ABSTRACT

POPE, JAREN C. Limited Attention, Asymmetric Information, and the Hedonic Model. (Under the direction of V. Kerry Smith.)

The broad objective of this research is to gauge the importance of relaxing the full information assumption in revealed preference models when decisions are made in complex, public information environments. This thesis focuses on housing markets. An information acquisition process is outlined that describes why homebuyers are often less informed than sellers for some housing attributes when they face more stringent information search and processing constraints. Adapting the hedonic model for the possibility that sellers are more informed than buyers suggests that estimates of the implicit price for a housing attribute may be attenuated towards zero if there is asymmetric information about the quantity of the attribute. The importance of the asymmetric information argument is gauged by applying the quasi-random experiment methodology to three applications involving exogenous information shocks for different housing attributes.

The first of these applications describes the impact of an airport noise disclosure on housing prices. The results indicate that the disclosure reduced housing prices near the airport by 2-3 percent. This suggests that an estimate of the implicit price for airport noise would have been attenuated towards zero by approximately 36 percent prior to the disclosure. The second application described the impact of a flood plain disclosure on housing prices. The results indicate that the disclosure reduced housing prices in designated flood zones by approximately 4 percent. Thus this application reconfirms the results from the airport noise application and the conceptual framework.

The third application describes the impact of information shocks related to the locations of registered sex offenders on housing prices. The results indicate that housing prices fall by 2 percent within one tenth of a mile of a registered sex offender when a sex offender moves into a neighborhood. However, this impact was not affected by increased media attention surrounding two child-abductions committed by registered sex offenders near the study area. These results are somewhat less conclusive about the role of asymmetric information on the estimated implicit price for proximity to sex offenders.

Limited Attention, Asymmetric Information, and the Hedonic Model

by

JAREN C. POPE

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

2006

APPROVED BY:

_____
V. Kerry Smith
(Chair of the Advisory Committee)

_____
Raymond B. Palmquist

_____
Daniel J. Phaneuf

_____
Walter N. Thurman

DEDICATION

This dissertation has been written in honor of my wonderful parents, my two lovely daughters, and my sweet and patient wife who has provided me the encouragement and support to complete this work.

BIOGRAPHY

Jaren C. Pope was born in Provo, Utah on June 18, 1978 to C. Arden Pope and Ronda G. Pope. He also lived in Iowa, Texas and Massachusetts while growing up, but returned to Utah and graduated from Springville High School in 1996. Upon completion of high school he attended Brigham Young University (BYU) for one year and then moved to Brazil for two years to serve a mission for his church. After returning from Brazil he continued his undergraduate training at BYU and received a bachelor's degree in economics with a minor in zoology in the summer of 2001. He then enrolled in the economics graduate program at North Carolina State University in the fall of 2001 and received a master's degree in economics in 2004. Upon completion of his Ph.D. he will begin his career as an assistant professor at Virginia Tech.

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the kindness, encouragement and support of many people. Foremost among these was my advisor V. Kerry Smith. The body of his academic work puts him in an elite echelon of economists. However, from my perspective what really sets him apart is his willingness to teach, mentor, and befriend those around him. I will always cherish the opportunity I had to sit in the office next to his and interact with him on a daily basis. Thank you Kerry for all the time you spent with me answering my questions, giving feedback on papers and chapters, and listening to my crazy ideas. You truly made my graduate school experience and the writing of this dissertation wonderful.

Another group of people I would like to thank are the members of my committee—Raymond B. Palmquist, Daniel J. Phaneuf, and Walter N. Thurman. It was a privilege to get to know them and work with them both on my dissertation and on other projects. I consider myself fortunate to have had each of them agree to serve on my committee. They constantly encouraged my work, but were also willing to tell me when they thought something needed to be corrected or changed. I thank each one of you.

I am grateful for the funding and support I received from the economics graduate program at North Carolina State University (NCSU), the Center for Environmental and Resource Economic Policy (CEnREP), and the Department of Housing and Urban Development (HUD). NCSU is a great school and has a fantastic economics graduate program. Many faculty members in both the economics and agricultural economics departments have provided me with excellent teaching and mentoring. CEnREP

provided an incredible collaborative research setting where undergraduate students, graduate students, post-docs, and faculty members worked together to understand interesting environmental problems.

Other graduate students, members of my church and co-workers have been good friends to my family and me. These people have made life and graduate work much more enjoyable. You know who you are and I thank you. I hesitate to start naming names for fear of missing someone. However, I would be remiss if I did not single out several people who had particularly meaningful impacts on my research and me. Jack Crawley patiently provided wise counsel on a broad range of topics and was a good friend and office companion at CEnREP. Nick Kuminoff has shared the same advisor, a similar path to dissertation completion, and we will also work together at the same university this coming year. I consider him a great friend and a bright and talented economist and look forward to continued interaction and collaboration. Carlos Carpio was my classmate and study partner through the early stages of graduate school. His quick mind and enthusiasm for economics made learning the "core" economic material easier and helped me to successfully prepare for the preliminary exams. McKay Curtis, who has a background in economics and is working on a Ph.D. in statistics, has been a good friend the past two years. His continual probing of my understanding of econometrics and the role of government in the economics framework has provided me with much thought and stimulating conversation. My brother Devin Pope, who has also been working on his Ph.D. in economics at another university, is one of my best friends. His creativity and passion for empirical research has fanned my own research flames. Thanks again to these individuals and all those that have been left unmentioned.

Perhaps the most important people to thank are my family. I have two incredible parents who have always encouraged me to do my best, to live according to correct principles, and to work to acquire an education. Much of the good and honorable in my life today is the fruit of your labors. Thank you for all your efforts to help me succeed Mom and Dad. I consider my brothers (Devin, Weston, Nolan, Bryson, Dallin and Collin) to be my friends. They reflect a wide range of ambitions, talents and attributes. I am proud to be associated with them and to be one of the "Pope boys". To my in-laws (both those who married into my family and also the family who adopted me since my marriage), I express thanks for your encouragement and support. A special thanks to "Mom-Laurel" who has more than once crossed the continent to help my family in our time of need.

Few things change your perspective and priorities like becoming a father. Nor are there many things as time consuming or challenging as raising children. However, there are also precious few things that bring as much joy as watching your children grow and develop. Savannah and Annika have blessed my life. In times of discouragement, realizing my responsibility to provide and care for them has given me the courage to press forward in both my schooling and life. I love you girls dearly.

My best friend and dearest companion who deserves much of the credit for seeing this dissertation completed is my sweet wife, Heather. Although in most ways her talents are vastly superior to my own, she has cheerfully taken a supporting role during this act in the play of our lives. Her kindness, encouragement, and unwavering confidence in my abilities has provided me with the solid foundation I needed to be successful in this endeavor. Heather, I love you and thank you from the bottom of my heart.

To have so many wonderful people to thank is a reflection of a benevolent Heavenly Father who has richly blessed me. To him I express my deepest gratitude.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

AIRPORT NOISE

FLOOD PLAINS

SEX OFFENDER LOCATIONS

LIST OF FIGURES

Page

CONCEPTUAL FRAMEWORK

AIRPORT NOISE

FLOOD PLAINS

SEX OFFENDER LOCATIONS

Chapter 1: Introduction

"*What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention.*" Herbert Simon (1971)

1.1 Thesis Motivation

In a society where accessible information is growing rapidly, human attention and cognitive effort required to process information are becoming scarcer. The amount of "publicly available" information, assumed known to decision makers in neoclassical economic models, is greater now than at any previous point in the history of modern economic analysis. This availability of information does not guarantee that it will be used. In many information-rich decision environments complex goods may significantly tax human attention and cognitive effort. The methodology of revealed preference typically relies on the assumption of optimizing behavior by economic agents that are "fully informed". This framework implies that *all* public information is reflected in the equilibrium prices of goods.

Differentiated product purchase decisions can be complex choices that may require significant information acquisition and processing. The hedonic model is an example of a revealed preference approach that accommodates the complexity of differentiated products but does not account for differences in agents' search or attention constraints. Rosen (1974) developed this model to describe the equilibrium implicit prices of the attributes of these differentiated products. Under ideal conditions, including

full information, the hedonic model suggests that an equilibrium price locus "reveals" the marginal willingness to pay (MWTP) for small changes in each attribute that distinguishes a heterogeneous good. However, it seems plausible that if the full information assumption is not satisfied, then MWTP estimates for an attribute of a differentiated product will be biased.

There has been a recent increase in government-mandated "seller" disclosure laws for differentiated products suggesting that buyers may not be fully informed. Examples include: nutritional labels, fuel economy ratings, restaurant grading systems, corporate financial reports, country of origin labeling for food, and seller provided disclosures in the housing market. Some of these disclosure laws have been created to correct asymmetric information caused by sellers being privately informed about product quality. This type of market failure has been well studied in the economic literature since Akerlof's (1970) seminal paper. However, other disclosure laws essentially re-release information that was already publicly available in a more "salient" format. Under these circumstances it is reasonable to conclude that regulators believe the public availability of this information does not assure all buyers are attentive to it.

Empirical work is surprisingly limited on the impact of programs requiring disclosure of public information to consumers in a more accessible format. Two notable exceptions include: (i) restaurant hygiene quality grades [see Jin and Leslie (2003)] and their impact on average restaurant hygiene and revenues, and (ii) nutritional labeling requirements [see Mathios (2000)] and their impact on sales of high fat salad dressings. Fung et al. (2004) provide an overview of disclosure laws and their impacts and conclude that part of the reason these and other information programs are successful in changing

behavior is that the information was "*made available at a time, place, and in a format that fit in with the way consumers...make choices as information users*." They go on to say that:

> "*Our analysis of cases suggests, however, that simply placing information in the public domain does not mean that it will be used, or used wisely. In practice information cannot be separated from its social context. Individuals and organizations simply ignore information that is costly to acquire or that lacks salience for decisions*."

The housing market is an example of a market that is currently being reshaped by disclosure laws. It is being transformed from a market where "caveat emptor" (buyer beware) ruled supreme to one where sellers are now required to provide various disclosures. Today most states have enacted some form of "seller disclosure" laws, apparently to reduce the information asymmetry between buyers and sellers.[1] Past research in the housing market has highlighted the key role of information for the interpretation of the hedonic price function. However there has not been work on the impact of sellers providing information to homebuyers. Asymmetric information has not been considered in this literature.

1.2 Thesis Objectives

This research gauges the importance of relaxing the full information assumption in revealed preference models when decisions are made in complex, public information environments. There are several objectives of the conceptual framework developed in chapter 2. First, it develops a framework that provides an explanation for why homebuyers are often less informed than sellers about some housing attributes. Second,

[1] See Lefcoe (2004) for a discussion of mandatory disclosure laws and their history.

the revealed preference approach most often used in conjunction with the housing market—the hedonic model—is adapted to allow for differences in information available to buyers and sellers. This adaptation of the hedonic model provides some suggestions as to how MWTP estimates may be affected when the full information assumption is relaxed. Finally, the chapter outlines a methodological framework for detecting the impact of asymmetric information on estimated implicit prices.

The primary objective of the three empirical applications is to identify the impact of asymmetric information on the estimated implicit prices for three different housing attributes. The goal is to find examples where: the choice involves a complex commodity; one or more attributes have publicly available information that is arguably known to sellers but may not be known to buyers without effort; and, most importantly, there is an exogenous change that triggers or induces recognition of the feature by buyers. A comparison of marginal implicit price estimates before and after an exogenous information event will provide a basis for gauging the importance of the asymmetry in knowledge between buyers and sellers.

1.3 Highlights of the Results

The information acquisition framework developed in chapter 2 suggests that buyers typically face stricter *search* and *attention* constraints for acquiring information for location specific attributes of housing than do sellers. The framework combines elements of search theory and the psychological concept of limited attention. "Search theory" assumes that a homebuyer considers all attributes of a house for which he is acquiring information. In this setting, the buyer determines the attributes that are most

important based on their contribution to the expected benefits and costs of the choice alternatives. Limited attention on the other hand suggests that consideration sets may not be full, especially for decisions involving a complex information environment. Psychological research suggests that attention is adaptive. It focuses on a subset of available information, and is drawn to "salient" or "immediate" stimuli. In combination, search and limited attention provide a basis for why some buyers likely remain uninformed about some housing attributes whose information is publicly available but costly to obtain or does not readily attract attention.

Chapter 2 adapts the hedonic model for asymmetric information between buyers and sellers using graphical descriptions of buyers' and sellers' bid and offer curves. It is argued that if sellers are informed about a location specific disamenity and a fraction of buyers are uninformed, that the implicit price of the disamenity will be attenuated towards zero. As the fraction of uninformed buyers in the market increases, it is more likely sellers will be unable to maintain a reservation price that relies on uninformed buyers' bids being insensitive to changes in the quality of the attribute.

Chapter 2 also describes how exogenous information shocks such as real estate disclosures can be treated as quasi-random experiments to identify the attenuating bias caused by asymmetric information. If market participants fully recognize publicly available information as assumed by the hedonic model, then an exogenous disclosure that re-provides that information should have no impact on housing prices. Identifying an effect provides evidence for the existence of asymmetric information.

Three information shocks provide the basis for the applications discussed in Chapters 3-5. Each application selects an attribute where it can be argued that sellers are

likely informed whereas buyers may not be. Moreover each identifies an exogenous information shock that may induce buyers to recognize and acquire pre-existing, publicly available information on the attribute in question. Although each of the application's events was selected to reflect situations where exogenous changes attract buyer attention, they vary in their informational content. If these information shocks can be treated as exogenous, and sellers are likely to have the information, then they offer the opportunity to consider the questions posed in this research.

Chapter 3 describes how an airport noise disclosure in Wake County, NC provides little new information, but rather directed buyer's attention to the noise issue. The disclosure altered the fraction of informed buyers. Using data on housing sales occurring near the airport that bracket the timing of the disclosure requirement, and controlling for confounding influences to the quasi-random experiment, results suggest that the disclosure reduces the selling price of homes in high noise zones by approximately 2-3 percent beyond the 4% impact that would have been attributed to airport noise prior to the disclosure. Therefore the noise disclosure impacted the implicit price for airport noise by approximately 36%.

Chapter 4 considers a flood plain disclosure. This disclosure influences *both* the attention and search costs of buyers. Using housing data that symmetrically bracketed the disclosure time period and after controlling for possible confounding influences to the quasi-random experiment, it appears that the disclosure reduces housing prices in FEMA designated flood zones by approximately 4 percent. Prior to the disclosure there was no detectable impact of flood zones on housing prices.

The results in chapter 5 are less conclusive about the role of asymmetric information. It focuses on the impact of publicly available information about the residential locations of sex offenders. It considers a discontinuous change in the coverage provided about the topic of registered sex offenders by the media as the result of two high-profile child abductions. This information is not a systematic information program directed at helping buyers in the housing market learn about the publicly available sex offender information, but was simply "news". Using housing data and the movements of sex offenders as exogenous neighborhood shocks, the results indicate that housing prices within one tenth of a mile of a registered sex offender are reduced by approximately 2 percent when a sex offender moves into a neighborhood. There is no evidence that an additional effect could be attributed to the information arising from media coverage.

These results provide support for the conceptual framework developed in chapter 2. They reinforce the overall argument that asymmetric information and limits in human attention can affect price equilibria in the marketplace.

Chapter 2: Conceptual Framework

2.1 Introduction

This chapter provides the conceptual framework for the three case studies. It highlights the importance of information and cognitive effort for understanding the behavioral responses that revealed preference models assume take place. In the hedonic framework buyers and sellers are taken to be fully informed about the quality of a differentiated product. Under these conditions the prices of a differentiated product will reflect any differences in observable attributes or quality of the product.[2] This research argues that the more complex the differentiated product purchase decision, the more likely the decision will entail large information and cognitive effort. Under these circumstances it is more likely that those buyers or sellers facing high search or cognitive costs will be incompletely informed even when information is "publicly available".

The housing market provides an ideal testing ground for understanding the impact of information search costs and cognitive effort on behavioral responses. Purchasing a house requires time and effort to make an informed choice. There are a host of structural and location specific characteristics to consider. There is some evidence in this market that sellers have lower search costs and are more informed about some attributes than homebuyers.[3] Sellers can acquire location specific information relevant to the house they are selling at low cost by virtue of living there. An information asymmetry between

[2] Another interpretation is given by Bajari and Benkard (2005) who assume an explicit form for utility and then estimate an aggregate unobserved heterogeneity term.
[3] An example of one such location specific disamenity where homebuyers appear less informed than sellers is landfills as pointed out in Shulze et al. (1986).

buyers and sellers is important because it affects the interpretation of estimates derived from models that rely on full information.

In the context of the housing market, the exact nature of the housing price equilibrium is difficult to identify. Using graphical descriptions of buyers' and sellers' bid and offer curves, it is argued that when sellers are informed about a location specific disamenity and a fraction of buyers are uninformed, the implicit price of the disamenity will be attenuated towards zero. As the fraction of uninformed buyers in the market increases, it is more likely sellers will set a reservation price that recognizes this incomplete information that implies buyers' bids will be insensitive to changes in the quality of the attribute.

The empirical strategy proposed in this chapter for detecting whether there is attenuating bias relies on the use of exogenous information shocks as quasi-random experiments. The price impact from real estate disclosures or other exogenous information events that re-release public information to market participants can be interpreted as the shift of the housing price equilibrium due to activities that induce agents to pay attention to the information that is available. The challenge in selecting quasi-random experiments (with a treatment and control group) is to identify situations where the choice involves one or more attributes with publicly available information that arguably is known to sellers but may not be known to buyers without effort. However, most important in this selection process is the need for an exogenous change that triggers or induces recognition of the feature by buyers. A comparison of marginal implicit prices before and after such an event provides a basis for gauging the importance of the information asymmetry. To some extent the analysis is an essay of persuasion,

accumulating diverse evidence to establish the relevance of the general point. There is no hypothesis test to determine if the "informational story" outlined in this chapter is the correct one.

The chapter develops the analysis in six sections. Section 2.2 introduces the problem of valuing the attributes describing the quality of a differentiated product and outlines the hedonic model under several different assumptions. Section 2.3 describes how search and attention costs can explain why a fraction of buyers are uninformed about publicly available information for a given attribute. In section 2.4 the hedonic model is modified to allow for asymmetric information about an attribute between buyers and sellers. Section 2.5 describes how the impact of asymmetric information on housing prices might be estimated using the logic of quasi-random experimental analysis. Finally, section 2.6 concludes the chapter.

2.2 Quality Differentiated Products and the Hedonic Model

Many products exhibit substantial heterogeneity. Houses, cars, computers and a myriad of other products are traded in ways that economists have come to describe as differentiated markets. These situations involve individual decisions selecting one variety of a type of product and then selecting an amount to consume. They do not fit the conventional "interior" solution description of choice because none of the other varieties is consumed in the time horizon relevant for the decision. The patterns are especially pronounced when the goods involve selection of a durable good such as a house or car. Because these varieties are substitutes, it is often suggested that there is a single market for different goods. This formulation implies people consider all the varieties before they

choose a type. Under these conditions there is a price schedule for these products even when the market in which they are traded is characterized as competitive. As early as the 1920's empirical analyses attempted to understand the contribution of quality to market prices of differentiated products.[4] These early analyses consisted of regressions of prices of differentiated products on quantifiable attributes of the products. This type of regression came to be known as a "hedonic price regression". Rosen (1974) is widely credited with developing the theoretical foundation for the interpretation of coefficients in properly specified hedonic price models. Since his seminal paper, the hedonic model has been extensively used in differentiated product markets including the housing market.[5]

2.2.1 The Hedonic Model with Long-Run Supply

When Rosen's hedonic model is applied to a housing choice, the framework maintains that a house along with the services conveyed by its location can be represented by a vector of observable attributes, $\mathbf{z} = \left(z^1, z^2, ... z^j\right)$. Some of these attributes describe the quality of the house. The model assumes that buyers and sellers know the prices of all houses for sale, the amount of each attribute for any given house, and cannot influence the market price schedule. It is also assumed that the market contains a "sufficiently large" number of houses, ensuring that a buyer's choice appears to have been made from a set of houses with continuously varying amounts of attributes.

[4] See Palmquist (2006) for additional background on some early examples of the use of the hedonic regression technique.
[5] The hedonic model has played a key role in informing public policy. For example, it has been used to adjust the CPI for quality change (see Abraham et al (1998)) and to estimate the value of improved air quality (see Smith and Huang (1995)).

That is, choice of a house can be represented as if it was a choice of a collection of attributes.

Buyer i maximizes utility by choosing $\mathbf{z}$ subject to his budget constraint as defined in equation (1),

(1) $$\underset{\max \mathbf{z}}{u_i} = U_i(\mathbf{z}, X) \quad \text{s.t.} \quad Y_i = X + P(\mathbf{z})$$

where $\mathbf{z}$ is defined as before, X is a composite good, Y_i is the consumer's income and $P(\mathbf{z})$ is the equilibrium price of a house with attributes $\mathbf{z}$. The relationship between the prices and the $\mathbf{z}$ is determined by the market and is exogenous to the individual's choice.

A buyer's maximum "bid" for a house would be implicitly defined by,

(3) $$U_i(Y_i - B_i, \mathbf{z}^*) = u_i^*$$

where $\mathbf{z}^*$ represents the optimally chosen quantities of all attributes and u_i^* is realized utility derived as the solution to equation (1). Inverting equation (3), a buyer's optimal bid function for a house would be defined as,

(4) $$B_i = B_i(Y_i, \mathbf{z}^*, u_i^*)$$

Equation (4) represents buyer $i's$ maximum willingness to pay (WTP) for the product.

On the supply side Rosen described the offer functions following the standard profit maximizing approach. This is done as if the commodities involved were produced over the same time horizon as the demand choices. Sellers' offer prices will depend on their profits which will be a function of the factor prices associated with producing a house with a given set of attributes. Inverting a seller's profit function one can define a similar function to (4) for sellers, that depends on profit maximization, of the following form,

(5) $$O_l = O_l(\tilde{\mathbf{z}}, \pi_l *)$$

where $\tilde{\mathbf{z}}$ represents the optimally chosen quantities of all attributes for sellers, and $\pi_l *$ represents the maximum attainable profit for seller l. When used in the context of housing, this view of the supply side of the market is best interpreted as arising from a long-run perspective.

In equilibrium the matching of buyers and sellers forms a locus of equilibrium transaction prices where the families of bid and offer curves are tangent. In a world of heterogeneity in preferences, income and production costs, the equilibrium price locus represents the upper envelopes of buyers' bids and the lower envelopes of sellers' offers. If households were identical then the equilibrium price locus would imply that the marginal price schedule revealed the marginal willingness to pay function for an attribute. Figure 2.1 provides an illustration of the hedonic price function for one attribute j with all other attributes being held constant at their optimal levels, where subscripts 1 and 2 represent different buyers and sellers and H and L correspond to high and low amounts of attribute j.

Under the assumption of an interior solution where consumers buy only one house and the price of all other goods (X) is normalized to 1, a key insight from Rosen's framework can be found in the first-order condition,

(6) $$\frac{\partial P}{\partial z^j} = \frac{\partial U_i / \partial z^j}{\partial U_i / \partial X} = \frac{\partial B_i^{\,j}}{\partial z^j} = \frac{\partial O_i^{\,j}}{\partial z^j}.$$

Equation (6) implies that the marginal price reveals the marginal willingness to pay (MWTP) or the marginal rate of substitution between the attribute and the composite

good X for buyers. Estimation of these implicit marginal prices has been referred to as the task of a "first stage" hedonic.[6]

2.2.2 The Hedonic Model applied to a Housing Market with Fixed Supply

Most hedonic analyses in the context of the housing market have assumed that the supply of homes is fixed, at least in the short-run, with the market composed primarily of the existing stock. "Sellers" in the short-run are simply owners of the existing housing stock who have decided to put their house up for sale. Sellers attempt to utility maximize just like buyers.

Similar to equations (3) and (4) for buyers, a seller's "offer" for their house would be governed by a constrained utility maximization process and implicitly defined by,

(7) $$U_l\left(Y_l + O_l, \mathbf{z}^*\right) = u_l^*$$

Inverting equation (7), a seller's optimal offer function for their house would be defined as,

(8) $$O_l = O_l\left(Y_l, \tilde{\mathbf{z}}, \tilde{u}_l\right)$$

Where Y is the seller's income and $\tilde{u}_l$ is the solution to the *sellers'* constrained utility maximization problem. Equation (8) represents the total willingness to accept (WTA) of a seller. In a competitive market, sellers viewed from this perspective are simply price takers just as buyers.

[6] Rosen also outlined a two-step method to estimate the hedonic demand and supply functions to enable estimation of non-marginal willingness to pay. However, this identification strategy has proved to be problematic. Bartik (1987), Epple (1987) and Ekeland et al (2004) discuss ways of overcoming these identification problems, but the so called "second stage" of the hedonic model remains difficult to implement in practice.

Rosen's framework does not explicitly model why buyers enter the housing market. Nor does it explain why sellers put their homes up for sale when supply is fixed. The hedonic model abstracts from the fact that buyer and seller motivations are likely influenced by various market conditions including interest rates, rental rates, life changing events, and other exogenous changes that induce realignment of an asset portfolio. It also abstracts from information search costs that will also influence buyer and sellers decisions.

2.2.3 The Hedonic Model with Incomplete Information on Prices

There are a number of reasons that the hedonic framework has proved attractive for applications in the housing market. Land is the ultimate "fixed" asset. In any particular location we cannot create more. As a result, location specific features conveyed thru a house and land purchase are capitalized. This logic is consistent with the hedonic framework. Greater access to data for housing sales has spurred interest in exploiting the capitalization logic. In addition the hedonic model provides a straight forward basis for estimating the MWTP of an attribute of a differentiated product. Nonetheless there has been concern about whether actual housing markets meet two key assumptions: 1) the range of differentiated product types are usually assumed continuously represented in the market, and 2) both buyers and sellers know all the prices in the market. Relaxing either of these assumptions changes the interpretation of what we learn from the hedonic price function.

Harding *et al.* (2003) provide an appealing treatment of the impact of relaxing assumption 1). They argue that when markets become "thin", buyers and sellers may

have some localized market power and thus can no longer be considered price takers. This condition creates a surplus area between buyers' and sellers' bid and offer curves. These authors use a graphical argument to suggest that bargaining likely ensues. Prices no longer lie on a single locus but somewhere in the area comprising the surplus to be bargained over between the bid and offer curves. Costly search for the best prices among a set of differentiated products will cause bids and offers to diverge in a similar manner.

If assumption 2) is relaxed, information about prices cannot be assumed to be widely known—learning the range of prices requires effort. From the perspective of the seller, searching for buyers and learning what they will bid for the house is costly. Conversely it is also costly for buyers to learn sellers' offers. An implication of these frictions is that potential exchange prices are not well known as implied by the hedonic price locus. For example, a seller may accept the bid from a buyer who does not have the highest WTP in the market *if the seller determines that it is optimal to stop searching for a higher WTP in the market*. Therefore the seller must determine the price at which he will accept any bid that exceeds that level. In other words he must determine an optimal "reservation offer". Buyers will also employ search stopping strategies to determine their "reservation bids".

There is a long history of search models in the context of the labor market to determine a "reservation wage".[7] There have also been some applications to the housing market.[8] The basic result of these search models for sellers in the housing market is that if the distribution of bids is known to the seller and he is risk neutral then the optimal

[7] See Lippman and McCall (1976) for a review.
[8] See Turnbull and Sirmans (1994) and Huang and Palmquist (2001).

reservation offer is implicitly defined by an equilibrium condition where marginal costs equal expected marginal benefits. That is,

$$C_l = \int_{O_l^{j*}}^{\infty} (B - O_l^*) f(B) dB, \tag{9}$$

where C_l is the incremental cost of searching for one period for seller l, B is a bid, O_l^* is the optimal reservation offer, and $f(B)$ is the probability density of potential bids. The left hand side of equation (9) is the marginal cost of search and the right hand side is the expected net benefit of continued search. The optimal reservation offer will be conditional on the distribution of potential bids, seller specific differences in search costs, and the attributes of the house.

Costly search ensures that sellers cannot find the "perfect match" and must be satisfied when a buyer is found that makes a bid that exceeds the minimum offer. Thus buyers' and sellers' reservation bids and offers can be pictured as diverging in price attribute space similar to the graphical argument made by Harding *et al.* for "thin" housing markets. In this setting, prices at which houses are transacted lie within a bounded portion of price-attribute space where the bounds are determined by the maximum reservation bids and the minimum reservation offers. Figure 2.2 shows this divergence graphically. Note that the price function is different than in Figure 2.1. Now the interaction between buyers and sellers also depends on who has more information—error is involved. In Figure 2.2 the equilibrium price function now has an error term

added to it (η). The error term represents costly search that arises due to the relaxation of the hedonic model assumption that prices are known by both buyers and sellers.[9]

2.2.4 The Hedonic Model with Incomplete Information on Attributes

Assuming that it is costly for buyers and sellers to learn *prices* is only a partial relaxation of the full information assumption in the hedonic model. The framework also assumes that buyers and sellers are fully informed about the existence and quality of *attributes*. Lack of information about attributes could also cause divergence in the bid and offer curves. Anecdotal and some formal evidence described in the next sections suggest that sellers are more informed than buyers about the existence and quality of attributes for the house they are selling. In other words, sellers "know what they have" and the primary information cost they incur is searching for information on buyers' bids, whereas buyers must not only search for sellers offers, they must also "learn what sellers have".

If information on housing attributes is asymmetric in this fashion between buyers and sellers, then sellers' actions to exploit their information advantage may impact housing prices. It appears there has been no published work considering the possible role of asymmetric information in the housing hedonic literature.[10] The next section provides

[9] In the Harding et al. model the error term would represent deviations caused by bargaining and market power. The error term also differs in interpretation from the unobserved heterogeneity in structural models such as Bajari and Benkard (2005). Here it explicitly represents differences in information rather than all possible types of differences. Later on the error term will also reflect the notion that some buyers may not adequately take into account the level or quality of an attribute.

[10] There are papers in the hedonic literature that have found that information can affect the hedonic price equilibrium. For example, Brookshire *et al.* (1985), Michaels and Smith (1990), Kask and Maani (1992), Kiel and McClain (1995), Gayer *et al.* (2000), and McCluskey and Rausser (2003). These papers however do not focus on the impact of asymmetric information.

a framework of information acquisition to provide a simple argument for why incomplete and asymmetric information may be an important feature of the housing market. This will serve as motivation for adapting the hedonic model in section 2.4 to reflect asymmetric information.

2.3 Sources of Asymmetric Information: Search and Attention Costs

It is difficult to know the information used by households at the time they purchase their homes. There is no routine mechanism for recording what they know or believe to be the attributes of the homes they purchase. As a result few economic studies have attempted to consider this issue. One exception is a study by Schulze *et al.* (1986). This study compared health risk beliefs of homeowners near a landfill site with expert judgments of the health risk and evaluated these differential risks in a hedonic model. To measure subjective health risks, Schulze *et al.* interviewed homeowners in the area of their study—neighborhoods in Southern California around a toxic landfill. An unexpected result from the survey was that only 35% of the survey respondents reported to have been aware of the landfill at the time they purchased their house even though there was significant coverage in the local media about the landfill and its potential health impacts over the relevant time period. This is in stark contrast to the 94% of survey respondents that reported to have become aware of the landfill *after buying a home in the area*.[11]

[11] In the survey, Schulze et al asked homeowners if they were aware of the landfill at the time of purchase and if they were currently aware of the landfill. Thus the percentages they report for awareness at the time of home purchase, and after having lived in the home are generated from the same sample of homeowners. Therefore the results also hinge on accurate recall of what homeowners knew at the time of home purchase.

The results of this survey illustrate the potential for large disparities of relevant property information between buyers and sellers. However it does not explain why this disparity arises in the housing market. Nor does it explain for which types of housing attributes one might expect potential information asymmetries of this sort. For example, why were so many buyers uninformed when information about the landfill and its risks had received significant coverage in the local media and could be considered publicly available? In a competitive market, information made publicly available would be immediately processed by both buyers and sellers and reflected in their housing choices. Search theory, combined with the hypothesis that buyers have limited attention, is used to provide an explanation for why information asymmetries like that in the Schulze et al. study might occur in housing markets.

2.3.1 Search Theory and Limited Attention Theory

The process of assembling information for a decision even when it is publicly available is costly; it requires time and effort. Moreover, what is often overlooked is that it requires another scarce resource—cognitive effort or attention—to identify what information should be searched for, and to process the information once it has been located. These resources associated with this effort can be thought of as fixed, or at least quasi-fixed for any given time period. Publicly available information on the other hand accumulates from period to period and has been growing at an exponential pace in modern societies. Nobel Prize laureate Herbert Simon suggested that: "*What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of*

information creates a poverty of attention."[12] A full model of information acquisition for an informationally demanding decision should consider both search and attention constraints

"Search theory" is used to explain why some economic agents are "under" informed.[13] It implicitly assumes that the individual considers all relevant objects for which he is acquiring information. For example, search theory suggests that homebuyers will identify all relevant attributes of a house and then gather information for the most important attributes.[14] In this framework agents may be ignorant about some types of information but that ignorance is thought to be rational. This is because agents *choose* not to know by employing an optimal stopping rule in their information collection process. Thus the structure of the search theory model relies on agents identifying all relevant attributes or in other words, having a "full" consideration set.

Limited attention on the other hand suggests that consideration sets may not be full, especially for decisions involving a complex information environment. Understanding the limits in humans' ability to focus attention has been one of the most studied topics by psychologists since the emergence of cognitive psychology in the late 1950's.[15] Most early work on limited attention was experimental and involved testing how humans focus attention in response to multiple audio and visual stimuli. Recent work has borrowed elements from computer science and neuroscience to attempt to

[12] See Simon (1971).
[13] For example, in section 2.2.3 search theory was used to describe how sellers weigh the benefits and costs from accepting the current buyers bid and waiting for a higher bid.
[14] As Stigler (1961) notes, the optimization process may occur in a sequential fashion by searching for a time period, then updating the value of continuing the search. However it still assumes that agents know all aspects of the market for which this sequential search technique can be applied.
[15] See Pashler (1998) for a thorough historical review of this literature.

understand the underlying brain circuitry involved in information processing and the focusing of attention. Although limited attention is far from being completely understood, the body of psychological research in this area does suggest that 1) humans must focus their attention on a subset of available information or stimuli in any given time period implying the possibility for incomplete choice sets, 2) where someone directs their attention while making a decision is adaptive and humans may choose a heuristic or employ an informal benefit-cost strategy to direct attention optimally[16], and 3) attention is drawn to information or stimuli that is "salient" or "immediate" in dimensions beyond economic benefits and costs (for example temporal, spatial, evolutionary and cultural dimensions).[17] These responses to the fact that attention effort is scarce may cause choices to appear to be "irrational' when compared to the logical process implied by a model based on search theory.

Recently economists have sought to integrate search theory and limited attention in explaining why economic agents may be uninformed about some of the attributes of a differentiated product.[18] Gabaix and Laibson (2005a) develop a model of cognitively costly search in which economic agents acquire information using approximately rational

[16] See Payne (1993) and related literature.

[17] Salience has been described by Fiske and Taylor (1984) as "stimuli that are statistically or contextually novel, negative or extreme, and relatively bright, moving, or complex". For this study this concept can be adapted so that salience can be treated as the degree to which an exogenous information shock about the quality of a housing attribute attracts buyer or seller attention. Additional discussion on what salience means in the context of this analysis will be taken up later on in this section and the following section.

[18] Other work that uses the limited attention paradigm to motivate empirical analysis includes: Mankiw and Reis (2002) and Sims (2003) who have evaluated how limited attention relates to monetary transmission mechanisms and learning, Hirshleifer and Teoh (2003) and Della Vigna and Pollet (2004) who consider the effect of limited attention on asset pricing and corporate disclosures, and Pope and Pope (2005) analyze how limited attention and sports success affect the complex decision of college choice.

benefit-cost calculations.[19] In their subsequent paper, Gabaix and Laibson (2005b) test the model using an experiment where subjects play games where they choose from a set of 8 goods with 10 attributes.

In their test, attributes are monetized and payoffs are the sum of the attributes of the chosen good. Subjects are time constrained and information is acquired about each good by clicking on the boxes in an 8X10 matrix on a computer screen. Only one box can be opened at a time. The expected value and variance of attributes (columns) decrease from left to right and the highest valued good is randomly assigned among rows. The question of interest in this game is when the subjects choose to stop revealing boxes in a given row for a single game, and when players can play multiple games, the number of boxes opened in any row before a row is chosen. Using a software program that records the clicks of subjects playing the game, Gabaix and Laibson found that their model of cognitively costly search was a good predictor of subjects stopping rules. In relation to the conceptual framework being developed in this chapter, it is important to notice that Gabaix and Laibson documented that players did stop. *None were fully informed*.

Of course Gabaix and Laibson had to simplify the decision problem in their experiment. Some important differences between their experiment and real world decisions are: (i) Monetizing the payoffs, (ii) ordering the expected value and precision of information from left to right, and (iii) letting the set of relevant attributes be deterministic which implied that consideration sets were automatically full. DeShazo and

[19] The approximation is in the form of an option value calculation of benefits where agents are partially myopic such that "At each decision point, agents act as if their next set of search operations were their last opportunity for search."

