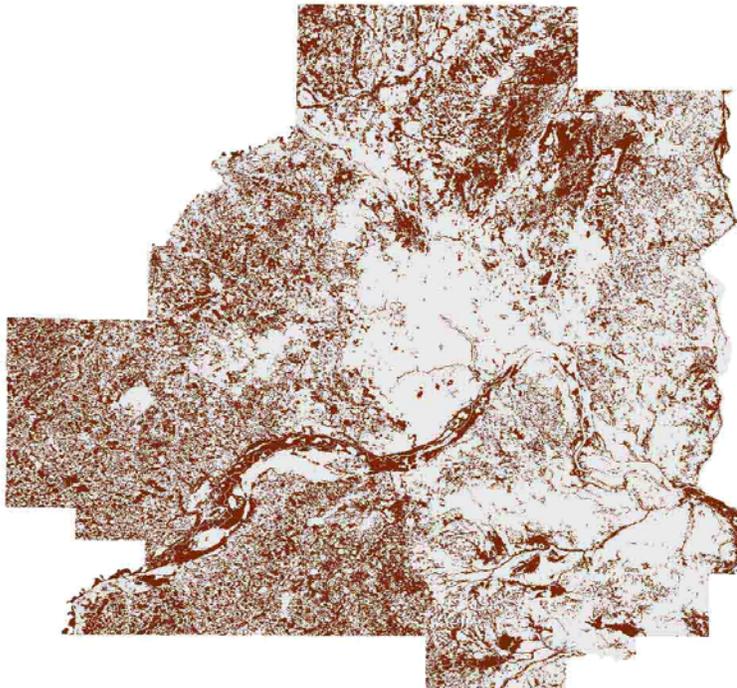
**Figure A.2: Price Per Acre for Land Transactions (2006 Dollars)**



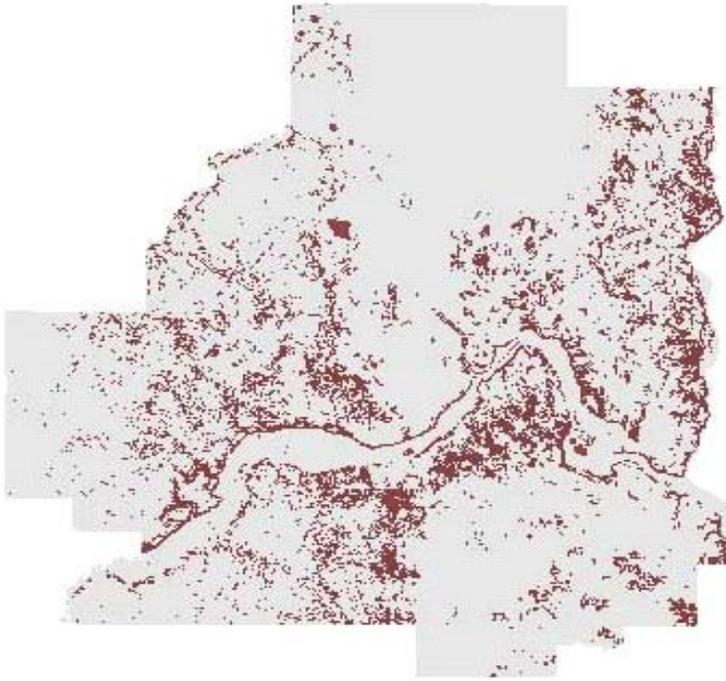
**Figure A.3: Areas of Poor Drainage**



**Figure A.4: Areas of Very Limited Agricultural Capability**



**Figure A.5: Areas of Very Limited Suitability for Dwellings**



**Figure A.6: Ares of Very High Slope**

# Appendix B

## Census Tract Spatial Units

### B.1 Introduction

An important contribution of the current research is the extremely small block group scale at which the choice set and data are defined. While the small spatial scale facilitates identification of an array of heterogeneous open space, it comes at a cost. Two drawbacks from using the block group level as the spatial scale are the lack of publicly available socio-economic data on race and income available at a smaller spatial unit needed for identification and the large number of non-chosen housing types present in supply estimation. Aggregating to a larger spatial scale alleviates both problems but comes with tradeoffs that are the subject of this appendix. Table B.1 shows the number of housing types available to both households and builders using census tracts.

