Essays on the Use of Hedonic Price Models to Measure Welfare for Quality Changes in the Public Goods

Congwen Zhang

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
In
Economics

Kevin J. Boyle, Chair
Nicolai V. Kuminoff
Jaren C. Pope
Wen You
Darrell Bosch

April 27, 2012
Blacksburg, Virginia

Three Essays on Hedonic Price Method

Congwen Zhang

(Abstract)

This dissertation consists of three essays on Hedonic price method which is widely used in non-market good evaluation. The first chapter outlines three topics involved and briefly discusses the motivations and methods, as well we some conclusions in each of the following chapters.

Chapter 2 uses a conventional first stage hedonic price method to estimate the effect of an aquatic invasive species (Eurasian watermilfoil) on lakefront property values at selected Vermont lakes. Results indicate that as the primary component of total aquatic macrophyte growth in a lake Eurasian watermilfoil significantly and substantially affects lakefront property values. As Eurasian watermilfoil infests a lake, adding to the total macrophyte growth, property values can diminish by <1% to 16% for incremental increases in the infestation level. Hence, policies that successfully prevent infestations have significant economic benefits to owners of lakefront properties and local communities.

Chapter 3 focused on a previously unexplored potential impact of 9/11—the impact it may have had on housing prices near mosques. Using a unique dataset that combines the locations of functioning mosques with housing transactions near the time of 9/11, combined with a generalized difference-in-differences framework, we find that housing prices decreased by approximately 7% ($10,559 for the average home) in areas near mosques along the east coast of the U.S. on average in the two years following the attacks.

However, on the west coast we find no evidence that 9/11 caused a systematic decrease in housing prices near mosques.

Chapter 4 begins from a conventional model of hedonic equilibrium where a nonmarket amenity is conveyed as an attribute of a differentiated traded good. Different metropolitan areas may have different equilibrium price functions due to geographic variation in consumer preferences, income, and production costs. We demonstrate that under relatively mild restrictions on the geographic extent of taste-based sorting, indicator variables for metro areas define "imperfect instruments" that can be used to identify bounds on demand curves. Bounds on demand curves correspond to ranges of partial equilibrium welfare measures for non-marginal changes in environmental quality. We find these ranges to be informative in a preliminary application to evaluating the benefits of reducing cultural eutrophication of lakes in Maine, New Hampshire, and Vermont.

The last chapter concludes and discusses the insights for future research.

Acknowledgments

I had a wonderful time in Blacksburg during the past five years. There are many memories I can never forget and many people I want to thank. I would like to take this opportunity to thank the people who helped me to make this happen.

Without the remarkable guidance of my advisors, Dr. Kevin J. Boyle, Dr. Nicolai V. Kuminoff, Dr. Jaren C. Pope, my research projects could not be finished. I am grateful to them for their help, patience, concerns and advice. They taught me how to be a good economist. I am also grateful to my committee members, Dr. You Wen, Dr. Darrell Bosch, for their help and valuable comments on my dissertation.

I would like to thank all of my colleagues. In particular, I would like to thank Sapna Kaul, Anthony Murray, Kristen Capogrossi, Shuwen Duan, Aditya Khanal, Xiaoshu Li, Yiran Li, Jeta Rudi, Di "Andy" Zeng, Jianfeng Gao, Ranju Baral for their help and suggestions in my research. I also would like to take this opportunity to thank all of my friends for bringing me lots of joys and happiness in my life.

I am particularly grateful to my husband, Jun Qi, for his love and support in my life. Thank him for sharing every day with me.

Finally I would like to give my deep gratitude to my parents, Lin Zhang and Yong Liu. They give their whole love to me. I dedicate this dissertation to my parents.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Valuing environmental goods is often difficult because improvements or decrements in goods like air and water are usually not exchanged in markets. Through the application of the concept of weak complementarity, one can use indirect methods to estimate values for these changes using observations of people's behavior in markets that are related to environmental quality. This is the basic idea of the hedonic-price method; by observing the price differential between properties that vary in areas with different air or water quality we indirectly observe household's willingness to pay for increments in environmental quality. Thus, this approach is appealing because it uses revealed behavior to infer values for changes in environmental quality to support environmental policy and the estimation results can be used to demonstrate to people that protecting environmental quality is in their own financial interest.

Following Rosen (1974), the hedonic price method helps us to estimate the "implicit price" of property attributes. This is typically accomplished by regressing property prices from arms-length transaction on the attributes of the house, which include the relevant environmental characteristic(s). By taking the partial derivative of the sale price with respect to the attribute under consideration, one can calculate the estimate willingness to pay for marginal changes in property environmental amenities and disamenities.

Despite the wide spread use of hedonic property values there are still a number of challenges to successful application of this method. These limitations may lead to biased estimates of implicit prices because of errors in measurements of environmental values and omitted relevant variables that confound an estimation of implicit prices and identification of the specific environmental effect being measured. In addition, the typical, first-stage hedonic model only measures people's willingness to pay for marginal changes in the environmental quality. However, policy makers are typically interested in nonmarginal changes in quality.