Fermo (2002) relaxed simplifications (i) and (ii) using a conjoint analysis that had been designed to assess the value of alternative types of services and infrastructure at newly created national park in Costa Rica.[20] Each subject was asked to choose a "park management scenario" with its associated attributes and price, from a set of scenarios. They then varied the complexity of this choice by 1) changing the number of scenarios, 2) changing the number of attributes and 3) changing the correlation structure of the information on attributes within a scenario. They found that as complexity increased in any of these three ways, the consistency of respondents choices decreased. In a subsequent study using the same dataset, DeShazo and Fermo (2004) found evidence that respondents direct their attention in a rationally adaptive, benefit-cost type manner, but the ordering of the scenarios from left to right matters. Scenarios further to the left seem to be more salient to respondents and receive more attention. DeShazo and Fermo also found that if attributes and scenarios are not weighted by the amount of attention spent on them, then estimates of an individual's MWTP of an attribute generated from respondents answers may be biased towards zero.

An economic decision outside the laboratory or survey setting does not provide the list of attributes to consider as did these studies. Determining the set of attributes to consider is a part of the choice process. It requires effort and attention. Thus limited attention allows for the possibility that an economic agent making a choice in a complex environment may never consider some attributes during the decision making period.

[20] The objective of conjoint analysis is to use multiple attributes and questions to determine which attribute or combination of attributes is most preferred by the survey respondent. However, the technique is not a controlled experiment like that of Gabaix and Laibson and does not have a set of payoffs for participants.

A buyer may well determine his consideration set by directing his attention to what he perceives as the most important attributes. Exogenous information shocks may also influence what is considered. As the Deshazo and Fermo (2004) study illustrates, even when all attributes are considered (since the attribute set was given in the survey), the ordering of the scenarios and attributes still influences choice by the amount of attention spent on a scenario or attribute. Exogenous information shocks (for example, a disclosure statement, media coverage, etc.) that occur during a homebuyers search process can influence choice both by causing the attribute to be considered, and by bumping the attribute up in the order in which attributes are considered.[21]

2.3.2 Search and Attention: A Unified Theoretical Approach

Acknowledging search and attention costs allows for a more unified theoretical framework to understand why some buyers may be uninformed in the housing market. A key implication of this information acquisition logic is that buyers may remain uninformed for two reasons. First they may never consider an attribute because the publicly available information never attracted their "initial attention". Second it may be their subjective expected costs are greater than the subjective expected benefits so they *choose* not to acquire additional information. This framework provides an explanation for why some types of attributes for which information is publicly available, are found to be insignificant in hedonic price regressions.[22]

[21] It is clear from the marketing literature that influencing/manipulating buyers' attention in this way is one of the primary purposes of advertising.
[22] An example of this is the study by Bui and Mayer (2003). They found no impact of polluting facilities listed on the TRI on housing prices in the state of Massachusetts and determined that this revealed that

Information acquisition when viewed through the search and attention lens provides some insights for the empirical analyses developed in chapters 3-5 of this thesis. An information shock for an important housing attribute that re-releases public information is more likely to cause uninformed buyers to become informed if the released information is salient and attracts buyers "initial attention". Furthermore the information shock is more likely to cause uninformed buyers to become informed if it reduces subjective search costs or increases buyer's subjective expected benefits from searching out additional information on the attribute.

There appears to be a conceptual basis for a nested relationship between initial attention and subjective expected benefits and costs (one must be initially attentive to form subjective expected benefits and costs from acquiring publicly available information). Nonetheless there is a difficult identification problem because most information shocks will affect both attention and search. Further discussion of this problem is left for section 2.5.3. Given the above discussion and evidence it is now assumed that buyers are more likely to be uninformed for certain housing attributes than sellers.

2.4 Hedonic Model with Asymmetric Information

As described in equation (9) of section 2.2.3, a seller's reservation price is a function of the distribution of bids and the search costs for obtaining information on a bid. Buyers' bids are also a function of the attributes of the house and therefore a seller's

---

homebuyers did not have preferences about this. An alternative hypothesis is that buyers were initially inattentive to the publicly available information.

reservation price will also depend on the attributes of his house (see equations 4 and 5 above). Equation (10) displays the marginal adjustment to the total reservation offer for a change in attribute j,

$$\frac{\partial\left[\int_{O_l*}^{\infty}(B-O_l*)\,f(B)dB\right]}{\partial z^j}-\frac{\partial C_l}{\partial z^j} \lesseqgtr 0 \tag{10}$$

where as in the initial definition the integral represents the expected net benefit from search, C_l is the cost of searching for one period for seller l, B is a buyer's bid, O_l* is the optimal reservation offer, and $f(B)$ is the probability density of bids in the market. The optimal reservation offer is conditional on the distribution of potential bids (assumed to be known to sellers), seller specific differences in search costs, and all other attributes of the house. The sign for equation (10) is indeterminate. It is not clear for a change in any given attribute how a seller's offer will change.

If buyers and sellers are fully informed about attribute j then equation (10) suggests that sellers will adjust their reservation offers according to the differences in $f(B)$ caused by the differences in the quality of attribute j. However, if some buyers are uninformed about an attribute of the house then the reservation offer will also depend on the fraction of informed buyers. To illustrate this point, consider again attribute j for which only a fraction of buyers (FI^j) are informed. Equation (10) would be rewritten as,

$$FI^{j} \cdot \frac{\partial \int_{O_l^*}^{\infty} (B_I - O_l^*) f(B_I) dB_I}{\partial z^{j}} + (1 - FI^{j}) \cdot \frac{\partial \int_{O_l^*}^{\infty} (B_{UI} - O_l^*) g(B_{UI}) dB_{UI}}{\partial z^{j}} - \frac{\partial C_l}{\partial z^{j}} \lesseqgtr 0 \quad (11)$$

Equation (11) differs from equation (10) because the seller now confronts two different probability density functions: $f(B_I{}^{j})$ for informed buyers and $g(B_{UI}{}^{j})$ for uninformed buyers. Uninformed buyers do not adjust their bids when attribute j changes. However, even though $g(B_{UI}{}^{j})$ does not change when there is a change in attribute j, adjustments to the reservation offer (determined by both terms on the LHS of (11)) will depend on the fraction of uninformed buyers who are insensitive to changes in attribute j. In this view of the housing market, a seller's reservation offer adjustment is a weighted average of the expected net benefits imposed by informed and uninformed buyers.

As the fraction of informed buyers decreases, equation (11) suggests that sellers would make smaller and smaller adjustments to their reservation offers for differences in attribute j. This arises due to increasing weight is placed on the uninformed buyers' bid distribution which is insensitive to changes in attribute j. An adaptation of figure 2.2 can be made to graphically illustrate how this situation might affect the hedonic price function. To be consistent with the applications of this thesis and later discussion, the horizontal axis in figure 2.3 has been changed so that the origin is labeled $\bar{z}$ for the low quality (major defect). Moving along the axis to the right increases the quality (or decreases the amount of "defect" towards zero).

In figure 2.3 it can be seen that uninformed buyers' bids are shown to lie along the horizontal axis for any quality level of the attribute. The maximum WTA offer

functions envelope has been shifted to also lie along this horizontal axis, tangent to the uninformed buyers' bids. This shift appears plausible if some sellers' search costs are low, so that they can always wait for an uninformed buyer to bid on their house. Equation (11) suggests that the lower the fraction of informed buyers the more likely that actual housing transaction prices will occur in the price attribute space near the horizontal axis and uninformed buyers bids. Thus a hedonic analysis that assumes full information may yield biased estimates of MWTP. The arrow from the $P(z^j \mid z*^{k \neq j}) + \eta$ equilibrium to the $P(z^j \mid FI^j, z*^{k \neq j}) + \eta$ equilibrium illustrates this possible attenuating bias in an estimate of the MWTP for attribute j. The inclusion of the FI^j in the second price equilibrium is a reminder that the hedonic price equilibrium should also be conditional on the fraction of informed buyers in the market when there is asymmetric information between buyers ands sellers.

Although the discussion above suggests that when the fraction of informed buyers for an attribute is low the hedonic price equilibrium is attenuated towards zero, the exact position of this price schedule (or even the direction in which the price schedule shifts) has not been established. Deriving the impact of information heterogeneity on equilibrium prices with a highly differentiated product is difficult and to the best of my knowledge has not been done. A paper by Fishman and Hagerty (2003) provides a game theoretic proof that as the fraction of informed buyers in a monopolistic market for a product with one attribute decreases, the implicit price of the attribute would reveal less and less about quality.

2.4.1 Impacts of Buyer Information Heterogeneity not Captured in Prices

In addition to the possibility of an attenuating impact of uninformed buyers on an attribute's implicit price, there are some secondary implications from attention and incomplete information. For example, in the traditional hedonic model sorting occurs in the housing market because of heterogeneity in buyers' income and preferences. When the full information assumption is relaxed in the hedonic model, then buyers may be sorting on information as well. Uninformed buyers are much more likely to buy "defective" properties even if they would have valued it much like informed buyers had they considered the defect before purchasing the property. If being informed is positively correlated with the intensity of preferences this may have an additional biasing impact on estimates of the MWTP for an attribute. Although, a model that develops a role for explicit sorting is beyond the scope of this thesis, it seems reasonable to expect that a policy that causes all buyers to pay attention and become informed about relevant housing attributes for which public information is available, may be welfare enhancing if implementation of the policy could be done at low cost.

A final implication from the model is that in the long run, producers of houses who are likely to be very well informed about the houses that they sell, may also take advantage of uninformed buyers. For example, if a significant fraction of buyers are uninformed about toxic landfills as in the Schulze *et al.* (1986) study, then producers will buy land near toxic landfills at a reduced price and then try to sell homes built on this land for a premium above what it would be worth in a full information environment. This type of producer action has the potential to significantly increase the amount of "defective" properties in the marketplace.

2.4.2 Impact of Housing Market Institutions

If there are significant information asymmetries in the housing market, then one would expect institutions to develop that assist buyers to become more informed. The housing market has a variety of institutions that could potentially relieve information asymmetries and thereby reduce the impact that they have on market prices. These include voluntary disclosures by sellers, seller reputation, real estate agents, mortgage lending institutions, inspections, and appraisals. These institutions are discussed below.[23]

Voluntary Disclosure and Reputation

The theoretical literature on disclosure of sellers' private information about a quality differentiated product, suggests that sellers will always voluntarily disclose if the disclosure can be performed and understood by buyers in a costless manner.[24] This argument is based on the suggestion that the highest quality sellers would willingly disclose putting pressure on lower quality sellers to disclose, until all but the lowest quality seller has an incentive to disclose.[25] This "unraveling" of disclosures causes all buyers to become informed. The unraveling result is reinforced in markets where sellers have repeat interactions with buyers and therefore have an incentive to maintain a good reputation by providing truthful disclosures.

[23] The hedonic model is typically implemented in a reduced form regression and abstracts from these types of institutional details by letting, for example, sellers embody the institutional affect of say a mortgage lender. Thus these institutions are embodied within the *application* of a hedonic price regression.
[24] Examples from this literature include Viscusi (1978), Grossman and Hart (1980), Milgrom (1981), Grossman (1981), Jovanovic (1983), Dye (1985), Shin (1994), Dye (1998) and Fishman and Hagerty (2003).
[25] The lowest quality seller is indifferent, but a sophisticated buyer can infer his quality from all the other disclosures in the market.

Quality in this literature is usually modeled as a single attribute. For this reason the assumption that disclosures are given and understood in a costless manner seems plausible. The housing market is more complex than a product with a single quality attribute. There are likely hundreds of attributes that a buyer may find consequential to the decision process. This complexity suggests that: sellers are unlikely to provide a disclosure for every attribute because disclosures involve substantial effort to make the information understandable. As a result, it becomes much more difficult for buyers to *infer* low quality when the disclosure set is not full even if they are initially attentive to all attributes.[26] In this setting a seller's disclosure or advertisement will only highlight the primary positive attributes of the house (new kitchen, close to downtown, etc.), but is unlikely to provide negative disclosures. As a result low quality or "defects" are rarely mentioned by sellers in the housing market. Under these conditions the burden to become informed falls on the buyer.

Real estate agents

Buyers could also rely on an intermediary to help them acquire information at low cost. For example, a real estate agent may perform this intermediary role through her familiarity with a market and thru directing a buyer's attention to specific issues that he might otherwise overlook. Real estate agents will have an incentive to protect reputational capital (unlike sellers who rarely sell homes and likely will never meet the same buyer twice) because they may have repeat interactions with their clients and their

[26] For example, in a for sale advertisement on a house there are only a handful of positive attributes that are included.

clients family and friends. However, real estate agents also have monetary incentives to sell homes quickly that potentially offset the reputation effects.

In a recent paper by Levitt and Syverson (2005), housing transactions that were aided by a real estate agent and transactions where real estate agents sold their own houses were compared. They found that when real estate agents sold their own houses, these houses stayed on the market longer and sold for a higher price than other comparable houses sold with the aid of a real estate agent. This result suggests that the monetary incentive of real estate agents to sell quickly may undermine their efficacy as an intermediary. In short, while intermediaries may help inattentive buyers, they are not likely to eliminate the effects of inattention entirely.

Mortgage lenders, Appraisals and Inspections

Another institution that may help protect uninformed buyers is the mortgage lender. A house is an investment to a mortgage company. It would seem that mortgage companies should have incentives to ensure that buyers do not overpay for a house. In fact, mortgage companies typically perform an appraisal on each house for which they loan money and encourage or require buyers to perform a home inspection.

However there are some perverse incentives at work in this system as well. First, many mortgages are indirectly insured by the Federal Government. If a mortgage meets Freddie Mac's or Fannie Mae's underwriting standards which includes an appraisal, then they can be sold to these two quasi-government agencies in the secondary mortgage market. This indirect insurance can induce a form of moral hazard in that lenders have less of an incentive to be diligent in correctly assessing the proper value of a house. Personal communication with a mortgage broker suggests appraisers can provide

appraisals to cover most transaction prices. Apparently some mortgage companies exert pressure to "bring in an appraisal above price".

Overall then, it is not clear that existing market institutions and disclosure incentives will eliminate information asymmetries in the marketplace. They may reduce the extent of the asymmetry. However it is also possible that, in some cases, they exacerbate the asymmetry when institutions face monetary incentives to provide incomplete information.

2.5 Testing the Impact of Asymmetric Information on Housing Prices

2.5.1 Developing an Empirical Strategy

A test of the impact of asymmetric information on housing prices could be conducted with two ideal datasets. The first dataset would be comprised of detailed housing characteristics over a period of time. The second dataset would be comprised of information on what buyers and sellers knew before a transaction occurred. With these datasets one could fully specify the hedonic model (there would be no omitted variable bias in this regression) by including the informational differences between buyers and sellers at the time of transaction. Knowing the fraction of informed buyers for an attribute would allow estimates of the impact of this fraction on the implicit price holding all other relevant attributes constant. However the compilation of these datasets is nearly impossible. Another empirical strategy would be to run a controlled experiment that would directly shift the fraction of informed buyers. However, controlled social experiments are notoriously costly to conduct. A viable alternative to these strategies is

to identify exogenous disclosures or information shocks that have occurred "naturally" in specific housing markets.[27]

2.5.2 Background on Quasi-Random Experiments

One of the key contributions of the natural or quasi-random experiment approach, in the words of Meyer (1995) is,

> *"The natural experiment approach emphasizes the importance of understanding the source of variation used to estimate key parameters. In my view, this is the primary lesson of recent work in the natural experiment mold. If one cannot experimentally control the variation one is using, one should understand its source."*

New government regulations, policy changes, natural disasters and other events can be the source of exogenous variation in information that can be used to estimate parameters of interest within the quasi-random experiment framework. With a quasi-random experiment, the researcher must document how the event is exogenous and how it provides information relevant to the test. This background "shoe-leather" component of the research is crucial in understanding the threats to the validity of the identification strategy.

There are two key steps in implementing an identification strategy that relies on a quasi-random experiment. First one must find an exogenous shock that provides variation in the explanatory variable of interest (i.e. an exogenous information shock in the housing market that causes some uninformed buyers to become informed). Then it is equally important to isolate a suitable comparison group or "untreated" group used as a

[27] See Meyer (1995), Rosenzweig and Wolpin (2000), and Angrist and Krueger (2001) for background and summaries of some of the work that has been done in this area.

baseline from which the impact of the treatment on the "treated" group can be determined. The identification strategy for natural experiments that occur at a point in time is often some variant of the following simple regression model outlined in Meyer (1995),

(12) $$P_{it}^{\ j} = \beta_0 + \beta_1 D_t + \beta_2 D^j + \beta_3 D_t^{\ j} + \varepsilon_{it}^{\ j}$$

where $P_{it}^{\ j}$ is the outcome variable (for example house price) for observation i in period t for group j, D_t is a dummy variable for observations that occurred before and after the treatment time period, D^j is a dummy variable for observations in the group that is to be treated ($j = 1$) during the treatment time period ($t = 1$), and $D_t^{\ j}$ is a dummy variable for observations in the treatment group during the treatment time period.

An unbiased estimate of the "average treatment effect" or parameter of primary interest is,

(13) $$\hat{\beta}_3 = \overline{p}_1^{\ 1} - \overline{p}_0^{\ 1} - (\overline{p}_1^{\ 0} - \overline{p}_0^{\ 0})$$

where the bar in equation (13) indicates an average over the observations in a group, if the identifying assumption $E\left[\varepsilon_{it}^{\ j} \mid D_t^{\ j}\right] = 0$ holds. The estimation of $\hat{\beta}_3$ in equation (13) is simply the differences in the mean values of the treated and non-treated groups and the condition $E\left[\varepsilon_{it}^{\ j} \mid D_t^{\ j}\right] = 0$ implies that if the treatment did not occur, then β_3 would be equal to zero. One might think of many reasons why the conditions surrounding a quasi-random experiment may invalidate the identifying assumption above. However, even when potential confounders exist, a detailed understanding of the institutional environment for the problem being studied can be used to augment the regression

framework to reflect potential confounders. For example, some of the problems that may occur because of omitted interactions, trends in outcomes or observable differences in the characteristics of the control and treatment groups can be overcome by including additional regression controls. Equation (12) can be updated to include such additional controls as follows,

$$P_{it}^{\ j} = \beta_0 + \beta_1 D_t + \beta_2 D^j + \beta_3 D_t^{\ j} + \mathbf{z}_{it}^{\ j} + \varepsilon_{it}^{\ j} \tag{14}$$

Where $\mathbf{z}_{it}^{\ j}$ is a vector of controls (i.e. structural and neighborhood controls in a housing market).

2.5.3 Using Quasi-Random Experiments to Alleviate Omitted Variable Bias

Quasi-random experiments can also provide a cost effective way for alleviating some of the estimation difficulties that have plagued traditional hedonic regressions. For example, the hedonic model provides little guidance on the appropriate functional form of the hedonic price function. There are two functional form difficulties: 1) what is the appropriate functional relationship between prices and the attributes in the hedonic price function, and 2) what attributes should be included in the hedonic price function specification to avoid omitted variable bias. Cropper *et al.* (1988) have provided some guidance on the first difficulty,[28] while research progress on the second difficulty has proceeded along three fronts.

[28] In their paper, Cropper *et al.* used an assignment model to construct the "true" hedonic equilibrium to analyze the properties of hedonic price functions. This was done by constructing sets of known utility functions and houses with given attributes, to simulate the matching that occurs in the market to generate equilibrium prices. They then used these prices and the housing characteristics to run regressions using a variety of functional forms that had been used in the hedonic literature from the simple linear OLS to the more complicated quadratic Box-Cox functional form. Their primary finding was that when all attributes

One method for dealing with the second difficulty has been to simply create more detailed housing datasets (see Fulcher (2003) as an example). Another method that has been employed to deal with unobserved spatial heterogeneity is to "correct" the error terms in a regression model directly for unobserved spatial heterogeneity using spatial regression techniques (see Kim *et al.* (2003) as an example). The third method used by researchers is to combine quasi-random experiments in conjunction with spatial and temporal fixed effects to parse out unobserved spatial heterogeneity over the same spatial and temporal extent as the potentially exogenous variation of interest. This can help to better identify an implicit price than in a cross-sectional analysis by increasing "internal validity" (see Chay and Greenstone (2005) as an example). The applications developed in the following chapters illustrate all three of these methods in an effort to obtain unbiased estimates for parameters of interest.

2.5.4 Exogenous Information Shocks as Quasi-Random Experiments

To use the quasi-random experimental design to test the implications of this theoretical framework, one would like to identify exogenous information shocks that shift the fraction of informed buyers in a housing market. Ideally these exogenous information shocks would involve (i) an attribute whose information on quality is publicly available, (ii) an attribute whose information on quality was known to sellers but remained unknown to some buyers, (iii) a situation where the information shock shifted the fraction

---

are observed by the researcher, the quadratic Box-Cox performed best. However, when there were omitted variables or proxy variables were used, the simpler functional forms such as the linear and semi-log forms estimated the marginal attribute prices most accurately.

of attentive and informed buyers, and (iv) an attribute whose quality can be determined by the researcher.

Distinguishing the impact of an information shock on the expansion in the consideration set versus its impacts on subjective expected benefits and costs is difficult. The contributions of consideration and search costs to the outcomes in a formal model will appear to be observationally equivalent. This is because exogenous information shocks not only direct buyer's initial attention to consider new attributes; they may also change buyer's perceptions of the expected benefits or costs of searching for publicly available information. To identify the impact of initial attention on consideration set expansion, ideally the exogenous information shock would not affect subjective expected benefits and costs. The application in chapter 3 comes closest to this ideal scenario. However, the estimates presented there and elsewhere in this thesis come from reduced-form regressions and therefore may provide evidence for attention, but this behavioral mechanism is only one of many "informational stories" that may be consistent with the results. Therefore, indirect evidence associated with the context of each example will be used to reinforce the persuasiveness of the attention and search argument.

2.6 Conclusion

This chapter described a simple framework to motivate and integrate the empirical applications in chapters three through five. Much of the analysis revolved around a key question—"what is an informed choice?" Neoclassical models simply assume that choices are informed. This assumption leads analysts to attribute effects of differences in information on prices to preference heterogeneity.

Limited attention and search theory suggest that information heterogeneity may be important for using the housing market to establish the MWTP of locationally delineated attributes. This analysis assumes sellers are more informed than buyers and argues we should expect a significant fraction of buyers to be uninformed for some housing attributes. When the fraction of uninformed buyers for a housing attribute is non-trivial, the basic logic of a search model suggests that estimates of MWTP for the attribute will likely be attenuated towards zero.

A quasi-random experiment offers an empirical strategy for gauging the importance of the argument. This methodology relies on an exogenous event to identify an outcome of concern. The information available to homebuyers is the primary focus. To be used, the event must be exogenous to the housing market and cause homebuyers to become more informed. The applications in chapters three through five all employ the quasi-random experiment framework in conjunction with an exogenous information shock. Chapter 3 discusses a real estate disclosure that is argued to have primarily affected buyer's attention rather than providing information. Chapter 4 discusses another real estate disclosure that likely attracted buyers' attention but also provided them with information—thereby affecting search costs and benefits. Chapter 5 discusses an information shock provided by the media that if received by buyers had the potential to impact initial attention, and both search costs and benefits. As a result, it is less clear that buyers would use the information for their home buying decisions.

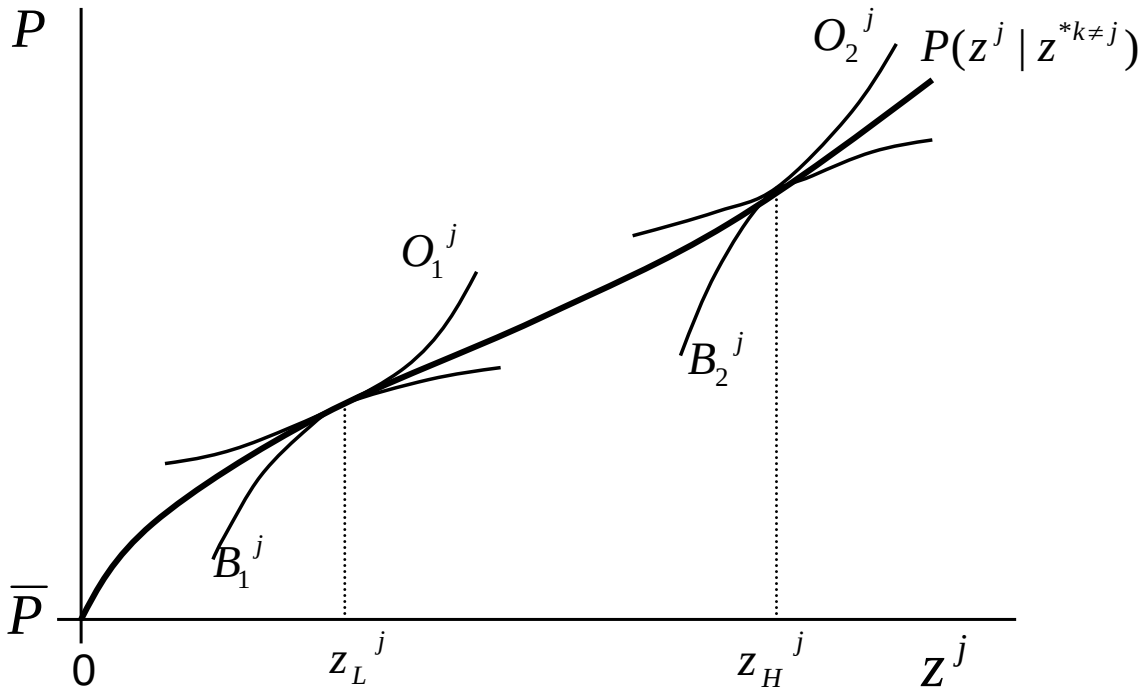


Figure 2.1: Long-Run Supply Price Locus for Attribute j

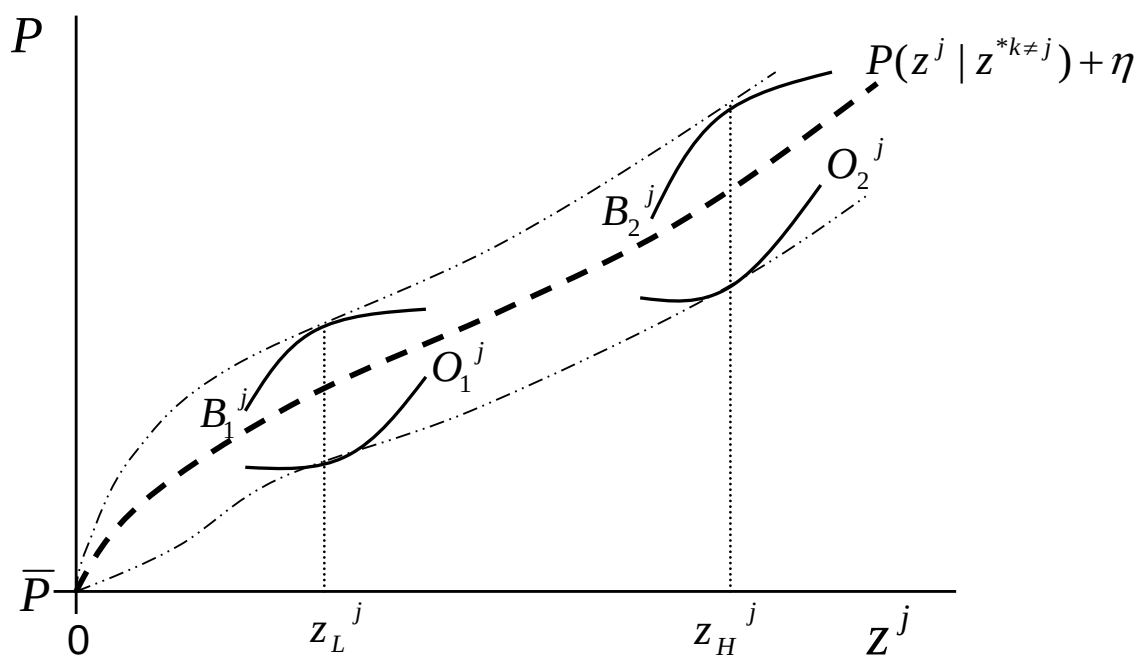


Figure 2.2: Impact of Incomplete information and Search Costs on Price Function

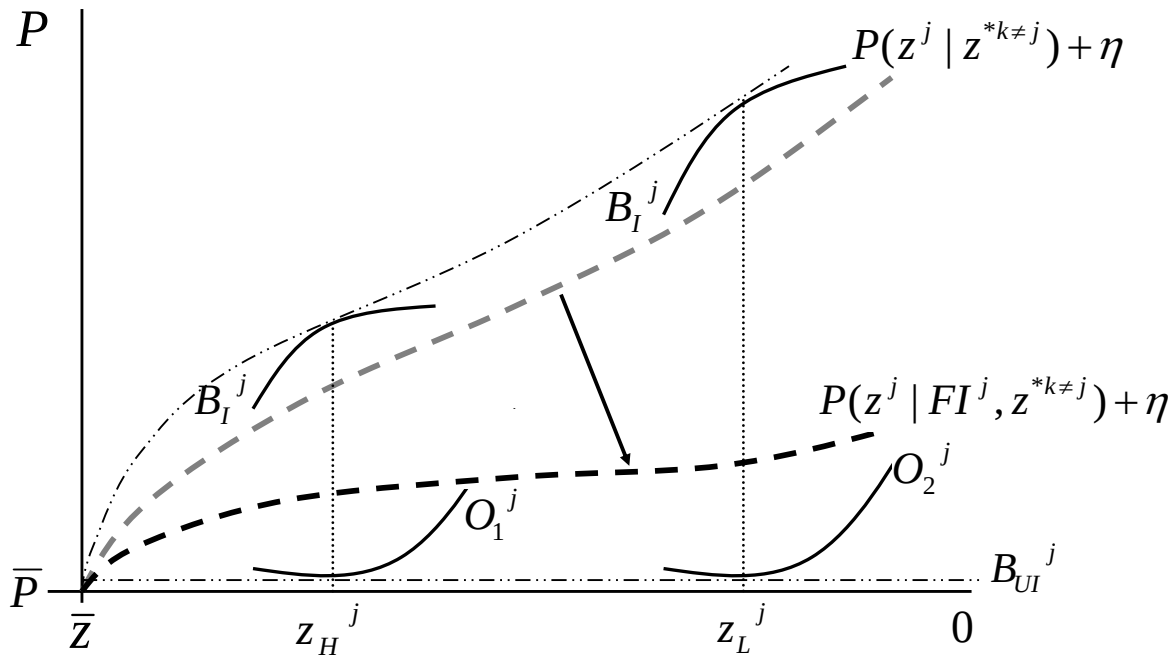


Figure 2.3: Impact of Inattentive Buyers on Equilibrium Prices

Chapter 3: Airport Noise

3.1 Introduction

This chapter describes how unanticipated source of information affects estimates of the marginal implicit price of airport noise. Few would dispute the argument that airport noise is undesirable. Major airports were historically located a reasonable distance from population centers to reduce the potential for noise externalities. However, as urban areas have developed, more housing has been built near airports in metropolitan areas. Also, some airports have dramatically increased aircraft traffic beyond what was expected when their locations were selected. For these reasons the extent of the airport noise problem has grown over time.

A common complaint by recent buyers of housing near airports is that they were unaware that the housing they purchased would be affected by airport noise.[29] This chapter deals with an example of this phenomenon where there is a clear record of the efforts to raise awareness of noise externalities for the Raleigh-Durham International Airport (RDU). In 1992 homeowners near RDU filed a lawsuit against the airport for the increased amount of noise since the opening of a new airline hub. This appears to be the impetus behind efforts by RDU to find a way to reduce the future liability posed by additional lawsuits by other homeowners who had bought their homes following the opening of the hub. After unsuccessfully trying to convince local municipalities to pass airport noise disclosure laws, in 1997 RDU was able to use a newly introduced state law to create their own noise disclosure and require sellers in the area to provide it to buyers.

[29] I have found this anecdotal evidence by searching newspapers in metropolitan areas, by reading logged complaints on airport websites, and talking with real estate agents.

Homeowners did not know about the disclosure requirement until just a few weeks before they were legally required to begin disclosing. The disclosure was provided to homes in a very generous area surrounding the airport as would be expected given the airport authority's concerns about reducing liability. As a result, some homes that are seldom or never affected by airport noise also received the disclosure.

It appears that the disclosure itself was not intended to provide information about the actual level of airport noise at a given house, but rather to alert buyers to the possibility that airport noise might have an impact on the property. The disclosure encourages buyers to obtain additional information. The costs of obtaining this noise related information did not change as a result of the disclosure. These costs would arise from activities associated with being at the property at different times of the day and perhaps interviewing neighbors. It could also entail calling the airport authority offices to obtain more specific information about the amount of airport noise that occurs at a given property.

The circumstances associated with the RDU disclosure, provides an opportunity to analyze the usefulness of the conceptual model in chapter 2. Choosing a house to buy is a complex decision where attention and search constraints will likely be relevant. Furthermore airport noise is an attribute for which information is publicly available. As a rule we can expect that sellers will be informed about the level of airport noise from having lived at the location. However, it is possible that buyers may be "initially inattentive" or uninformed for several reasons.[30] First airport noise is not constant. It changes throughout the day and throughout the week. In fact, anecdotal evidence

[30] See section 2.3 for how "initial inattention" is defined.

suggests that prior to the introduction of the noise disclosure real estate agents exploited this temporal variation in noise by showing homes near the airport during low noise time periods. Furthermore the area surrounding the airport is wooded and it is difficult to see the location of the airport while visiting a property. Thus a fraction of buyers may remain initially inattentive even after having visited the house several times. The conceptual framework implies that the higher the fraction of uninformed buyers the more the implicit price for airport noise may be attenuated towards zero.

The disclosure requirement introduced by the RDU authority can be treated as an exogenous information shock that changed the fraction of informed buyers in the housing market surrounding the airport. The primary channels for an information shock to shift the fraction of informed were described in chapter 2 as: (i) causing buyers who were "initially inattentive" to consider an attribute, (ii) shifting the subjective expected benefits of buyers to acquire information about the attribute above subjective expected costs, and (iii) shifting the subjective expected costs of buyers to acquire information about the attribute, below subjective expected benefits.

Since information about airport noise was publicly available and did not change as a result of the disclosure, it is unlikely that channel (iii) had a role in causing changes in the fraction of informed buyers after the disclosure. However, it is possible that the disclosure changed the subjective expected benefits. When buyers receive a disclosure indicating a public authority believes that the noise *necessitates* a disclosure, then a buyer may perceive a greater likelihood of airport noise problems. This reaction could be greater than what would actually be warranted. Therefore the disclosure may have some impact on channel (ii) but this is likely minimal. It is clear that the disclosure will affect

information acquisition through channel (i). If inattention is relevant then the disclosure policy should cause those buyers that were "initially inattentive" to consider the airport noise issue.

As was mentioned in chapter 2, distinguishing the impact of an information shock on expansions in the consideration set versus its effects on subjective expected benefits and costs is difficult. The effects of the consideration set and the search costs on the observed outcomes will be observationally equivalent. Under ideal circumstances identifying the impact of attention on consideration set expansion would require that the information from the disclosure have no impact on subjective expected benefits and costs. The RDU airport noise disclosure is arguably an example where this appears to be the case. The disclosure has minimal impacts on subjective expected benefits and costs.

Using data on housing sales occurring near the RDU airport that bracket the timing of the disclosure, it is possible to compare implicit prices for airport noise before and after this information shock. The results suggest that the disclosure reduces the selling price of homes in high noise areas by approximately 2-3 percent. This reduction is in addition to the approximate 4 percent impact that would be attributed to airport noise in these areas prior to the disclosure. Thus these findings suggest that a hedonic price regression using the housing sales the year before disclosures began taking place would provide an estimate of the implicit price of airport noise that understates MWTP by approximately 36 percent.

The remainder of the chapter will proceed as follows. Section 3.2 provides additional background on RDU's airport noise disclosure program. Section 3.3 describes the data used in the analysis. Section 3.4 outlines the identification strategy used for the

hedonic price regression analysis. Section 3.5 describes the results of the hedonic price regressions. Section 3.6 concludes the chapter.

3.2 Background on RDU's Airport Noise Disclosure Program

Airport noise is a leading environmental concern at major airports (US GAO, 2000a, 2000b). Previous papers on the impact of airport noise on housing prices have found that airport noise can have a substantial effect on the price of homes.[31] Over the past twenty five years efforts have been made to limit the number of households exposed to significant noise levels of 65 decibels (dB) or more. However airports receive numerous complaints from new residents indicating that they were unaware of the presence or extent of the airport noise before purchasing their homes. This suggests that although airport noise is considered public information, there are buyers that may not pay attention to this attribute before buying a house.

A set of unique events occurred at RDU that allows a test of whether or not buyers were aware and fully informed about airport noise. In June of 1987 American Airlines opened a hub at RDU causing the number of flights at the airport to approximately double. This change increased the amount of aircraft related noise experienced by nearby homeowners. After this expansion, 125 homeowners sued the Raleigh-Durham Airport Authority (the governing body for the airport which will be referred to as RDU) complaining that the increased noise had reduced their property values. In 1992 RDU settled the lawsuit by agreeing to pay 1.8 million dollars to the 125 homeowners. Most homeowners were compensated in the range of $11,000 to $23,000

[31] See Nelson (2004) for a thorough review of this literature.

depending on the amount of noise experienced at the location of the house and the value of the house. The average compensation received by the homeowners participating in the lawsuit was $14,400.[32]

The lawsuit prompted RDU to consider a number of ways to reduce their future liability. For example, in 1994 RDU made recommendations to the towns affected by airport noise on how to pass ordinances that among other things would require homeowners to notify buyers of potential aircraft noise problems. However, the ordinances were strongly opposed by developers and homeowners who did not want to provide the noise disclosure and risk further reducing their property values. Without its constituents support, towns were unable to pass the ordinances recommended by RDU.