The purpose of this appendix is to present comparable estimation results to those presented in the demand and supply chapters, but to use year 2000 census tracts as the spatial unit of the choice set as opposed to the block group units used in chapters 5 and 6. Recall

that the choice set facing households and builders consists of housing types broken down by location, house size, and time period. While the estimation presented here does not incorporate income or race into the model, it serves as a baseline for future work that could incorporate those elements. In addition, examining the ability to identify open space at a much coarser resolution provides important information on the spatial scale of data needed to identify a variety of heterogeneous open space.

In order to estimate the housing demand and supply models using census tracts as the spatial unit, all housing characteristics, land use characteristics, price indexes, supply characteristics, and census characteristics are recalculated at the tract level. The individual characteristics consisting of age and family size obtained from the block level data were not changed. Similarly, builder characteristics on the number of houses built are also unchanged. It would also be possible to include median income at the block group as an additional individual characteristic as multiple block groups make up a single tract thereby providing data variability needed for identification. The results presented in this appendix do not incorporate income and are thus a more direct comparison with the block group level demand estimation from chapter 5. The next two sections present estimation results followed by a discussion of the differences between using data at the block group level and the tract level.

## **B.2 Demand Estimation**

The first estimation step is the creation of price indexes at the census tract level. The results from price index estimation are reported in table B.2 and are as expected. There are virtually no qualitative differences in these results compared with the results obtained at the block

group level. As a result, attention is focused on differences in the sorting estimation results from both first and second stage estimation.

Estimation of the sorting model proceeds using the identical horizontal sorting framework used for demand estimation with block group spatial units. The only difference in included variables is the omission of percent owner occupied from both first stage and second stage estimation as a result of missing data at the tract level. By using census tracts as opposed to block groups, the total number of housing types with observable housing transactions is reduced to 7,613 from 20,444. Table B.3 provides a breakdown of housing types by time period and house size using census tracts. While the total number of housing types is roughly a third of the number of housing types using block groups, the uniformity in terms of number of housing types within each size and time period division is improved. This result is expected given that larger spatial areas encompass more housing transactions.

### **B.2.1 First stage estimation results**

First stage estimation results are presented in table B.4 and appear similar to the estimates reported using block groups. The only qualitative difference between the two sets of results is the negative and insignificant coefficient on the interaction between number of children and percent of local parks. The estimation results using block group spatial units reported a positive and significant interaction effect. This difference is likely the result of local parks losing significance when averaged over a larger spatial dimension.

Aside from the difference in local parks, there are no other differences in first stage estimation either in terms of sign or significance from using census tracts as opposed to

census block groups for the spatial unit of the choice set. However, there are differences in magnitudes as shown by examining the coefficient on the interaction between retired people and number of stories. For the block group results, this coefficient takes on an estimated value of -5.9 compared with the smaller magnitude -4.1 using census tracts. For other interaction estimates, the magnitudes are larger using tracts than block groups.

### **B.2.2 Second stage estimation results**

Ordinary least squares estimation results are reported in table B.5. As is the case with the block group based estimation, the coefficient on price takes on an unexpected positive and significant sign. The positive OLS estimate for price is likely the result of endogeneity resulting in the need to perform instrumental variables estimation. The instrumentation strategy employed is identical to that employed for demand estimation using block group spatial units.

Second stage IV results are presented in table B.6 where price is shown to take on a negative and significant sign, as expected. As is the case with estimation using block group spatial units, the reversals of sign for price as well as other variables such as golf indicate that the instrument for price appears appropriate and is correcting for the endogeneity problems inherent in naïve OLS estimation. Despite this encouraging result, many of the land use variables lose significance compared to the estimation results using block group spatial units.

The following variables become insignificant using tracts as the spatial unit of the choice set but are significant using block groups: commercial, highways, local parks, and RIM sites. In addition to the loss of significance, the coefficient on regional parks becomes

negative and significant as opposed to negative and insignificant using block group spatial units. For housing attributes, the only substantial change is a switch in sign for the coefficient on bathrooms from positive and significant to negative and significant.

Table B.7 combines both first and second stage results by reporting marginal willingness to pay values for a variety of household compositions. Compared with results from block group estimation, the most noticeable differences in open space variables are the much smaller marginal willingness to pay values for RIM conservation easements and the smaller marginal willingness to pay for regional parks. The only other substantial difference is the reversal of sign associated with the commercial land use category.