In the following chapters, I first provide three extensions and applications of the hedonic, property-value model. The "traditional" hedonic price method is applied in Chapter 2 to investigate a property-specific environmental disamenity, an aquatic invasive species. In Chapter 3, I apply a quasi-experimental design to investigate the effect that the 9/11 terrorist attacks had on sale prices of properties located near mosques. Chapter 4 provides a new approach to estimating the second-stage demand for water quality using a novel partial-identification strategy.

In Chapter 2, I provide a unique application of the "traditional" hedonic-price method, using property-specific data to investigate if an aquatic invasive species (Eurasian watermilfoil) affects property values on lakefront properties. This unique data may allow a more accurate implicit-price estimate than the common data that only indicates the presence of an environmental amenity or disamenity in the area of a property (Halstead et al., 2003; Horsch and Lewis, 2009).

Two measures of aquatic macrophyte (plant) growth are investigated. First, I investigate if the infestation of Eurasian watermilfoil reduces lakefront property sale prices. Second, I investigate if total aquatic macrophyte growth (including both Eurasian watermilfoil and native plants) reduces lakefront property sale prices. Native plants have been in place and are not likely to spread substantially, while milfoil adds to plant growth and it expands the area of plant growth rapidly. There are several reasons to investigate two measures of aquatic plant growth in lakes. Eurasian watermilfoil looks like some of the native aquatic plant species, and some people may not be able to distinguish between Eurasian watermilfoil and native aquatic plants. In addition, total plant growth, including natives and invasives, can combine to potentially reduce the desirability of lakes for recreation activities and diminish the esthetic appeal of the waters. If people can identify milfoil, then they may be purely concerned with the expansion of areas of lakes colonized by milfoil. This study is unique because the plant-growth variables are specific to each property, measuring the aerial extent of plant growth in a specific are in front of each property. Results indicate that total aquatic macrophyte growth significantly reduces property values; Eurasian milfoil does not solely reduce property values. Eurasian milfoil, however, is the major component of total plant growth and affects property values through total plant growth. As Eurasian watermilfoil infests a lake, adding to the total macrophyte growth, property values can diminish by <1% to 16% for incremental increases in the infestation level. Hence, policies that successfully prevent infestations have significant economic benefits to owners of lakefront properties and local communities by protecting property-tax revenues.

The “traditional” hedonic price method, used above, may lead to biased parameter estimates because of omitted variables that are correlated with the policy variable of interest, which confounds a cross-sectional estimation strategy. For this reason, a number of studies have begun to combine the “traditional” hedonic model with quasi-experiments that have occurred from nature or public policies to identify a causal impact of the change in an amenity or disamenity on property prices. Recent examples that have used the quasi-experimental approach include air quality (Chay and Greestone, 2005), cancer risk (Davis, 2004), school quality (Black, 1999), and water quality (Horsch and Lewis, 2009).

Chapter 3 provides an application of the quasi-experimental hedonic price model to the previously unexplored potential impact of 9/11 on property prices near mosques. In addition, using a national data set I am able to exploit the use of an internal meta-analysis (Banzhaf and Smith, 2007) of our implicit price estimates to confirm the effect on 9/11 on property prices.

Our identification strategy exploits 9/11 as an unanticipated and exogenous event. Specifically, I compare prices of properties that sold within a quarter of a mile of a mosque (the treatment group) and properties that sold between a quarter of a mile and a half mile (the control group) before and after 9/11 to properties located between a half mile and a mile away from the nearest mosque. Results from this generalized differences-in-differences strategy suggest that property prices in the treatment group in areas along the east coast of the United States decreased by approximately 7% ($10,559 for the average home) in the two years following 9/11. Properties in the control group saw no change in their property prices relative to properties between one half and one mile from a mosque. On the west coast I find no evidence that 9/11 caused decreases in property prices near mosques. Furthermore, the price effects I find in the east are largely driven by a relatively small percentage of the mosques. The internal meta-analysis of the results suggests that significant property price impacts were more likely to occur in densely populated areas with higher concentrations of white households. These results suggest where policy makers may want to focus actions to reduce racial/religious tensions after significant societal events.

The “traditional” hedonic price method, even using a quasi-experimental design, measures people’s willingness to pay for a marginal change in environmental quality and is usually referred to as the 1st-stage hedonic analysis. The limitation of this approach is that welfare measurements are limited to marginal changes, while policy makers typically are interested in nonmarginal changes in quality. Therefore, developing an approach to estimate the WTP for non-marginal changes is especially important to policy evaluation.

Estimating people’s non-marginal WTP requires economic researchers to extend hedonic analyses to estimate underlying demands for environmental goods. This process is referred to as the 2nd stage hedonic analysis. Compared to the 1st stage, 2nd stage analysis requires significantly more data and sophisticated statistical analysis to estimate demand for environmental goods and services.

Bartik (1987) and Epple (1987) demonstrate that the nonlinearity of the hedonic price function in the 1st stage leads to an endogeneity problem when 1st stage implicit prices are used in the 2nd stage demand estimation. Over of the past decade, researchers have sought to address this endogeneity problem by adding more information to the model. One strategy is to add assumptions about the shape of utility functions (e.g., Driscoll, Dietz and Alwang, 1994; Chattopadhyay, 1999; Bajari and Benkard, 2005; Bajari and Kahn, 2005; Sieg et al., 2002, 2004; Smith et al., 2004; Klaiber and Phaneuf, 2010). Another strategy is to add more data (e.g., Palmquist, 1984; Parsons, 1986; Bartik, 1987; Cheshire and Sheppard, 1998; Boyle, Poor and Taylor, 1999; Palmquist and Israngkura, 1999; Zabel and Kiel, 2000; and Kuminoff and Pope, 2010). A third strategy is to combine the first two: provide more structure and more data. Bishop and Timmins (2008) track the same individuals over time and space. They also write down a parametric specification for an individual's indirect utility function.