In 1995 North Carolina General Statute 47E known as the "Residential Property Disclosure Act", was passed requiring owners of residential real estate to provide prospective buyers with a property disclosure statement after January 1, 1996. Although there was nothing about airport noise on the disclosure statement, part of the statute required that any notification from a government agency affecting the property must also be furnished to prospective buyers. RDU which is recognized as a government agency in the state of North Carolina used this opportunity to require homeowners to provide a separate airport noise disclosure. RDU developed a notification letter and map that was sent out in March of 1997 to all homeowners within specified noise contours (see appendix 3.A for copies of the letter and map).[33] Additionally the information was sent

[32] These dollar amounts are in 1992 dollars and based on reports in the primary local paper "The News and Observer".

[33] The letter and map have been mailed two times since the initial mailing—once in April 2000, and again in May 2004. Furthermore, the letter and map is available on RDU's website and many real estate agents in the area provide a link to these digital copies on their websites.

to real estate agents in the area and a "noise officer" from RDU made spot checks to ensure that both sellers and their agents were using the disclosure letter and map for each housing transaction that occurred in the area. The airport noise disclosure developed by RDU appears to have occurred unexpectedly.[34]

In the time since airport noise disclosure has been mandatory, RDU has supported plaintiffs in two Wake County Superior Court cases where sellers failed to present buyers with an airport noise disclosure. In these cases the buyers sued the sellers because they claimed they were unaware of the airport noise issue before buying the houses. These court cases illustrate both the potential for buyers to be inattentive to the airport noise attribute of housing, and the firm stance that RDU has taken on airport noise disclosure. Personal conversations with the RDU noise officer suggest that compliance has been exceptional since the airport noise disclosures have been required.

3.3 Data Used in the Analysis

There are several sources of data that are needed to estimate the impact of the airport noise disclosure on housing prices. The primary dataset used in this analysis consists of single-family housing transactions occurring between 1992 and 2000 in Wake County North Carolina.[35] Data on sale prices and property characteristics were compiled from information provided by the Wake County Revenue Department. Sales were

[34] A search of the primary local paper "The News and Observer" shows no articles revealing that RDU had developed or was going to implement their noise disclosure program leading up to the mailing of the letters and maps.

[35] The airport is located on the western side of Wake County near the Durham County line. Most of the residential properties affected by noise from the airport are located in Wake County. However, there are a small number of properties located within Durham County that are potentially impacted by airport noise that are excluded from the analysis.

screened to ensure that they were arms-length and to eliminate outlying observations. A GIS shapefile was acquired from the Wake County GIS department that georeferenced each of the house locations in the above dataset of housing transactions. Using this GIS shapefile a variety of other variables were linked spatially to the housing locations using ArcView GIS.

RDU provided a GIS shapefile that spatially referenced noise contour zones of 55Ldn or greater which required the noise disclosure after April 1, 1997.[36] These noise contour zones are a composite of present and future expectations about noise levels around the airport.[37] The housing dataset was then screened to include only those housing transactions that were in noise zones requiring disclosure, or housing transactions within a one mile buffer around this area of disclosure. The houses within this one mile buffer act as a natural control group. Figure 3.1 shows housing locations in relation to the airport noise zones and table 3.1 provides descriptions and summary statistics for the variables used in the analysis.

Structural control variables include age of the house, number of bathrooms, acreage of the lot, heated area other than attic and basement, detached garage dummy, number of fireplaces, deck area, sewer availability dummy, hardwood floors dummy, screen porch area, brick walls dummy, attic heated area, basement heated area, garage area, pool dummy, full basement dummy, partial basement dummy, enclosed porch area and a set of dummies denoting an assessment of housing condition. Unlike housing data commonly used for hedonic analysis, variables such as garage and basement heated area

[36] Ldn stands for "level day night" which is a day-night weighted average noise level in decibels where nighttime noise is weighted more heavily than daytime noise because it is considered more obtrusive.
[37] The noise contour map states that the contours are based on "current flight operations at the airport and those projected through approximately the year 2010."

have the square feet calculated rather than simply denoted by a dummy variable. Thus this dataset provides the analyst with a very detailed set of structural control variables.[38]

Neighborhood control variables derived from census block group level data include percent non-white, median house values, median time to work, percent of population under 18 years of age and percent owner occupied housing. In addition to these census variables, variables were created to measure the distance to the nearest park, the distance to the nearest shopping center, and the property tax rate for the area.

The estimation strategy, discussed in detail below, utilizes a fixed-effects model in time and space. Therefore yearly dummy variables were created for the years 1992-2000.[39] For example, yrdum93_apr in table 3.1 is a dummy variable equal to 1 for all houses transacted from April 1, 1993 to April 1, 1994. Years are broken up from April to April to facilitate analyzing the impact of airport noise disclosures which began taking place on April 1, 1997.

Four spatial variables were created to represent the spatial relationship between each house and the airport. First, the variable airport_dist is the linear distance between each house and the main entrance to the airport terminals. The variable noise65_dum is a dummy variable equal to 1 if the house lies in the 65-70Ldn noise contour where disclosure was mandatory after April 1, 1997. The variable noise55_dum is a similar dummy variable which is equal to 1 if the house lies in the 55-65Ldn noise contour where disclosure was also mandatory after April 1, 1997. Finally, the last dummy variable is set equal to 1 if the house lies in the 1 mile buffer zone just outside the 55-65Ldn noise

[38] As discussed in section 2.5.3, one method used by researchers to overcome omitted variable bias in hedonic regressions is to compile more complete housing datasets.
[39] The 1992 dummy is the omitted variable in the estimation.

contour where there was no disclosure requirement after April 1, 1997.[40] Two additional dummy variables, noise55_post and noise65_post are created that are equal to 1 for houses that lie in the 55-65Ldn and 65-70Ldn zones respectively and were *transacted* after April 1, 1997. Table 3.2 provides the counts of housing transactions for noise55_dum, noise65_dum, noise55_post and noise65_post over different time horizons used in the analysis.

To appreciate the consequences of living in a 55-65Ldn noise contour zone or a 65-70Ldn noise contour zone, it is useful to consider how human perception of noise disturbance is related to the decibel scale which is used to create the measures of Ldn.[41] Humans perceive noise on a log scale rather than a linear scale. Noise is perceived to be twice as loud for every increase of 10dB. Typical background noise levels in urban areas have been estimated to be between 50-60dB during the day and 40dB at night. For example, light traffic a hundred feet away would likely register at over 50dB's. Houses in the 65-70Ldn zone are likely to experience aircraft noise that is perceived as roughly twice as loud as the noise experienced on average by those living in the 55-65Ldn contour and roughly 3-4 times as loud as houses in the 1 mile buffer. Federal Aviation Administration (FAA) regulations only pertain to housing in areas of 65Ldn's or above.

[40] The dummy variable representing houses in the 1 mile buffer area is omitted in the estimation.
[41] As footnoted earlier, Ldn stands for "level day night" which is a day-night weighted average noise level in decibels where nighttime noise is weighted more heavily than daytime noise because it is considered more obtrusive.

3.4 Identification Strategy for the Hedonic Model

The temporal source for variation in information about airport noise arises from disclosures that began at the end of March 1997. Spatial variation on noise impact zones stems from the noise contours provided by RDU. A simple estimation approach using the compiled housing dataset and these sources of temporal and spatial variation using a fixed-effects approach and a semi-log form, is as follows[42],

$$
\begin{aligned}
\mathbf{lprice} = \beta + \alpha\mathbf{Structural} + \phi\mathbf{Neighborhood} \\
+ \delta\mathbf{yrdum93\_apr}...\mathbf{yrdum00\_apr} \\
+ \theta\mathbf{noise55\_dum} + \gamma\mathbf{noise65\_dum} \\
+ \lambda\mathbf{noise55\_post} + \psi\mathbf{noise65\_post} + \varepsilon
\end{aligned}
\tag{1}
$$

where the variables are defined as in table 3.1, ε is an error term and β, θ, γ, λ, and ψ are parameters to be estimated and α, ϕ, and δ represent vectors of parameters to be estimated. In experimental terminology houses in the 55-65Ldn and 65-70Ldn contours have been "treated" by the airport noise disclosure while the houses in the one mile buffer area are "untreated" and act as a control. Using the estimating equation above, the coefficients on noise55_post and noise65_post would provide estimates of the impact of the airport noise disclosure on housing prices in these disclosure zones relative to houses in the one mile buffer control area.

[42] In their paper, Cropper *et al*. used an assignment model to construct the "true" hedonic equilibrium to analyze the properties of hedonic price functions. This was done by constructing sets of known utility functions and houses with given attributes, to simulate the matching that occurs in the market to generate equilibrium prices. They then used these prices and the housing characteristics to run regressions using a variety of functional forms that had been used in the hedonic literature from the simple linear OLS to the more complicated quadratic Box-Cox functional form. Their primary finding was that when all attributes are observed by the researcher, the quadratic Box-Cox performed best. However, when there were omitted variables or proxy variables were used, the simpler functional forms such as the linear and semi-log forms estimated the marginal attribute prices most accurately.

Controlling for Potential Confounding Influences

As discussed in Meyer (1995), there are a number of potential confounders when applying the quasi-random experiment methodology. Foremost among these is omitted variables bias which in this context could occur from other factors that change over time or over space and lead to a mis-measurement of the disclosure effect. One potential spatial confounder that is acknowledged by previous studies on the impact of airport noise on housing prices is that proximity to an airport can also be a desirable attribute for households because it provides convenient access to airport services. If airport noise were evenly distributed in space around an airport it would be very difficult to identify the positive and negative effects due to living near an airport. However, airport noise is typically greatest extending out from the ends of the runways with much less noise extending outwards from the sides of the runways. Thus by including a distance measure to the main entrance of the airport, it is possible to account for this effect and be reasonably confident the variable being used is uncorrelated with the noise measures.

Another potential spatial confounder can be seen in figure 3.1. Houses in the vicinity of the airport are grouped primarily in a northern set and a southern set. If the block group level controls, tax rate information and other spatial variables included in the regression are not adequately controlling for differences between the northern and southern groups, then the noise contours may be picking up some of these differences. To address this possibility, a dummy variable was created for the northern grouping and included in the analysis.

There were a number of new homes built in the vicinity of the airport during the 1992-2000 period included in the dataset. If there is something unique about this new

housing stock built over this time period in the area near the airport relative to the existing housing stock, this could potentially bias the results. To avoid this bias, a dummy variable was created for houses that sold in the same year that they were built. Although it is possible that houses sold more than once in the year after they were built, this variable should act as a reasonably good proxy for newly built homes.[43]

An interstate (I-540) that passes through the northern sample of houses has been under construction since 1992 and is still ongoing. If the benefits of this interstate are correlated with airport noise, this could potentially bias empirical estimates as well. To overcome this bias, a GIS street file was acquired and a point was created for each intersection of the new interstate (see figure 3.1). Linear distances from each house to the nearest intersection were then calculated. Because the benefits associated with the interstate likely dissipate the further a home is from an intersection (for example the houses in the southern sample are likely to derive almost no benefit due to their house location) the following function was used to relate houses to benefits from the interstate:

(2) $$\textbf{interstate} = \max\left[1-(d/d_{\max})^{1/2}, 0\right],$$

where d is distance in miles and $d_{\max}$ is a benefit cutoff point set to 4 miles for this application. This function creates a convex index between zero and one.[44]

Another concern for this analysis is that expectations about airport noise changed over the study period, thereby biasing the results. To analyze the potential for changes in

[43] This variable is also important because the NC property disclosure law does not apply to brand new housing. However, while RDU airport authority cannot enforce compliance they have asked all developers to also provide the disclosure and make checks on whether or not the disclosure occurs for new housing. Discussions with the noise officer suggest that compliance for new houses is nearly as high as for houses that resale. Nonetheless the new sale dummy is important for capturing any differences in propensity to disclose.

[44] Of course houses extremely close to an intersection may be negatively affected by traffic noise. However this is a very small percentage of houses and so should not bias the results.

airport noise expectations to confound the ability to use the variation in time to identify the impact of noise disclosure, statistics were gathered on the number of travelers passing through RDU by year which are presented in table 3.3. The number of travelers using RDU increased considerably in the late 1980's coinciding with American Airlines designation of RDU as one of its hubs. During the early 1990's the traffic passing through RDU was relatively constant. However, a dip in traffic can be seen between 1994 and 1995 coinciding with American Airlines closing their hub during this time period. From 1995-2002 it can be seen that traffic slowly increases and then plateaus at levels coinciding with the early 1990's. Overall these statistics provide little reason to believe that there were dramatic shifts in expectations about airport noise over the relevant time horizon.

Although considerable effort was devoted to discover spatial and temporal confounders, it is difficult to be certain that there were no other potential temporal-spatial interactions during the 1992-2000 time period. Thus some form of sensitivity analysis for the time period chosen in the empirical analysis is desirable. This strategy reduces the likelihood that omitted spatial factors correlated with the timing of the airport noise disclosure are responsible for the findings. One approach to undertake this analysis would be to introduce a more restrictive sample selection bound around April 1, 1997 when disclosures began. Therefore this analysis is performed to determine whether the smaller sample affects the estimates of the impact of the noise disclosure.

When equation (1) is updated to include the variables used to control for spatial influences and to restrict the time interval to the year before and after disclosures began taking place, the result is equation (3),

$$
\begin{aligned}
\mathbf{lprice} = \beta + \alpha\mathbf{Structural} + \phi\mathbf{Neighborhood} \\
+ \delta\mathbf{yrdum97\_apr} \\
+ \theta\mathbf{noise55\_dum} + \gamma\mathbf{noise65\_dum} \\
+ \lambda\mathbf{noise55\_post} + \psi\mathbf{noise65\_post} \\
+ \vartheta\mathbf{airport\_dist} + \rho\mathbf{interstate} \\
+ \omega\mathbf{new\_sale} + \pi\mathbf{north} + \varepsilon
\end{aligned} \tag{3}
$$

where airport_dist is the linear distance from each house to the airport entrance, interstate is the distance measure described by equation (2), new_sale is a dummy variable denoting houses that were transacted for the first time and north is a dummy for the houses north of the airport. In this specification the 1997 year dummy is included and the 1996 year dummy is omitted because only two years of housing data from April 1996 to April 1998 are used.

Although the empirical specification in equation (3) represents a detailed analysis of potential unrelated features varying in time and space that could bias estimates, there nonetheless remains a possibility that spatial autocorrelation could potentially bias the estimates of the impact of the airport noise disclosure program. The most common techniques for taking account of spatial autocorrelation are the spatial-lag model and the spatial error models (see Anselin (2002) for a review). The spatial lag model would include the term in equation (4) along with the other determinants of price given in equation (3),

$$\rho\mathbf{W}(l\mathbf{price}), \tag{4}$$

where ρ is a spatial autocorrelation parameter and $\mathbf{W}$ is a n x n spatial weight matrix. The spatial-lag model is analogous to an auto-regressive time-series model. However in

the spatial case, interactions are in multiple directions in that "neighbors" influence each other where in time this interaction is in a single direction.

The spatial error model adjusts the error term of equation (3) as in equation (5),

(5) $$\varepsilon = \lambda \mathbf{W} \varepsilon + u \,,$$

where λ is the spatial autoregressive coefficient, $\mathbf{W}$ is the spatial weight matrix, and u is assumed to be a vector of i.i.d. errors. In this model the errors are corrected directly for spatial autocorrelation. While the lag and error models are subject to slightly different interpretations (see Kim et al. (2003) for a discussion), in practice they are both used to control for spatial autocorrelation.[45] Deciding which of the two models to use in applied work is typically determined by using robust Lagrange multiplier test statistics.

Two types of weight matrices were constructed. These are the "rook" contiguity weight matrix and the "queen" contiguity weight matrix. These were chosen because the sample size is large for spatial regression analysis and these matrices are particularly amenable to sparse matrix routines. Other types of weight matrices require that the full variance covariance matrix be inverted which becomes infeasible when n is greater than approximately 800. The rook weight matrix defines a house as a neighbor if its Thiesen polygon shares a common boundary. The queen weight matrix defines neighbors as those whose Thiesen polygons share either a common boundary or a common vertice.[46] Because these weight matrices require that each observation be unique in space, 109

[45] Use of the spatial lag model in this paper can be interpreted as an attempt to "filter out" spatial autocorrelation rather than an explicit model of neighborhood interactions.

[46] A Thiesen polygon is a region assigned uniquely to each house. Every point within that polygon is closer to the assigned house location than the location of any other house. Therefore the borders between Thiesen polygons are the collection of points that are equidistant between two houses. See Anselin (2002) for further discussion.

repeat sales and 10 spatial outliers were dropped so that each transaction could be represented by a unique latitude and longitude coordinate.

3.5 Results

3.5.1 Primary Results

Columns [1]-[3] in table 3.4 contain estimation results from the semi-log least squares specification corresponding with equation (3). Column [1] uses the full temporal sample of housing transactions between 1992 and 2000. Column [2] temporally restricts the sample to the years 1995 through 1998. Column [3] further restricts the sample to the years 1996 and 1997. Estimates of the parameters for the structural and neighborhood characteristics and temporal fixed effects have been omitted from table 3.4, but were consistent with a priori expectations and other hedonic studies using similar variables. These results are reported in appendix 3.B.[47]

As can be seen in column [1] of table 3.4, the estimates for the dummy variables new_sale and north suggest that any differences between new and existing houses and houses in the north and the south groupings are being controlled with the structural and neighborhood characteristics included in the regressions. The coefficient on airport_dist is negative and highly significant in columns [1]-[3] suggesting that homeowners value access to the airport. The coefficient on interstate is also negative and highly significant indicating access to the newly built interstate is valued by homeowners. The coefficient

[47] The estimated parameters for the yearly fixed effects are all of the expected sign and magnitude. Prices are increasing for housing in this area of Wake County over the time period. Also note that although the yearly fixed effects in the columns labeled 1995-1998 and 1996-1997 are reported in the same table in appendix 3.B, they are not directly comparable because the omitted dummy is not the same.

on noise55_dum suggests that houses in the 55-65Ldn aircraft noise zone sold for approximately 2.3% less than comparable homes in the 1 mile buffer zone.[48] The coefficient on noise65_dum suggests a more substantial price impact of approximately 5.1% on housing in the 65-70Ldn noise contour. These estimates are consistent with the existing estimates of the impact of airport noise in the literature.[49]

The variables of primary interest in table 3.4 are those labeled noise55_post and noise65_post denoting those housing transactions in the 55-65Ldn and 65-70Ldn zones respectively that took place after April 1, 1997. As discussed in section 3.2, if the coefficients on noise55_post and especially noise65_post are insensitive to the temporal bounding restriction around when disclosures began taking place, then this strengthens the credibility of the identification strategy. The coefficient on noise55_post is small, positive, and statistically significant in columns [1] and [2], but in column [3] where housing transactions have been tightly bounded around the disclosure date, it becomes negative and is not statistically significant. The coefficient on noise65_post however is consistently negative and statistically significant for all three temporal sample types. Therefore assuming that the tightly bounded third regression is most likely to correctly identify the relationship between airport disclosure and housing prices, the coefficients

[48] As noted by Halvorsen and Palmquist (1980), coefficients on dummy variables in a semi-log regression cannot be interpreted as 1/100th of the percentage effect of that variable on the variable being explained like the coefficients on continuous variables. However the closer a dummy variable coefficient is to zero, the reported coefficients multiplied by 100 are likely to be more accurate estimates of the percentage effect of that variable on the variable being explained than if the coefficient is large (i.e. the more accurate approximation of a coefficient of -0.3 is -2.96%). Percentages reported in the text of this paper are the more accurate Halvorsen and Palmquist adapted estimates rounded to the second decimal place.

[49] For example, the meta analysis by Nelson (2004) used estimates of the impact of airport noise from 33 published and unpublished studies and concluded that a change of 10 dB of noise would impact the value of a house by between 5 and 6 percent.

from this regression suggest that the airport noise disclosure further reduced property values of houses in the 65-70Ldn zone by approximately 2.9%.

To gauge the robustness of the estimated impact of noise disclosures on housing prices, a spatial regression analysis was performed on the most restrictive 1996-1997 sample based on the timing of the sales. To construct the weights matrix, 109 repeat sales and 10 spatial outliers were dropped from the sample of 4,369. To ensure that dropping these observations did not affect the analysis, a model was fit to the remaining 4,250 observations and is presented in column [4]. Comparing columns [3] and [4] suggests that there was virtually no change in the coefficients due to the dropping of the observations. Using the errors from this regression, a Moran's I statistic was calculated that suggested the presence of spatial autocorrelation and the need for spatial regression analysis (see table 3.5). To determine whether the spatial-lag or the spatial error model was most appropriate, robust LM statistics were performed (see table 3.5). These statistics are consistent with selecting the spatial-lag model.

Columns [5] and [6] of table 3.4 present the spatial-lag regression results using maximum likelihood and the rook and queen contiguity weight matrices.[50] The reported standard errors are the asymptotic standard errors. As one might expect, the spatial regressions do have an impact on the spatial variables. For example, the size of the noise65_dum is reduced by approximately one half. The sign on noise55_post is slightly negative but insignificant. The coefficient on the noise65_post although reduced by approximately a percentage point, is still negative and statistically significant in both

[50] Although the LM statistics suggest the spatial-lag model to be chosen, the analysis was also performed using the spatial error model. The results of the spatial-error model were very similar to the spatial-lag results reported in table 3.4.

regressions. It suggests that the airport noise disclosure that began taking place on April 1, 1997 impacted the price of houses in the 65 Ldn contour by approximately 2.1%.

3.5.2 Discussion of Results

These results indicate that the airport noise disclosure impacted housing prices near the RDU airport. It appears that buyer's "initial attention" matters for the estimated MWTP for airport noise. In the most tightly controlled regression, using spatial regression analysis and a limited temporal sample of houses that bracket the RDU airport noise disclosure, the estimates imply that the disclosure reduced housing prices by approximately 2.1% in the high noise zone. The total impact of airport noise in this zone is approximately 5.8 % (obtained from column [6] by adding the coefficients on noise65_dum and noise65_post together). Therefore, the impact of the airport noise disclosure represents an approximate 36% adjustment in the implicit price estimate for the high noise zone. This may signal that there was a large fraction of inattentive and uninformed buyers in this area prior to the initiation of the disclosure rule.

It does not appear that the disclosure impacted housing prices in the low noise zone. However it is important to acknowledge that this zone was defined using a generous delineation of the low noise area. It includes houses that are seldom, if ever impacted by airport noise. One interpretation of the lack of an effect from the disclosure in the low noise zone is that the disclosure itself is not causing an "overreaction" effect through the information acquisition channel of expected benefits. In other words one might argue that the low noise zone received the disclosure "placebo". As a result it

serves to provide a comparison zone for the high noise zone which was significantly impacted by the disclosure.

These results are directly relevant to the previous literature that has attempted to quantify the disamenity cost of airport noise for two reasons. First, no published work on the impact of airport noise on housing prices has used spatial regression analysis. The results from this paper suggest that previous studies that have not adequately controlled for unobserved spatial heterogeneity may over-estimate the average MWTP for airport noise ceteris paribus. Second, this study has also demonstrated that the implicit price for airport noise is impacted by the information available to buyers. In contrast to the impact of unobserved spatial heterogeneity on previous results in the airport noise literature, the information impact suggests previous results may be biased downwards.[51]

3.6 Conclusion

The conceptual framework in chapter 2 suggests that for a complex commodity such as a house, attribute quality may not be fully capitalized even when information about the quality of the attribute is considered publicly available. This outcome arises because some buyers may not be fully informed due to search and attention costs. In the housing market sellers are often more informed than buyers because of the information advantages that ensue from having lived in the house that is being transacted. If sellers

[51] Poor control of unobserved spatial heterogeneity and information available to buyers partially offset each other in how they bias hedonic estimates of the MWTP for airport noise. Some airport noise papers seem to recognize that this biasing affect may occur, but do not try to articulate the mechanism by which it occurs. For example, Espey and Lopez (2000) estimated the impact of airport noise on houses surrounding the Reno-Tahoe International Airport and found houses in the 65 Ldn contour sold for approximately 2% less than comparable houses. Yet they admit that their estimates may be biased downward because the airport was expanding rapidly and “such growth may not have been anticipated at the time of purchase”.

understand their information advantage, then they can change their offers to reflect the fraction of uninformed buyers in the market. Depending on the fraction of uninformed or under-informed buyers, these strategies may lead to an estimated implicit price that is attenuated towards zero; not capturing the MWTP of agents in the market.

The set of circumstances surrounding the RDU airport noise disclosure program provides a way to analyze the potential relevance of the conceptual framework. Sellers are likely informed about the airport noise around their homes whereas buyers may be initially inattentive and uninformed. After disclosures began taking place it is reasonable to assume that there was a reduction in the fraction of uninformed buyers. This shift was argued to have occurred primarily because buyers paid more attention to the noise issue. Using data on housing sales occurring near the airport that bracket the timing of disclosures, and after controlling for possible confounding influences to the quasi-random experiment, results suggest that the disclosure reduces the selling price of homes in the high noise zone by approximately 2-3 percent. This reduction *is in addition* to the approximate 4 percent decline that would be attributed to airport noise in the area *prior* to the disclosure.

Overall these results provide support for the intuition developed in chapter 2. Airport noise is a readily perceptible location specific disamenity and information about differences in noise is publicly available. The structure of the conventional hedonic would imply it should be capitalized in the price of a house even before the disclosure law. The robust estimate for the differential effect of the new disclosure policy is difficult to reconcile with search models. The acquisition of airport noise information was relatively easy before and after the disclosure. Thus it appears that an information

release that primarily causes market participants to pay attention to a relevant feature of the market can affect the price equilibrium.

Although the hedonic estimates developed in this chapter provide support for the primary prediction of the conceptual framework, these results do not provide a test of the explanation developed in chapter 2. Rather they are suggestive that buyer's attention and search constraints can impact market prices. Additional analysis on other information shocks that had large impacts on both initial attention *as well as* subjective expected benefits and costs will be necessary to further understand the impact of search and attention constraints on housing prices. Chapters 4 and 5 present two additional empirical examples using other quasi-random experiments of this sort.

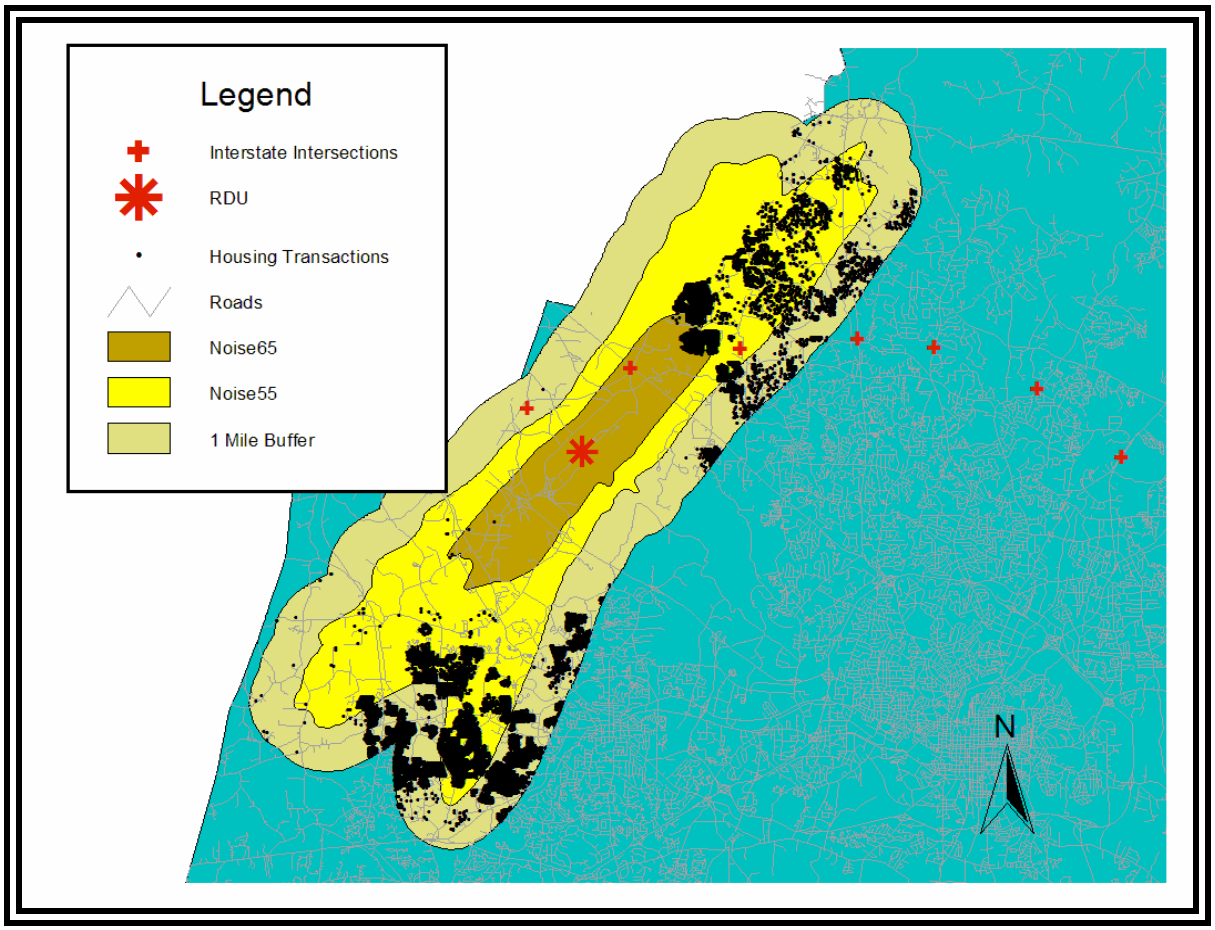


Figure 3.1: Housing transactions and disclosure zones

Table 3.1: Summary Statistics

Variable	Description	Mean	Median	Standard Deviation	Minimum	Maximum	Observations
lprice	Log of sale price of property	12.16	12.11	0.34	9.68	14.82	16900
age	Age of house in years	3.54	1	5.80	0	97	16900
baths	Number of bathrooms	2.85	2.50	0.65	1	7	16900
acreage	Lot size in acres	0.41	0.27	0.60	0.06	37.53	16900
regheatarea	Main heated living area in sq. ft.	2295.39	2243	646.83	620	5704	16900
detgarage	Detached garage dummy	0.01	0	0.12	0	1	16900
fireplaces	Number of fireplaces	0.98	1	0.21	0	4	16900
deck	Deck area in sq. ft.	186.26	168	142.10	0	1545	16900
sewer	Sewer availability dummy	0.902964	1	0.2960151	0	1	16901
flordum1	Hardwood floors dummy	0.01	0	0.10	0	1	16900
scrporch	Screened porch area in sq. ft.	16.91	0	58.03	0	1050	16900
walldum1	Brick walls dummy	0.05	0	0.22	0	1	16900
atticheat	Attic heated area in sq. ft.	56.83	0	163.32	0	1743.33	16900
bsmtheat	Basement heated area in sq. ft.	24.44	0	163.29	0	2690	16900
garage	Garage area in sq. ft.	432.82	462	184.79	0	1268	16900
poolres	Pool dummy	0.01	0	0.08	0	1	16900
bsmtdum1	Full basement dummy	0.04	0	0.20	0	1	16900
bsmtdum2	Partial basement dummy	0.02	0	0.14	0	1	16900
encporch	Enclosed porch area in sq. ft.	1.52	0	20.01	0	990	16900
opnporch	Open porch area in sq. ft.	54.43	30	74.83	0	1161	16900
condadum	House of "A" condition dummy	0.02	0	0.15	0	1	16900
condcdum	House of "C" condition dummy	0.01	0	0.08	0	1	16900
condddum	House of "D" condition dummy	0.00	0	0.03	0	1	16900
perc_no~1990	percent non-white	9.94	7.97	5.32	0	69.64	16900
medianvalu~t	Median house values	185838.00	183920	37864.09	99240	559810	16900
medttw_int	Median time to work	21.88	22	1.79	17	27	16900
perc_under~t	Percent population < 18	29.45	30.54	4.16	7.38	38.51	16900
perc_owner~t	Percent owner occupied housing	78.90	85.05	16.59	14.62	95.56	16900
nearestpark	Distance to nearest park	3.62	3.53	1.47	1.08	8.37	16900
nearestsc	Distance to nearest shopping cent	6.49	6.13	1.88	2.92	13.95	16900
taxrate	Property tax rate for area	0.38	0.54	0.24	0	0.60	16900
yrdum93_apr	April 1993 to April 1994 dummy	0.12	0	0.32	0	1	16900
yrdum94_apr	April 1994 to April 1995 dummy	0.12	0	0.32	0	1	16900
yrdum95_apr	April 1995 to April 1996 dummy	0.12	0	0.33	0	1	16900
yrdum96_apr	April 1996 to April 1997 dummy	0.13	0	0.33	0	1	16900
yrdum97_apr	April 1997 to April 1998 dummy	0.13	0	0.34	0	1	16900
yrdum98_apr	April 1998 to April 1999 dummy	0.14	0	0.35	0	1	16900
yrdum99_apr	April 1999 to April 2000 dummy	0.09	0	0.29	0	1	16900
yrdum00_apr	April 2000 to January 2001 dumm	0.05	0	0.23	0	1	16900
airport_dist	Linear distance to Airport	6.21	6.46	1.62	1.66	9.65	16900
new_sale	New property sale dummy	0.42	0	0.49	0	1	16901
north	Northern Region dummy	0.33	0	0.47	0	1	16901
interstate	Distance to interstate function	0.14	0	0.21	0	0.74	16901
noiseless6~m	In 55-65 dB zone dummy	0.47	0	0.50	0	1	16900
noise65_dum	In 65-70 dB zone dummy	0.05	0	0.22	0	1	16900
noiseless6~t	In 55-65 dB zone & sold with discl	0.19	0	0.39	0	1	16900
noise65_post	In 65-70 dB zone & sold with discl	0.03	0	0.16	0	1	16900

Table 3.2: Number of Observations for Key Dummy Variables

Variables	Number of Observations [1] (1992-2000)	[2] (1995-1998)	[3] (1996-1997)
noise65_dum	854	513	292
noise65_post	472	347	190
noiseless65_dum	7,866	3,790	1,857
noiseless65_post	3,174	2,035	973
Total # of obs. In Sample	16,900	8,873	4369

Table 3.3: Annual Passengers Traveling Through RDU

Year	Enplaned	Deplaned	Total
1985	1,381,798	1,389,211	2,771,009
1986	1,555,362	1,544,640	3,100,002
1987	2,428,101	2,425,972	4,854,073
1988	3,696,712	3,655,295	7,352,007
1989	4,318,325	4,276,346	8,594,671
1990	4,650,872	4,614,793	9,265,665
1991	4,698,543	4,683,043	9,381,586
1992	4,977,071	4,948,293	9,925,364
1993	4,862,285	4,833,601	9,695,886
1994	4,498,837	4,500,654	8,999,491
1995	2,962,701	2,974,434	5,937,135
1996	3,206,136	3,211,735	6,417,871
1997	3,360,478	3,364,396	6,724,874
1998	3,615,439	3,613,214	7,228,653
1999	4,471,065	4,470,710	8,941,775
2000	5,210,297	5,228,288	10,438,585
2001	4,806,513	4,777,474	9,584,087
2002	4,252,453	4,221,044	8,473,497

Notes: Statistics were acquired from the RDU website.