### **B.3 Supply Estimation**

A concern using the block group spatial units for supply analysis is the large number of non-chosen supply alternatives included in the choice set facing builders. Aggregating to the tract level helps alleviate the problem, but as shown in table B.1 more than 50% of the available supply alternatives remain non-chosen. Despite this, estimation at the tract level sheds important insights into the level of spatial detail required to maintain data variation needed to tease out the drivers of builder housing decisions.

As is the case for demand analysis at the tract level, the estimation strategy and variables used for supply are nearly identical to those used in chapter 6. Housing prices used in supply estimation are the same as created for tract level demand estimation. Because the original land price indexes were calculated at the tract level, no new estimation is required to

obtain new versions of those indexes. As a result, the price indexes for land discussed in chapter 3 are used during estimation discussed in the following sections.

### **B.3.1 First stage estimation results**

Table B.7 reports first stage estimation results for the supply model using census tracts. Compared with the results using block group spatial units in chapter 6, all statistically significant coefficients have the same signs, but their magnitudes are smaller. For the interaction between size of builder and limited development potential, the positive coefficient is not statistically significant. These results indicate that enough data variability exists at the census tract level to identify heterogeneous preferences for home builders.

### **B.3.2 Second stage estimation results**

Quantile regression results for the median quantile are reported in table B.8 and are similar to the estimates recovered using block group spatial units. To overcome potential endogeneity problems from retail price being correlated with unobservable components of utility, instrumental variables quantile regression is performed and the coefficient estimates for the median quantile are reported in Table B.9. The price instrument used is created using the identical process as for block group spatial unit supply estimation described in chapter 6.

Examining the second stage coefficient estimates, a loss of significance is evident for the tract results compared with the block group spatial unit results. Coefficients that are significant at the block group level but not at the tract level include distance to central business district and poor drainage. For variables that remain significant using both spatial

units, all coefficient signs remain the same including the sign on retail price and also the sign on land price.

The coefficients that are not statistically significant using tract level spatial units take on unexpected signs as compared with the coefficients reported for the block group spatial units. Both the coefficients on agricultural preserves and RIM conservation easements have positive and insignificant coefficients as compared with negative and insignificant coefficients for block group estimation. In addition, the coefficients for poor drainage, poor agricultural potential, and very high slope take on positive and insignificant signs as opposed to negative signs for block group estimation.

The implications of the different coefficient signs are evident when examining the marginal willingness to pay measures reported in table B.10. The unintuitive signs coupled with larger magnitudes results in larger willingness to pay values with the opposite sign as expected for many of the supply variables. This result indicates that aggregation to the tract level has a much greater impact on supply estimation than on demand estimation. The implication of this result is that supply estimation requires better spatially resolved data than demand and possibly explains why there exists much less work on the supply side of the housing market than on the demand side using micro-level data.

## **B.4 Conclusions**

The purpose of this appendix was to show the impacts of increasing the size of the spatial unit of the choice set on estimation results for housing demand and housing supply. While increasing the spatial dimension has several potential benefits with regards to the added

ability to use block group data for individual characteristics and the increased estimation speed resulting from a smaller choice set, there are significant drawbacks in terms of the ability to identify the effects of several land use types and supply variables.

The inability to recover statistically significant coefficients for several open space land use categories is not problematic in general, but takes away from one of the goals of the current work which is to treat open space as a heterogeneous good in order to provide policymakers with a greater understanding of the impacts from a variety of proposed open space policies. For supply, the loss of statistical significance proves more problematic in that fewer variables are available to identify supply compared with demand, making estimation more challenging. Despite this, it is still possible to estimate a very parsimonious model of supply using a subset of variables that remain significant when aggregating to the tract level.

For future work with slightly different objectives, combining open space categories into more general groupings such as permanent and temporary, public and private, or large and small would likely remove many of the identification challenges experienced on the demand side of the market while allowing estimation of models using additional individual characteristics. For supply, a more parsimonious specification would allow integration of the demand and supply sides of the housing market and permit estimation of comparable general equilibrium welfare measures to the very flexible general equilibrium welfare measures reported using block group spatial units.