In Chapter 4, I propose a new solution to the endogeneity problem. Instead of providing more structure or data, I propose a fundamentally different approach. The key observation I make is that what we call the “endogeneity problem" in hedonic demand estimation is only a problem if we limit ourselves to the extreme of point identification of demand parameters, in which identification is viewed as an all-or-nothing concept. If we take a broader perspective on identification, we can partially identify demand welfare measures.

Specifically, I propose a partial identification strategy to reveal consistent bounds on demand parameters based on the concept of “imperfect instruments" proposed by Nevo and Rosen (2010). Furthermore, I extend the Nevo and Rosen approach to demonstrate that partially identified demand curves can be used to bound Marshallian welfare measures for non-marginal changes in environmental quality. Marshallian Consumer Surplus can be single bounded if the price parameter is single bounded and can be either double bounded or point identified if the key parameter is double bounded. I present an application to water quality where the welfare effect is single bounded, providing an upper bound on the welfare measure.

The following chapters make contributions to the estimation and use of hedonic price model by: (1) providing an application of the “traditional” hedonic model to demonstrate how property-specific changes in the quality of environmental goods affects property prices; (2) providing an application of the quasi-experimental approach with the “traditional” hedonic model to investigate a neighborhood quality effect on property prices; and (3) proposing a new method to overcome the endogeneity problem in the hedonic demand estimation, which provides bounds on welfare estimates.

Chapter 2

The Effect of an Aquatic Invasive Species (Eurasian Watermilfoil) on Lakefront Property Values

ABSTRACT

Invasive species are one of the major threats to ecosystems. One of these "invaders", Eurasian watermilfoil, can crowd out important native aquatic plants, decrease habitat and diversity of native species in a lake, and interfere with water-based recreation. This study uses a hedonic property value method to estimate the effect of Eurasian watermilfoil on lakefront property values at selected Vermont lakes. Results indicate that as the primary component of total aquatic macrophyte growth in a lake Eurasian watermilfoil significantly and substantially affects lakefront property values. As Eurasian watermilfoil infests a lake, adding to the total macrophyte growth, property values can diminish by <1% to 16% for incremental increases in the infestation level. Hence, policies that successfully prevent infestations have significant economic benefits to owners of lakefront properties and local communities.

KEYWORDS: Aquatic Invasive Species; Eurasian Watermilfoil; Economic Costs; Hedonic Property Values

2.1 Introduction

"Invasive species" are non-indigenous animals or plants that adversely affect the ecology of native habitats and can have adverse impacts on economic welfare (e.g., Halstead et al., 2003; Holmes et al., 2006; Horsch and Lewis, 2009; Kaiser and Burnett, 2006). Wilcove et al. (1998) argued the invasive species are second only to habitat loss as the greatest threats to biological diversity. Invasive species damage the lands and waters that native plants and animals need to survive. The estimated costs of invasive species worldwide total more than $1.4 trillion – five percent of the global economy (Pimentel et al., 2001). In the U.S. alone, estimated economic damages (welfare and production losses) and control costs associated with invasive species amount to approximately $120 billion annually (Pimentel et al., 2005).

The economic costs of invasive species estimated in Pimentel's work are best estimates based on available data and appear to focus on the direct costs through production losses in agriculture, forestry and other segments of the U.S. economy, and costs to manage invasive species. Welfare losses are more difficult to estimate due to their "non-market" nature (Lovell et al., 2006). If a full accounting of welfare losses were available the economic costs of invasive species would be much larger than the figure reported by Pimentel (2005).

One species of aquatic invasive plant, Eurasian watermilfoil (*Myriophyllum Spicatum*), was introduced into North America in the mid 1940s and has spread to at least

45 states.[1] The primary means of transport between lakes is on boats, boat trailers, water skis, scuba gear and waterfowl.[2] Throughout Vermont, in particular, Eurasian watermilfoil infests about 60 lakes and several rivers, including the Connecticut River (Figure 2.1). Eurasian watermilfoil is highly invasive and competes aggressively with native aquatic plant species, thereby reducing biodiversity. Dense milfoil infestations can severely impair human uses such as swimming, boating, and fishing. Water quality and fish abundance and distribution can also be affected when the plants grow into dense mats on the water surface.[3]

This study investigates if Eurasian watermilfoil (milfoil hereafter) affects property values on selected lakes in Vermont, and two measures of aquatic macrophyte (plant) growth are investigated.[4] First, we investigate if the infestation of milfoil reduces lakefront property sale prices. Second, we investigate if total aquatic macrophyte growth (including both milfoil and native plants) reduces lakefront property sale prices. There are several reasons to investigate two measures of aquatic plant growth in lakes. Milfoil looks like some of the native aquatic plant species.[5] This means that some people may not be able to distinguish between milfoil and these native aquatic plants. In addition, total plant growth, including natives and invasives, can combine to potentially reduce the