Table 3.4: Airport Noise Regression Results

Dep. Var. = lprice	OLS (1992-2000)	OLS (1995-1998)	OLS (1996-1997)	OLS using spatial dataset (1996-1997)	ML Rook Contiguity Lag (1996-1997)	ML Queen Contiguity Lag (1996-1997)
	[1]	[2]	[3]	[4]	[5]	[6]
Variable						
new_sale	-0.001	0.008	-0.002	-0.003	-0.006	-0.006
	[0.002]	[0.003]**	[0.004]	[0.004]	[0.003]	[0.003]
north	0.004	-0.010	-0.015	-0.029	-0.024	-0.024
	[0.008]	[0.011]	[0.015]	[0.015]*	[0.012]**	[0.012]**
five40_convex	-0.151	-0.140	-0.095	-0.092	-0.055	-0.055
	[0.015]**	[0.021]**	[0.028]**	[0.028]**	[0.023]**	[0.023]**
airport_dist	-0.011	-0.014	-0.014	-0.017	-0.012	-0.012
	[0.001]**	[0.002]**	[0.003]**	[0.003]**	[0.002]**	[0.002]**
noise55_dum	-0.023	-0.029	-0.026	-0.027	-0.019	-0.019
	[0.003]**	[0.004]**	[0.006]**	[0.006]**	[0.005]**	[0.005]**
noise65_dum	-0.051	-0.063	-0.078	-0.079	-0.038	-0.038
	[0.007]**	[0.010]**	[0.012]**	[0.012]**	[0.010]**	[0.010]**
noise55_post	0.012	0.013	-0.001	-0.002	-0.0002	-0.0002
	[0.004]**	[0.005]**	[0.007]	[0.007]	[0.006]	[0.006]
noise65_post	**-0.031**	**-0.030**	**-0.029**	**-0.029**	**-0.021**	**-0.021**
	[0.008]**	[0.012]**	[0.014]**	[0.014]**	[0.011]*	[0.011]*
Constant	11.110	11.348	11.339	11.375	6.694	6.694
	[0.019]**	[0.030]**	[0.042]**	[0.042]**	[0.109]**	[0.109]**
lag model parameter					0.403	0.403
					[0.009]**	[0.009]**
Structural & Neigh. Controls	X	X	X	X	X	X
Year Fixed Effects	X	X	X	X	X	X
Observations	16900	8873	4369	4250	4250	4250
R2 or Pseudo R2	0.89	0.88	0.89	0.89	0.93	0.93

Notes: Standard errors in brackets.
* significant at 10%; ** significant at 5%

Table 3.5: Spatial Statistics from Regression [4]

Statistic	Result	P-value
Rook Moran's I	37.37	2.2E-16
Queen Moran's I	37.37	2.2E-16
Robust LM (for error model)	405.48	2.2E-16
Robust LM (for lag model)	725.26	2.2E-16

Chapter 4: Flood Plains

4.1 Introduction

This chapter reports the effect of a flood disclosure on housing prices. The form of the disclosures allows consideration of both the attention and search dimensions of the conceptual framework. Flooding has caused more economic loss than any other natural hazard in the United States.[52] During the 1980's the average annual property loss from flood damage was 3.85 billion dollars. This aggregate loss increased to 6.44 billion dollars per year during the 1990's.[53] With other recent large flooding events, it appears flood damages may continue to increase for the current decade. There are two main types of floods: inland stream flooding and coastal flooding. Inland stream flooding accounts for the majority of the economic losses, whereas coastal flooding accounts for the majority of deaths. The focus of this chapter is on inland stream flooding which is a widespread phenomenon that affects properties throughout the United States.

Private homeowner insurance policies typically do not provide coverage for flooding. This omission is due largely to the fact that private insurance companies, facing adverse selection and an inability to spatially diversify their risks, have found it unprofitable to provide flood insurance services. Historically, as the number of homes being built in flood plains increased over the years, local governments and the Federal government began bearing more and more of the social costs related to floods. Those costs included: (i) emergency services costs, (ii) loss of taxes due to damaged and lost

[52] GAO Report (2003).
[53] Dollar amounts have been adjusted to 2002 dollars. The information was obtained from the National Weather Service at http://www.nws.noaa.gov/oh/hic/flood_stats/Flood_loss_time_series.shtml.

businesses and homes, and (iii) other post-disaster relief spending such as loans, emergency housing, food and clothing.[54] Recognizing these costs and the lack of a private market solution, Congress enacted the "National Flood Insurance Act" (NFIP) in 1968. One of the primary goals of the act was to provide homeowners in participating communities that met certain standards, the option of buying flood insurance.[55] However, because of low participation rates, the "Flood Disaster Protection Act" of 1973 was enacted requiring purchase of flood insurance for property owners taking out loans with federally regulated lending institutions.

Since 1973, the NFIP along with the Federal Emergency Management Agency (FEMA), who has responsibility to map the flood zones, has become one of the biggest domestic programs (and liabilities) in the United States.[56] As of fiscal year 2004, there were approximately 4.5 million flood insurance policies in over 20,000 participating communities representing approximately 764 billion of insured property.[57] Yet even with the mandatory insurance requirement, many property owners have been able to drop their policies after purchasing insurance at the time of financing a home. The lenders involved often fail to ensure that insurance is maintained.[58] As a result, it appears that potential buyers of homes in flood zones may often be unaware that a house has

[54] See King (2005) for more details.
[55] A community must adopt the NFIP's minimum floodplain management standards in its local ordinances. They must also have Flood Insurance Rate Maps (FIRM) created by FEMA and use the maps to guide the community's land-use development plan.
[56] This is along with Social Security and Federal Health programs such as Medicare and Medicaid. See King (2005).
[57] See the statistics provided at the FEMA website: www.fema.gov/nfip/pcstat.shtm.
[58] In 1994 the Riegel Community Development Regulatory Improvement Act was passed that required mortgage lenders to ensure that flood insurance policies were taken out for homes in certain FEMA flood plain zones. The final regulations were issued in October of 1996. Since this time period participation rates in NFIP have risen but it does not appear that there is anywhere near full compliance (see again King (2005).

substantial flood risks and is subject to mandatory flood insurance requirements. Writing before flood insurance became mandatory in 1973, Krutilla (1966) noted that one of the potential benefits of compulsory flood insurance would be that,

> "*…premiums proportional to risk and equal to both the private and social cost of flood plain occupance will serve as a rationing device eliminating economically unwarranted uses of flood plain lands on one hand, while not prohibiting uses for which flood plain location has merit on the other hand.*"

Yet for Krutilla's efficiency scenario to take place via the NFIP, buyers need to be informed about the flood risks and insurance premiums for a property they are considering purchasing.

A survey of buyers' knowledge about flood risk and insurance premiums was conducted by Chivers and Flores (2002) in Colorado. They asked homeowners that had recently purchased homes in a FEMA flood zone that required flood insurance: "When did you first learn of the potential flood risk associated with your home?" They found that 8% learned prior to making an offer, 6% prior to closing, 60% during closing, 4% after moving, 6% after being flooded, and 16% at some other time. This survey indicates that most buyers find out about flood risk and insurance premiums during closing, well after they have made an offer on a house. Chivers and Flores note that it is very difficult for buyers who learn this information during closing to back out of the offer because they face intense pressure from the other parties involved (real estate agents, sellers, mortgage lenders, etc.) who have a financial stake in assuring the transaction is completed.

The results from this survey are quite striking since the information about flood insurance rates and the maps designating flood zones are publicly available. The survey suggests that when faced with the complexity of the home buying decision, buyers may

not fully consider the flood plain attribute of a house. Furthermore it suggests that the housing institutions have incentives not to provide buyers with the flood information early on in the home buying process because it may affect the buyer's purchase decision, imposing unwanted costs on the housing institutions.[59] Sellers on the other hand are likely better informed about flood risk after living in the house for a time because of information given them by neighbors or local government. They may also be better informed or have more accurate perceptions of the importance of flood risk for their property due to observing floods near their homes. However, the influence of previous flooding on seller's information and perceptions of flood risk will depend on the frequency of flooding in the area.

Given the discussion above, the flood plain housing attribute appears to be another good candidate for testing the predictions of the conceptual framework developed in chapter 2. A disclosure provided to buyers by sellers in Wake County North Carolina about whether or not their home lies in a flood plain is used as the "information shock" for this application. This information shock is similar to that of the airport noise disclosure in many ways. They both: (i) represent plausibly exogenous information shocks, (ii) sellers are more informed than buyers, (iii) housing market institutions do not seem to fully correct the asymmetry, (iv) the disclosure is given at a "salient" time and place—to buyers before they make an offer on a house, (v) the attribute is moderately well defined—both noise and flood risk have agreed upon measures of severity (although risk is not as well defined as noise), and (vi) if buyers did not consider the attribute before, the disclosures seem likely to affect their "initial attention." Despite these many

[59] For lenders this suggests that the value of completing the transaction overrides the default risk cost.

similarities in the information environment surrounding both disclosure programs, they likely differ in one key regard. While the airport noise disclosure had no impact on subjective expected cost calculations and perhaps some impact on subjective expected benefit calculations of information search, the flood plain disclosure likely had substantial impact on both subjective expected costs and benefits.

When a seller marks "yes" on the disclosure form that his house is in a flood zone, this certainly changes a buyer's subjective expected costs. In fact, it partially eliminates these costs because the seller *gave* the information to the buyer. Furthermore, by seeing that the seller marked "yes" on the form, this should immediately revise a buyer's subjective expected benefits to search out any additional information. Additional information that a buyer may search for include: flood insurance premium rates and information that would help to form a more accurate measure of the probability and consequences of the flood risk.[60] Thus the attribute is different from the airport noise disclosure application whose primary focus was to cause buyers to pay attention to the issue but not provide concrete information. In that case it was argued that the disclosure had little impact on subjective expected benefits and costs. The flood plain disclosure on the other hand *does* provide concrete information about the housing attribute in question and will therefore have a greater impact on buyer's information acquisition process.[61]

---

[60] Although the airport noise attribute and this flood risk attribute of housing are considered "well-defined", the flood attribute being a risk is certainly more complex than airport noise. Both the probability of flooding and the consequences of flooding must be assessed by a buyer.

[61] Of course a direct comparison cannot be made because the airport noise and flood plain applications focus on two different attributes. A true comparison of the impact of information disclosures that affect information acquisition differently would need an application where different types of information disclosures occurred *for the same housing attribute*.

The findings for Wake County suggest that houses sold for approximately 4.5% less in the post-disclosure time period. This is after temporally bounding housing data used in the analysis to a short interval before and after the disclosure began taking place, and after controlling for other potential spatial confounders. The analysis suggests that houses in FEMA flood zones were not being discounted by the housing market in the pre-disclosure time period. Thus a "naïve" regression using all of the data but not taking into account the information shift would have estimated the flood plain impact to be approximately half of the "full information" estimate after disclosures began taking place.

The remainder of the chapter will proceed as follows. Section 3.2 provides a literature review of the impact of flood plains on property values, and additional background on the North Carolina disclosure program. Section 3.3 describes the data used in the analysis. Section 3.4 outlines the identification strategy used for the hedonic price regression analysis. Section 3.5 describes the results of the hedonic price regressions. Section 3.6 concludes the chapter.

4.2 Background

4.2.1 Background on Previous Flood Plain Hedonic Literature

There is a small body of research that has examined the impact of flood plain designation on housing prices.[62] One would expect there to be a price discount associated with a house being located in a flood plain relative to comparable houses. Nonetheless the available studies provide mixed results. As was suggested by Chivers

[62] Early examples include Muckleston (1983), Holway and Burby (1990), Speyrer and Ragas (1991) and Tobin and Montz (1994).

and Flores (2002), one reason for these findings is incomplete information about flood risks and insurance premiums. Two recent articles by Bin and Polasky (2004) and Hallstrom and Smith (2005) have explored whether or not flooding events caused by hurricanes convey this needed information to buyers.

In their study Bin and Polasky used Hurricane Floyd as a quasi-random experiment. Hurricane Floyd caused substantial flooding in Pitt County, North Carolina in September of 1999. Using data on single-family residential sales between 1992 and 2002 in the county, Bin and Polasky statistically compared houses in FEMA designated flood zones before and after hurricane Floyd using hedonic price regressions. The results from these regressions suggested a price discount for houses located in flood zones. They also found that this price discount increased substantially after hurricane Floyd.

While their analysis takes advantage of the potential for a quasi-random experiment provided by nature (which is presumably exogenous), there may be confounders to their identification strategy. Bin and Polasky's primary regression specification takes the following form,

$$\begin{aligned}\mathbf{lprice} = \beta + \alpha\mathbf{Structural} + \phi\mathbf{Townships} + \psi\mathbf{Distances} \\ + \theta\mathbf{Flood} + \delta\mathbf{Floyd} + \omega\mathbf{Flood * Floyd} + \varepsilon\end{aligned} \tag{1}$$

where "lprice" is the log of the housing transaction prices deflated by a consumer price index over the 10 year time period, "Structural" is a set of structural characteristics of houses in the sample, "Townships" are 15 dummy variables for the townships in the county, "Distances" are distances to various neighborhood features including nearest

creek or stream[63], "Flood" indicates if a house is in a flood zone, "Floyd" indicates if the house sold after Hurricane Floyd, and "Flood*Floyd" indicates houses in a flood zone that sold after Hurricane Floyd. Also, ε is an error term and β, θ, δ, and ω are parameters to be estimated and α and ϕ represent vectors of parameters to be estimated.

A concern with Bin and Polasky's identification strategy is that it provides limited temporal control. Instead of including dummy variables to control for unobserved year specific events, housing prices were deflated using an unreported consumer price index. One concern is that omitted trends or other events correlated with the timing of the hurricane over this 10 year period may lead to violations of the key identifying assumption that is central to their measure of the impact attributed to the flood risk information conveyed by Floyd. Another potential confounder is that after the substantial flooding caused by the hurricane, buyers are likely to respond to flood damaged houses and this would also be reflected in the information impact estimates if unaccounted for in the identification strategy.

Hallstrom and Smith (2005) use Hurricane Andrew and a different identification strategy to estimate the impact of the risk information conveyed by the hurricane on housing prices. Hurricane Andrew, one of the strongest hurricanes ever to hit the United States, passed through the lower part of the Florida panhandle in 1992. The areas directly hit by the hurricane experienced significant damage. Lee County, Florida just missed being hit by the storm. For approximately 20 years prior to Andrew, the Florida Gulf coast which includes Lee County had had below normal hurricane activity. Hallstrom

[63] The distance to nearest creek or stream variable was used to distinguish the amenity affect of creeks and streams from the flood risk affect on housing prices.

and Smith hypothesized that although not hit directly, the housing market in Lee County might adjust after observing the devastation of Hurricane Andrew in nearby counties.

Using a housing database of transactions occurring between 1993 and 2000 that also provided each house's previous transaction, Hallstrom and Smith were able to construct a set of price differences for these houses and conduct a repeat sales analysis.[64] By exploiting the variation in the timing of the bracketed housing prices, they estimated that the information acquired by buyers and sellers in the housing market in Lee county from hurricane Andrew lead to a 19 percent decline in housing prices in FEMA flood zones. Although their identification strategy avoids some of the temporal and damage related issues of Bin and Polasky, it does rely on the assumption that there were no other omitted trends or events that might drive any difference in flood plain housing prices otherwise attributed to Hurricane Andrew over a very lengthy time period.[65] As will be seen, the identification strategy in this paper relies on a much shorter time horizon.

4.2.2 Background on the Study Area

As with the airport noise application, Wake County North Carolina is the geographic setting for the analysis in this chapter.[66] Approximately 44% of the land in the county is forest, 32% developed residential areas, 12% agriculture, 8% miscellaneous other uses and 4% water surface area. The county lies between North Carolina's coastal

[64] One advantage of the repeat-sales approach in this context is that the amenity affects of the coast and the risk affects from floods can be more clearly separated.
[65] Hallstrom and Smith dropped houses whose first sale occurred before 1982 from their sample. This suggests that their housing dataset is composed of transactions that occurred between 1982 and 2000. Although they included two time trends (trend since hurricane and trend between sales), whether or not the hedonic is stable or if there are other temporal confounders not explicitly controlled for in their analysis that occur during this 18 year window are concerns for their identification strategy.
[66] However, the housing samples used in these two applications are different and this will be discussed in more detail in section 4.3.

zone to the east and the "piedmont uplands" to the west. The lowest point in the county is 160 feet above sea level whereas the highest point is 540 feet above sea level. Approximately 80% of the county is drained by the Neuse River and its tributaries. The other 20% of the county is drained by the tributaries of the Cape Fear River. Figure 4.1 shows the location of Wake County in relation to North Carolina and the coast.

As with other inland stream networks throughout the U.S., there are periodic floods that occur in Wake County. Two of the most severe flooding episodes occurred when hurricanes Fran and Floyd past through the area on September 5, 1996 and September 16, 1999 respectively.[67] However, flooding also occurs when there is substantial rainfall in the area over a short period of time. Unlike in coastal areas, it is very difficult to visually determine whether or not a house lies in a flood plain. Therefore one must rely primarily on FEMA mapping information to determine flood risk for a given house.[68]

Recognizing the complexity and cost of information acquisition to buyers in the real estate market, the state of North Carolina and other states have recently decided to make seller disclosure of flood plain information and other housing attributes mandatory. In 1995 North Carolina General Statute 47E known as the "Residential Property Disclosure Act", was passed requiring owners of residential real estate to provide prospective buyers with a property disclosure statement after January 1, 1996.[69] The last of 20 questions on the disclosure form that sellers were required to provide buyers was

[67] The timing of hurricane Fran in relation to the information shock used in this study is important and will be discussed in section 3.4.
[68] This comes from personal communication with real estate agents in the county.
[69] There are some exempted residential properties such as those sold for the first time and those that are leased with an option to purchase.

the question "do you know of any FLOOD HAZARD or that the property is in a FEDERALLY-DESIGNATED FLOOD PLAIN?" (Capitalized words are the same as in the disclosure). Appendix 4.A shows the complete 2 page disclosure. This disclosure offers the potential for a quasi-random experiment to identify the impact of the release of information on flood plain zones on the selling prices of houses in these zones.

Much like the airport noise disclosure discussed in Chapter 3, it does not appear that sellers knew in advance about the residential disclosures that would be required of them. A search of the primary newspaper in the region found no articles describing the residential property disclosure prior to its implementation.[70] This suggests that from the perspective of buyers and sellers that the seller disclosure was an exogenous shock to their decision environment. Thus sellers whose houses were located in a flood plain were likely unable to put their house on the market "early" to avoid the disclosure.

Buyers who failed to receive a disclosure statement before making their offer on a house after January 1, 1996 have the opportunity to cancel the contract and receive a refund for their deposited money. Conversations with real estate agents in Wake County, suggests that the disclosure is in fact given before an offer is made. Furthermore, it appears to be the convention in Wake County to provide the disclosure to a buyer when they visit the house for the first time with their agent.

As was noted in the introduction, it was possible for buyers to obtain the FEMA flood zone designation and insurance premium information on their own without the aid of this disclosure. The primary ways of doing this was by contacting their mortgage

[70] Electronic searches in "The News and Observer" were conducted for the 2 years prior to the implementation of the residential disclosure.

lender or by looking for the information on the surveyor's map of the property. But as suggested by the survey in Chivers and Flores (2002) as well as in conversations with the real estate agents in the area, many buyers fail to make use of these publicly available resources.

4.3 Data Used in the Analysis

There are three primary datasets used in the analysis. The first is a dataset of single-family housing transactions that overlap the beginning of the seller disclosures on January 1, 1996 in Wake County North Carolina. This is the same housing dataset previously described and used in chapter 3. Among other things, this data contain a detailed set of structural and neighborhood characteristics derived from the Wake County Revenue Department, The U.S. Census Bureau, and other GIS shapefiles at the county level.[71] Although the names of the structural and neighborhood characteristics and the compilation of the housing dataset are the same as that used in chapter 3, different spatial and temporal dimensions of the data are used more directly for the present analysis. Here transactions throughout the county are used whereas in chapter 3 only the housing transactions in the vicinity of the airport were used in the analysis. Also, in this chapter only transactions that occurred between Jan 1, 1995 and September 5, 1995 and transactions occurring between Jan 1, 1996 and September 5, 1996 are used for reasons to be detailed in section 4.4.

The second dataset is a GIS shapefile of the FEMA flood zone designations. FEMA issues flood maps for each county that participates in the NFIP. For Wake

[71] For more detail on the formation of this dataset refer to chapter 3.

County there are two primary flood designations that are used in the analysis. Houses in "AE" zones are those that FEMA has determined have a 1% or greater annual probability of flooding.[72] Houses in "X500" zones are those that FEMA has determined have between 0.2% and 1% annual probability of flooding. These flood zone designations were spatially matched to the houses falling in these zones and dummy variables labeled "zoneAE" and "zoneX500" were created. Figure 4.2 shows these zones in relation to the housing transactions in Wake County.

The third dataset provides information on the major hydrologic units in the county. Using this dataset the distance from each house to the edge of the nearest major stream or lake was calculated. This new variable which was named "hydro_ft" will be used to help control for the amenity affect of water bodies in the area. Figure 4.3 shows the major hydrologic lines in the county. These boundaries lie in the same hydrologic regions as the FEMA flood zones.

Table 4.1 provides descriptions and summary statistics for the housing variables used in the present analysis.[73] These are different from what is reported in chapter 3 because the sample of transactions encompasses the entire county rather than just the transactions in the vicinity of the airport. The average house in the dataset was approximately ten years old, had two and a half baths, and sold for approximately $150,000. Notice that approximately 1.2% of houses (178 observations) transacted

[72] There are also a small number of houses that fall in FEMA designated "A" zones. These are similar to the "AE" zones in that they have a similar probability of flooding. In the analysis these zones are combined with the "AE" zone designation.

[73] The block-group and census tract dummies have been excluded and a few of the listed variables will be explained in section 4.4.

occurred in an AE flood zone and less than 1% in the X500 zone (68 observations).[74] Post1996 in table 4.1 is a dummy variable equal to 1 for all houses transacted from January 1, 1996 to September 5, 1996. Approximately 53% of the transactions in the dataset took place during this time period. Dummy variables were also created for two different census neighborhood designations—block groups and census tracts—but are omitted from table 4.1 because of space constraints.

4.4 Identification Strategy for the Hedonic

4.4.1 Initial Identification Strategy

The strategy to identify the impact of flood plain disclosure on housing prices relies on two key sources of variation. The first is a discrete temporal shift in flood plain information that occurred when the NC Residential Disclosure Statement began being used on January 1, 1996. The second source of variation is spatial variation in flood risk derived from the FEMA flood plain map in Wake County that identifies areas where sellers are required to provide flood plain disclosure. Using these two sources of variation together with the housing dataset, a simple fixed-effects estimation strategy using a semi-log specification for the price function yields equation (2),

$$\begin{aligned} \mathbf{lprice} = \beta + \alpha\mathbf{Structural} + \phi\mathbf{Neighborhood} \\ + \delta\mathbf{Postdisclosure} + \omega\mathbf{Spatial\_dums} \\ + \theta\mathbf{ZoneAE\_dum} + \gamma\mathbf{ZoneX500\_dum} \\ + \lambda\mathbf{ZoneAE\_post} + \psi\mathbf{ZoneX500\_post} + \varepsilon \end{aligned} \tag{2}$$

where the “Structural”, “Neighborhood”, “Zone__” variables are defined as in table 4.1, “Postdisclosure” is a dummy variable designating houses that sold after January 1, 1996,

[74] 101 of the AE observations and 44 of the X500 observations occur in the post-disclosure time period.

"Spatial_dums" are dummy variables for spatial neighborhoods such as census tracts or block groups, ε is an error term, β, δ, θ, γ, λ, and ψ are parameters to be estimated and α, ϕ, and ω represent vectors of parameters to be estimated.

In experimental terminology, houses in the AE flood plain zone have been "treated" by the disclosure. Houses in the X500 flood plain zone likely received no disclosure but still have flood risk and therefore act as a potentially interesting comparison group.[75] Other houses in the county outside of the flood risk zones (but perhaps near streams and lakes) serve as the primary control group. Using the estimating equation above, the coefficient on ZoneAE_post (houses in the AE zone that sold after disclosures began) is intended to provide estimates of the impact of the flood plain disclosure on housing prices relative to houses in the X500 zone and other houses in the county.

4.4.2 Identification Strategy that Controls for Known Confounding Influences

As discussed in Meyer (1995) there is always the prospect for confounders in applying the quasi-random experiment methodology. Omitted variables bias can arise from both the time and space dimensions. One confounding temporal event that occurred in Wake County that likely affected both the housing market and perceptions about flood risks was Hurricane Fran. Hurricane Fran passed directly over the area on September 5, 1996 with wind gusts between 80 and 90 mph and with over 10 inches of rain for most of the county. This storm caused substantial damage to developed and undeveloped

[75] "Being in a FEMA flood zone" as asked by the NC residential disclosure statement is a designation typically given only the "A" and "AE" zones, not the X500 zone.

properties throughout the area. In fact, there was over 900 million dollars (in 1996 dollars) of reported damage to residential and commercial property due to the hurricane in the county.

It seems reasonable that homes in flood plain zones were most likely to have been damaged by the severe flooding caused by the hurricane. Without any record of which homes received structural damage from floods, this could potentially bias estimates generated from equation (2) if the "Postdisclosure" time period were to overlap September 5, 1996. The hurricane also acts as another type of information treatment that could confound the disclosure treatment. To avoid these confounding influences of Hurricane Fran, one could limit the analysis to a symmetric set of housing transactions that occurred between Jan 1, 1995 and September 5, 1995 (pre-disclosure time period) and transactions occurring between Jan 1, 1996 and September 5, 1996 (post-disclosure time period).[76] This sample restriction avoids the confounding influence of the hurricane and it may also eliminate other potential temporal confounders that could affect analyses utilizing a larger temporal window (as in Hallstrom and Smith (2005)).

Another confounder, highlighted in the Bin and Polasky (2004) study, is the amenity effects of proximity to streams and lakes. Thus location in a flood plain may actually partially proxy for this amenity. To avoid a bias due to this possibility, the major hydrologic units in the county were obtained and the linear distance from each house to the nearest major lake or stream was calculated. Because the benefits associated with being near streams and lakes likely dissipate the further the location of a house, the

[76] January 1, 1995-September 5, 1995 are used instead of May 1, 1995-January 1, 1996 to control for seasonal variation in housing prices.

function given in equation (3) was used to control for proximity effects for the house in relation to streams and lakes:

(3) $$\textbf{hydro\_dist} = \max\left[1 - (d / d_{\max})^{1/2}, 0\right],$$

where d is distance in miles and $d_{\max}$ is a benefit cutoff point set to 1/4 mile for this application. This function creates a convex index between zero and one.

Another potential omitted variable problem may occur if for some reason the type of housing sold during the pre-disclosure and post-disclosure time periods differed substantially. This source of error is also potentially important because housing types may be correlated with other structural disclosures included on the residential disclosure form. A confounder of this type might be new houses built in flood plain zones during the time period included in the dataset. If there is something unique about this new housing stock built in flood zones relative to the "control" houses over this time period, this change in the housing stock could lead to misinterpreting the effects attributed to flood plains. To avoid this bias, a dummy variable was created for houses that sold in the same year that they were built. Although it is possible that houses sold more than once in the year after they were built, this should act as a reasonable proxy for newly built homes.

Table 4.2 provides summary statistics of the structural and neighborhood variables for houses inside the AE zone and outside of the AE zone for the pre and post-disclosure time periods. Although housing in the AE zone is on average slightly smaller and less expensive, there do not appear to be any major changes across the pre and post disclosure time periods. To test this more formally a multinomial logit regression was performed on the three main sample types: 1) housing transacted inside the AE zone

before the disclosure, 2) housing transacted inside the AE zone after the disclosure and 3) housing transacted outside of the AE zone. Included in the regression were the primary structural characteristics commonly used in hedonic analyses and the omitted category was sample type 3) above.

Table 4.3 reports the results of the regression. Columns [1] and [2] correspond with sample types 1) and 2) above. Column [3] reports the results of Chi-squared tests for differences between the models in columns [1] and [2]. It appears that houses in the AE zone have larger acreages relative to houses outside the AE zone. Also, houses inside the AE zone transacted after the disclosure have more bathrooms and smaller garage space and this is statistically different from sample 1) as suggested by the chi-squared tests. However, as long as there is sufficient variation in baths and garage space in the full sample of transactions to identify their marginal impacts, these differences should pose no serious problems to the analysis. An overall test of differences between models [1] and [2] is not significant at the 5% level. There appears to be no difference among the three samples for the "new_sale" variable. This result suggests that the concern about changes in housing composition do not appear to be warranted.

Equation (2) can be updated to include the variables used to control for spatial influences and to restrict the time interval to the year before and after disclosures began taking place in the following way,

$$
\begin{aligned}
\mathbf{lprice} = \beta + \alpha\mathbf{Structural} + \phi\mathbf{Neighborhood} \\
+ \delta\mathbf{Post}1996 + \omega\mathbf{Spatial\_dums} \\
+ \theta\mathbf{ZoneAE\_dum} + \gamma\mathbf{ZoneX500\_dum} \\
+ \lambda\mathbf{ZoneAE\_post} + \psi\mathbf{ZoneX500\_post} \\
+ \chi\mathbf{New\_sale} + \vartheta\mathbf{Hydro\_dist} + \varepsilon
\end{aligned}
\tag{4}
$$

where “Post1996” is a dummy for houses that transacted between January 1, 1996 and September 5, 1996, “New_sale” is a dummy for new housing and “Hydro_dist” is the distance measure described by equation (3).

Although the empirical specification in equation (4) represents a detailed analysis of potential unrelated features varying in time and space that could bias estimates, there nonetheless may still be unobserved spatial heterogeneity that could affect estimates of the impact of the flood disclosure on housing prices. In the airport noise application a spatial regression analysis was performed that took into account spatial correlation among nearby neighbors. That analysis seemed reasonable for the application given the spatial extent of airport noise. However, a spatial regression analysis appears to be less reasonable for the case of flood zones which are much smaller than the noise contours examined in the previous application. This is because a spatial regression analysis may substantially alter an estimate by blurring the disclosure impact across the more discontinuous flood zone lines.

An alternative procedure to address unobserved spatial correlation is used for this application. This is accomplished by changing the “control” group from all the housing observations in the county outside of the X500 and AE zones, to only those observations that are “near” the AE zones. Regressions can then be performed following the specification developed in equation (4) where the housing transactions are limited to those that are within 0.3, 0.2 or 0.1 miles of an AE flood zone boundary. As a final precaution, the standard errors from estimates of equation (4) are clustered at the census tract or block group level. These specifications provide the greatest degree of control for features in time and space that could potentially bias estimates of the impact of flood

plain disclosure on housing prices. The next section describes estimated results from this hedonic regression specification.

4.5 Results

4.5.1 Primary Results

Table 4.4 presents estimates of the key parameters generated from three hedonic regression specifications. All three specifications use housing transactions that occurred between Jan 1, 1995 and September 5, 1995 (pre-disclosure time period) and transactions occurring between Jan 1, 1996 and September 5, 1996 (post-disclosure time period), to avoid the confounding bias of Hurricane Fran and to focus on a limited time span that brackets the start of disclosures on January 1, 1996. Each regression also controls for structural and neighborhood characteristics. Parameter estimates for the structural and neighborhood characteristics were consistent with a priori expectations and are reported in appendix 4.B.

The estimation results reported in Column [1] correspond with the specification developed in equation (4) except that the "ZoneAE_post" and "ZoneX500_post" variables are omitted and the "Spatial_dums" are 103 census tract dummy variables.[77] The estimate on "Post1996" is positive and statistically significant suggesting that nominal housing prices have increased by approximately 3% in Wake County in the year between the two temporal groupings of housing data.[78] The coefficient on "Hydro_dist"

[77] In the 2000 census, there are 104 designated census tracts in Wake County.
[78] These parameter estimates and all other parameter estimates on dummy variables reported in this section have been corrected using the Halvorsen and Palmquist (1980) correction for interpreting dummy variables in semilogarithmic equations.

is positive but not statistically significant. The coefficient on "zoneAE_dum" suggests that houses in the FEMA designated AE flood zone sold for approximately 2% less than other comparable homes outside of the flood zone area but is not statistically significant at conventional levels. It appears there is no price impact on houses located in the X500 flood zone.

Column [2] reports results of a regression that follows the same specification as Column [1] with the addition of the "zoneAE_post" and "zoneX500_post" variables. It is important to note that sellers likely marked "yes" on the disclosure statement if their home fell in an "AE" zone, but were not as likely that they marked "yes" if the home fell in an "X500" zone. This is because houses in the X500 zone are not required to purchase insurance and are typically not considered to be a FEMA flood zone. The coefficient on "zoneAE_post" suggests that the disclosure did impact housing prices in the flood zone by approximately 4.3%. The coefficient on "zoneAE_dum" in column [2] is approximately zero and insignificant suggesting that all of the negative impact of the AE flood zone on housing prices that was estimated in Column [1] is due to houses that sold in the January 1, 1996 to September 5, 1996 time period. Although the parameter estimate on "zoneX500_post" is negative, it is statistically insignificant. The estimates for "Post1996", "Hydro_dist", as well as the estimates for other structural and neighborhood characteristics reported in appendix 4.B are virtually unchanged between columns [1] and [2].

As a check on the sensitivity of the estimates reported in column [2] to the census tract dummies used as spatial controls, column [3] re-estimates the same specification but in this case includes 252 block-group dummy variables instead of the census tract

dummies.[79] If the estimate on "zoneAE_post" was weakened due to disaggregating the spatial fixed effects from 104 (census tracts) to 253 (block-group) zones, then this change would cause concern about the identification strategy. However, as can be seen in table 4.4 the estimate for zoneAE_post is essentially the same and other coefficients are also stable.

Columns [4]-[6] present estimates from the same specification as column [3] except that housing transactions are limited to those that are within 0.3, 0.2 or 0.1 miles of an AE flood zone boundary in an effort to control for omitted spatial heterogeneity. The estimate on zoneAE_post in these regressions is basically unchanged. The estimate corresponding to the most restrictive zone is given in column [6]. In this case only housing transactions within a tenth of a mile are used as the control group to the X500 and AE zones. These results suggest a 4.9% decrease in housing prices in the AE zone after the flood zone disclosure.

4.5.2 Discussion and Comparison of Results with Flood Insurance Costs

The conceptual framework developed in chapter 2 argued that when the fraction of uninformed buyers for a housing attribute is non-trivial, estimates of MWTP for the attribute generated from a hedonic price regression will likely be attenuated towards zero. The results from this analysis suggest that the fraction of uninformed buyers was substantial prior to the flood plain disclosure. The flood plain disclosure likely affected all three hypothesized information acquisition pathways—initial attention, subjective expected benefits, and subjective expected costs. The housing price impact, after sellers

[79] In the 2000 census, there are 253 designated block-groups in Wake County.

were required to provide a disclosure about FEMA flood zones, is consistent with the conceptual framework. Prior to the disclosure the price impact of a house in an AE flood plain was approximately zero, afterwards the estimate was in the range of 3.8% to 4.5% (percentages are derived by adding the coefficients on the zoneAE_dum and zoneAE_post variables). This post-disclosure estimate may be a closer approximation to the "true" value to avoid flood zones when there is full information in the housing market.[80]

The estimated price impact of 3.8% to 4.5% to avoid the AE flood plain represents a price discount in the range of $5,434 to $6,435 for an average priced house in an AE zone.[81] To put this price discount in context, information on typical flood insurance rates during the 1996 time period was gathered. FEMA produces detailed flood insurance manuals to help those that write flood insurance policies. A host of variables enter into a flood insurance rate for a given house such as (i) whether or not the structure was built before or after the flood insurance rate maps were established for the area (called Pre-FIRM and Post-FIRM properties), (ii) the structures elevation in relation to the base flood elevation, and (iii) whether or not there is an enclosure at the base of the property or other built-in flood mitigators.

Abstracting from the multitude of factors that determine a flood insurance rate, an approximation of insurance rates can be obtained using sample flood insurance rates for a "typical house" provided by FEMA. The rates are for flood insurance coverage of a house valued at $100,000 and $25,000 of contents. The annual policy costs in 1996 for a

[80] MWTP is not used here to describe estimates because there is not really a margin of flood risk, rather a discrete shift of being in or out of a flood zone.
[81] The average priced house in an AE zone was $143,000.

Post-FIRM house in the North Carolina Region, in an "AE" zone, whose elevation of its lowest floor is three feet above the base flood elevation estimated by FEMA, was $211.00.[82] The annual policy cost for the same house whose lowest floor is one foot *below* the base flood elevation is $1,114.50. The present value of these two policies assuming a 6% discount rate in perpetuity is $3,517 and $18,575 respectively. Therefore, the estimated flood insurance risk premium of $5,434 to $6,435 falls on the lower end of the flood insurance cost range, but appears to be reasonable.

4.6 Conclusion

Inland stream flooding is one of the most costly reoccurring natural disasters in the United States. One of the primary costs of these floods is the damage they cause residential housing that has been built in flood plains. In this chapter a flood plain disclosure that likely influenced all three information acquisition channels was used as a quasi-random experiment to test the impact of buyer's information about flood zones and flood risk on housing prices. Beginning January 1, 1996 homeowners in Wake County, North Carolina were required to provide a seller disclosure to buyers which included a question about whether or not the house fell in a FEMA designated flood zone. Using housing data that symmetrically bracketed the disclosure time period but omitted transactions that would have been impacted by Hurricane Fran, it was estimated that the disclosure reduced housing prices in FEMA designated "AE" zones by approximately 3.8% to 4.5%. For the average priced house in "AE" zones this discount represents a

[82] These dollar amounts are in 1996 dollars. Historical flood insurance rates are extremely difficult to acquire. I am indebted to Philip Letsinger of the NC Floodplain Management Branch for finding and providing me with these historical sample flood insurance rates.

$5,434 to $6,435 reduction. When compared with the costs of insuring properties in these flood prone areas, this range of estimates falls within reasonable bounds of insurance costs.

Like the airport noise application presented in chapter 3, it appears that a simple information disclosure that re-provides information that is already publicly available, can impact housing prices by increasing the fraction of informed buyers in the market. These examples are supportive of the conceptual framework developed in chapter 2. It should be noted that both the airport noise and flood plain examples represent housing attributes where quality is at least moderately well defined and measured. Furthermore the disclosures used in these two examples were consistently given to homebuyers during a salient time and at a salient place—to homebuyers before an offer is made on a house. Chapter 5 describes an application about the residential locations of sex offenders, that further explores these two dimensions of public information as it relates to the hedonic price equilibrium.