**Table B.1: Housing Types at the Census Tract**

Time Period	House Size	Agent		
		Household	Builder (chosen)	Builder (all)
1990 - 1996	Small	638	267	635
	Medium	651	380	648
	Large	629	395	626
	All	1918	1042	1909
1997 - 1999	Small	621	151	618
	Medium	637	266	634
	Large	603	313	600
	All	1861	730	1852
2000 - 2004	Small	636	179	632
	Medium	654	332	650
	Large	644	389	640
	All	1934	900	1922
2005 - 2006	Small	617	44	598
	Medium	651	103	633
	Large	632	137	614
	All	1900	284	1845
All	Small	2512	641	2483
	Medium	2593	1081	2565
	Large	2508	1234	2480
	All	7613	2956	7528

**Table B.2: Price Index Regressions for 2000 Census Tracts by Time Period**

Variables	Time Period							
	1990-1996		1997-1999		2000-2004		2005-2006	
	Estimate	t-statistic	Estimate	t-statistic	Estimate	t-statistic	Estimate	t-statistic
<b>Housing Characteristics</b>								
acreage	0.022	21.670	0.027	21.760	0.031	37.130	0.042	33.820
bedrooms	0.024	21.980	0.024	18.360	0.021	22.580	0.020	15.640
square feet	0.000	87.990	0.000	75.530	0.000	99.680	0.000	76.210
stories	0.017	6.710	0.017	5.530	0.012	5.520	0.011	3.650
garage	0.107	30.130	0.139	33.600	0.114	44.300	0.067	18.380
bathrooms	0.066	44.440	0.060	33.480	0.059	47.140	0.062	35.430
age	-0.003	-57.430	-0.003	-48.490	-0.003	-65.090	-0.003	-45.190
<b>Views</b>								
park	0.032	7.340	0.034	7.100	0.022	6.420	0.014	3.150
golf	0.085	8.130	0.152	14.740	0.121	18.000	0.092	9.470
cemetery	-0.059	-3.240	-0.052	-2.320	-0.035	-2.350	-0.075	-3.350
open space (non-park)	0.031	7.270	0.023	5.200	0.013	4.450	0.011	3.230
water	0.204	46.910	0.228	42.200	0.218	55.370	0.222	38.140
<b>Year Dummy Variables</b>								
Year 1990	n/a	n/a						
Year 1991	-0.033	-10.420						
Year 1992	-0.037	-12.820						
Year 1993	-0.024	-8.510						
Year 1994	-0.002	-0.560						
Year 1995	0.012	4.120						
Year 1996	0.047	16.190						
Year 1997			n/a	n/a				
Year 1998			0.053	22.960				
Year 1999			0.148	63.340				
Year 2000					n/a	n/a		
Year 2001					0.117	50.490		
Year 2002					0.198	89.210		
Year 2003					0.273	123.460		
Year 2004					0.344	155.030		
Year 2005							n/a	n/a
Year 2006							-0.007	-4.110
<b>Model Information</b>								
N	160344		89445		155871		59567	
R <sup>2</sup>	0.9994		0.9995		0.9996		0.9997	

**Table B.3 First Stage Interaction Parameters**

Variables (Neighborhood-X-Household)	Estimate	Std Err <sup>a</sup>	t-stat
% golf-X-# of retirees	<b>8.0656</b>	0.1583	50.9666
% industrial-X-# of children	<b>-1.0804</b>	0.0449	-24.0388
% local park-X-# of children	-0.1059	0.0648	-1.6337
% regional park-X-# of working aged adults	1.7206	0.0523	32.8780
% water-X-# of retirees	<b>1.9668</b>	0.0893	22.0235
% born US-X-# of retirees	<b>5.4297</b>	0.1902	28.5404
acreage-X-# of children	<b>0.0799</b>	0.0071	11.3210
# of baths-X-household size	<b>0.4632</b>	0.0045	101.9454
# of stories-X-# of retirees	<b>-4.0763</b>	0.0283	-143.7891
# of bedrooms-X-# of children	<b>0.2802</b>	0.0079	35.6884
distance to CBD-X-# of working aged adults	<b>0.1393</b>	0.0004	317.9773

<sup>a</sup>Standard errors calculated using the inverse of numerically approximated hessian