---

[1] http://www.iisgcp.org/exoticsp/watermilfoil.htm, last accessed on May 31, 2010.
[2] http://www.miseagrant.umich.edu/downloads/ais/fs_EWM-milfoil.pdf, last accessed on May 31, 2010.
[3] http://www.vtfishandwildlife.com/library/factsheets/NonGame_and_Natural_Heritage/Invasive_Exotic_Plant_FactSheet.pdf, last accessed on April 20, 2010.
[4] "Macrophytes are aquatic plants, growing in or near water that are either emergent, submergent, or floating. Macrophytes are beneficial to lakes because they provide cover for fish and substrate for aquatic invertebrates. … However, an overabundance of macrophytes can result from high nutrient levels and may interfere with lake processing, recreational activities (e.g., swimming, fishing, and boating), and detract from the aesthetic appeal of the system" (http://www.epa.gov/bioiweb1/html/macrophytes.html, last accessed on February 5, 2010).
[5] http://www.seagrant.umn.edu/exotics/eurasian.html, last accessed on May 31, 2010.

desirability of lakes for recreation activities and diminish the esthetic appeal of the waters.

2. 2 Application

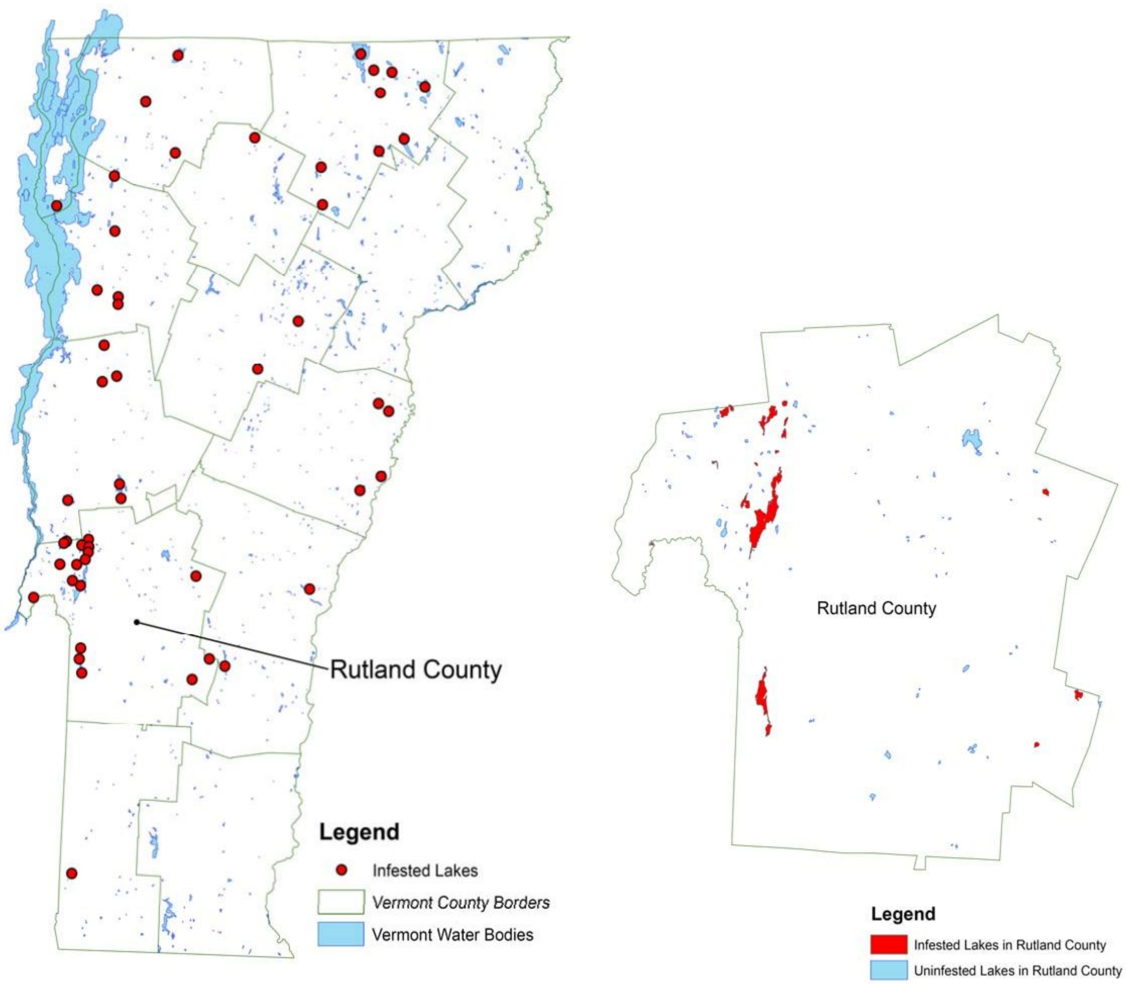
The presence of milfoil often brings changes in the natural lake environment (Madsen et al., 1991; Smith and Barko, 1990). Over time, milfoil may out compete or eliminate more beneficial native aquatic plants, reducing natural plant diversity within a lake. Commonly found in shallow bays and along the shoreline, milfoil can grow quickly from the lake bottom to the surface, forming very dense mats of vegetation on the surface of the water. These mats interfere with recreational activities such as swimming, fishing, water skiing, and boating (Eiswerth et al., 2000; Eiswerth et al., 2005). The dense mats on the water surface may all reduce the esthetic qualities of lakes.