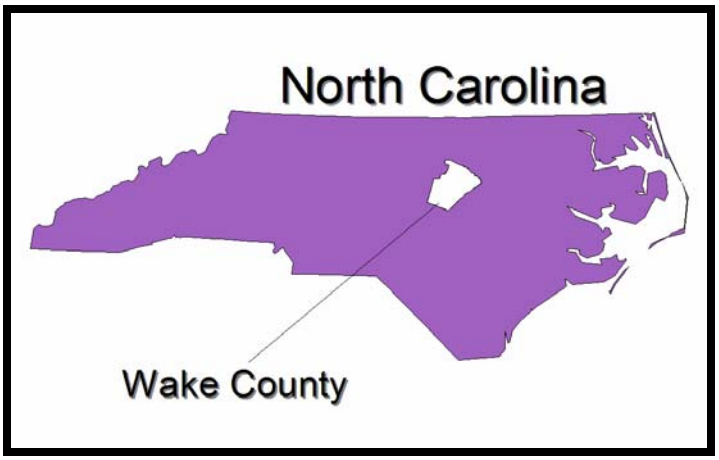


Figure 4.1: Location of Wake County in Relation to N.C.

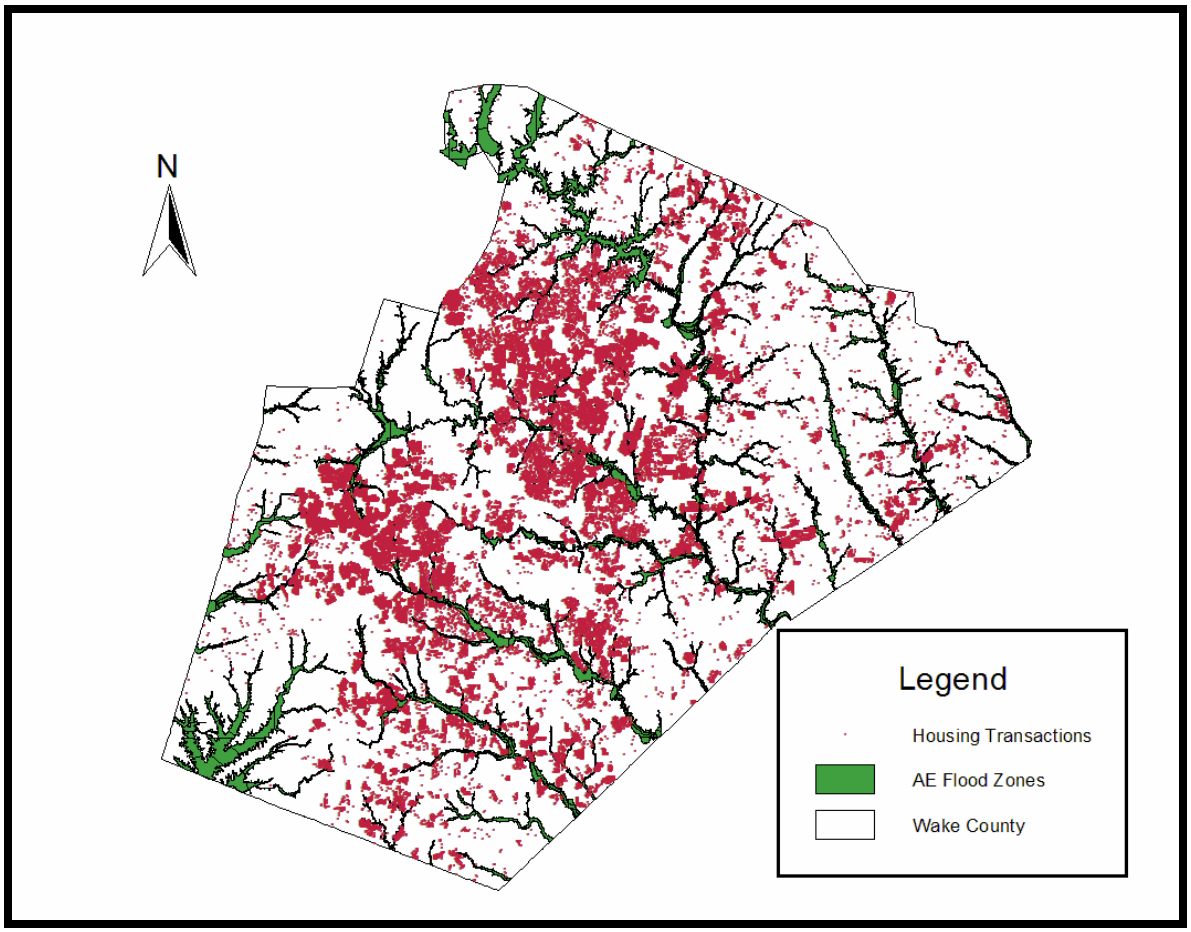


Figure 4.2: Spatial Layout of Flood Zones and Housing Transactions

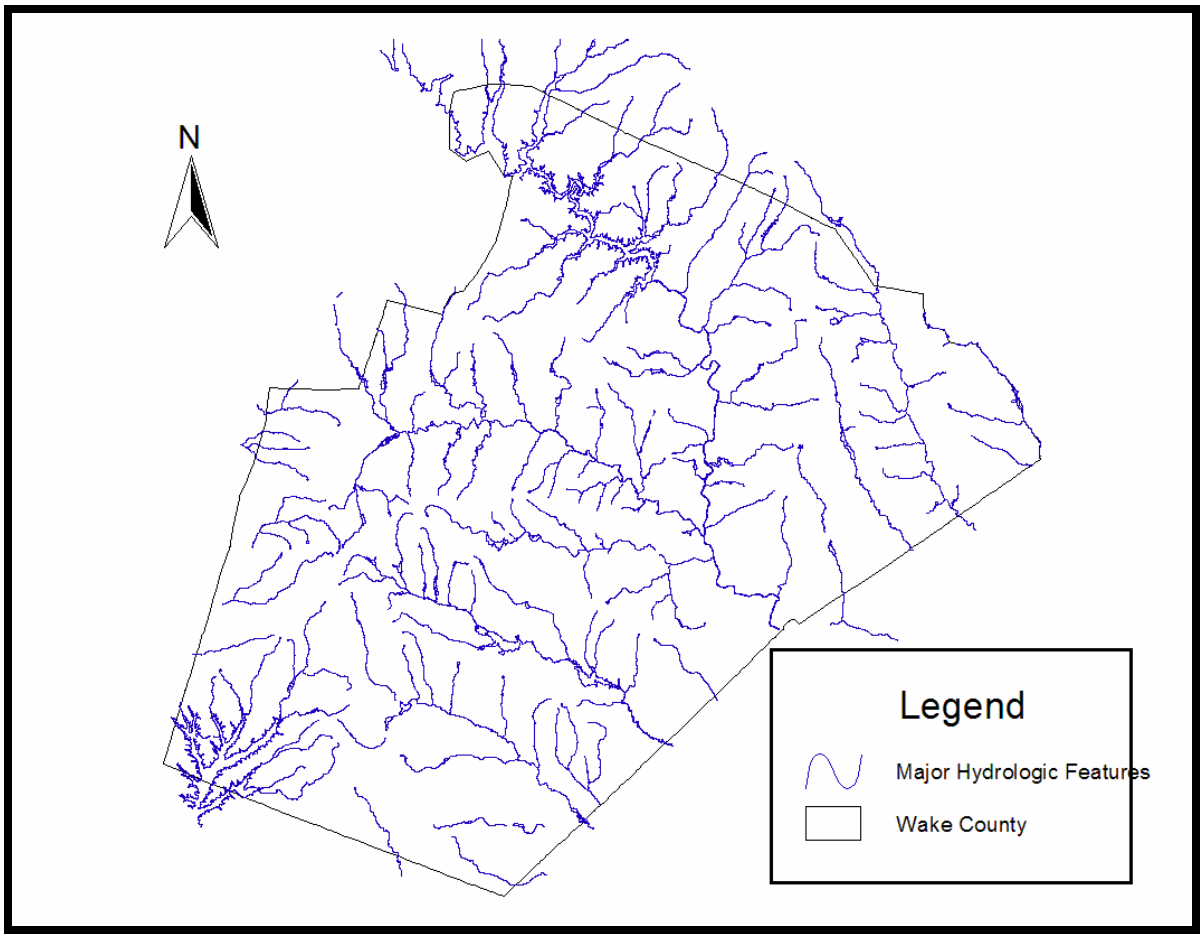


Figure 4.3 Hydrologic Features in Wake County

Table 4.1: Summary Statistics

Variable	Description	Mean	Median	Standard Deviation	Minimum	Maximum	Observations
lprice	Log of sale price of property	11.93	11.91	0.43	9.90	14.15	15514
age	Age of house in years	9.45	3	14.54	0	98	15514
baths	Number of bathrooms	2.45	2.50	0.69	1	8	15514
acreage	Lot size in acres	0.50	0.30	1.01	0.03	75.66	15514
regheatarea	Main heated living area in sq. ft.	1941.58	1800	725.02	400	7593	15514
detgarage	Detached garage dummy	0.03	0	0.18	0	1	15514
fireplaces	Number of fireplaces	0.89	1	0.37	0	3	15514
deck	Deck area in sq. ft.	172.30	152	150.97	0	1486	15514
sewer	Sewer availability dummy	0.80	1	0.40	0	1	15514
flordum1	Hardwood floors dummy	0.10	0	0.29	0	1	15514
scrporch	Screened porch area in sq. ft.	18.41	0	60.17	0	904	15514
walldum1	Brick walls dummy	0.12	0	0.32	0	1	15514
atticheat	Attic heated area in sq. ft.	48.01	0	153.64	0	1875.33	15514
bsmtheat	Basement heated area in sq. ft.	51.78	0	209.75	0	4788	15514
garage	Garage area in sq. ft.	290.85	364	255.35	0	1440	15514
poolres	Pool dummy	0.01	0	0.11	0	1	15514
bsmtdum1	Full basement dummy	0.05	0	0.22	0	1	15514
bsmtdum2	Partial basement dummy	0.07	0	0.25	0	1	15514
encporch	Enclosed porch area in sq. ft.	4.38	0	31.74	0	848	15514
opnporch	Open porch area in sq. ft.	64.13	35	83.89	0	1145	15514
condadum	House of "A" condition dummy	0.07	0	0.25	0	1	15514
condcdum	House of "C" condition dummy	0.03	0	0.16	0	1	15514
condddum	House of "D" condition dummy	0.00	0	0.06	0	1	15514
perc_no~1990	percent non-white	17.10	13.06	15.44	0	100	15514
medianvalu~t	Median house values	149196.10	138340	51689.62	30350	548000	15514
medttw_int	Median time to work	22.62	22	4.07	12	35	15514
perc_under~t	Percent population < 18	26.69	27.88	5.05	2.71	49.21	15514
perc_owner~t	Percent owner occupied housing	72.25	77.92	16.81	7.58	95.65	15514
nearestpark	Distance to nearest park	4.36	3.42	2.93	0.70	18.20	15514
nearestsc	Distance to nearest shopping center	7.89	6.97	4.81	0.41	25.68	15514
taxrate	Property tax rate for area	0.39	0.54	0.25	0	0.64	15514
post1996	(see section 3.4)	0.53	1	0.50	0	1	15514
zoneAE_dum	In AE Flood zone dummy	0.0115	0	0.11	0	1	15514
zoneX500_dum	In X500 Flood zone dummy	0.0044	0	0.07	0	1	15514
zoneAE_post	In AE zone & sold with disclosure	0.0065	0	0.08	0	1	15514
zoneX500_post	In X500 zone & sold with disclosure	0.0029	0	0.05	0	1	15514
new_sale	(see section 3.4)	0.29	0	0.45	0	1	15514
hydro_dist	(see section 3.4)	0.10	0	0.18	0	0.98	15514

Table 4.2: A comparison of housing characteristic means inside and outside of zoneAE before and after disclosure

variable	outside zoneAE	inside zoneAE	pre outside zoneAE	pre inside zoneAE	post outside zoneAE	post inside zoneAE
price	152,580.87	142,573.08	150,234.10	132,280.51	154,673.23	150,953.93
lprice	11.94	11.87	11.92	11.79	11.95	11.92
age	9.43	11.04	9.38	13.96	9.46	8.82
baths	2.45	2.41	2.46	2.21	2.45	2.56
acreage	0.50	0.75	0.51	0.71	0.50	0.78
regheatarea	1,942.90	1,828.32	1,947.00	1,703.21	1,939.02	1,923.69
detgarage	0.03	0.02	0.03	0.01	0.03	0.02
fireplaces	0.89	0.85	0.90	0.83	0.89	0.87
deck	172.23	178.22	174.28	189.60	170.38	169.55
sewer	0.80	0.81	0.79	0.84	0.80	0.78
flordum1	0.10	0.18	0.10	0.22	0.09	0.15
scrporch	18.43	16.93	18.57	22.45	18.24	12.71
walldum1	0.12	0.20	0.12	0.21	0.11	0.19
atticheat	47.81	65.80	47.65	60.71	48.22	69.68
bsmtheat	51.70	58.88	55.34	70.42	48.45	50.09
garage	291.46	238.88	289.20	229.31	292.90	246.18
poolres	0.01	0.03	0.01	0.04	0.01	0.02
bsmtdum1	0.05	0.11	0.05	0.14	0.05	0.08
bsmtdum2	0.07	0.06	0.07	0.04	0.06	0.08
encporch	4.39	3.33	5.02	4.81	3.81	2.21
opnporch	64.05	70.84	62.74	59.30	65.42	79.63
condadum	0.07	0.11	0.07	0.09	0.06	0.13
condcdum	0.03	0.02	0.03	0.03	0.03	0.02
condddum	0.00	0.01	0.00	0.01	0.00	0.00
perc_no~1990	17.11	16.29	16.82	17.23	17.35	15.57
medianvalu~t	149,319.60	138,515.70	146,399.20	131,916.90	151,841.10	143,546.50
medttw_int	22.61	22.77	22.50	22.09	22.73	23.29
perc_under~t	26.70	26.20	26.48	25.91	26.89	26.42
perc_owner~t	72.22	74.84	71.78	75.99	72.63	73.96
nearestpark	4.36	4.56	4.29	4.13	4.42	4.90
nearestsc	7.88	8.86	7.72	7.45	8.05	9.93
taxrate	0.39	0.38	0.39	0.36	0.40	0.41
post1996	0.53	0.57	0.00	0.00	1.00	1.00
zoneAE_dum	0.00	1.00	0.00	1.00	0.01	1.00
zoneX500_dum	0.00	0.00	0.00	0.00	0.01	0.00
zoneAE_post	0.00	0.57	0.00	0.00	0.01	1.00
zoneX500_post	0.00	0.00	0.00	0.00	0.01	0.00
new_sale	0.29	0.26	0.30	0.25	0.27	0.28
hydro_ft	5,500.01	6,007.11	5,409.13	6,271.79	5,584.47	5,805.33

Table 4.3: Multinomial Logit Test of Comparability of Samples

variables	AE Zone Pre-Disclosure [1]	AE Zone Post-Disclosure [2]	Chi2 Test that [1]=[2] [3]
age	0.011	-0.006	1.84
	[0.008]	[0.009]	
baths	-0.211	0.473	4.63**
	[0.264]	[0.179]**	
acreage	0.052	0.058	0.03
	[0.032]	[0.028]**	
regheatarea	0.000	0.000	1.01
	[0.000]	[0.000]	
detgarage	-1.214	-0.617	0.23
	[1.016]	[0.724]	
fireplaces	-0.029	-0.302	0.44
	[0.306]	[0.276]	
garage	0.000	-0.001	3.78*
	[0.001]	[0.001]**	
new_sale	0.034	0.081	0.01
	[0.300]	[0.256]	
Constant	-4.269	-5.482	
	[0.569]**	[0.438]**	
Observations	15514	15514	

Notes: Standard errors are presented in brackets. The * represents significance at the 10% level and the ** represents significance at the 5% level. Omitted category are houses not in an AE zone in either the pre-disclosure or post-disclosure time periods

Table 4.4: Flood Disclosure Regression Results

Dep. Var. = lprice	Full Sample	Full Sample	Full Sample	Within 0.3 Miles of AE zone	Within 0.2 Miles of AE zone	Within 0.1 Miles of AE zone
	[1]	[2]	[3]	[4]	[5]	[6]
Variable						
Post1996	0.028	0.028	0.027	0.045	0.049	0.038
	[0.003]**	[0.004]**	[0.006]**	[0.009]**	[0.011]**	[0.014]**
Hydro_dist	0.015	0.015	0.011	0.006	0.017	0.036
	[0.014]	[0.014]	[0.013]	[0.016]	[0.019]	[0.029]
zoneAE_dum	-0.020	0.005	0.003	0.012	0.007	0.010
	[0.013]	[0.014]	[0.015]	[0.014]	[0.014]	[0.016]
zoneX500_dum	0.010	0.027	0.029	0.038	0.021	0.016
	[0.016]	[0.025]	[0.023]	[0.025]	[0.027]	[0.022]
zoneAE_post		**-0.043**	**-0.042**	**-0.053**	**-0.052**	**-0.048**
		[0.022]**	[0.023]*	[0.025]**	[0.026]**	[0.028]*
zoneX500_post		-0.027	-0.016	-0.029	-0.027	-0.030
		[0.031]	[0.029]	[0.029]	[0.029]	[0.028]
Constant	10.851	10.850	11.485	11.583	11.626	11.585
	[0.074]**	[0.074]**	[0.247]**	[0.224]**	[0.232]**	[0.357]**
Census Tract Dummies	X	X				
Block Group Dummies			X	X	X	X
Structural & Neigh. Controls	X	X	X	X	X	X
Clustering at Census Tract	X	X				
Clustering at Block Group			X	X	X	X
# of spatial Dummies	104	104	253	207	192	163
Observations	15514	15514	15514	6243	4253	2136
R-squared	0.92	0.92	0.92	0.93	0.93	0.93

Notes: Robust clustered standard errors are presented in brackets. The * represents significance at the 10% level and the ** represents significance at the 5% level.

Chapter 5: Sex Offender Locations

5.1 Introduction

This chapter describes how information on the residential locations of sex offenders impacts housing prices. The analysis is possible because sex offender information was recently made publicly available due to public outrage over high profile sex related crimes. In October of 1989 an eleven-year-old boy named Jacob Wetterling was kidnapped and never found. As a result of his abduction, the "Jacob Wetterling Crimes Against Children and Sexually Violent Offender Registration Act" was enacted in 1994, which required every state to create a sex offender registry. The brutal murder of Megan Kanka by a neighbor who was a twice-convicted child molester was the impetus behind congress enacting the controversial 1996 "Megan's Law". Megan's Law amended the 1994 Jacob Wetterling Act by requiring dissemination of information from the sex offender registry to the public. Currently every state has complied with the legislation and most states have websites that provide access to the sex offender registry over the internet. Accessible information typically includes a picture of the offender, information on the offence(s), whether or not the offender is classified as a "predator", and the current address of the sex offender. [83]

What this publicly available information conveys to homeowners and homebuyers is difficult to characterize. Defining what it means to live "near" a registered sex offender is one of the primary differences between this application and the previous applications. Considering how information and attentiveness affect the performance of

[83] See www.klaaskids.com for information and links to state websites.

models for differentiated products, the "sex offender location" attribute of housing likely represents a complex combination of subjective risk estimates of possible future sex offenses, perceptions of neighborhood quality and even the potential emotional reactions to living in close proximity to an individual considered to be repulsive. These intangible features contrast with the relatively more definable attributes of airport noise and flood risk.

Although defining what it means to live near a sex offender is difficult, it does seem clear that an informed household would likely consider close proximity to a registered sex offender to be a negative attribute of a house. However, whether or not homebuyers are attentive to the registry and use the information for their home buying decisions is not clear.[84]

Sex offenders are generally not required to post visible signs in their yards warning potential homebuyers in the neighborhood of their presence.[85] Thus if a buyer is inattentive to the sex offender registry or the perceived search costs of finding the sex offender information are greater than the perceived benefits, he may remain uninformed of the residential locations of sex offenders in a neighborhood where he is considering purchasing a house. Furthermore, other housing institutions such as real estate agents are

[84] There has been surprisingly little empirical research on this important question. Only one published study, Larsen et al. (2003), has attempted to analyze the impact of residential locations of sex offenders on housing prices. During the writing of this chapter an NBER working paper was also released on this topic by Linden and Rockoff (2006). These two papers will be discussed in relation to the present work later on in the chapter.

[85] The words "generally not required" are used above in light of the fact that several judges in Texas did in fact recently order some sex offenders to post signs in their front yards warning the neighborhood of their sex offender status.

generally not required to disclose information about sex offenders to prospective buyers so the information acquisition burden falls primarily on the buyer.[86]

Sellers on the other hand are often informed directly without being required to actively search for the information because they are *currently* living in the area. For example, in Hillsborough County Florida—the study area for this chapter—sellers are likely to be more informed than buyers because: the sheriff's office conducts "reverse-911" calls to households and passes out fliers when high-risk sex offenders move into a neighborhood, there are requirements for rental companies to provide disclosures to nearby residents when renting a property to a sex offender and sellers are more likely to receive information about nearby sex offenders from their neighbors through word-of-mouth and neighborhood watch groups. Thus in the county studied here sellers are *given* information on nearby sex offenders. As a result they are not required to actively search for this information. In this respect there is an information asymmetry between sellers and buyers comparable to the airport noise and flood plain examples.

Unlike these previous applications, this case does not analyze a disclosure about sex offenders provided *directly to the housing market*. It analyzes two different types of information shocks. The first is caused by sex offender movements and the second is caused by two child abductions committed by sex offenders. Sex offender locations, unlike airport noise and flood zones, can change and they often do. This provides an opportunity to exploit the variation in sex offender movements in time and space to identify the impact of registry information on housing prices. Nonetheless, while sex offender movements allow increased spatial and temporal control in the identification

[86] This is the case in the state of Florida where the empirical analysis of this chapter is conducted.

strategy as suggested by the conceptual chapter of this thesis, estimates may be subject to bias due to asymmetric information. Two child abductions in the state of Florida brought increased awareness to sex offender issues and may have attracted buyers' attention. This public information is used to test for differential price impacts before and after the abduction.

The child abductions information shock is similar to the flood zone application in that if buyers paid attention to the information provided by the media, it may have influenced their information acquisition process by changing both the subjective expected benefits and costs. Nonetheless this information shock differs from both the airport noise and flood zone applications in two important dimensions. First, the disclosure was not given consistently at a salient time and place to buyers. The airport noise and flood zone disclosures were given from seller (or seller's agent) to homebuyers before they made an offer on a house. The information shock associated with the abductions of specific children on the other hand was less controlled and systematic. The media reports were not intended as an information policy. They are "news" and thus are not directly linked to the housing market or the sex offender registry. A concern with this information "treatment" is that it was not targeted at homebuyers to help them acquire information from the registry. Therefore, it is not clear that they actually received information that would alter their home buying decisions.

Using housing data purchased from the property appraiser's office in Hillsborough County, Florida and information on sex offender residential locations obtained from the Florida Department of Law Enforcement (FDLE), an analysis of the impact of sex offender residential locations on housing prices was conducted. The

findings suggest that houses within a tenth of a mile of a registered sex offender living in a single family residence on average sold for approximately 2.3% less after the sex offender moved into the neighborhood. This is about a $3,500 reduction for the average priced house in the sample. Moreover, housing prices in the tenth of a mile area surrounding a sex offender residence appear to rebound shortly after the sex offender leaves the neighborhood suggesting robustness of the sex offender neighborhood "entry" estimate. Further analysis indicates that the price impact appears to be invariant to whether or not the sex offender is labeled as a "predator" or if the housing transaction took place after the highly publicized abductions. The predator result suggests that household reactions may be based more on perceived neighborhood impacts from sex offenders rather than objective risk assessments. The abductions result is less conclusive about the impact of asymmetric information than the airport noise and flood plain results because we cannot be sure buyers actually received the pertinent information on the registry.

The remainder of the chapter will proceed as follows. Section 5.2 provides a discussion on how buyers might perceive the information on residential locations of sex offenders, and background on the study area used for the empirical section of the chapter. Section 5.3 describes the data used in the analysis. Section 5.4 outlines the identification strategy used for the hedonic price regressions. Section 5.5 describes the results of the hedonic price regressions. Section 5.6 concludes the chapter.

5.2 Background

5.2.1 Residential Locations of Sex Offenders: A Complex Housing Attribute

Why does society choose to make information about sex offenders publicly available? There are no public registries for murderers who are subsequently released from prison for example. By choosing to make information about sex offenders publicly available, society has chosen to brand these individuals with a type of "scarlet letter". This process may appear to punish sex offenders twice. The first punishment occurs when they serve probation or time in jail, and the second punishment occurs when they are released and are punished through public humiliation and animosity from being listed on the registry. Whether or not this double punishment of sex offenders is justified, has been the subject of intense public debate.

There are likely two primary reasons why society has chosen to make sex offender information publicly available. The first reason is that there is some evidence that convicted sex offenders have a substantial probability of re-offending. Using rates of reconviction for various types of sex offenders from the largest dataset of its kind (4724 offenders), Hanson et al. (2003) found that sexual recidivism rates are approximately 14% after five years, 20% after 10 years, and 30-40% after 20 years. However, the Hanson study admits that these observed recidivism rates may substantially underestimate the actual rates because some sex offenders re-offend and are not caught. They argue based on what they view as plausible assumptions that the true recidivism rates are likely 10-15% higher than the observed rates suggesting that on average

approximately half of sex offenders re-offend within 20 years.[87] In related work, Hanson and Bussiere (1998) have shown that certain characteristics are strong predictors of recidivism risk. For example sexual interest in children, deviant sexual preferences, prior sexual offenses, and whether or not the victims were strangers are some of the strongest predictors of re-offense. These predictors are often used by states to label riskier offenders as "predators".

Other evidence suggests that sex offenders are more likely to commit their offenses locally than other types of criminals. Using statistics of a nationally representative survey conducted with prison inmates by the United States Department of Justice, Larsen et al. (2003) report that 85.1% of sex offenders committed their offense in the same city in which they resided at the time of their arrest. This is compared to 80.2% for other criminal offenses. Furthermore they report that 64.9% of sex offenders committed their offense in their own neighborhoods compared to 44.6% for other criminal offenses. Although these statistics are rather vague in the definition of neighborhood and the type of offenses used to calculate the statistics, they are suggestive that sex offenders are apt to commit their offenses locally. Given sex offenders likelihood for re-offending on a local scale, the policy rationale for a registry was that providing this information to households would enable them to self-protect by warning their children and increasing neighborhood vigilance.

Stewardship of children is a widely recognized social responsibility. As a result, the second reason for this policy no doubt stems from this shared commitment and the emotional response of the lay public to crimes against children. Feelings of shock,

[87] Also see Doren (1998) for a discussion of why recidivism base rates are generally underestimated.

revulsion, fear, and finally outrage likely describe the emotional reaction that society has towards highly publicized heinous sex crimes. This emotional response has made it politically feasible to pass laws requiring registration of sex offenders and public dissemination of the information.

Just as sex offense risk and emotional reactions to the offenses were likely the impetus for the passing of sex offender legislation, they are also likely to be important drivers of households' responses to sex offender information. Some evidence of emotional and risk based responses is provided by a study conducted by Beck and Travis (2004). They examined the relationship between "fear of victimization" and receiving a sex offender notification. Surveys were conducted with heads of households that had received notification that a sex offender lived on the adjacent property (their treated group) and other nearby households that had not received any such notification (their control group). They found that households that had received a notification were more likely to fear sexual victimization for themselves and for their family members than households not living near an offender and not having received a notification. This result was strongest when the survey respondent was female or had less education.

The Beck and Travis study highlights the importance of households' *perceptions* on living near a sex offender. These perceptions are certainly influenced by subjective risk calculations and emotion. Thus the object of choice or attribute conveyed by a home's location to a buyer—proximity to a sex offender—is less clear-cut than airport noise or flood zone information. It is more difficult to quantify the emotional response that some people may feel upon learning of a sex offender in their neighborhood. These

reactions seem likely to contribute to an informed buyer's perceptions of neighborhood quality.

One can consider other locational attributes that may affect the perceived quality of a neighborhood in a similar fashion. For example, group homes for the mentally impaired, halfway houses, certain commercial facilities, and low income housing can change perceptions about crime risk and the quality of a neighborhood. However sex offender locations differ in a key way. Their locations are less permanent. Since a house is an asset, the impact of a sex offender location on housing prices will also depend on the perceived permanence of the offender. The frequent movements of sex offenders however, does provide a unique opportunity to test the impact of this locational disamenity on housing prices that is not provided by other more permanent features.

Without direct evidence of household reactions to publicly available sex offender information for the study area, the subsequent empirical analysis uses distance to a sex offender's residence as a proxy for the emotional and subjective risk and neighborhood perception responses of households. Distance is likely highly correlated with households' emotional responses to being "near" an offender and may also reflect some elements of risk. It is difficult *a priori* to determine what households' perceptions of "near" might be. However, the results presented later in this chapter suggest that only houses that transacted within a tenth of a mile *after a sex offender entered* a neighborhood are impacted by the residential location of sex offenders. If a substantial fraction of buyers are uninformed, then this estimate may under-represent the spatial magnitude of the changes in this object of choice.

5.2.2 Background on the Study Area

There were multiple criteria involved in choosing the geographic location for this the empirical analysis. A number of criteria were used to find a geographic location that would provide a suitable "natural laboratory" in which to test the impact of the residential location of sex offenders on housing prices and the impact of asymmetric information for this housing attribute. Ideally the location would be in a state that: (i) makes sex offender information publicly available through the internet or through a toll free number, (ii) has additional disclosure requirements that would cause sellers to be more informed than buyers, (iii) records and archives sex offender move-in and move-out dates, and (iv) experiences an information shock that may have caused inattentive or uninformed buyers to acquire the publicly available information on the residential locations of sex offenders.

Florida provided the best match to these selection criteria.[88] Florida was the first state to list sex offenders on a publicly accessible website beginning October 14, 1997.[89] In addition to the website, the same information was made available through a 24 hour a day hotline.[90] Therefore, the residential locations of sex offenders have been publicly available and easily accessible for a substantial period of time.

Florida legislation also requires county sheriff's offices to provide additional public disclosure as "deemed necessary". This is typically done in several ways including (i) verifying addresses of offenders by knocking on the offenders door, and if not home, checking with neighbors to confirm the residential location of the offender, (ii)

[88] Another state that would have been a good place to conduct the analysis was Ohio. This state has unique disclosure requirements that could have been used as the "information shock". However, although the state has the move-in and move-out dates archived, efforts to acquire this information were unsuccessful.
[89] This website is maintained by the Florida Department of Law Enforcement (FDLE) and can be accessed at www.fdle.state.fl.us/.
[90] This number is 1-888-FL-PREDATOR

providing prominent links to the FDLE website and state sex offender registry on their websites, (iii) speaking to community and crime watch groups and updating them about predators and offenders in their neighborhoods and (iv) using an automated phone system (called reverse-911) to notify schools, day-care centers and the public when predators and some offenders move into a neighborhood. These measures provide information to current owners of houses in these neighborhoods but not prospective buyers. Furthermore, there is no requirement that sellers or real estate agents disclose knowledge of sex offenders to buyers. Thus the disclosure actions of police may actually help to create an information asymmetry between buyers and sellers.

Florida is one of the few states that archive the residential locations of sex offenders *over time*. Most other state registries update their websites to keep the posted information current[91], but the update destroys any previous information on an offender such as their previous residential addresses. This practice is unfortunate for researchers who would like to use the temporal variation of sex offender locations in their research. Therefore, access to information on changes in sex offender locations was an important consideration in selecting Florida to do this analysis.

Florida was also chosen because recent events occurring in the state focused the media spotlight on registered sex offenders. On February 23, 2005 nine-year-old Jessica Lunsford was reported missing. A search for the girl which included extensive media coverage was unsuccessful until on March 19, 2005 it was discovered that Jessica Lunsford was killed by a registered sex offender that was living near her residence at the

[91] Posted information includes a picture of the offender, the offenders designation (predator or offender), name, supervision status, date of birth, race, sex, hair and eye color, weight and height, aliases, any scars/marks/tattoos, address information, adjudication and date, crime description and limited victim information.

time of the abduction. Households in Florida as well as more generally were outraged. The number of hits on the FDLE website increased from 19,000 the week before to 209,000 the week after the event. Three weeks later on April 10, 2005 thirteen-year-old Michelle Lunde was also reported missing. On April 17, 2005 it was announced that Michelle Lunde had been abducted and killed by a registered sex offender. These abductions resulted in a significant shift in the amount of media coverage given to sex offenders in Florida. Table 5.1 provides statistics on the number of newspaper articles on the issue of sex offenders and registration. The change in the number of relevant articles from the 3/1/2004 – 3/1/2005 period to the 3/1/2005 – 3/1/2006 both for the national coverage (number of articles written by the Associated Press (AP)) and the Florida Newspapers coverage (number of articles in Miami Herald, St. Petersburg Times and The Tampa Tribune) is quite dramatic.

Once Florida was determined to be the state in which the empirical analysis would be conducted, the next task was to select a county. The selection criteria included: (i) a county with a large number of houses, (ii) a county with a large number of sex offenders, (iii) a county that provides the above described additional public disclosure to existing homeowners, and (iv) a county where housing prices, attributes and geographic locations of housing could be acquired.

Hillsborough County, Florida was found to meet these criteria.[92] Based on the 2000 census, the county has approximately 1 million people. The largest city in the county is Tampa with approximately 300,000 people. Figure 5.1 shows the location of

[92] Alachua County was initially used because the housing data was available from a previous project. However, upon matching the housing data to the sex offender information, it was determined to have an insufficient number of housing transactions and sex offender locations for the identification strategy used in this chapter.

Hillsborough County in relation to the rest of Florida. There are a large number of housing transactions that have occurred in the county over the relevant time period. In addition, there have been a large number of sex offenders that have lived in this county since the sex offender registry was developed. The county also actively provides the additional public disclosures described earlier. Finally, the Michelle Lunde abduction and murder occurred in the town of Ruskin which lies in the southern part of Hillsborough County.

5.3 Data Used in the Analysis

There are two primary sources of data used in the analysis. The first is a dataset on all single-family housing transactions occurring between October 1996 and April 2006 in Hillsborough County. Data on sales prices and property characteristics were purchased from the Hillsborough County Property Appraiser's Office. The data was screened to drop "unqualified" sales and outlying observations.[93] A detailed GIS parcel map was also acquired from the appraiser's office. From this map the centroid of each parcel was calculated and this geographic reference point was linked to each of the houses in the database. This process allows information from the sex offender dataset described later, to be spatially merged to the housing dataset.

In addition to the sale prices, the appraiser's database also provides a set of important structural control variables that can be used in the analysis. These include

[93] House sales below $5,700 and above $2,900,000 which correspond with the approximate 1st and 99th percentiles were dropped. Unqualified sales are those that involve one of the following: multiple parcels, a court order, developer, an addition, business Inc. or personal property Inc., improvement incomplete, agricultural sale, death certificate, bank sale, government subsidized, or distressed sale.

information on the age of the house, the acreage of the lot on which the house is built, the number of bedrooms, the number of fixtures, the source for heating and cooling of the house, the architectural type, and the "effective" area of the house. This last variable takes the gross area of the house and multiplies each one of the sub areas by an "effective rate factor" that takes into account if it is a heated area and the area "type". This adjustment provides a more comprehensive measure of house size than heated square footage. Panel A in table 5.2 provides summary statistics for the single family transactions in Hillsborough County from 1996-2006.

The second dataset is the information on the sex offenders that have resided in Hillsborough County. The registry was made publicly available in 1997. The Florida Department of Law Enforcement provided this archived information from November 1997 through May 2006. These data include addresses, latitude and longitude coordinates of the residence, whether or not the offender was listed as a "predator", gender of the offender, and the date when the offender listed the address as his/her new residence. Sex offenders since 1997 have been required to register their new address with the FDLE within 2-10 days of moving to a new residence. Using the dates when a new address was listed for an offender in the archive, approximate move-in and move-out dates were derived. Table 5.3, panel A, provides summary statistics for the 2,824 sex offenders that resided at some point between 1997 and 2006 in Hillsborough County, Florida. Approximately 10% of these sex offenders were listed as "predators"[94], 76% were white and 98% were male. Offenses that warrant listing on the registry include

[94] In the state of Florida the "predator" designation is issued by the court for certain types of heinous offenses, or if an offender re-offends for certain types of offenses. The predator designation signifies a greater risk that the offender will re-offend. See the FDLE website and their bulletin entitled "2004 guidelines to Florida Sex Offender Laws" for more details.

those that commit sexual offenses against either children or adults. However, the FDLE did not provide the offense-specific information for the offenders in this dataset for this research although the information is available on the internet.