**Table B.4: Second Stage Results Using Naïve OLS**

Variables	Estimate	Std Err	t-statistic	Variables	Estimate	Std Err	t-statistic
Constant	<b>9.8848</b>	0.5635	17.5404	% born in US	<b>-10.5160</b>	0.5218	-20.1519
% Agricultural Preserves	<b>1.6760</b>	0.7835	2.1390	% Vacant houses	<b>-20.7463</b>	1.1316	-18.3329
% Agricultural/Undeveloped	<b>-0.7309</b>	0.2039	-3.5857	Acreage	<b>-0.4135</b>	0.0436	-9.4927
% Cemetary	<b>3.3945</b>	0.9804	3.4624	Age of house	<b>-0.0070</b>	0.0015	-4.7217
% Commercial	<b>-2.7891</b>	0.5160	-5.4054	# of baths	<b>-1.6789</b>	0.0641	-26.1895
% Golf	<b>-4.4490</b>	0.7081	-6.2828	Garage	<b>1.8202</b>	0.1178	15.4545
% 4-lane highways	0.5880	0.7473	0.7869	# of stories	<b>0.5458</b>	0.0944	5.7845
% Industrial	<b>1.1795</b>	0.2941	4.0102	# of bedrooms	<b>-0.2870</b>	0.0560	-5.1246
% Open Space (non-park)	<b>2.0443</b>	0.8891	2.2993	Square feet (in 100s)	0.0117	0.0073	1.6159
% Local Parks	1.2551	0.7364	1.7045	Distance to CBD	<b>-0.1852</b>	0.0077	-24.1890
% Regional Parks	<b>-4.6645</b>	0.3759	-12.4102	Price (in \$1000s)	<b>0.0339</b>	0.0006	52.7112
% Water	-0.2360	0.3408	-0.6924				
# of RIM Sites	0.0065	0.0255	0.2541				

**Table B.5: Second Stage Estimation Results Using IV**

Variables	Estimate	Std Err	t-statistic	Variables	Estimate	Std Err	t-statistic
Constant	<b>20.6358</b>	1.4890	13.8585	% born in US	<b>-13.4029</b>	1.1480	-11.6750
% Agricultural Preserves	1.5854	1.6179	0.9799	% Vacant houses	<b>-66.0831</b>	3.8400	-17.2089
% Agricultural/Undeveloped	<b>-2.4842</b>	0.4124	-6.0238	Acreage	<b>-0.4341</b>	0.1051	-4.1291
% Cemetary	-1.2787	1.8105	-0.7063	Age of house	<b>0.0320</b>	0.0037	8.7195
% Commercial	-1.3889	1.0833	-1.2821	# of baths	<b>-0.5385</b>	0.1673	-3.2184
% Golf	<b>7.8028</b>	1.7794	4.3850	Garage	<b>0.7665</b>	0.2999	2.5559
% 4-lane highways	1.6398	1.5381	1.0661	# of stories	<b>0.9451</b>	0.1990	4.7485
% Industrial	-1.1129	0.5980	-1.8610	# of bedrooms	<b>-0.5727</b>	0.1269	-4.5121
% Open Space (non-park)	<b>26.3071</b>	2.5857	10.1742	Square feet (in 100s)	<b>0.0670</b>	0.0165	4.0655
% Local Parks	1.7893	1.4732	1.2146	Distance to CBD	<b>-0.1196</b>	0.0152	-7.8789
% Regional Parks	<b>-2.9137</b>	0.7711	-3.7788	Price (in \$1000s)	<b>-0.0675</b>	0.0062	-10.8319
% Water	<b>5.1510</b>	0.8000	6.4384				
# of RIM Sites	0.0057	0.0502	0.1137				

**Table B.6: Marginal Willingness to Pay Heterogeneity**

Household Attribute	Household Structure				
	Mean	A	B	C	D
# children	0.83	0	2	0	0
# working age	1.78	2	2	0	1
# retirees	0.22	0	0	2	1
household size	2.83	2	4	2	2
Variables (Change)	Marginal WTP (\$)				
% Agricultural Preserves (1%)	235	235	235	235	235
% Agricultural/Undeveloped (1%)	-368	-368	-368	-368	-368
% Cemetary (1%)	-189	-189	-189	-189	-189
% Commercial (1%)	-206	-206	-206	-206	-206
% Golf (1%)	1,419	1,156	1,156	3,546	2,351
% 4-lane highways (1%)	243	243	243	243	243
% Industrial (1%)	-298	-165	-485	-165	-165
% Open Space (non-park) (1%)	3,897	3,897	3,897	3,897	3,897
% Local Parks (1%)	252	265	234	265	265
% Regional Parks (1%)	22	78	78	-432	-177
% Water (1%)	827	763	763	1,346	1,054
# of RIM Sites (1)	1	1	1	1	1
% born in US (1%)	-1,809	-1,986	-1,986	-377	-1,181
% Vacant houses (1%)	-9,790	-9,790	-9,790	-9,790	-9,790
Acreage (0.1)	-545	-643	-406	-643	-643
Age of house (1)	474	474	474	474	474
# of baths (.5)	5,721	2,873	9,736	2,873	2,873
Garage (0.1)	1,136	1,136	1,136	1,136	1,136
# of stories (0.5)	358	7,001	7,001	-53,389	-23,194
# of bedrooms (1)	-5,039	-8,484	-182	-8,484	-8,484
Square feet (in 100s) (1)	993	993	993	993	993
Distance to CBD (1)	1,902	2,356	2,356	-1,772	292
Price (in \$1000s) (1)	-1,000	-1,000	-1,000	-1,000	-1,000