Barko and Smith (1990) state that eradication of milfoil is rarely ever successful because of the ability of this plant to reproduce from small fragments. Thus, from policy and management perspectives, a high priority should be placed on protecting lakes without milfoil from infestation and, if a lake is infested, continuous control efforts are required. The persistence of milfoil suggests that lakes that become infested may demonstrate reduced lakefront property values because of the perceived permanent reduction in the quality of lakes for recreation activities and the diminished esthetics of the lake surfaces.

2.2.1 Study Area

Milfoil currently infests a number of Vermont lakes, including the state's largest lakes, Champlain, Memphremagog, and Bomoseen (Figure 2.1). Local populations of milfoil in Vermont were first documented in Lake Champlain in 1962 and it has since spread to about 60 lakes and several rivers throughout Vermont.[6]

Figure 2.1 Eurasian Watermilfoil Distribution in Vermont, 2004

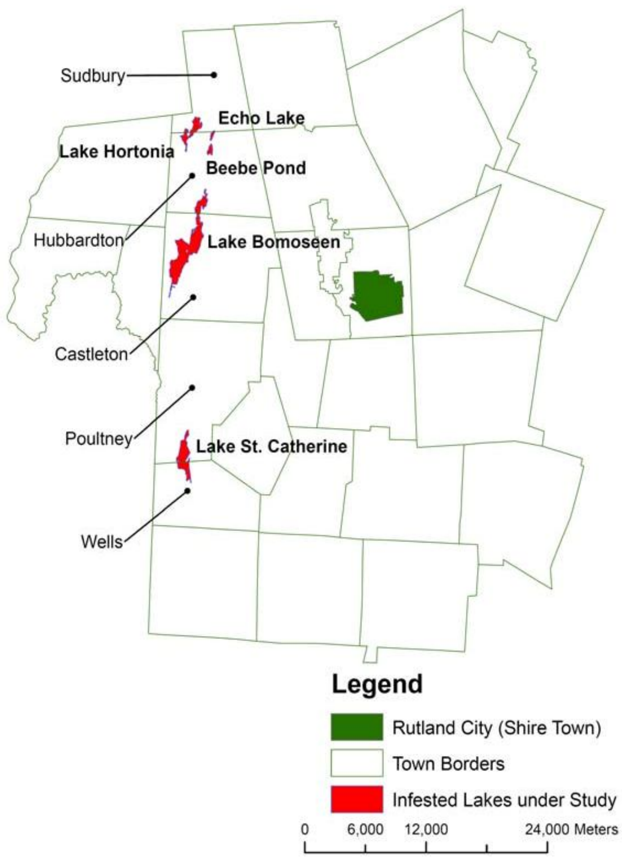


The region for this study is four lakes and a pond in Rutland County, Vermont (Figure

[6] http://www.lcbp.org/nuissum.htm, last accessed on May 31, 2010.

2.2). These waters were selected because they have established milfoil infestations and the Vermont Department of Environmental Conservation could provide data on the extent of milfoil coverage on the lake surface in front of each sold property.[7] The milfoil infestations in these waters, which occurred as early as 1982, are sufficient that each water has experienced multiple types of control actions (Table 2.1).

Figure 2.2 Selected Lakes with Eurasian Watermilfoi Infestation in Rutland County



These unique, property-specific data on milfoil infestations provide the opportunity to

[7] The four lakes and the pond in Rutland County, Vermont are the only waters where the Vermont Department of Environmental Conservation was able to provide property-specific data on milfoil and total aquatic plant growth. As shown in Figure 2. 1, many other Vermont lakes and ponds also have identified infestation of milfoil, but property-specific infestation data are not available for these other waters.

estimate a hedonic, property-value model to better understand the economic costs of this invasive species. The model can be used to estimate the marginal benefits of preventing infestations or reducing the extent of infestations, which can be used to help justify the management costs of preventing milfoil infestations or reducing existing milfoil infestations.

Table 2.1 Rutland County, Vermont Water Bodies Investigated[a]

Waters	Location	Year Eurasian Watermilfoil Found	Eurasian Watermilfoil Management Actions[b]	Size (acres)
Beebe Pond	Hubbardton	1991	BB, HB, HP, SH	111
Lake Bomoseen	Castleton, Hubbardton	1982	BB, DD, H, HP, HR, W	2,360
Echo Lake	Hubbardton, Sudbury	1989	BB, HP	54
Lake Hortonia	Hubbardton, Sudbury	1984	BB, DD, H, HB, HP	479
Lake St. Catherine	Wells, Poultney	1983	BB, H, HB, HP, HR, SH	904

[a] The information in this table is taken from VTDEC, Water Quality Division's website; it was last updated at September, 2009 (http://www.anr.state.vt.us/dec/waterq/lakes/docs/ans/lp_aismapmajorspecies2009.pdf, last accessed on June 30, 2010).