To identify the causal impact of sex offender's residential locations on nearby housing prices, it was necessary to drop some of the sex offender location observations before this information was linked to the housing dataset. All sex offender residential locations where the offender did not live in the residence for at least 6 months were dropped.[95] Panel B in table 5.3 gives summary statistics of the sex offenders that lived for at least six months in their residence. Using the addresses provided by the FDLE for this screened sample of sex offenders, it was then possible to match approximately 85% to a parcel in the housing dataset.[96] The housing dataset provides information on what types of buildings and activities occur on each parcel. All sex offenders from the sample who were not capable of being linked to a parcel designated as "single-family-residential" were dropped from the sample. Including only sex offenders on single-family-residential parcels that lived in the residence for at least six months, reduces the likelihood of multiple sex offender treatments (the halfway house affect) or some of the

[95] It is very difficult to define the experimental treatment and determine an identification strategy for temporary sex offender locations because there may be insufficient housing observations while the sex offender resided in the location and because temporary locations are often in areas with multiple sex offenders.

[96] This was an involved process. It was found that the latitude and longitude coordinates provided by the FDLE did not provide sufficient spatial resolution for the empirical analysis. As a result, the sex offender information had to be matched to the exact parcel to ensure locational accuracy. First using a GIS software called Arcview and an electronic road file, the addresses of both the sex offenders and the parcels were geocoded. Using the lat/longs from the geocoding as a unique id, approximately 70% of the sex offenders were matched directly to their corresponding parcel layer. The other 30% were assigned to the nearest 25 parcel centroids. Further matches were determined by hand, comparing the sex offender address with these nearest parcel addresses. Approximately half of the remaining 30% were either matched to the correct parcel or a "suitably correct" parcel which consisted of the exact same address but the number being within 5. So an address of "301 Oak Lane could be matched to 303 or 305 Oak Lane if 301 did not exist, but it could not be matched to 307 Oak Lane.

unobserved impacts from trailer courts, apartments, etc. that can affect residential neighborhoods.

After linking the sex offender location information to the parcel map, the distances from each parcel centroid where a housing transaction occurred to the nearest 25 sex offender locations were calculated. Using this "distance to sex offender locations" information with the full housing transaction dataset described above, two subsets of housing data were created. Housing subset #1 was created by dropping all housing locations that occurred within 0.15 miles *of more than one offender location* at any time between 1997-2006. In other words this sample is composed of the houses that were never "near" a sex offender location and houses that were only "near" an offender's location once. Panel B in table 5.2 provides summary statistics for housing subset #1. Housing subset #2 was created by dropping all housing locations in housing subset #1 that were further than 0.3 miles from a sex offender location. Panel C in table 5.2 provides summary statistics for housing subset #2. Panel C in table 5.3 provides summary statistics of the 322 sex offenders that are relevant to both of the housing subsets and figure 5.2 shows their residential locations in relation to Hillsborough County.. When compared with panel A of table 5.3 it can be seen that there is a lower fraction of offenders listed as "predators" (6%) and a higher fraction of white offenders in this sub-sample.

Using the timing of when these sex offenders moved in and out of their residence in relation to the timing of the housing sale as well as the linear distance between sex offender residences and neighboring housing, four key variables were created: (i) *one_tenth_mile* is a dummy variable equal to 1 for houses that sold within one tenth of a

mile of a sex offender the year before or the year after the sex offender moved into the neighborhood, (ii) *two_tenth_mile* is a dummy variable equal to 1 for houses that sold within two tenths of a mile of a sex offender the year before or the year after the sex offender moved into the neighborhood, (iii) *one_tenth_post* is a dummy variable equal to 1 for houses that sold within one tenth of a mile of a sex offender the year *after* the sex offender moved into the neighborhood, and (iv) *two_tenth_post* is a dummy variable equal to 1 for houses that sold within two tenths of a mile of a sex offender the year *after* the sex offender moved into the neighborhood, and 0 otherwise. Panels A, B, and C in table 5.2 also provide summary statistics for these key variables. Figure 5.3 shows an example of a sex offender residence in relation to residential parcels within 0.1, 0.2 and 0.3 miles.

5.4 Identification Strategy for the Hedonic Analysis

The spatial-temporal nature of the collected housing and sex offender data allows an identification strategy that exploits both space and time to isolate the impact of the publicly available sex offender information on housing prices. Sex offender movements provide one type of "information shock" that can be exploited in an identification strategy. Publicly available information provided by the registry is relatively constant over the study period's time frame, while its spatial relevance changes with sex offender movements (as do the information disclosures to existing homeowners that may cause sellers to be more informed). This allows for a test of the differences in housing prices before and after a sex offender "enters" a neighborhood. The child abductions occurring in Florida provide another "information shock", allowing a test for differential housing

price impacts before and after it became widely known that the abductions occurred by registered sex offenders. Such a change may signal a shift in the fraction of informed buyers caused by the increased media attention. Identification strategies for these two types of information shocks are developed in this section. However, prior to developing these identification strategies a cross-sectional approach to determining the causal impact of sex offenders on housing prices is discussed in an effort to gauge the importance of exploiting the temporal nature of the collected data.

5.4.1 Cross-Sectional Strategy

An identification strategy that uses the distance of houses to the nearest sex offender together with a cross-section of housing data to estimate the impact of sex offenders on housing prices could take the following form,

$$\begin{aligned}\textbf{lprice} = \beta + \alpha\textbf{Structural} + \phi\textbf{Spatial\_dums} \\ + \theta\textbf{Proximity\_dums} + \varepsilon\end{aligned} \tag{1}$$

where "lprice" is the log of housing transaction prices over one time period (a year for example), "Structural" is a set of structural characteristics of houses in the sample, "Spatial_dums" are dummy variables for spatial subsets of the study area (such as census tracts or block groups), and "Proximity_dums" are dummy variables for houses within specified distances of the nearest sex offender location listed on the registry at the end of the time period.[97] Also, ε is an error term and β is a parameter to be estimated and α,

[97] As discussed in section 5.2.1, informed buyers' responses to the information that an offender lives in close proximity to a house are likely to be a mixture of subjective risk assessments of a sexual offense against the household, changes in perceptions of the neighborhood and other emotional responses. In the identification strategies described in this section the distance to a sex offender's residence is used as a proxy for the emotional and risk responses made by households.

ϕ, and θ represent vectors of parameters to be estimated. The primary issue with this type of cross-sectional analysis is omitted variable bias. Estimates of θ may overstate the true impact of sex offender locations if the "Proximity_dums" variables reflect other negative locational attributes that are correlated with sex offender locations and are not being adequately controlled for by the "Structural" and "Spatial_dums" variables.[98] Determining the appropriate spatial aggregation for the "Spatial_dums" variables is therefore especially important in a cross-sectional identification strategy.

Another difficulty with specification (1) is that since the "Proximity_dums" are distance dummies to the *nearest* sex offender, if sex offender locations are correlated with each other in space then the estimates may reflect the impact of multiple offenders (for example, ten offenders living in the same half-way house) rather than the estimate of a single sex offender on housing prices. An added complication arises if some sex offenders moved during the period of the analysis, then some housing transactions may be inappropriately indicated by the "Proximity_dums" since the transactions may have occurred before the sex offender moved-in or after the sex offender moved-out.

5.4.2 Defining the Sex Offender Treatment

To overcome these potential biases a specification with both cross-sectional variation *and temporal variation* for sex offenders and housing locations offers the opportunity to increase the resolution in the estimated impact of proximity to a registered sex offender on housing prices. Using the datasets described earlier, the exact timing of a

[98] See Angrist and Krueger (1999) for a more thorough discussion of the difficulties inherent in using cross-sectional analysis to determine causal impacts.

sex offender moving into a neighborhood (or being first registered) in conjunction with temporally comparable housing sales can be exploited as a quasi-random experiment. A sex offender entering a neighborhood can be thought of as the experimental "treatment". Ideally this treatment on transaction prices in the neighborhood can be compared to similar houses transacted in neighborhoods that did not receive an offender.

Sex offenders as a group are highly transient. The "stay_length" variable in panel A table 5.3 shows that the average length of stay for all locations where sex offenders lived between 1997 and 2006 in Hillsborough County was approximately 297 days. However the distribution is highly skewed. The median number of days is only 100. In areas that have sex offenders moving in and out frequently, it would be difficult to estimate the "sex offender treatment effect" when multiple treatments are occurring over a short time horizon.[99] Housing subsets #1 and #2 were created to address this issue. The average stay length for the set of sex offenders used to create these datasets is approximately 3 years (reported in Panel C of table 5.3).

By excluding the most transient offenders and limiting the analysis to residential areas that were only "treated" by a sex offender one time (or never), this subset of sex offenders and housing allows for a clear definition of the sex offender treatment. This definition being: *the introduction of one sex offender into a residential neighborhood*. Restricting the data in this way is not costless. Estimates of the average treatment effect

[99] Residential areas that have had more than 1 sex offender during the period of analysis appear to be substantial. Using the sex offender residential locations where the sex offender lived in the residence for at least 6 months, the table in appendix 5.A was derived. It provides summary statistics for some of the housing variables of houses transacted with different numbers of sex offenders having lived within 0.3 miles of the house at some point over the time period of this analysis. As can be seen, approximately 32% of the transactions in the database were in areas where more than 2 sex offenders had lived. 12% were in areas where 5 or more sex offenders had lived and 3.5% were in areas where 10 or more sex offenders had lived. Houses in areas with more sex offenders on average are smaller, older, located on smaller lots and have lower transaction prices.

for this selected portion of the data may not provide an accurate measure of the housing price effects of areas with multiple sex offenders or areas with a high percentage of apartments and trailer courts.[100] Estimating the differential impact of sex offenders that dwell in other types of housing or in areas with a highly transient population of sex offenders on housing prices is an interesting avenue for future research, but is not considered here.

5.4.3 Spatial-Temporal Identification Strategy

An initial specification that exploits both the cross-sectional and temporal variation could take the following form,

$$\begin{aligned} \mathbf{lprice} = \beta + \alpha\mathbf{Structural} + \phi\mathbf{Spatial\_dums} \\ + \delta\mathbf{Year\_dums} + \theta\mathbf{Proximity\_dums} \\ + \omega\mathbf{Proximity\_dums\_post} + \varepsilon \end{aligned} \tag{2}$$

The difference between the cross-sectional specification shown in equation (1) and the specification developed in equation (2) is the inclusion of the "Year_dums" to control for the longer time horizon of the data and also the addition of the "Proximity_dums_post" variables and the parameter vectors δ and ω. These latter dummy variables indicate houses that sold within a certain distance of a sex offender *after* the sex offender moved into the residence. The coefficients on these dummy variables provide estimates of the impacts of a registered sex offender on housing prices relative to some omitted proximity

[100] Appendix 5.B provides the results of a multinomial logit that compares the differences in housing attributes among areas with different numbers of sex offenders. Chi-squared tests suggest that the housing attributes in the different areas are significantly different.

dummy variable.[101] The task of developing an adequate "control" for this "treatment" is discussed in the remainder of this section.

Controlling for Potential Temporal Confounders

While the identification strategy described by equation (2) extends a simple cross-sectional analysis, it may nonetheless yield biased estimates. If there is one or more omitted temporal or spatial variables that invalidate what is being used as the control group, then these factors can lead to biased estimates and confound the ability of the model to detect these types of subtle influences. For example, Hillsborough County and many other counties in Florida have seen rapid price appreciation relative to many other parts of the country over the past few years. If the dummy variables for annual effects do not adequately control for differential trends in price appreciation over the spatial extent of the county during the 1996-2006 time period of the housing dataset, this limitation could bias estimates. To mitigate this possibility a set of linear time trends, specific to each of the "Spatial_dums" used in the analysis were created.

There may be other temporal confounders not controlled for by these linear time trends over such a long time horizon. To help mitigate other potential temporal confounders, only housing transactions near a sex offender that took place the year before or the year after the sex offender moved into the neighborhood were retained for the final analysis. Equation (2) can be updated to include the measures used to mitigate potential temporal confounding influences in the following way,

[101] The primary regressions presented later on in section 5.6 use proximity dummies of 0.1, 0.2, 0.3 miles.

$$
\begin{aligned}
\textbf{lprice} = \beta + \alpha\textbf{Structural} + \phi\textbf{Spatial\_dums} \\
+ \delta\textbf{Year\_dums} + \theta\textbf{Proximity\_dums}^{R} \\
+ \omega\textbf{Proximity\_dums\_post}^{R} \\
+ \psi\textbf{Spatial\_linear\_trends} + \varepsilon
\end{aligned} \tag{3}
$$

where "Proximity_dums" and "Proximity_dums_post" have been superscripted with an "R" to denote that housing transactions that are proximate to sex offenders are "restricted" to the year before or the year after the sex offender moved into the residential location, and "Spatial_linear_trends" are the set of linear time trends for each of the "Spatial_dums".

Controlling for Potential Spatial Confounders

Unobserved spatial heterogeneity is always a concern in hedonic analyses of the housing market. It is not clear what would be the appropriate spatial controls to include in the specification. A valid control group from the perspective of identifying the impact of a registered sex offender on housing prices is one where the spatial variables statistically render the treated and untreated groups comparable. Several different sets of "Spatial_dums" variables were developed including tax districts, census tracts and census block groups in Hillsborough County.[102] An attractive alternative to these political boundary dummy variables would be sex offender specific spatial dummies. To this end, spatial dummy variables for all houses within each 0.3 mile area of each sex offender were created. These are the "Spatial_dums" used in conjunction with housing subset #2 in the most tightly bounded (in time and space) regressions reported in the next section. However, this reduces the housing sample substantially so specifications using housing

[102] These various sets of spatial dummies will be compared in a cross-sectional analysis to illustrate the importance of adequate spatial control.

subset #1 and block group spatial dummies (the smallest available political units) are also considered.[103]

A related spatial concern is that the location sex offenders choose to live in, even within the 0.3 mile spatial controls or block group controls, is not random. The summary statistics for the housing data in table 5.2 suggest that the sex offenders in single-family residential houses tend to be in areas with lower housing prices. If sex offenders live in the housing that is of lower quality in terms of unobserved attributes relative to other housing in the block-group level or 0.3 mile spatial control areas, this would bias a cross-sectional analysis. It may also bias an analysis that exploits the timing of sex offender movements if the introduction of an unobserved low quality attribute is correlated with the sex offenders' arrival in the neighborhood. For example if sex offenders move into newly built low income housing, estimates of the impact of sex offenders on housing prices may be influenced by the perceptions about the low income housing as well as the sex offenders.

In an effort to determine if the specifications used in the primary regressions are being confounded by unobserved temporal or spatial variables, a specification check and a robustness check are considered. The specification check involves estimating models using housing subset #1 with the block group controls and housing subset #2 with the 0.3 mile controls excluding observations that are "near" a sex offender and occurred after the sex offender moved into the neighborhood. This strategy would imply a model that takes the following form,

[103] If block groups provide adequate spatial control then the additional observations available in the housing subset #1 will help to identify the other relevant parameters in the model.

$$\begin{aligned}\textbf{lprice} = \beta + \alpha\textbf{Structural} + \phi\textbf{Spatial\_dums} \\ + \delta\textbf{Year\_dums} + \theta\textbf{Proximity\_dums}^{RN} \\ + \psi\textbf{Spatial\_linear\_trends} + \varepsilon\end{aligned} \tag{4}$$

where the "R" in the "RN" superscript on "Proximity_dums" is a reminder that observations are again "Restricted" to the year before and after a sex offender's arrival. The "N" is a reminder that those observations that are "Near" a sex offender and occurred the year after the sex offender moved into the neighborhood have been removed.[104] If the results from such a regression suggest no difference between the control group and observations in the same area as the "treated observations" that were excluded, then the specification would appear to be providing adequate control for any unobserved temporal and spatial elements of the housing environment.

The robustness check exploits another treatment that the temporal nature of the archived sex offender information in Florida allows. Not only do these data provide an approximate date for when sex offenders enter a neighborhood, they also provide an approximate exit date. If sex offender locations do in fact have a causal impact on housing prices, then it might be expected that once sex offenders leave a neighborhood, housing prices would rebound.[105] Such a result would lend credibility to the causal interpretation of the sex offender "entry treatment" effect. The specification in equation (3) can be redefined to analyze the sex offender "exit treatment". The housing observations in this specification would be restricted to those that occurred the year before or the year after the sex offender *exited* a neighborhood. Thus the estimate on the

[104] The regression results reported in section 5.6 suggest that it is the observations within 0.1 miles after the sex offender entered the neighborhood that should be excluded from this specification check.
[105] This assumes there are no path dependencies such as a stigma effect on a previous sex offender residence.

coefficient for “Proximity_dums” would be negative if sex offenders lower housing prices in these distances and the estimate on the coefficient for “Proximity_dums_post” would be approximately zero if housing prices immediately rebound once the sex offender exits the neighborhood.

5.4.4 Heterogeneity of the Sex Offender Treatment

Sex Offenders Labeled as “Predators”

The impact of the “type” of sex offender on housing prices may provide clues as to how households use the publicly available sex offender information and what they consider to be the sex offender housing attribute. If buyers are truly concerned about sexual offense risk to their household and they search out all publicly available information on the sex offender registry, then one would expect to see differential housing price impacts near offenders labeled as “predators”. However, if buyers find it too costly to understand the different risks imposed by those offenders labeled as “predators”, or if the choice to live near an offender is governed by a simple emotional yes/no heuristic based solely on the person being listed on the registry irregardless of the offender type, then there may not be a differential impact.

To test for heterogeneity in the treatment effect based on sex offender “type”, a dummy variable for houses near sex offenders labeled as “predators” was created. It was interacted with the “Proximity_dums_post” variables in equation (3) to create the following specification,

$$
\begin{aligned}
\textbf{lprice} = \beta + \alpha\textbf{Structural} + \phi\textbf{Spatial\_dums} \\
+ \delta\textbf{Year\_dums} + \theta\textbf{Proximity\_dums}^{R} \\
+ \omega\textbf{Proximity\_dums\_post}^{R} \\
+ \psi\textbf{Spatial\_linear\_trends} + \\
+ \xi\textbf{Proximity\_dums\_post\_interact} + \varepsilon
\end{aligned} \tag{5}
$$

where "Proximity_dums_post_interact" is the appropriate interactions of the proximity dummies with the predator dummy.

Post Abduction Time Period

It was hypothesized that sellers in the study area are on average more informed than buyers since they often receive information about sex offenders living near their homes without actively searching for the information. Asymmetric information of this sort can lead to an estimated implicit price for a housing attribute that is attenuated towards zero. Unlike the airport noise and flood zone applications, there was no disclosure for the sex offender attribute provided directly to the housing market. However, as noted at the outset there were two abductions that were widely covered by the media that possibly caused buyers to pay more attention to the sex offender attribute of housing.

If the child abductions caused buyers to become more informed, then it would be expected that the estimated impact of sex offender information on housing prices would be larger after this information shock occurred. To check for a differential impact of this sort, a dummy variable for housing transactions that occurred after March 19, 2005 (the date it was discovered Jessica Lunsford was abducted and killed by a registered sex offender) was created and interacted with the "Proximity_dums_post" variables. Equation (5) can be simply adapted to perform this test by redefining

"Proximity_dums_post_interact" to be the appropriate interactions of the proximity dummies with the post abductions dummy.

5.5 Summary of Previous Results

The identification strategy concerns raised in the previous section relate to two existing papers on this topic. Larsen et al. (2003) is the only published paper on the impact of sex offenders on housing prices. Their paper uses the cross-sectional hedonic methodology similar to that presented in equation (1). To develop their estimates, they used single-family housing transactions that occurred during the year 2000 in Montgomery County, Ohio and the registered sex offenders that were listed as living in the county *at years end*. However they note that a substantial number of offenders had moved into or within the county during the year of their housing data. Because they did not have this "move in" and "move out" information, they report the results of a simple regression that considered the effects of "end-of-year" sex offender locations on housing prices controlling for various housing characteristics of the following form,

$$\begin{aligned}\mathbf{lprice} = \beta + \alpha\mathbf{Structural} + \phi\mathbf{Spatial\_dums} \\ + \delta\mathbf{Season\_dum} + \theta\mathbf{Proximity\_dums} + \varepsilon\end{aligned} \tag{6}$$

where "Structural" represents a set of structural characteristics of houses[106], "Spatial_dums" are dummy variables for 26 tax districts in the county, "Season_dum" is a dummy variable to delineate between the summer months and the winter months, and "Proximity_dums" are dummy variables for houses within 1, 2, 3, 4, and 5 tenths of a mile of a sex offender.

[106] Structural characteristics included in their analysis were: living space, house age, lot size, number of fireplaces, owner-occupied dummy, full basement dummy, and 3 or more bathrooms dummy.

They concluded that proximity to offenders labeled as having a high risk of re-offending (predators), reduce housing prices. More specifically, for homes within .1, .2, and .3 miles of the registered offender these effects were: 17.4%, 10.2% and 9.3% respectively (and statistically significant at the 5% level). For less risky offenders the estimates for the same distances were 7.5%, 5.0% and 3.8% (only the first 2 distances were statistically significant at the 5% level). As discussed in section 5.4.1, whether or not these estimates can be interpreted as causal, hinges on any omitted variable bias introduced by the specification they used. Of particular concern is how well their 26 tax districts control for unobserved spatial heterogeneity.

A recent NBER working paper by Linden and Rockoff (2006) also considers the impact of registered sex offenders on housing prices. Their identification strategy is similar to what was developed for the present analysis since it exploits both the cross sectional and temporal variation in their sex offender and housing datasets. To perform their analysis they acquired sex offender information from the North Carolina Sex Offender Registry and housing data from 1994-2004 for Mecklenburg County, North Carolina. The sex offender data used by Linden and Rockoff provided the move-in dates for the current locations of all offenders living in the county as of January 1, 2005. However, it did not provide information on the previous locations of these offenders or any other offenders that had moved out of the county, or had been placed in jail or a half-way house before January 1, 2005.

It appears that North Carolina does not archive the sex offender registry data but simply updates an offender's information once he/she moves. This pattern of record keeping does not allow the same temporal resolution of sex offender location information

that the sex offender data in Hillsborough County provides. As a result, Linden and Rockoff cannot eliminate the possibility that their estimates may be impacted by multiple sex offenders that lived near a currently listed offender, but had moved out before January 1, 2005. Whereas the analysis in this chapter uses the information on all sex offender locations (not just the current locations) to select a sample of housing transactions in areas where only one sex offender had ever resided. This eliminates the possibility of multiple treatment effects. The summary statistics provided in appendix 5.A suggest that a significant fraction of the housing transactions in Hillsborough County are located near multiple offenders rendering this screening potentially important.[107] It is possible that a similar spatial pattern of sex offenders exists in Linden and Rockoff's study area as well.

Using a difference-in-difference specification similar to that developed in equation (3) above and the approximate move-in date for all sex offenders living in Mecklenberg County *as of January 1, 2005*, Linden and Rockoff found that houses within a tenth of a mile of mile decreased in value by approximately 4% after the sex offender entered the neighborhood, whereas housing prices between 0.1 and 0.3 miles distance showed no impact. Because the sex offender data available to them is only for currently listed offenders, it is impossible for them to derive move-out dates, and to conduct the robustness check of analyzing the "exit treatment".

[107] Appendix 5.A shows that approximately 64% of transacted houses, located within 0.3 miles of where a sex offender resided during the study time period, were located within 0.3 miles of more than one offender. 24% were located within 0.3 miles of 5 or more offender locations.

5.6 Results

5.6.1 Cross-Sectional Results

To illustrate the problem of unobserved spatial heterogeneity in an application of this type, table 5.4 presents estimates of a set of cross-sectional regressions. These regressions use the full sample of housing data (1996-2006) and the most proximate location of any sex offender that lived in a single family residence for at least six months. The results in Column [1] are generated from a regression of the log of housing price on yearly dummy variables and variables that indicate if a house transaction occurred within 0.1 ("onetenth_mile") or 0.2 ("twotenth_mile") miles of the nearest sex offender residential location. The estimates for the parameters on the onetenth_mile and twotenth_mile dummy variables suggest that sex offenders are clustered in areas with significantly lower housing prices. Column [2] reports the results for the same regression as in column [1] with housing characteristics available from the appraiser's database that were described in section 5.3 included. Controlling for observed differences in housing characteristics reduces the size of the parameter estimates of the two proximity dummies significantly—from -0.41 to -0.12 for the "onetenth_mile" dummy and from -0.30 to -0.07 for the "twotenth_mile" dummy.[108]

Of course there are likely other spatially unobserved characteristics of housing that may be correlated with sex offender locations that should be included. Columns [3]-[5] provide the results of three regressions analogous to equation (1), each including a different set of "Spatial_dums" to control for unobserved spatial heterogeneity.

[108] These parameter estimates and all other parameter estimates on dummy variables reported in this section have been corrected using the Halvorsen and Palmquist (1980) correction for interpreting dummy variables in semilogarithmic equations.

Interestingly the estimates of the proximity to sex offender dummies shown in column [3] where 81 tax district dummies were included in the specification are approximately equal to those in column [2] and similar in magnitude to the estimates provided by Larsen et al. (2003). It appears that the sex offender clustering is occurring at a much lower spatial resolution than that provided by the 81 tax districts. The results in Columns [4] and [5] indicate that the magnitude of these same coefficients are reduced substantially when the 247 census tract dummies or the 759 census block group dummies are included to control for unobserved spatial heterogeneity.

5.6.2 Specification Check

To determine a specification that exploits both the temporal and spatial movements of sex offenders and provides adequate spatial control, a specification check described by equation (4) is reported in Columns [1] and [2] of table 5.5. Column [1] provides the results from this specification using housing subset #1 and block groups as the "Spatial_dums" (notice the house controls, census block-group dummies, and linear time trends for each block-group have been added). The housing transactions within 0.1 miles that occurred after the sex offender entered the neighborhood were temporarily dropped. If this specification provides adequate spatial control then the coefficients on the onetenth_mile and twotenth_mile variables should be close to zero.[109] However, the coefficients on "onetenth_mile" and "twotenth_mile" suggest that there remains unobserved heterogeneity that reduces housing prices by approximately 2.2% for areas within 0.1 miles of where sex offenders will move to *in the future*.

[109] As will be seen later on, it appears that sex offender impacts are restricted to this localized area.

In an effort to better control for unobserved heterogeneity, Column [2] also presents regression results following equation (4). Now the model is based on the sample labeled housing subset #2 (only housing transactions within 0.3 miles of a sex offender) and the spatial dummies for housing observations within 0.3 miles of each individual sex offender location. The results from this regression (after again removing housing transactions within 0.1 miles of an offender, and adding the housing, 0.3 mile dummies and time trends) suggests that prices of houses between 0 and 0.1 miles distance from the future sex offender residence are no different from the houses that are between 0.1 and 0.2 or 0.2 and 0.3 distance from the future sex offender location. This appears to adequately remove unobserved spatial heterogeneity and provide a better specification to explain the causal impact of sex offender relocations on housing prices.

5.6.3 Primary Results

Table 5.5, columns [3]-[5], presents regression results (using the preferred 0.3 mile control specification) that exploit both the temporal and spatial movements of sex offenders. Column [3] corresponds with the specification developed in equation (3). The data used in this specification includes housing transactions within 0.1 miles of the sex offender location. The results suggest that after including the various controls and prior to the sex offender entering the neighborhood, there is no difference between housing prices in the 0-0.1, 0.1-0.2, and 0.2-0.3 bands surrounding the sex offender location. However after the sex offender enters the neighborhood, houses within 0.1 miles of the sex offender location appear to decline in price by approximately 2.3%. This is

significant at the 6% level even after clustering the standard errors for the houses within 0.3 miles of each individual sex offender.[110]

Column [4] evaluates the potential for heterogeneity of the sex offender impact for houses near sex offenders labeled as "predators" using the specification in equation (5). The coefficient on "onetenth_mile_post_interact" is positive but insignificant. The coefficient on "twotenth_mile_post_interact" is negative and also insignificant. Therefore it does not appear that there is a differential impact for those houses near a sex offender that is labeled as a predator.

Column [5] provides the results from the specification in equation (5) that checks for a differential impact of sex offenders on housing prices after the abductions of the Florida children. Both interaction variables are insignificant. This suggests that the abductions and the ensuing media coverage did not have a significant influence on the asymmetric information that likely exists between buyers and sellers.

5.6.4 Robustness Check: The Exit Treatment

As discussed earlier, the interpretation of the "entry treatment" estimate as the causal impact of sex offender locations on housing prices would be enhanced if housing prices rebounded after the sex offender *exited* the neighborhood. Column [6] presents the estimates of a regression that again uses the preferred specification and housing sample #2. The sample was restricted to those housing transactions that occurred the year before or the year after the offender moved *out* of the house. As a result, in this specification the

[110] Clustering allows the regression errors for observations within 0.3 miles of each sex offender location to be correlated.

“onetenth_mile_post” indicates the houses within a tenth of a mile of a sex offender location the year after the sex offender *exited* the neighborhood. The estimated coefficient for “onetenth_mile” is statistically significant at the 10% level and suggests that the year before the sex offender exited the neighborhood houses within a tenth of a mile were 1.8% lower than houses further away. The magnitude of this estimate is similar to the estimate of 2.3% reduction for the “entry treatment”. The estimate on twotenth_mile is also negative but is not significant by conventional standards. The estimates on “onetenth_mile_post” suggests that after the sex offender exited the neighborhood, housing prices rebounded to where they appear to be no different than the prices of houses in the 0.1 to 0.2 mile range (see coefficient on “twotenth_mile_post”) or the 0.2 to 0.3 mile range which is the omitted distance category.

5.6.5 Discussion of Results

An identification strategy that exploits the temporal nature of sex offender movements was developed to estimate the impact of the residential locations of sex offenders on housing prices. Prior to conducting the analysis, cross-sectional results suggested that sex offender location estimates could be inflated upwards if the spatial dummies used to control for unobserved spatial heterogeneity were highly aggregated. Using sex offender specific spatial dummies of 0.3 miles appears to adequately reduce the unobserved spatial heterogeneity. By exploiting this spatial specification together with a housing sample that is bounded to the year before and the year after the first (and only) sex offender ever to have lived in a neighborhood arrives, it is estimated that housing prices within a tenth of a mile are reduced by approximately 2.3%. This

percentage change translates into a $3,500 reduction for the average priced house in the sample. The robustness of this estimate is enhanced by the fact that the year prior to the offender exiting the neighborhood, housing prices in this zone are still depressed by 1.8% but they appear to *immediately rebound* after the sex offender exits the neighborhood.

The estimated impact does not appear to differ based on whether or not the offender near the housing transactions is labeled a "predator". One would expect if households were fully informed and cared about sex offender risk information provided by the registry that a higher premium would be placed on living far from predators than from sex offenders. One interpretation of this result is that emotional responses and neighborhood perceptions influence housing prices more than objective risk. It could also signal high search costs of obtaining and understanding available objective risk information.

If sellers are more informed than buyers, then these estimates of the MWTP to avoid living in close proximity to a sex offender may be attenuated towards zero as argued in the conceptual chapter of this thesis. To gain insight into the impact of information on the estimated treatment effect, housing transactions that occurred before and after the two Florida abductions were compared. The estimated treatment effect was not statistically different in the pre and post abduction periods.

Why do the estimates appear to be insensitive to the child abductions? One possibility is that unlike the airport noise and flood zone disclosures, the broadly released media information was not targeted enough to home buyers and the housing market to cause a difference in their decisions through changes in attention and expected benefits and expected costs. In other words, receiving specific information from the seller or

seller's agent while searching for a house is likely to be more salient to buyers than information provided more generally by the media. A test of this hypothesis could be conducted if Hillsborough County were to require sellers to disclose if they have neighbors who are registered sex offenders. Unfortunately (from a researcher's perspective) this has not occurred.

5.7 Conclusion

This chapter analyzed households' reactions to the residential locations of sex offenders after this information was made publicly available in Hillsborough County, Florida. The analysis contributes to the overall theme in several ways. The example involved a complex commodity (a house) and focused on an attribute for which information is publicly available. This information was arguably better known to sellers than buyers because of programs that provided information to current homeowners but not prospective buyers. Therefore, an estimate of the implicit price for sex offender proximity may be attenuated if the fraction of uninformed buyers is large. The example also focused on exogenous changes in the information environment.

There are also several differences between the sex offender application and the previous two case studies. The sex offender attribute of housing is complex and more difficult to define than airport noise or flood zones. The object of choice for a homebuyer involves subjective perceptions of risk and emotional responses to the sex offender's proximity. Another key difference relates to the exogenous changes in the information environment—sex offender movements and two Florida abductions. Unlike the other case studies, information for buyers was constant during the exogenous sex

offender movements, whereas sellers may have received information from local law enforcement or neighbors. Although the abductions information shock may have attracted some buyers' attention to the sex offender registry, it was less clear how the information shock would effect buyers' information acquisition compared to the housing market-specific airport noise and flood plain disclosures.

A final difference between this and the previous case studies is the spatial and temporal control in the identification strategy that is allowed for by sex offender movements. Unlike the airport noise and flood zone applications where the spatial disamenity was geographically fixed over time and the primary focus was the impact of the information disclosures, the natural occurrence of sex offender movements allowed for a type of spatial experimental control that is not typically encountered in a natural experiment.

Focusing on the two types of information shocks, the estimates presented above indicate that housing prices within one tenth of a mile of a registered sex offender are reduced when a sex offender moves into a neighborhood. Thus it appears that publicly available information on sex offenders does impact housing prices. However, this impact was not affected by the two Florida abductions. This result suggests that buyers and sellers were either already fully informed so that additional information would necessarily have no impact, or the media provided public information does not provide a clear signal to buyers about sex offender information and risks. Although anecdotal evidence suggests that buyers are not fully informed, more research is needed to understand how they acquire and interpret information provided by the media and sex offender registries.

Whether or not providing sex offender information to the public is an appropriate use of government resources is a difficult question. It also raises important philosophical issues about punishing individuals twice by placing them to wear a “scarlet” letter. Some argue that providing the information will help households to self-protect by being more vigilant with themselves and their children. However, the issues raised in this thesis suggest that more proactive disclosures that help to inform potential sellers may have unwanted consequences as well. Informed sellers near a registered sex offender who have flexibility in when they must sell their house, may be more likely to wait for an uninformed buyer that is willing to pay “full price”. If this is a frequently occurring phenomenon, then this represents a real societal inefficiency of Megan’s Law because households that are best prepared to self-protect *after buying a house near a sex offender* are not necessarily the ones that end up purchasing the house. It could be that the single mom with six children, who did not think to check the registry, ends up living next door to the child-molester.