**Table B.7: First Stage Supply Interaction Parameters**

Variables (Characteristic-X-Builder)	Estimate	Std Err <sup>a</sup>	t-stat
% agricultural/undeveloped-X-# of houses	<b>0.2235</b>	0.0054	41.6693
% poor drainage-X-# of houses	<b>-0.1066</b>	0.0062	-17.2927
% very limited dwelling-X-# of houses	0.0005	0.0083	0.0569
% urban services-X-# of houses	<b>0.1170</b>	0.0030	38.5342

<sup>a</sup>Standard errors calculated using analytical score

**Table B.8: Second Stage Supply Results Using Naïve Quantile Regression (0.5)**

Variables	Estimate	Std Err	t-statistic
constant	<b>-25.5953</b>	1.6438	-15.5707
% Agricultural Preserves	7.7231	5.1280	1.5061
% Agricultural/Undeveloped	<b>36.8095</b>	1.8532	19.8629
% Open Space (non-park)	<b>-6.0872</b>	2.5081	-2.4270
# of RIM Sites	<b>-0.1335</b>	0.0600	-2.2245
Distance to CBD	<b>0.0535</b>	0.0223	2.3959
% Poor Drainage	0.8458	1.4987	0.5644
% Very Limited Development	-2.5020	1.6040	-1.5599
% Urban Services	-0.9462	1.5999	-0.5914
% Very Poor Ag Potential	-0.0526	0.9244	-0.0569
% Very High Slope	2.0797	1.5998	1.3000
Price (1000s)	<b>0.0161</b>	0.0014	11.4228
Land Price (1000s)	-0.0007	0.0005	-1.4222

**Table B.9: Second Stage Supply Results Using Instrumental Variables Quantile Regression (0.5)**

Variables	Estimate	Std Err	t-statistic
constant	<b>-67.6035</b>	3.1136	-21.7126
% Agricultural Preserves	12.8874	11.0457	1.1667
% Agricultural/Undeveloped	<b>55.7558</b>	2.2278	25.0275
% Open Space (non-park)	<b>-106.6441</b>	13.6446	-7.8159
# of RIM Sites	0.1142	0.1226	0.9318
Distance to CBD	0.0857	0.0709	1.2090
% Poor Drainage	4.6382	3.4691	1.3370
% Very Limited Development	<b>-16.3970</b>	3.9031	-4.2010
% Urban Services	<b>13.7594</b>	1.8618	7.3905
% Very Poor Ag Potential	5.7928	3.5673	1.6239
% Very High Slope	7.1878	4.8044	1.4961
Price (1000s)	<b>0.3594</b>	0.0240	15.0020
Land Price (1000s)	<b>-0.0392</b>	0.0070	-5.6217

**Table B.10: Marginal Willingness to Pay Heterogeneity for Builders**

Builder Attribute	Builder Size Structure				
	Mean	A	B	C	D
# of houses built	5.4059	1	50	100	1000
Variables (Change)	Marginal WTP (\$)				
% Agricultural Preserves (1%)	359	359	359	359	359
% Agricultural/Undeveloped (1%)	1,552	1,551	1,554	1,558	1,614
% Open Space (non-park) (1%)	-2,967	-2,967	-2,967	-2,967	-2,967
# of RIM Sites (1)	318	318	318	318	318
Distance to CBD (1)	238	238	238	238	238
% Poor Drainage (1%)	129	129	128	126	99
% Very Limited Development (1%)	-456	-456	-456	-456	-456
% Urban Services (1%)	383	383	384	386	415
% Very Poor Ag Potential (1%)	161	161	161	161	161
% Very High Slope (1%)	200	200	200	200	200
Price (1000s) (1)	1,000	1,000	1,000	1,000	1,000
Land Price (1000s) (1)	-109	-109	-109	-109	-109