[b] Key to abbreviations: BB--bottom barrier, DD--drawdown, H--mechanical harvesting, HB--aquatic herbicide HP--hand pulling, HR--hydro-raking, SH--diver operated suction harvesting, W--weevil introduction or augmentation

2.2.2 Management Methods

The spread of milfoil is largely due to human uses (e.g., transporting recreational boats from one lake to another) that are difficult to monitor and efforts to control the spread of this invasive plant are largely dependent on public education efforts and voluntary cooperation of lake users.[8] These efforts have included educational brochures distributed when people purchase fishing or boat licenses, signs at public access points to lakes, and greeters at public access points that educate people how to check their boats

[8] http://www.seagrant.umn.edu/exotics/eurasian.html; http://www.anr.state.vt.us/dec/waterq/lakes/htm/ans/lp_ans-index.htm#help; http://dnr.wi.gov/invasives/publications/pdfs/EWMbrochure.pdf; last accessed on May 31, 2010.

and trailers before and after they enter the lake. Because these voluntary compliance programs have met with limited success, milfoil has continued to spread to new lakes and states are left with attempts to control milfoil infestations. For example, the number of Vermont lakes infested with milfoil grew from a very small number in 1970 to about 65 in 2008 (Vermont Agency of Natural Resources, 2010).

Physical, mechanical, chemical and biological control methods (Table 2.2) are used in attempts to manage infestations of milfoil (e.g., Boylen et al., 1996; Unmuth et al., 1999; Wagner et al., 2007; Sheldon and Creed, 1995) and all four methods have been used on at least one of the five waters included in the present study (Table 2.1). Some methods are more appropriate for well-established populations, while others are better suited for those that are recent introductions. Pulling milfoil plants by hand, when done properly, can be somewhat effective for controlling newly introduced populations. However, this solution is tedious and it is virtually impossible to remove all of the plants in this manner. Machine cutting improves the lake for human uses, but does not remove milfoil colonies. Chemical control has been effective in temporary reductions in milfoil, but chemical applications are expensive and can be harmful to native aquatic plants. Natural predators, biological controls, of milfoil can also be introduced and the Vermont Department of Environmental Conservation has been working with the watermilfoil weevil (*Euhrychiopsis lecontei*) since 1989.

Table 2.2 Eurasian Watermilfoil Control Methods Comparison

Class	Method	Advantages	Disadvantages	Costs
Physical	Bottom Barriers	Effective at treating very dense beds; control growth in localized areas	Eliminates some non-target species; may interrupt spawning of some warm-water fish; may eliminate some benthic invertebrates	$10,000-$20,000 per acre for professional installation
	Suction Harvesting	Removes only target plants; More effective in medium density beds	Labor intensive; Added equipment costs; some difficulty with very dense beds	$20,000-$30,000 for equipment and $1,000-$25,000 per acre for operations and disposal of harvested plants
	Hand Harvesting	Removes only target plants; Low equipment costs	Very labor intensive; Harvesting dense beds is inefficient	$400-$1,000 per acre
	Drawdown	Can be very effective for smaller water bodies with control structures	Negatively impact the ecosystem and recreational use of the lake	N/A
Mechanical	Rotovating	Both stem and roots are removed	Severe disturbance to sediments can lead to recolonization by invasive species; fragmentation of EWM can lead to colonization of new areas	$100,000-$200,000 for equipment and $200-$300 per acre for operations; or $1,500 per acre to hire professional service
	Mechanical Harvesting	Provide habitat for fish; leaves benthic community intact	May have to be repeated more than once a year; fragmentation of EWM can lead to colonization of new areas	$100,000-$200,000 for equipment and $200-$300 per acre for operations
Biological	Herbivorous Insects	Milfoil weevil the aquatic moth target only EWM and are native species; slow reduction in plant biomass; minimizes chance of increased eutrophication	Slow method; results from introduction are inconsistent	Stocking costs approximately $1,000 per acre
	Grass Carp	Very little labor involved; very effective at removing vegetation given time	Removal of non-target species; grass carp prefer moving water and are very likely to migrate from the lake; highly regulated	Stocking costs $50-$100 per acre
Chemical	Aquatic Herbicides	Effective on EWM; can provide short and long term control	Removal of non-target species; decomposing vegetation can reduce dissolved oxygen and cause algal blooms; use restrictions may be place on the lake after application	$200-$400 per acre

Source: All Information is replicated from "Black Lake Eurasian Watermilfoil Management Plan" prepared by Quantitative Environmental Analysis, LLC , Liverpool, NY.
http://www.weedinfo.blacklakeny.com/FINAL_Black_Lake_milfoil_plan_07_14_08-1.pdf, last accessed on June 15, 2010.

The effectiveness of each method depends on a suite of factors including the extent of the infestation, availability of funding, volunteer time and effort, follow-up efforts, and physical/environmental conditions in each lake. However, there is no way to completely eradicate milfoil from a lake once it has been introduced. Therefore, control efforts focus on controlling new infestations, preventing further spread of milfoil in established infestations, and reducing the nuisance level of well-established infestations.

2. 3 Previous Research

A number of researchers have examined the relationship between water quality measures and housing prices using hedonic models of property values (e.g., Mendelsohn et al., 1992; Michael et al., 2000; Leggett and Bockstael, 2000; and Poor et al., 2007). Only four hedonic studies have investigated an invasive species, but two investigated watermilfoil.