Figure 5.1: Location of Hillsborough County, Florida

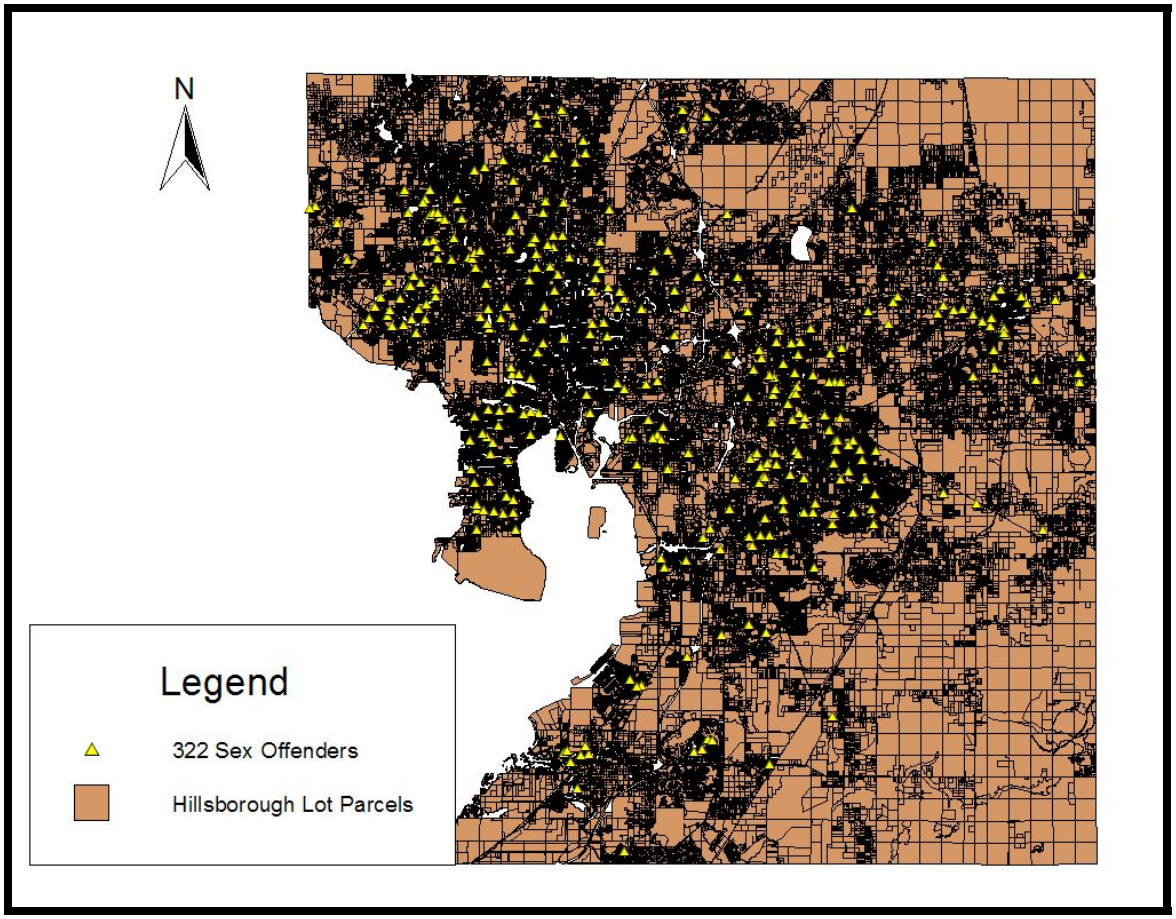


Figure 5.2: The 322 Sex Offender Locations Used in the Analysis

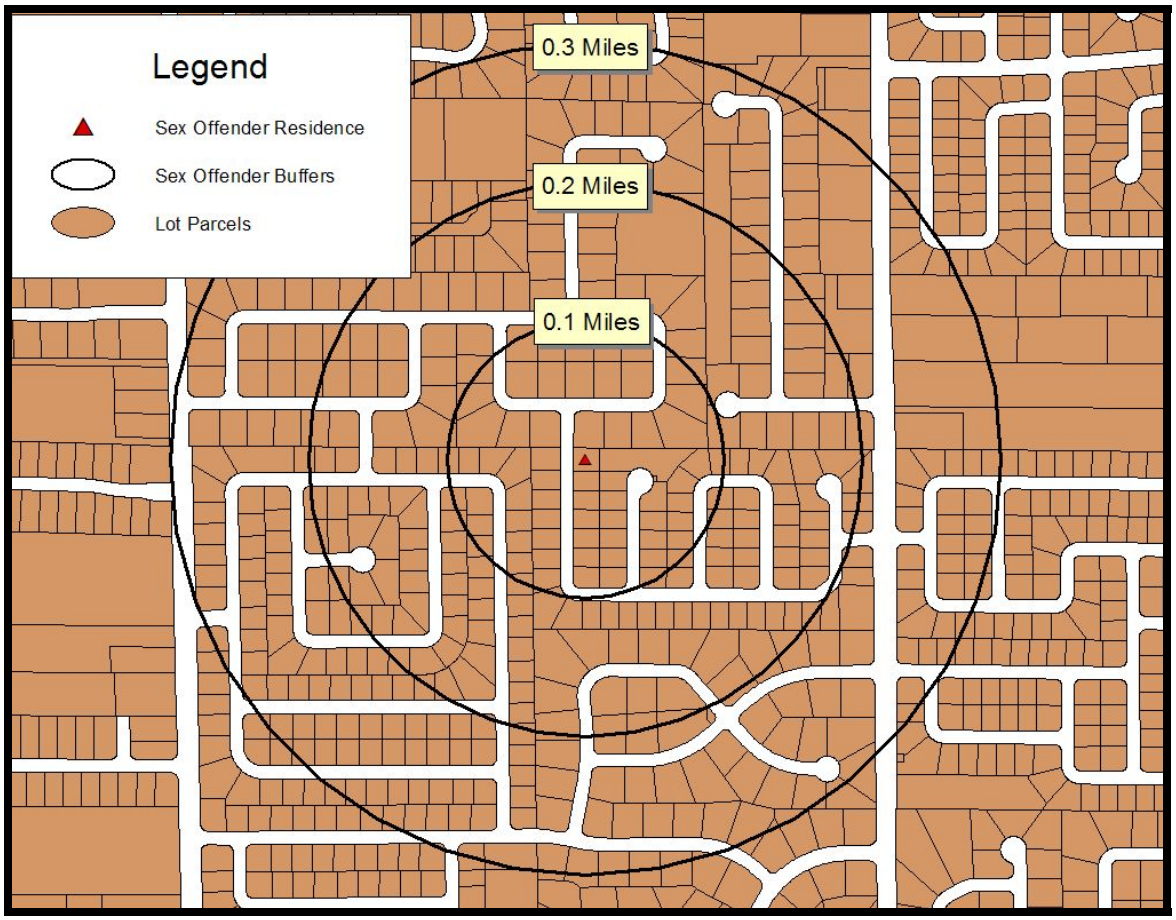


Figure 5.3: Example of Parcels Within 0.1, 0.2, and 0.3 Miles

Table 5.1: Sex Offender Newspaper Articles Statistics

Year (March 1rst-March 1rst)	Florida (broad)	Florida (refined)	AP (broad)	AP (refined)
1990-1991	6	0	-	-
1991-1992	13	0	-	-
1992-1993	17	1	-	-
1993-1994	11	0	-	-
1994-1995	37	2	-	-
1995-1996	48	4	-	-
1996-1997	46	4	-	-
1997-1998	43	6	-	-
1998-1999	35	3	71	32
1999-2000	60	14	42	18
2000-2001	70	28	43	18
2001-2002	45	12	87	43
2002-2003	52	28	65	30
2003-2004	49	30	92	45
2004-2005	66	41	86	38
2005-2006	264	133	233	141
change from 2004 to 2006	400%	324%	271%	371%

Notes: "Florida" is a composite of three major Florida papers: Miami Herald, St. Petersburg Times, and The Tampa Tribune. Statistics for this category were generated using Lexis Nexis's Academic News Service. The (AP) Associated Press statistics were generated using the Associated Press Digital Archive that can be accessed at: http://www.apdigitalnews.com/. "Broad" refers to a more general search using "sex offender" in the headline, lead paragraph and terms as the search criterion. "Refined" requires "registered" or "register" to also appear in addition to "sex offender". A "-" denotes that data was not available in an archive during the time period

Table 5.2: Summary Statistics of Housing Variables Used in Analysis

Variable	Description	Mean	Median	Standard Deviation	Minimum	Maximum	Observations
Panel A: All Single Family Transactions from 1996-2006							
price	sale price of property	172,183.80	143,900.00	114,771.90	5,700.00	2,863,299.00	189,491
lprice	ln(price)	11.89	11.88	0.56	8.65	14.87	189,491
age	age of property in years	19.37	13.00	20.71	0.00	106.00	189,491
acreage	acreage of lot	0.30	0.19	0.83	0.00	123.97	189,491
utsBR	number of bedrooms	3.21	3.00	0.85	0.00	12.00	189,488
utsBR_2	utsBR squared	11.00	9.00	5.61	0.00	144.00	189,488
utsBT	number of fixtures	6.40	6.00	2.13	0.00	24.00	189,491
eff_ar	the "effective" square footage	2,065.35	1,953.00	780.07	105.00	10,749.00	189,491
AC_dum2	has central AC	0.98	1.00	0.15	0.00	1.00	189,491
onetenth_mile	within .1 miles of S.O.	0.16	0.00	0.37	0.00	1.00	189,491
twotenth_mile	between .1 and .2 miles of S.O.	0.20	0.00	0.40	0.00	1.00	189,491
Panel B (housing subset #1): Transactions in Areas with no S.O.'s or within .3 Miles of Non-Transient S.O.'s							
price	sale price of property	198,977.80	166,000.00	124,532.70	5,700.00	2,863,299.00	118,558
lprice	ln(price)	12.06	12.02	0.52	8.65	14.87	118,558
age	age of property in years	13.39	7.00	16.62	0.00	106.00	118,558
acreage	acreage of lot	0.35	0.20	1.00	0.00	123.97	118,558
utsBR	number of bedrooms	3.42	3.00	0.80	0.00	12.00	118,557
utsBR_2	utsBR squared	12.33	9.00	5.59	0.00	144.00	118,557
utsBT	number of fixtures	6.99	6.00	2.04	0.00	24.00	118,558
eff_ar	the "effective" square footage	2,302.65	2,171.00	771.22	191.00	8,687.00	118,558
AC_dum2	has central AC	0.99	1.00	0.10	0.00	1.00	118,558
onetenth_mile	within .1 miles of S.O.	0.03	0.00	0.17	0.00	1.00	118,558
twotenth_mile	between .1 and .2 miles of S.O.	0.07	0.00	0.26	0.00	1.00	118,558
Panel C (housing subset #2): Transactions in Areas within .3 Miles of Non-Transient S.O.'s							
price	sale price of property	155,574.20	138,500.00	79,968.85	11,500.00	880,000.20	5,924
lprice	ln(price)	11.85	11.84	0.46	9.35	13.69	5,924
age	age of property in years	15.80	10.00	17.73	0.00	101.00	5,924
acreage	acreage of lot	0.24	0.19	0.38	0.00	18.87	5,924
utsBR	number of bedrooms	3.26	3.00	0.79	0.00	9.00	5,923
utsBR_2	utsBR squared	11.27	9.00	5.46	0.00	81.00	5,923
utsBT	number of fixtures	6.47	6.00	1.81	0.00	15.00	5,924
eff_ar	the "effective" square footage	2,041.95	1,910.00	689.66	394.00	6,453.00	5,924
AC_dum2	has central AC	0.99	1.00	0.10	0.00	1.00	5,924
onetenth_mile	within .1 miles of S.O.	0.15	0.00	0.36	0.00	1.00	5,924
twotenth_mile	between .1 and .2 miles of S.O.	0.36	0.00	0.48	0.00	1.00	5,924
onetenth_mile_post	within .1 miles and transacted after S.O. enters	0.07	0.00	0.25	0.00	1.00	5,924
twotenth_mile_post	between .1 and .2 miles and transacted after S.O. enters	0.19	0.00	0.39	0.00	1.00	5,924

Notes: Year dummies, block group dummies, sex offender area dummies, fuel and heat dummies and Architecture type dummies are not included in the summary statistics above.

Table 5.3: Sex Offender Summary Statistics

Variable	Mean	Median	Standard Deviation	Minimum	Maximum	Observations
Panel A: All S.O.'s living in Hillsborough sometime between 1996 and 2006						
Offender	0.90	1	0.31	0	1	2,824
Predator	0.10	0	0.31	0	1	2,824
White	0.76	1	0.43	0	1	2,824
Black	0.24	0	0.43	0	1	2,824
Other	0.00	0	0.05	0	1	2,824
Male	0.98	1	0.13	0	1	2,824
Female	0.02	0	0.13	0	1	2,824
stay_length	291.70	100	499.86	0	3,139	11,472
Panel B: S.O.'s living in a single family residence for at least 6 months						
Offender	0.91	1	0.28	0	1	1,130
Predator	0.09	0	0.28	0	1	1,130
White	0.74	1	0.44	0	1	1,130
Black	0.26	0	0.44	0	1	1,130
Other	0.00	0	0.05	0	1	1,130
Male	0.98	1	0.14	0	1	1,130
Female	0.02	0	0.14	0	1	1,130
stay_lengt	826.20	504.5	746.74	183	3,139	1,660
Panel C: S.O.'s living in a single family residence for at least 1 year and not within 0.3 miles of any other offender						
Offender	0.94	1	0.24	0	1	322
Predator	0.06	0	0.24	0	1	322
White	0.87	1	0.34	0	1	322
Black	0.13	0	0.33	0	1	322
Other	0.00	0	0.06	0	1	322
Male	0.98	1	0.16	0	1	322
Female	0.02	0	0.16	0	1	322
stay_length	1,137.89	820.5	828.64	366	3,139	322

Table 5.4: Cross-Sectional Regression Results Using Full Sample of Housing Data and Sex Offenders

Dep. Var. = lprice					
Variable	[1]	[2]	[3]	[4]	[5]
onetenth_mile	-0.411	-0.119	-0.120	-0.044	-0.029
	[0.003]**	[0.002]**	[0.002]**	[0.002]**	[0.002]**
twotenth_mile	-0.302	-0.069	-0.071	-0.028	-0.017
	[0.003]**	[0.002]**	[0.002]**	[0.001]**	[0.001]**
Constant	11.629	9.955	9.967	10.236	10.303
	[0.008]**	[0.010]**	[0.010]**	[0.011]**	[0.015]**
Year Dummies	X	X	X	X	X
House Controls		X	X	X	X
Tax District Dummies			X		
Census Tract Dummies				X	
Block Group Dummies					X
# of Spatial Dummies			81	247	759
Observations	189,491	189,488	189,485	189,488	189,488
R-squared	0.36	0.81	0.82	0.88	0.89

Notes: standard errors are presented in brackets.
* significant at 10%; ** significant at 5%

Table 5.5: Primary Regression Results

Dep. Var. = lprice Variable	Pre-Entry 0.3 Mile & No S.O. [1]	Pre-Entry 0.3 Mile Only [2]	Entry Treatment [3]	Entry Treatment Interact: Predator [4]	Entry Treatment Interact: Abduct [5]	Exit Treatment [6]
onetenth_mile	-0.022	0.001	0.000	0.001	0.000	**-0.018**
	[0.007]**	[0.008]	[0.009]	[0.009]	[0.009]	[0.010]*
twotenth_mile	-0.017	-0.007	-0.008	-0.010	-0.009	-0.011
	[0.006]**	[0.006]	[0.008]	[0.007]	[0.007]	[0.008]
onetenth_mile_post			**-0.023**	**-0.026**	**-0.025**	0.009
			[0.012]*	[0.013]**	[0.013]*	[0.014]
twotenth_mile_post			0.001	0.003	0.000	-0.006
			[0.006]	[0.007]	[0.007]	[0.011]
onetenth_mile_interact				-0.011		
				[0.031]		
twotenth_mile_interact				0.023		
				[0.049]		
onetenth_mile_post_interact				0.036	0.008	
				[0.026]	[0.021]	
twotenth_mile_post_interact				-0.033	0.011	
				[0.040]	[0.015]	
Constant	6.299	13.578	12.935	12.935	12.923	11.695
	[0.140]**	[1.246]**	[1.149]**	[1.149]**	[1.149]**	[0.027]**
Year Dummies	X	X	X	X	X	X
House Controls	X	X	X	X	X	X
Block Group Dummies	X					
Block Group Linear Trends	X					
Clustering at Block Group	X					
.3 Mile Dummies		X	X	X	X	X
.3 Mile Linear Trends		X	X	X	X	X
Clustering at .3 Mile		X	X	X	X	X
# of Spatial Dummies	619	275	275	275	275	268
Observations	115,532	5,518	5,923	5,923	5,923	4,278
R-squared	0.90	0.93	0.93	0.93	0.93	0.92

Notes: Robust clustered standard errors are presented in brackets.
* significant at 10%; ** significant at 5%

Chapter 6: Conclusion

6.1 Conclusions Drawn from this Thesis

This research gauges the importance of relaxing the full information assumption in applications of revealed preference models in complex, public information environments. An information acquisition process described why homebuyers are frequently less informed than sellers for some housing attributes. Often they face more stringent search and cognitive processing constraints. Adapting the hedonic model for the possibility that sellers are more informed than buyers suggests that estimates of the implicit price for a housing attribute may be attenuated towards zero if there is asymmetric information about the quantity of the attribute. The importance of the asymmetric information argument was gauged by applying the quasi-random experiment methodology to three applications involving exogenous information shocks for different housing attributes. These applications differed somewhat in their identification strategies, the complexity of the attribute, and also how the information event likely affected buyers' information acquisition.

The first of these applications described the impact of an airport noise disclosure on housing prices. The noise disclosure, given to buyers by sellers, was designed to cause buyers to "pay attention" to this housing market feature. Using spatial regression analysis and a limited temporal sample of houses that bracket the RDU airport noise disclosure to control for temporal and spatial confounders, the results indicate that the disclosure reduced housing prices near the airport by 2-3 percent. This price reduction

increased the *total* impact of airport noise on housing values in the area by 36%. Thus this application confirmed that in the absence of the disclosure hedonic estimates of the MWTP for airport noise would be attenuated towards zero.

The second application described the impact of a flood plain disclosure on housing prices. The flood plain disclosure attracted buyer's attention and may have changed the expected benefits and costs from acquiring flood plain designation information. Using a spatial fixed-effects strategy and a limited temporal and spatial sample of houses that bracketed both the timing of the disclosure and the flood zone areas, the estimates indicate that the disclosure reduced housing prices in FEMA designated flood zones by approximately 4 percent. Prior to the disclosure there appeared to be no impact of flood plain designation on housing prices in the area. Thus this application also suggests that estimates of marginal values for flood plains would be attenuated towards zero prior to the disclosure.

The third application described the impact of two types of information shocks on housing prices. The first information shock was related to sex offender movements and the second looked at the increased news coverage on registered sex offenders after two abductions in Florida. The first shock provided variation used to estimate the impact of registry information on housing prices whereas the second was hypothesized to have reduced the information asymmetry between buyers and sellers about sex offender locations. The results indicated that housing prices within one tenth of a mile of a registered sex offender are reduced by approximately 2 percent when a sex offender moves into a neighborhood. However, there was not a differential price impact after the two Florida abductions. It was argued that this finding was the result of the media

provided public information not providing a clear signal to buyers about sex offender information and risks.

Overall these findings support the concern identified at the outset with the full information assumption used in applying hedonic models to complex choice environments. This is potentially important because this model has been a central method for estimating the MWTP for environmental attributes. The same logic may also be relevant to other complex, information-rich environments and the corresponding economic models that are based on rational choice. Choosing where to go to college, determining which car to purchase, choosing a vocation, deciding where to go on vacation, selecting among job offers, and many other decisions may involve similar attention and search constraints for the information acquisition process of the consumer.

The remainder of this chapter discusses two extensions that that arise from this research. The first describes how one might attempt to isolate a good quasi-random experiment. The second describes future research opportunities related to the thesis. Although there are many possible specific extensions to the applications of this thesis (i.e. using nonparametric estimation, exploring the impact of multiple sex offender treatments on housing prices, etc.), this second set of topics focuses on broader areas for future research.

6.2 Important Elements of a Good Quasi-Random Experiment

Developing the evaluation of choices involving complex commodities required locating information events in the housing market that met three criteria: (i) The

information shock purported by a law change or other event actually took place, (ii) The information shock was exogenous to the housing market (and one could argue that there was a discrete information discontinuity over time) and (iii) data were available to create a control group allowing comparisons to the treatment group to be made. Two information events that were initially considered for use in this thesis but did not meet the criteria above are discussed here to provide insight into the selection process of quasi-random experiments.

Occurrence of the Quasi-Random Experiment

The task of verifying that the event took place may seem trivial. However, this verification is critical. One of the initial applications considered for this thesis involved evaluation the impact of an underground storage tank disclosure on housing prices in the state of New York. New York had passed a seller disclosure law that required homeowners to fill out a disclosure form (or face a $500 fine) which included several questions about underground storage tanks and fuel spills. It was hypothesized that this disclosure would reduce an information asymmetry between sellers and buyers about these events by reducing buyers' search and attention costs. After telephoning a small sample of real estate agents in the state of New York, it appeared that the disclosure form was being presented to buyers by sellers.[111] Two datasets were gathered to test the hypothesis: (i) a housing dataset for the entire state of New York and (ii) A spills database that recorded all of the spills that had been reported to the state department of environmental conservation (NYDEC).

[111] Approximately 10 real estate agents were called in various counties in New York.

After evaluating these data, the Long Island area (Nassau, Suffolk and Westchester counties) appeared to offer the only location in the state with sufficient observations to conduct the analysis and that was likely to comprise a single housing market. The results from an empirical analysis evaluating the disclosure suggested that there was *no impact* on the housing prices of homes that had previously had a leaky UST. This result rejected the primary hypothesis.

To confirm its plausibility, several regional NYDEC officers and real estate agents in the Long Island area were contacted to confirm the finding. These follow up interviews were surprising. Unlike other parts of New York, most sellers in the Long Island area never provided the disclosure form to buyers.[112] With the average house selling for approximately $400,000 in the Long Island area during the study time period (1995-2005), the risk of filling out the disclosure form appeared to have been outweighed by the $500 penalty from not providing the form.[113] Therefore, it appears no additional information was reaching buyers. This outcome suggested that a legal mandate does not guarantee it will yield the intended outcome. Compliance is an economic choice and penalties related to non-compliance influence the results.

Exogeneity of the Quasi-Random Experiment

Another information event that appeared promising was the Federal Lead Paint Disclosure law. Numerous scientific studies have linked lead to health problems in both adults and children. As a result lead was banned from paint in 1978, from house pipes in 1986 and from gasoline in 1995. Based on a 1990 study conducted by HUD, it was

[112] This suggests that my "random" sample of phone calls to real estate agents in New York must not have included real estate agents in this area.
[113] An interpretation of this behavior is that sellers were revealing that the disclosure had value through their $500 payment to opt-out of disclosing.

estimated that approximately 84 percent of homes built before 1980 had some lead based paint.

In 1992 Congress passed the Residential Lead-Based Paint Hazard Reduction Act. This legislation stipulated that the U.S. Environmental Protection Agency (EPA) and HUD should develop a strategy whereby information on lead-based paint was disclosed prior to a housing transaction for houses built before 1978. The EPA and HUD developed a Real Estate Notification and Disclosure Rule that among other things requires sellers and rental property owners disclose all known lead-based paint and lead-based paint hazards in the housing they own. This rule went into effect for owners of more than four dwelling units (rental property owners) on September 6, 1996 and for owners of four or fewer dwelling units (sellers) the effective date was December 6, 1996.

This disclosure rule appears to be an interesting quasi-random information experiment. However, the disclosure law fails the "plausibly exogenous" criteria. It is difficult to argue that September 6, 1996 represents a bright line for when information about lead paint began being released to buyers in the housing market. Since lead paint risks were being widely discussed during the late 1980's and early 1990's, the housing market was already beginning to adapt before the federal disclosure law was implemented. Many of the principal real estate companies developed their own "in-house" disclosure forms for lead paint that they began using between 1992 and 1996. Thus it is reasonable to assume the housing market was adjusting to lead paint information prior to 1996 which invalidates the "exogeneity" criteria for a good quasi-random experiment.

6.3 Research Issues that Emerge from this Thesis

6.3.1 Pros and Cons of Quasi-Random Experiments

Typically the application of the quasi-random experiment methodology is via reduced form regressions. The behavioral mechanism for the effect of the exogenous event on the outcome variable is not directly estimated. The goal is to estimate some "average treatment effect" for the population in the study. The value of this strategy (relative to more structural modeling) and the light it can shed on economic behavior is a current topic of debate.[114]

Regarding the difference between the structural approach and the "experimentalist" approach Angrist and Krueger (1999) state:

> "*An alternative to structural modeling, often called the quasi-experimental or simply the 'experimentalist' approach, also uses economic theory to frame causal questions. But this approach puts front and center the problem of identifying the causal effects from specific events or situations. The problem of generalization of findings is often left to be tackled later, perhaps with the aid of economic theory or informal reasoning……Of course there is considerable overlap between structural and quasi-experimental approaches to causal modeling, especially when it comes to data and measurement issues. The difference is primarily one of emphasis, because structural modeling generally incorporates some assumptions about exogenous variability in certain variables and quasi –experimental analyses require some theoretical assumptions*."

Whether imposing a statistical or theoretical restriction on the data places greater limitations on ones ability to identify a causal effect is unclear in many situations. However, what is clear is that whatever the exclusion restriction assumption, it must be valid in the data sample/population for a causal effect to be estimated. Thus the burden lies on the researcher to provide outside evidence that the identification restriction holds.

[114] For an intriguing look at this debate, see Heckman (1997) and the associated comment and response papers: Angrist and Imbens (1999) and Heckman (1999).

The quasi-random events studied in this thesis were each argued to have been exogenous to housing transactions prior to the event. Outside evidence was presented to make these arguments. However, even if the causal effects of these exogenous information shocks on housing prices were correctly identified in the study areas, one cannot necessarily argue that the effect took place via the mechanisms of attention and search as hypothesized. The most one can do is present supplementary evidence outside of the data used for estimation that would support one behavioral claim over other claims associated with other "information stories". The three examples in this thesis used a variety of anecdotal and other types of confirming information to appeal to a "weight of evidence" criterion. However, a fully parameterized model of housing choice was not specified nor was an explicit estimate for an "information parameter" provided.

Although the experimental approach, in conjunction with reduced-form models, is restricted in its ability to recover primitives of economic models, it does have the advantage of limiting the number of alternative explanations that can be given for the obtained estimates. The emphasis in the regressions in this thesis, was to bound the sample in time and space and include the covariates that would reduce threats to "internal" validity (i.e. concerns about endogeneity caused by selection bias) of the estimates.[115] This in conjunction with the "shoe-leather" research and anecdotal evidence that suggests the information shocks were not anticipated, is helpful in concluding that the estimates are causal for the treatment effects in these samples of data.

[115] See Meyer (1995) and Moffitt (2005) for informative discussions of internal and external validity as they relate to quasi-random experiments.

One possible consequence of this emphasis on internal validity is a loss of "external" validity (i.e. generalizability of the findings to other situations or populations). The spatial and temporal restrictions placed on the data in an effort to find an appropriate control may cause "investigator induced selection effects". This was perhaps most glaring in the sex offender application where the sample of sex offender relocations and housing data were restricted to just those areas where sex offenders moved to a single family residential house. Although this was done in order to eliminate those areas with multiple sex offender treatments, it also limits the generalizability of the results to areas with primarily rental units or areas with frequent sex offender movements. More work on the tradeoffs imposed by the quasi-random experiment methodology for these types of analyses is warranted in the future.

6.3.2 Data Availability: Opportunities for Quasi-Random Experiments

Increasing hedonic data accessibility, quantity and resolution provides opportunities for quasi-random experimental analysis. The applications in this thesis used housing datasets with large numbers of observations and excellent spatial resolution on the geographic location of homes. A danger of having the statistical power of hundreds of thousands or even millions of housing transaction observations is that a variable is much more likely to be statistically significant in a regression analysis. Rather than viewing the data as an opportunity to substantially adjust the degrees of freedom in a

regression, the applications in this thesis used the data as an opportunity to define control groups for the quasi-random experiment at hand.[116]

Given a housing sample that was argued to have been exogenously treated by an information shock, determining a useful control group involved constructing a rationale for isolating other houses in time and space that were thought to be comparable to the treated group. This was typically done by narrowing the spatial and temporal "windows" around the housing observations that were deemed "close" to the treated group. An important concern with limiting the number of housing observations is whether or not one can still estimate the hedonic price function for the market. Using all of the observations in a housing market will provide a more efficient estimate of implicit prices, but may not provide a reasonable control group. However, as long as the housing observations selected in the bounded control sample overlap with other segments of the housing market that are not included, then one can still consistently estimate the hedonic price function. The advantage to bounding the housing observations in time and space is that it typically provides a more reasonable control group by ruling out other explanations for the regression results that might be caused by omitted spatial or temporal factors.

This focus on the appropriate selection of a control group in large datasets is likely to be applicable to many other quasi-random experiment situations where micro-level data is extensive. This is related conceptually to the idea of determining an appropriate bound around the discontinuity in a regression discontinuity design. The

[116] The spatial resolution of the micro-level housing data also provided the opportunity to devise identification strategies that explicitly accounted for spatial autocorrelation. The applications in this thesis reemphasize the work of Mendelsohn *et al.* (1992), Black (1999), Poulos and Smith (2002), and Chay and Greenstone (2005), who point out the importance of devising appropriate spatial controls in quasi-random experimental analysis depending on the spatial extent of the attribute under consideration.

burden of proof falls on the researcher to provide an argument for the location of the bounds. Sensitivity analysis on the temporal and spatial bounds in quasi-random experimental work is essential. Additional work on systematically determining the appropriate spatial and temporal bounds is an area for future research.

6.3.3 Isolating the Impact of "Initial Attention"

Although the empirical work in this thesis was unable to formally isolate the behavioral mechanism underlying the influence of the information disclosures on housing prices, gaining a clearer understanding of that mechanism would be an important avenue for future research. It was hypothesized that both attention and search played a role in whether or not a buyer was informed about the housing attribute in question. "Initial Attention" was defined as that "spark" that caused a buyer to add the attribute to his consideration set of attributes.[117] From this set of attributes the buyer would calculate the expected benefits and costs of obtaining additional information about the attributes.

Delineating between initial attention and the search process was impossible in the reduced form regressions of this thesis. However, through richer datasets, experimental analysis, or some composite strategy that combines elements of richer data and experimentation, providing this delineation might be feasible.[118] One way to do this would be to conduct more refined experiments. The Gabaix and Laibson (2005b) paper that was discussed in chapter 2 (see section 2.3.1 for more detail) is an example of this in a controlled experimental setting. They found in their study that even though all of the

[117] This "spark" is also sometimes described as a "cue". See for example Bernheim and Rangel (2004).
[118] Another option might be to develop a structural model and then simulate the results of asymmetric information, for example, in an agent-based model. However, this may require some highly restrictive assumptions about agent interactions.

information relevant to choice was publicly available in their experiment, none of the subjects were fully informed when they made their decisions. This illustrated the importance of cognitive costs and time constraints for subjects' decision making. However, their analysis relied on several simplifications of their experiment. A very restrictive simplification was that the experiment automatically made the subjects consideration sets complete by constraining the set of relevant attributes to those in the experimental exercise. It would be desirable to further test their results in a real market setting where consideration sets are not necessarily "full".

A field experiment in the housing market joint with a follow-up survey may provide a solution to understanding the behavioral mechanism at work in a complex information environment like the housing market. An example would be if a real estate agency would allow the researcher to randomly assign their clients to information treatment groups. Some buyers could be randomly picked to receive an information treatment to ensure they were initially attentive to a certain housing disamenity, and others would receive an information treatment that not only caused them to be attentive, but also provided information at low cost about the disamenity. Finally some buyers would receive neither treatment. A follow-up survey could then be sent to these groups of buyers receiving the two treatments and a control group of buyers that did not receive either information treatment. The survey would ask all buyers about the house they purchased and what attributes were considered and what information was used to make

the purchase decision. With a housing transaction database and this information it might be possible to delineate the impacts of search and initial attention.[119]

6.3.4 Benefits and Costs of Re-Providing Public Information

Understanding the welfare effects of information programs that re-provide publicly available information such as the airport noise and flood zone disclosures is difficult. It is difficult because the revealed preference doctrine that underlies welfare economics and the hedonic model would predict that it is unbeneficial to re-provide, or cause agents to pay attention to information that is already publicly available. Yet this contrasts with policy makers "revealing" that they believe that these information programs have value. The empirical applications of this thesis provide evidence that re-released public information can impact equilibrium prices in the housing market suggesting that some households found the information beneficial in making their home purchase decisions.[120] These results suggest that buyers' choices diverged from their preferences prior to the disclosures because of a lack of information.

Consistency between choices and preferences simplifies welfare analysis of governmental programs. A divergence in choices and preferences such as in the pre-disclosure time periods in the applications complicates ones ability to measure a program's benefits. In order to determine a program's benefits, one must make a normative judgment about which "state" choices and preferences of agents most closely align. In the applications it was argued that choices in the "post-disclosure" state were

[119] A less costly exercise would be to perform a similar analysis on a hypothetical basis in a controlled lab experiment with couples who are in the process of buying a house.
[120] Buyers benefit from the information if it allows them to find a house that more closely matches their preferences. More efficient sorting in the housing market due to buyers being more fully informed

most closely aligned with buyers' preferences and therefore a more efficient matching of households to their preferred houses occurred. If this normative judgment is true, then it may be possible to measure the benefits from the increase in market efficiency and compare it to the costs of providing the information as a type of welfare measurement.

However, relaxing the doctrine of revealed preference is a slippery slope. It is possible that the information disclosures actually cause households to over-react to the information or causes them to divert their attention away from other housing attributes. Thus the validity of any such benefit-cost analysis hinges on the normative judgment made about the state in which choices and preferences were most closely aligned. By relaxing the revealed preference doctrine and applying it selectively it opens up the possibility of governmental abuse in deciding the "correct" welfare criteria for society [see Bernheim and Rangel (2005)]. Additional work on understanding the benefits from programs such as those studied in this thesis is likely an important but difficult task that is left for the future.

References

Akerlof, George A. (1970), "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84, 488-500.

Angrist, Joshua D. and Guido W. Imbens. (1999), "Comment on James J. Heckman, 'Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations'," *Journal of Human Resources*, 34, 823-827.

Angrist, Joshua D. and Alan B. Krueger. (1999), "Empirical Strategies in Labor Economics," Elsevier Science B.V., Handbook of Labor Economics, Volume 3, 1277-1366.

Angrist, Joshua D. and Alan B. Krueger. (2001), "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," *Journal of Economic Perspectives*, 15, 69–85.

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

Bajari, Patrick and Matthew E. Kahn. (2005), "Estimating Housing Demand With an Application to Explaining Racial Segregation in Cities," *Journal of Business and Economic Statistics*, 23, 20-33.

Bartik, Timothy J. (1987), "The Estimation of Demand Parameters in Hedonic Price Models," *Journal of Political Economy*, 95, 81-88.

Bernheim, B. Douglas and Antonio Rangel. (2004), "Addiction and Cue-Triggered Decision Processes," *American Economic Review*, 94, 1558-1590.

Bernheim, B. Douglas and Antonio Rangel. (2005), "Behavioral Public Economics: Welfare and Policy Analysis with Non-Standard Decision Makers," NBER Working Paper 11518.

Bin, Okmyung and Stephen Polasky. (2004), "Effects of Flood Hazards on Property Values: Evidence Before and After Hurricane Floyd," *Land Economics*, 80, 490-500.

Black, Sandra E. (1999), "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics*, 114, 577-599.

Brookshire, Mark Thayer, John Tschirhart and William Schulze. (1985), "A Test of the Expected Utility Model: Evidence from Earthquake Risks," *Journal of Political Economy*, 93, 369-389.

Bui, T. M. and C. Mayer. (2003), "Regulation and Capitalization on Environmental Amenities: Evidence from the Toxic Release Inventory in Massachusetts," *Review of Economics and Statistics*, 83, 281-302.

Chay, Kenneth Y. and Michael Greenstone. (2005), "Does Air Quality Matter? Evidence from the Housing Market," *Journal of Political Economy*, 113, 376-424.

Chivers, James and Nicholas E. Flores. (2002), "Market Failure in Information: The National Flood Insurance Program," *Land Economics*, 78, 515-521.

Cropper, M.L., L. Deck, and K.E. McConnell. (1988), "On the choice of functional forms for hedonic price functions," *Review of Economics and Statistics*, 70, 668-675.

Cropper, M.L., L. Deck, N. Kishor and K.E. McConnell. (1993), "Valuing product attributes using single market data: a comparison of hedonic and discrete choice approaches," *Review of Economics and Statistics*, 75, 225-232.

Della Vigna, S. and J. Pollet. (2004), "Strategic Release of Information on Friday: Evidence from Earnings Announcements," mimeo.

DeShazo, J.R. and G. Fermo. (2002) "Designing Choice Sets for Stated Preference Methods: The Effects of Complexity on Choice Consistency," *Journal of Environmental Economics and Management*, 43, 360-385.

DeShazo, J.R. and G. Fermo. (2004) "Implications of Rationally-Adaptive Pre-choice Behavior for the Design and Estimation of Choice Models." Mimeo.

Doren, Dennis M. (1998), "Recidivism Base Rates, Predictions of Sex Offender Recidivism, and the "Sexual Predator" Commitment Laws," *Behavioral Sciences and the Law*, 16, 97-114.

Dye, Ronald A. (1985), "Disclosure of Nonproprietary Information," *Journal of Accounting Research*, 23, 123-145.

Dye, Ronald A. (1998), "Investor Sophistication and Voluntary Disclosures," *Review of accounting studies*, 3, 261-187.

Ekeland, Ivar, James J. Heckman, and Lars Nesheim. (2004), "Identification and Estimation of Hedonic Models," *Journal of Political Economy*, 112, S60-S109.

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

Epple, Dennis and Holger Sieg. (1999), "Estimating Equilibrium Models of Local Jurisdiction," *Journal of Political Economy*, 107, 645-681.

Fishman, M. J. and K. M. Hagerty. (2003), "Mandatory Versus Voluntary Disclosure in Markets with Informed and Uninformed Customers," *Journal of Law Economics and Organization*, 19, 45-63.

Fulcher, Charles M. (2003), "Spatial Aggregation and Prediction in the Hedonic Model," Dissertation, North Carolina State University.

Fung, A, D. Weil, M. Graham and E. Fagatto. (2004), "The Political Economy of Transparency: What Makes Disclosure Policies Effective?" Ash Institute for Democratic Governance and Innovation, Occasional Paper OP-03-04.

Fiske, S. T. and S. E. Taylor. (1984), "Social Cognition," (Addison-Wesley, London).

Gabaix, X. and D. Laibson. (2005a), "Bounded Rationality and Directed Cognition," mimeo.

Gabaix, X. and D. Laibson. (2005b), "Costly Information Acquisition: Experimental Analysis of a Boundedly Rational Model," Forthcoming in *American Economic Review*.

Gayer, Ted, James Hamilton and W. Kip Viscusi. (2000), "Private Values of Risk Tradeoffs at Superfund Sites: Housing Market Evidence on Learning about Risk," *Review of Economics and Statistics*, 82, 439-451.

Grossman, S. (1981), "The Informational Role of Warranties and Private Disclosure About Information Quality," *Journal of Law and Economics*, 24, 461-483.

Grossman, S. and O. Hart. (1980), "Disclosure Laws and Takeover Bids," *Journal of Law and Economics*, 35, 323-334.

Hallstrom, Daniel G. and V. Kerry Smith. (2005), "Market Responses to Hurricanes," *Journal of Environmental Economics and Management*, 50, 541-561.

Halvorsen, R. and R. Palmquist. (1980), "The Interpretation of Dummy Variables in Semilogarithmic Equations," *American Economic Review*, 70, 474-475.

Hanson, R. Karl and M.T. Bussiere. (1998), "Predicting Relapse: A Meta-Analysis of Sexual Offender Recidivism Studies. *Journal of Consulting and Clinical Psychology*, 66, 348-362.