Holmes et al. (2006) examined the impact of the hemlock woolly adelgid (*Adelges tsuga*), an exotic forest pest, on the value of residential properties in Sparta, New Jersey. Land areas were classified according to four different categories of hemlock tree conditions: (1) lightly defoliated (<25%), (2) moderately defoliated (25-50%), (3) severely defoliated (50-75%), and (4) dead (>75%). The percentage land coverage in each of these four classes were included in the hedonic model as independent variables for buffers of 0.1, 0.5 and 1 kilometers around sold properties. The lightly defoliated variable was significant and positive in all model specifications, indicating that the presence of hemlocks enhances property values. The moderately defoliated variable was significant and negative in all equations, indicating that defoliation of hemlocks by the invasive species diminishes property values. Moderate defoliation on subject properties reduced property values by 1%, and moderate defoliation in the 0.1, 0.5 and 1 kilometer buffers reduced property values by 1.7%, 3.0% and 4.8%, respectively.

Kaiser and Burnett (2006) investigated reductions in property values due to the infestation by the coqui (*Eleutherodactylus coqui*), a species of small, noisy tree frogs in Hawaii. Two indicator variables were included to measure the presence of coqui; whether a property is within 500m of a previous complaint and whether a property is between 500-800m of a previous complaint. The results showed that a noise complaint within 500m reduces property values 0.16% and a complaint between 500m and 800m reduces property values by an additional 0.12%.

Halstead et al. (2003) analyzed the effects of variable milfoil (*Myriophyllum heterophyllum*) on shoreline property values of selected New Hampshire lakes. Two milfoil variables were included in the hedonic equation: (1) a dummy variable indicating whether milfoil was present in the lake at the time of house purchase and (2) an interaction term between the size of the lake and the presence of milfoil. The interaction term is included because the presence of milfoil concentrated somewhere in a large lake may have less of an effect on properties than in a smaller lake. The results indicate that the presence of milfoil in a water body has a substantial deleterious effect on shoreline property values, with reduction of 21% (linear dependent variable) to 43% (natural log of dependent variable). The authors note that a 40% reduction in property values is rather steep. There is every reason to question whether the 21% and 43% price diminutions are accurate. A binary variable (presence of milfoil) captures all differences between lakes that are not represented by the other explanatory variables in the model. These other effects could lead to over- or underestimation of the average effect on individual properties. If a property is not in a milfoil area, then the average effect overestimates the effect. Conversely, if a property has an extensive milfoil infestation in the lake immediately in front of the property, then underestimation may be present. In addition, if lakes with milfoil infestations are less desirable than lakes without milfoil for reasons in addition to the presence of milfoil, then a binary variable indicating the presence of milfoil will overestimate the average price diminution due solely to milfoil. While Halstead et al. did not identify how many and which New Hampshire lakes had milfoil infestations during their study period (1990-95), available data indicate that only two of the 10 lakes in their data had variable milfoil infestations.[9] Thus, it is difficult to interpret the estimates presented in the Halstead et al. study.

Horsch and Lewis (2009) investigated the effect of Eurasian watermilfoil (*Myriophyllum Spicatum*) on property values over 170 lakes in Vilas County, Wisconsin where 20% of the lakes had milfoil infestations. These researchers use an identification strategy based on a spatial difference-in-differences specification, instead of a

[9] http://des.nh.gov/organization/divisions/water/wmb/exoticspecies/documents/milfoil_map_list.pdf, last accessed on May 31, 2010.

conventional cross-sectional hedonic model, to investigate how a milfoil infestation affects property values. The difference-in-differences method accounts for both bias and inefficiency problems associated unobserved neighborhood effects that may be spatially correlated with milfoil infestations. The key milfoil variable is whether a property was or was not purchased before the occurrence of a milfoil infestation. Results indicate that a milfoil infestation reduces average property values by approximately 8% and reduces average land values, net of the value of any structures on the property, by approximately 13%. Horsch and Lewis avoid the Halstead et al. problem of simply observing whether milfoil is or is not present in a lake. They retain the Halstead et al. problem that their study says nothing about the effect of the level of milfoil on individual properties. Again, the average price diminution form purchasing property after an infestation likely overstates the price effect for a property with no milfoil in the water in front of the property and likely understates the effect on a property with a heavy infestation in the water immediately in front of the property. In addition, while the difference-in-differences approach controls for neighborhood effects, it does not capture other lake characteristics that might be changing concurrent with the infestation and the binary variable (sold before or after the infestation) could have an omitted variable bias that could lead to over- or underestimation.

These studies collectively indicate that invasive species reduce property values, and that the presence of milfoil in a lake can result in a substantial reduction in property values. However, neither the Halstead nor the Horsch studies had data on the coverage of milfoil infestations that are specific to individual properties. The unique contribution of the research reported here is that the Vermont Department of Environmental Conservation staff was able to provide data on a qualitative scale that indicates the percent the lake surface covered by milfoil immediately in front of each lakefront property for a small group of four lakes and one pond in Rutland County, Vermont. The scale ranges from 1 (less than 1% coverage) to 6 (80-100% coverage). Thus, the coverage of milfoil variable in the hedonic equation is unique to each individual property sale, and these data are better able to capture the effect of milfoil on individual property values than the crude presence or absence or before and after measures used in the Halstead and Horsch studies.