Hanson, R. Karl, Kelly E. Morton and Andrew J. R. Harris. (2003) "Sexual Offender Recidivism Risk: What We Know and What We Need to Know," *Annals New York Academy of Sciences*, 989, 154-166.

Harding, J.P., S.R. Rosenthal and C.F. Sirmans. (2003), "Estimating Bargaining Power in. the Market for Existing Homes," *Review of Economics and Statistics*, 85, 178-188.

Heckman, James J. (1997), "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations," *Journal of Human Resources*, 32, 441-462.

Heckman, James J. (1999), "Instrumental Variables: Response to Angrist and Imbens," *Journal of Human Resources*, 34, 828-837.

Hirshleifer, D. and S. Teoh. (2003), "Limited Attention, Information Disclosure, and Financial Reporting", *Journal of Accounting and Economics*, 36, 337-386.

Holoway, James M., and Raymond Burby. (1990), "The Effects of Floodplain Development Controls on Residential Land Values," *Land Economics*, 66, 259-270.

Huang, Ju-Chin and Raymond B. Palmquist. (2001), "Environmental Conditions, Reservation Prices, and Time on the Market for Housing," *Journal of Real Estate Finance and Economics*, 22, 203-219.

Jin, Ginger and Phillip Leslie. (2003), "The Effects of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards," *Quarterly Journal of Economics*.

Jovanovic, B. (1983), "Truthful Disclosure of Partially Verifiable Information," *Bell Journal of Economics*, 13, 36-44.

Kask, S. and S. Maani. (1992), "Uncertainty, Information, and Hedonic Pricing," *Land Economics*, 68, 170-184.

Kiel, Katherine and Katherine McClain. (1995), "House Prices during Siting Decision Stages: The Case of an Incinerator from Rumor through Operation," *Journal of Environmental Economics and Management*, 28, 241-255.

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

King, Rawle O. (2005), "Federal Flood Insurance: The Repetitive Loss Problem," CRS Report for Congress, Order Code RL32972.

Krutilla, John V. (1966), "An Economic Approach to Coping with Flood Damage," *Water Resources Research*, 2, 183-190.

Larsen, James E., Kenneth J. Lowrey and Joseph W. Coleman. (2003), "The Effect of Proximity to a Registered Sex Offender's Residence on Single-Family House Selling Price," *The Appraisal Journal*, 71, 253-265.

Lefcoe, G. (2004), "Property Condition Disclosure Forms: How the Real Estate Industry eased the Transition from *Caveat Emptor* to "Seller Tell All"," *Real Property, Probate and Trust Journal*, 39, 193-250.

Levitt, S. and C. Syverson. (2005), "Market Distortions when Agents are Better Informed: The Value of Information in Real Estate Transactions," NBER working paper 11053.

Linden, Leigh L. and Johah E. Rockoff. (2006), "There Goes the Neighborhood? Estimates of the Impact of Crime Risk on Property Values from Megan's Laws," NBER Working Paper 12253, May 2006.

Lippman, S. A. and J. J. McCall. (1976), "The Economics of Job Search: A Survey," *Economic Inquiry*, 14, 155-190.

Logan, Wayne A. (2003, "Sex Offender Registration and Community Notification: Emerging Legal and Research Issues," *Annals New York Academy of Sciences*, 989, 337-351.

Mankiw, N. and R. Reis. (2002), "Sticky Information Versus Sticky Prices: a Proposal to Replace the New Keynesian Phillips Curve," *Quarterly Journal of Economics*, 117, 1295-1328.

Mathios, Alan D. (2000), "The Impact of Mandatory disclosure Laws on Product Choices: An Analysis of the Salad Dressing Market," *Journal of Law and Economics*, XVIII, October.

McCluskey, Jill and Gordon C. Rausser. (2003), "Estimation of Perceived Risk and Its Effect on Property Values," *Land Economics*, 77, 42-55.

Mendelsohn, Robert, Daniel Hellerstein, Michael Huguenin, Robert Unsworth and Richard Brazee. (1992), "Measuring Hazardous Waste Damages with Panel Data Models," *Journal of Environmental Economics and Management*, 22, 259-271.

Meyer, Bruce D. (1995), "Natural and Quasi-Experiments in Economics," *Journal of Business & Economic Statistics*, 13, 151-161.

Michaels, R. Gregory and V. Kerry Smith Bruce D. (1990), "Market Segmentation and Valuing Amenities with Hedonic Models: The Case of Hazardous Waste Sites," *Journal of Urban Economics*, 28, 223-242.

Milgrom, Paul. (1981), "Good News and Bad News: Representation Theorems and Applications," *Rand Journal of Economics*, 12, 380-391.

Moffitt, Robert. (2005), "Remarks on the Analysis of Causal Relationships in Population Research," *Demography*, 42, 91-108.

Muckleston, K. W. (1983), "The Impact of Flood-plain Regulations on Residential Land Values in Oregon," *Water Resources Bulletin*, 19, 1-7.

Nelson, Jon P. (2004), "Meta-Analysis of Airport Noise and Hedonic Property Values," *Journal of Transport Economics and Policy*, 38, 1-28.

Palmquist, R. B. (2005), "Property Value Models," *Handbook of Environmental Economics*, edited by Karl-Göran Mäler and Jeffery R. Vincent. Amsterdam: North-Holland.

Pashler, H. (1998), "The Psychology of Attention," MIT Press: Cambridge, Mass.

Payne, J., J. Bettman, and E. Johnson. (1993), "The Adaptive Decision Maker," Cambridge, UK: Cambridge University Press.

Pope, D. and J. Pope. (2005), "Understanding College Choice Decisions: How Sports Success Garners Attention and Provides Information," mimeo.

Poulos, Christine and V. Kerry Smith. (2002), "Transparency and Takings: Applying an RD Design to measure compensation," unpublished working paper, CEnREP, North Carolina State University.

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

Rosenzweig, Mark R. and Kenneth I. Wolpin. (2000), "Natural 'Natural Experiments' in Economics," *Journal of Economic Literature*, 38, 827–74.

Schulze, W., G. McClelland, B. Hurd, and J. Smith. (1986), "A Case Study of Hazardous Waste Site: Perspectives from Economics and Psychology," Draft report, U.S. Environmental Protection Agency, Contract CR812054-01-1.

Shin, Hyun Song. (1994), "News Management and the Value of Firms," *Rand Journal of Economics*, 25, 58-71.

Sims, C. (2003), "Implications of Rational Inattention," *Journal of Monetary Economics*, 50, 665-690.

Simon, Herbert A., "Designing Organizations for an Information-Rich World," in Martin Greenberger, ed., *Computers, communications and the public interest*, Baltimore: Johns Hopkins University Press, 1971, 40-41

Slovic, P., M. Finucane, E. Peters and D. G. MacGregor. (2004) "Risk as Analysis and Risk as Feelings: Some Thoughts about Affect, Reason, Risk, and Rationality," *Risk Analysis*, 24, 311-322.

Speyrer, J. and W. Ragas. (1991), "Housing Prices and Flood Risk: An Examination Using Spline Regression" *Journal of Real estate Finance and Economics*, 4, 395-407.

Stigler, G. (1961), "The Economics of Information," *Journal of Political Economy*, 69, 213-225.

Tobin, G. A. and B. E. Montz. (1994), "The Flood Hazard and Dynamics of the Urban Residential Land Market," *Water Resources Research*, 2, 183-190.

Turnbull, Geoffrey K. and C. F. Sirmans. (1993), "Information, search, and house prices," *Regional Science and Urban Economics*, 23, 545-557.

US General Accounting Office, (2000a), "Aviation and the Environment: Airport Operations and Future Growth Present Environmental Challenges," GAO/RCED-00-153, GAO, Washington, D.C.

US General Accounting Office, (2000b), "Aviation and the Environment: Results from a Survey of the Nations's 50 Busiest Commercial Airports," GAO/RCED-00-222, GAO, Washington, D.C.

U.S. General Accounting Office. (2003), "Challenges Facing the National Flood Insurance Program," GAO Report GAO-03-606T, p. 16.

Viscusi, W. Kip. (1978), "A Note on 'Lemons' Markets with Quality Certification," *Bell Journal of Economics*, 9, 277-279.

Appendices

Appendix 3.A

AIRCRAFT NOISE NOTIFICATION

Dear Property Owner:

You are listed by the Wake County Tax Office as the owner of a parcel of land located within the general area surrounding Raleigh-Durham International Airport (RDU) that is exposed to average aircraft noise levels which exceed typical ground-based, or background, noise. The map displays that area and shows contours of equal average aircraft noise exposure associated with current flight operations at the airport and those projected through approximately the year 2010. Sites closer to the airport are exposed to higher average noise levels than those farther away.

The purpose of this notice is to advise you that exposure to aircraft noise may affect the usability of some land for certain types of noise sensitive uses, including residential use. Persons who are sensitive to aircraft noise should satisfy themselves before buying the property that exposure to such noise will not materially affect their ability to use and enjoy land whose purchase they may be considering.

The Raleigh-Durham Airport Authority has and, upon request, will provide information which may be helpful to property owners and prospective purchasers in assessing the likely effect of aircraft noise on the use of land they own or are considering purchasing.

You also are advised that the "Residential Property Disclosure Act" (N.C.G.S. Chapter 47E) was enacted by the North Carolina General Assembly and became effective January 1, 1996. That law requires the owners of residential real property to disclose to prospective purchasers the existence of certain conditions associated with the property no later than the time an offer to purchase, exchange or option the property is made, or an option to purchase the property pursuant to a lease with an option to purchase is exercised.

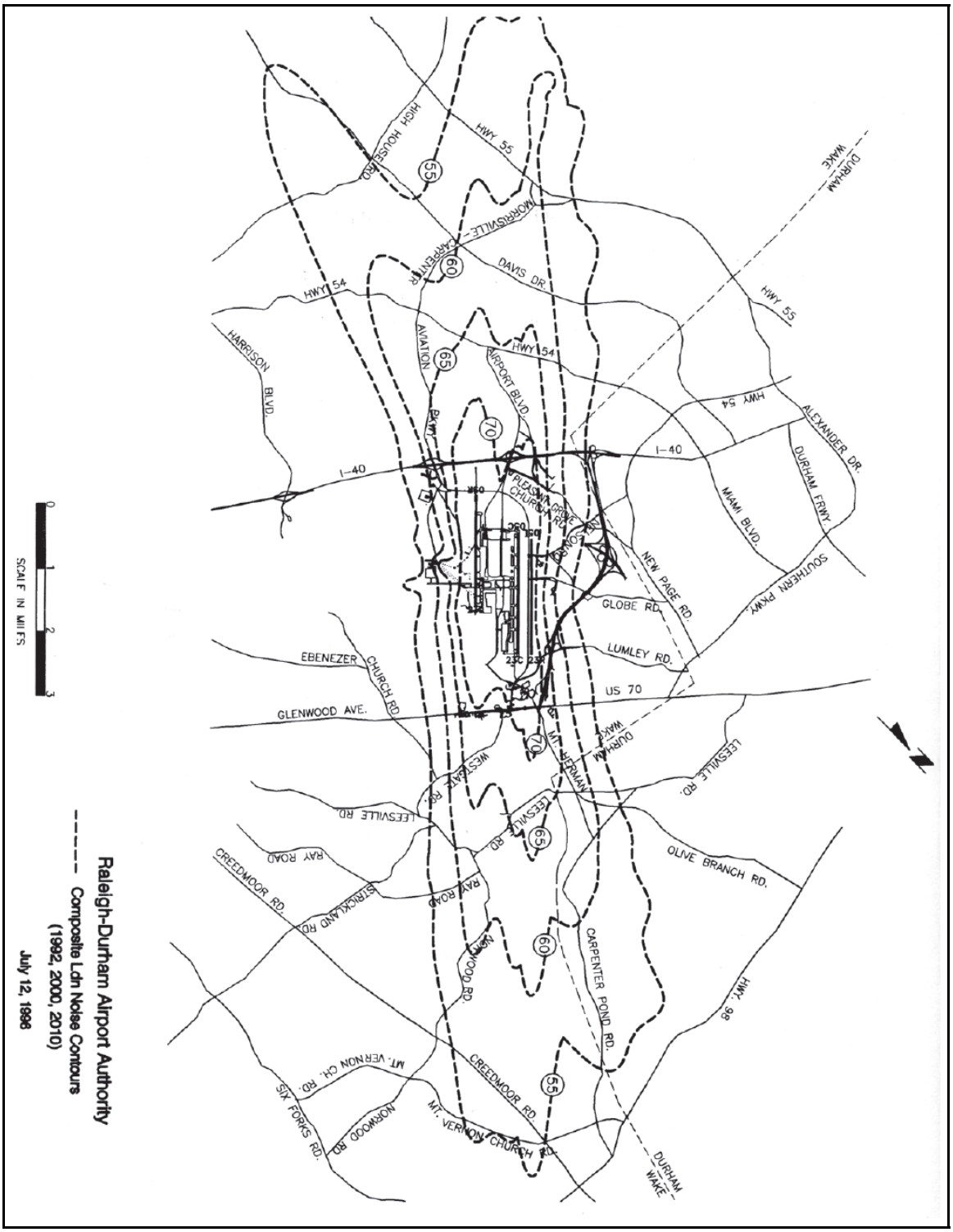
Among the conditions that must be disclosed to and acknowledged by the prospective purchaser are any notice from any governmental agency affecting the property. The Airport Authority is a governmental agency. THIS NOTICE SERVES AS YOUR NOTICE OF POTENTIAL AIRCRAFT NOISE IMPACT UPON YOUR PROPERTY AND SHOULD BE DISCLOSED TO ALL PROSPECTIVE PURCHASERS WHO MAY BE CONSIDERING USE OF THE PROPERTY FOR A RESIDENTIAL PURPOSE.

For additional information or if you have questions or need assistance, please call the RDU Noise Officer at 919-840-2100 between 9:00 a.m. and 5:00 p.m. Monday-Friday or write to:

Noise Officer
Raleigh-Durham Airport Authority
P. O. Box 80001
RDU Airport, North Carolina 27623-0001

The Raleigh-Durham Airport Authority

Appendix 3.A (Continued)



Appendix 3.B: Additional Regression Results

Dep. Var. = lprice	OLS (1992-2000)	OLS (1995-1998)	OLS (1996-1997)	OLS using spatial dataset (1996-1997)	ML Rook Contiguity Lag (1996-1997)	ML Queen Contiguity Lag (1996-1997)
	[1]	[2]	[3]	[4]	[5]	[6]
Variable						
yrdum93_apr	0.054					
	[0.004]**					
yrdum94_apr	0.138					
	[0.004]**					
yrdum95_apr	0.181					
	[0.004]**					
yrdum96_apr	0.199	0.018				
	[0.004]**	[0.004]**				
yrdum97_apr	0.231	0.050	0.036	0.036	0.034	0.034
	[0.005]**	[0.005]**	[0.005]**	[0.005]**	[0.004]**	[0.004]**
yrdum98_apr	0.258	0.077				
	[0.005]**	[0.005]**				
yrdum99_apr	0.290					
	[0.005]**					
yrdum00_apr	0.327					
	[0.006]**					
age	-0.006	-0.006	-0.008	-0.008	-0.006	-0.006
	[0.000]**	[0.000]**	[0.000]**	[0.001]**	[0.000]**	[0.000]**
baths	0.028	0.025	0.03	0.03	0.024	0.024
	[0.002]**	[0.003]**	[0.003]**	[0.003]**	[0.003]**	[0.003]**
acreage	0.075	0.071	0.08	0.081	0.063	0.063
	[0.002]**	[0.003]**	[0.004]**	[0.004]**	[0.003]**	[0.003]**
regheatarea	2.75E-04	2.65E-04	2.65E-04	2.64E-04	1.96E-04	1.96E-04
	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**
detgarage	0.08	0.066	0.091	0.087	0.062	0.062
	[0.008]**	[0.011]**	[0.015]**	[0.015]**	[0.012]**	[0.012]**
fireplaces	0.005	0.006	0.012	0.009	0.015	0.015
	[0.004]	[0.006]	[0.008]	[0.008]	[0.007]**	[0.007]**
deck	1.81E-04	1.92E-04	1.95E-04	1.96E-04	1.30E-04	1.30E-04
	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**
sewer	-0.024	-0.033	-0.045	-0.052	-0.018	-0.018
	[0.004]**	[0.006]**	[0.008]**	[0.008]**	[0.006]**	[0.006]**
flordum1	-0.028	-0.04	-0.006	-0.004	-0.010	-0.010
	[0.010]**	[0.015]**	[0.019]	[0.020]	[0.016]	[0.016]
scrporch	2.82E-04	3.39E-04	3.16E-04	3.17E-04	2.11E-04	2.11E-04
	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**
walldum1	0.099	0.122	0.135	0.134	0.087	0.087
	[0.004]**	[0.006]**	[0.008]**	[0.008]**	[0.006]**	[0.006]**
atticheat	1.89E-04	1.84E-04	1.91E-04	1.93E-04	1.53E-04	1.53E-04
	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**

Appendix 3.B (Continued)

bsmtheat	8.44E-05	7.16E-05	6.29E-05	6.44E-05	4.22E-05	4.22E-05
	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**
garage	3.15E-04	3.19E-04	2.88E-04	2.83E-04	1.79E-04	1.79E-04
	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**
poolres	0.008	0.004	-0.014	-0.017	0.014	0.014
	[0.011]	[0.015]	[0.018]	[0.018]	[0.015]	[0.015]
bsmtdum1	0.09	0.09	0.101	0.101	0.079	0.079
	[0.005]**	[0.007]**	[0.010]**	[0.010]**	[0.008]**	[0.008]**
bsmtdum2	0.073	0.089	0.095	0.093	0.082	0.082
	[0.007]**	[0.010]**	[0.014]**	[0.014]**	[0.011]**	[0.011]**
encporch	4.71E-05	1.93E-05	-1.73E-04	-1.71E-04	3.21E-05	3.21E-05
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]**	[0.000]**
opnporch	1.31E-04	1.63E-04	9.33E-05	9.23E-05	7.04E-05	7.04E-05
	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**
condadum	0.065	0.057	0.055	0.064	0.022	0.022
	[0.007]**	[0.009]**	[0.013]**	[0.013]**	[0.010]**	[0.010]**
condcdum	-0.066	-0.064	-0.14	-0.091	-0.110	-0.110
	[0.012]**	[0.019]**	[0.028]**	[0.030]**	[0.024]**	[0.024]**
condddum	-0.101	-0.139	-0.151	-0.233	-0.109	-0.109
	[0.030]**	[0.045]**	[0.076]**	[0.100]**	[0.081]	[0.081]
perc_nonwhite_1990	-4.91E-04	-0.001	-0.001	-0.001	-0.002	-0.002
	[0.000]**	[0.000]**	[0.000]*	[0.000]**	[0.000]**	[0.000]**
medianvalue_int	-5.98E-03	7.82E-07	8.01E-07	7.75E-07	-0.003	-0.003
	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.001]**	[0.001]**
medttw_int	-0.006	-0.007	-0.006	-0.006	-0.003	-0.003
	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**
perc_under18_int	0.002	0.002	0.002	0.003	0.002	0.002
	[0.000]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**	[0.001]**
perc_owner_occ_int	-0.001	-0.001	-0.001	-0.001	0.000	0.000
	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**	[0.000]**
nearestpark	0.006	0.004	0.005	0.004	0.003	0.003
	[0.001]**	[0.002]**	[0.002]**	[0.002]*	[0.002]*	[0.002]*
nearestsc	0.003	0.006	0.005	0.007	0.002	0.002
	[0.001]**	[0.001]**	[0.002]**	[0.002]**	[0.001]	[0.001]
taxrate	0.066	0.056	0.073	0.07	0.048	0.048
	[0.008]**	[0.011]**	[0.014]**	[0.014]**	[0.011]**	[0.011]**
Observations	16856	8846	4353	4250	4250	4250
R-squared / pseudo	0.89	0.88	0.89	0.89	0.93	0.93

Notes: Standard errors in brackets.
* significant at 10%; ** significant at 5%

Appendix 4.A

STATE OF NORTH CAROLINA
RESIDENTIAL PROPERTY DISCLOSURE STATEMENT

INSTRUCTIONS TO PROPERTY OWNERS

1. G.S. 47E requires owners of residential real estate (single-family homes and buildings with up to four dwelling units) to furnish purchasers a property disclosure statement. This form is the only one approved for this purpose. A disclosure statement must be furnished in connection with the sale, exchange, option and sale under a lease with option to purchase (unless the tenant is already occupying or intends to occupy the dwelling). A disclosure statement is not required for some transactions, including the first sale of a dwelling which has never been inhabited and transactions of residential property made pursuant to a lease with option to purchase where the lessee occupies or intends to occupy the dwelling. For a complete list of exemptions, see G.S. 47E-2.
2. You must check one of the boxes for each of the 20 questions on the reverse side of this form.
   a. If you check "Yes" for any question, you must describe the problem or attach a report from an engineer, contractor, pest control operator or other expert or public agency describing it. If you attach a report, you will not be liable for any inaccurate or incomplete information contained in it so long as you were not grossly negligent in obtaining or transmitting the information.
   b. If you check "No", you are stating that you have no actual knowledge of any problem. If you check "No" and you know there is a problem, you may be liable for making an intentional misstatement.
   c. If you check "No Representation", you have no duty to disclose the conditions or characteristics of the property, even if you should have known of them.
   * If you check "Yes" or "No" and something happens to the property to make your Statement incorrect or inaccurate (for example, the roof begins to leak), you must promptly give the purchaser a corrected Statement or correct the problem.
3. If you are assisted in the sale of your property by a licensed real estate broker, you are still responsible for completing and delivering the Statement to the purchasers; and the broker must disclose any material facts about your property which they know or reasonably should know, regardless of your responses on the Statement.
4. You must give the completed Statement to the purchaser no later than the time the purchaser makes an offer to purchase your property. If you do not, the purchaser can, under certain conditions, cancel any resulting contract (See **"Note to Purchasers"** below). You should give the purchaser a copy of the Statement containing your signature and keep a copy signed by the purchaser for your records.

Note to Purchasers: If the owner does not give you a Residential Property Disclosure Statement by the time you make your offer to purchase the property, you may under certain conditions cancel any resulting contract and be entitled to a refund of any deposit monies you may have paid. To cancel the contract, you must personally deliver or mail written notice of your decision to cancel to the owner or the owner's agent within three calendar days following your receipt of the Statement, or three calendar days following the date of the contract, whichever occurs first. However, in no event does the Disclosure Act permit you to cancel a contract after settlement of the transaction or (in the case of a sale or exchange) after you have occupied the property, whichever occurs first.

5. In the space below, type or print in ink the address of the property (sufficient to identify it) and your name. Then sign and date.

Property Address: _____

Owner's Name(s): _____

Owner(s) acknowledge having examined this Statement before signing and that all information is true and correct as of the date signed.

Owner Signature: _____ Date _____

Owner Signature: _____ Date _____

Purchaser(s) acknowledge receipt of a copy of this disclosure statement; that they have examined it before signing; that they understand that this is not a warranty by owner or owner's agent; that it is not a substitute for any inspections they may wish to obtain; and that the representations are made by the owner and not the owner's agent(s) or subagent(s). Purchaser(s) are encouraged to obtain their own inspection from a licensed home inspector or other professional.

Purchaser Signature: _____ Date _____

Purchaser Signature: _____ Date _____

(OVER)

REC 4.22
REV 4/06

Appendix 4.A (Continued)

Property Address/Description: _____

[Note: In this form, "property" refers only to dwelling unit(s) and not sheds, detached garages or other buildings.]

Regarding the property identified above, do you know of any problem (malfunction or defect) with any of the following:

	Yes*	No	No Representation
1. FOUNDATION, SLAB, FIREPLACES/CHIMNEYS, FLOORS, WINDOWS (INCLUDING STORM WINDOWS AND SCREENS), DOORS, CEILINGS, INTERIOR AND EXTERIOR WALLS, ATTACHED GARAGE, PATIO, DECK OR OTHER STRUCTURAL COMPONENTS including any modifications to them?	☐	☐	☐
a. Siding is ☐ Masonry ☐ Wood ☐ Composition/Hardboard ☐ Vinyl ☐ Synthetic Stucco ☐ Other _____			☐
b. Approximate age of structure? _____			☐
2. ROOF (leakage or other problem)?	☐	☐	☐
a. Approximate age of roof covering? _____			☐
3. WATER SEEPAGE, LEAKAGE, DAMPNESS OR STANDING WATER in the basement, crawl space or slab?	☐	☐	☐
4. ELECTRICAL SYSTEM (outlets, wiring, panel, switches, fixtures etc.)?	☐	☐	☐
5. PLUMBING SYSTEM (pipes, fixtures, water heater, etc.)?	☐	☐	☐
6. HEATING AND/OR AIR CONDITIONING?	☐	☐	☐
a. Heat Source is: ☐ Furnace ☐ Heat Pump ☐ Baseboard ☐ Other_____			☐
b. Cooling Source is: ☐ Central Forced Air ☐ Wall/Window Unit(s) ☐ Other_____			☐
c. Fuel Source is: ☐ Electricity ☐ Natural Gas ☐ Propane ☐ Oil ☐ Other _____			☐
7. WATER SUPPLY (including water quality, quantity and water pressure)?	☐	☐	☐
a. Water supply is: ☐ City/County ☐ Community System ☐ Private Well ☐ Other _____			☐
b. Water pipes are: ☐ Copper ☐ Galvanized ☐ Plastic ☐ Other _____ ☐ Unknown			☐
8. SEWER AND/OR SEPTIC SYSTEM?	☐	☐	☐
a. Sewage disposal system is: ☐ Septic Tank ☐ Septic Tank with Pump ☐ Community System ☐ Connected to City/County System ☐ City/County System available ☐ Straight pipe (wastewater does not go into a septic or other sewer system [note: use of this type of system violates state law]) ☐ Other _____			☐
9. BUILT-IN APPLIANCES (RANGE/OVEN, ATTACHED MICROWAVE, HOOD/FAN, DISHWASHER, DISPOSAL, etc.)?	☐	☐	☐

Also regarding the property identified above, including the lot, other improvements, and fixtures located thereon, do you know of any:

10. PROBLEMS WITH PRESENT INFESTATION, OR DAMAGE FROM PAST INFESTATION OF WOOD DESTROYING INSECTS OR ORGANISMS which has not been repaired?	☐	☐	☐
11. PROBLEMS WITH DRAINAGE, GRADING OR SOIL STABILITY OF LOT?	☐	☐	☐
12. PROBLEMS WITH OTHER SYSTEMS AND FIXTURES: CENTRAL VACUUM, POOL, HOT TUB, SPA, ATTIC FAN, EXHAUST FAN, CEILING FAN, SUMP PUMP, IRRIGATION SYSTEM, TV CABLE WIRING OR SATELLITE DISH, OR OTHER SYSTEMS?	☐	☐	☐
13. ROOM ADDITIONS OR OTHER STRUCTURAL CHANGES ?	☐	☐	☐
14. ENVIRONMENTAL HAZARDS (substances, materials or products) including asbestos, formaldehyde, radon gas, methane gas, lead-based paint, underground storage tank, or other hazardous or toxic material (whether buried or covered), contaminated soil or water, or other environmental contamination)?	☐	☐	☐
15. COMMERCIAL OR INDUSTRIAL NUISANCES (noise, odor, smoke, etc.) affecting the property?	☐	☐	☐
16. VIOLATIONS OF BUILDING CODES, ZONING ORDINANCES, RESTRICTIVE COVENANTS OR OTHER LAND-USE RESTRICTIONS?	☐	☐	☐
17. UTILITY OR OTHER EASEMENTS, SHARED DRIVEWAYS, PARTY WALLS OR ENCROACHMENTS FROM OR ON ADJACENT PROPERTY?	☐	☐	☐
18. LAWSUITS, FORECLOSURES, BANKRUPTCY, TENANCIES, JUDGMENTS, TAX LIENS, PROPOSED ASSESSMENTS, MECHANICS' LIENS, MATERIALMENS' LIENS, OR NOTICE FROM ANY GOVERNMENTAL AGENCY that could affect title to the property?	☐	☐	☐
19. OWNERS' ASSOCIATION OR "COMMON AREA" EXPENSES OR ASSESSMENTS?	☐	☐	☐
20. FLOOD HAZARD or that the property is in a FEDERALLY-DESIGNATED FLOOD PLAIN?	☐	☐	☐

* **If you answered "Yes" to any of the above questions, please explain (Attach additional sheets, if necessary):** _____

_____

_____

_____

Appendix 4.B: Additional Regression Results

Dep. Var. = lprice Variable	*(1)	*(2)	*(3)
age	-0.004	-0.004	-0.004
	[0.000]***	[0.000]***	[0.000]***
baths	0.042	0.042	0.04
	[0.003]***	[0.003]***	[0.003]***
acreage	0.026	0.027	0.027
	[0.005]***	[0.005]***	[0.005]***
regheatarea	0	0	0
	[0.000]***	[0.000]***	[0.000]***
detgarage	0.086	0.086	0.078
	[0.008]***	[0.008]***	[0.008]***
fireplaces	0.041	0.041	0.039
	[0.004]***	[0.004]***	[0.004]***
deck	0	0	0
	[0.000]***	[0.000]***	[0.000]***
sewer	-0.015	-0.015	-0.016
	[0.005]***	[0.005]***	[0.005]***
flordum1	-0.004	-0.004	-0.001
	[0.007]	[0.007]	[0.007]
scrporch	0	0	0
	[0.000]***	[0.000]***	[0.000]***
walldum1	0.039	0.039	0.041
	[0.005]***	[0.005]***	[0.005]***
atticheat	0	0	0
	[0.000]***	[0.000]***	[0.000]***
bsmtheat	0	0	0
	[0.000]***	[0.000]***	[0.000]***
garage	0	0	0
	[0.000]***	[0.000]***	[0.000]***
poolres	0.014	0.014	0.015
	[0.013]	[0.013]	[0.013]
bsmtdum1	0.107	0.107	0.104
	[0.008]***	[0.008]***	[0.008]***
bsmtdum2	0.105	0.105	0.098
	[0.007]***	[0.007]***	[0.007]***
encporch	0	0	0
	[0.000]***	[0.000]***	[0.000]***
opnporch	0	0	0
	[0.000]***	[0.000]***	[0.000]***
condadum	0.071	0.071	0.08
	[0.008]***	[0.008]***	[0.008]***
condcdum	-0.116	-0.116	-0.12
	[0.010]***	[0.010]***	[0.010]***
condddum	-0.403	-0.403	-0.371
	[0.036]***	[0.036]***	[0.038]***
new_sale	0.015	0.015	0.017
	[0.003]***	[0.003]***	[0.003]***
perc_nonwhite_1990	-0.001	-0.001	-0.004
	[0.000]***	[0.000]***	[0.001]***
medianvalue_int	0	0	0
	[0.000]***	[0.000]***	[0.000]***
medttw_int	-0.003	-0.003	-0.011
	[0.001]***	[0.001]***	[0.003]***
perc_under18_int	0	0	-0.01
	[0.001]	[0.001]	[0.003]***
perc_owner_occ_int	0	0	0
	[0.000]	[0.000]	[0.001]
nearestpark	-0.001	-0.001	-0.002
	[0.001]	[0.001]	[0.001]*
nearestsc	-0.001	-0.001	-0.003
	[0.001]	[0.001]	[0.001]**
taxrate	0.012	0.012	0.011
	[0.009]	[0.009]	[0.009]
Observations	15514	15514	15514
R-squared	0.92	0.92	0.92

Notes: **Standard errors in brackets.**
*** significant at 10%; ** significant at 5%; *** significant at 1%**

Appendix 5.A: Housing Summary Statistics for Areas with different #'s of Sex Offenders

Variable	Description	Mean	Median	Standard Deviation	Minimum	Maximum	Observations
Panel A: All Single Family Transactions from 1996-2006							
price	sale price of property	172,123.50	143,900.00	114,601.60	5,700.00	2,863,299.00	189,285
age	age of property in years	19.36	13.00	20.71	0.00	106.00	189,285
acreage	acreage of lot	0.30	0.19	0.82	0.00	123.97	189,285
eff_ar	the "effective" square footage	2,065.16	1,953.00	779.88	105.00	10,749.00	189,285
onetenth_mile	within .1 miles of S.O.	0.16	0.00	0.37	0.00	1.00	189,285
twotenth_mile	between .1 and .2 miles of S.O.	0.20	0.00	0.40	0.00	1.00	189,285
Panel B: Transactions in Areas with 0 S.O.'s within .3 Miles							
price	sale price of property	209,584.00	175,300.00	129,145.30	5,700.00	2,863,299.00	94,450
age	age of property in years	12.08	5.00	15.99	0.00	105.00	94,450
acreage	acreage of lot	0.37	0.21	1.09	0.00	123.97	94,450
eff_ar	the "effective" square footage	2,386.00	2,264.00	775.43	191.00	8,687.00	94,450
onetenth_mile	within .1 miles of S.O.	0.00	0.00	0.00	0.00	0.00	94,450
twotenth_mile	between .1 and .2 miles of S.O.	0.00	0.00	0.00	0.00	0.00	94,450
Panel C: Transactions in Areas with only 1 S.O. within .3 Miles							
price	sale price of property	161,061.90	138,500.00	96,467.63	6,000.00	2,200,000.00	33,947
age	age of property in years	18.21	14.00	18.28	0.00	106.00	33,947
acreage	acreage of lot	0.26	0.19	0.46	0.00	25.07	33,947
eff_ar	the "effective" square footage	2,001.31	1,891.00	664.20	317.00	6,818.00	33,947
onetenth_mile	within .1 miles of S.O.	0.15	0.00	0.36	0.00	1.00	33,947
twotenth_mile	between .1 and .2 miles of S.O.	0.36	0.00	0.48	0.00	1.00	33,947
Panel D: Transactions in Areas with 2-5 S.O.'s within .3 Miles							
price	sale price of property	134,018.40	120,000.00	75,487.02	7,000.00	1,600,000.00	37,711
age	age of property in years	25.41	22.00	20.33	0.00	106.00	37,711
acreage	acreage of lot	0.24	0.18	0.41	0.01	50.00	37,711
eff_ar	the "effective" square footage	1,769.37	1,672.00	612.33	105.00	10,749.00	37,711
onetenth_mile	within .1 miles of S.O.	0.31	0.00	0.46	0.00	1.00	37,711
twotenth_mile	between .1 and .2 miles of S.O.	0.44	0.00	0.50	0.00	1.00	37,711
Panel E: Transactions in Areas with 5-10 S.O.'s within .3 Miles							
price	sale price of property	102,350.40	90,000.02	52,941.40	6,500.00	924,999.80	16,769
age	age of property in years	38.16	40.00	22.40	0.00	105.00	16,769
acreage	acreage of lot	0.18	0.16	0.12	0.02	4.20	16,769
eff_ar	the "effective" square footage	1,374.62	1,310.00	406.66	263.00	8,386.00	16,769
onetenth_mile	within .1 miles of S.O.	0.53	1.00	0.50	0.00	1.00	16,769
twotenth_mile	between .1 and .2 miles of S.O.	0.42	0.00	0.49	0.00	1.00	16,769
Panel F: Transactions in Areas more than 10 S.O.'s within .3 Miles							
price	sale price of property	85,414.37	78,000.02	42,643.82	6,000.00	425,000.10	6,408
age	age of property in years	48.05	48.00	25.14	0.00	106.00	6,408
acreage	acreage of lot	0.15	0.14	0.07	0.01	1.86	6,408
eff_ar	the "effective" square footage	1,222.07	1,202.00	338.85	256.00	4,235.00	6,408
onetenth_mile	within .1 miles of S.O.	0.75	1.00	0.43	0.00	1.00	6,408
twotenth_mile	between .1 and .2 miles of S.O.	0.24	0.00	0.43	0.00	1.00	6,408

Appendix 5.B: Multinomial Logit Test of Comparability of Areas with Different #'s of Sex Offenders

variables	Only 1 Sex Offender [1]	2-5 Sex Offenders [2]	5-10 Sex Offenders [3]	>10 Sex Offenders [4]
age	0.014	0.025	0.04	0.054
	[0.000]**	[0.000]**	[0.001]**	[0.001]**
acreage	-0.295	-0.582	-2.742	-8.116
	[0.018]**	[0.024]**	[0.094]**	[0.265]**
utsBR	0.335	0.122	0.813	0.543
	[0.050]**	[0.047]**	[0.073]**	[0.085]**
utsBR_2	-0.038	-0.008	-0.101	0.008
	[0.007]**	[0.007]	[0.013]**	[0.015]
utsBT	0.002	-0.038	-0.046	-0.096
	[0.006]	[0.006]**	[0.009]**	[0.013]**
eff_ar	-0.001	-0.001	-0.002	-0.003
	[0.000]**	[0.000]**	[0.000]**	[0.000]**
AC_dum2	0.21	0.195	0.12	-0.229
	[0.063]**	[0.053]**	[0.057]**	[0.063]**
Constant	-0.568	0.614	0.957	1.727
	[0.100]**	[0.090]**	[0.117]**	[0.141]**
Observations	189,282	189,282	189,282	189,282

Notes: Standard errors are presented in brackets. The * represents significance at the 10% level and the ** represents significance at the 5% level. Omitted category are houses that have never had a sex offender live within 0.3 miles. Chi Squared tests for differences across 4 models were significant for all but AC_dum2