2.4 Hedonic Model

In a 1966 paper Lancaster developed what he called a "new theory of consumer demand" where consumers derive utility from characteristics of a good. One such characteristic of a lakefront property would be the extent of milfoil growth in lake water in front of the property. Rosen (1974) demonstrated that the hedonic price function is simply an envelope of equilibrium of transactions between buyers and sellers of a good that is differentiated by its characteristics. Under the assumptions of this model, the marginal values (implicit prices) consumers place on individual characteristics can be recovered by regressing sale prices on the characteristics of the good.

The hedonic price model is widely applied to study the housing market transactions because of the number and variation of housing characteristics. A house is seen as a bundle of characteristics and the sale price is a function of these characteristics. The value that a characteristic adds to the price of a house can be thought of as an implicit price for that characteristic. Such implicit price estimation is particularly helpful for characteristics that are not priced independently in markets, e.g., proximity to environmental amenities and disamenities.

2.4.1 Hedonic Model

To formalize the basic idea of the hedonic model, let the sale price of a lakefront property, *SP*, be expressed as a function of the property's attributes:

$$SP = f(S, P, L, Q) \qquad (1)$$

where *SP* is a vector of property sale prices, *S* represents structural characteristics, *P* represents lot characteristics, *L* represents location characteristics, and *Q* represents environmental characteristics. In previous hedonic studies, examples of structural characteristics commonly included are square feet of living area, type of heating system, number of bedrooms and number of bathrooms. Lot characteristics of lake applications might include feet frontage on a lake and lot size. Location characteristics describe the area surrounding the property such as distance to the nearest large town or business district, property tax rates and neighborhood demographic characteristics. Environmental characteristics include environmental amenities (disamenities) that would contribute to (depreciate) the value of the property, e.g., water quality. Selection of independent variables, property characteristics, is based on knowledge from previous studies, intuition about the specific application, and data availability.

Hedonic theory does not guide the functional form for equation (1). The only restriction is that the first derivative for the environmental attribute of concern be positive if it is an amenity and negative if it is a disamenity (Freeman, 2003). Therefore, it is necessary that the functional form for the hedonic price function be determined empirically (Cropper et al., 1988; Palmquist, 2003; Taylor, 2003). Following the majority of previous hedonic studies of water quality, and the two aquatic invasive species studies cited above, this study utilizes a nonlinear specification where the dependent variable is the natural log of sale prices. Thus, base specification of the hedonic model is:

$$
\begin{aligned}
ln(SP) = {} & \beta_0+\beta_1 UNIMP+\beta_2 ln(LVAREA)+\beta_3 HEAT+\beta_4 FULLBATH \\
& +\beta_5 LAKEWATER+\beta_6 LOT+\beta_7 FF+\beta_8 DIST+\beta_9 INTWC \\
& +h\ (MC)+g(lakes)+u
\end{aligned}
\qquad (2)
$$

where the *βs* are parameters to be estimated and the independent variables are defined in Table 2.3. The independent variables in the hedonic model are additive with nonlinear specifications of living area (*LVAREA*), water clarity (*INTWC* – an indicator of eutrophication that is visually observable to property owners) and macrophyte coverage (*MC*). The natural log of the total square feet of living area is used because of a presumed nonlinear relationship between property price and house size. Marginal increases in living space will provide less utility to residents as initial total living area increases. Water clarity (*WATERCLARITY*), a component of *INTWC*, is logged because it is difficult for people to see changes in water clarity at deeper levels of clarity (Smeltzer and Heiskary, 1990). *ln*(*WATERCLARITY*) is then interacted with lake area based on the results of previous hedonic studies of water clarity (Gibbs et al., 2002; Michael et al., 2000).[10]

The function *h*(*MC*) is modeled using two aquatic plant variables, either *EWM* (Eurasian watermilfoil coverage rating) or *TOTAL* (total aquatic macrophyte coverage rating). Intuitively, one would expect that an increase in aquatic plant coverage would lead to a decrease in property prices. As aquatic plant coverage increases, each succeeding increment of increase may have a larger detrimental effect on property values. If the aquatic plant coverage increases from 10 to 30 percent of an area, this change may be noticeable to people, but they still have plenty of area for aquatic recreation activities. If the aquatic plant coverage increases from 70 to 90 percent, then it may be almost impossible for the lake area to support aquatic recreation. Thus, two nonlinear specifications of the aquatic plant coverage variable are considered in the hedonic price function, a quadratic form and an exponential form, which both allow the marginal price to increase in absolute value as aquatic plant coverage increases. These specifications are:

$$h_{QUA}(MC) = \beta_{10}MC + \beta_{11}MC^2 \qquad (3a)$$

and

$$h_{EXP}(MC) = \beta_{10}MC + \beta_{11}exp(MC) \qquad (3b)$$

where *MC* is the percent coverage rating of macrophytes (either milfoil – *EWM* or total aquatic plant growth – *TOTAL*).

[10] *INTWC* = *LAKEAREA* · *ln*(*WATERCLARITY*). *LAKEAREA* and *WATERCLARITY* are defined in Table 2. 3.

Table 2.3 Names and Descriptions of Variables Used in Hedonic Model (N=65)

Variable Name	Description	Mean	S.E.	Min	Max
SP	Actual sale price of property (1995 dollars)	108660.60	57179.46	18000	270000
UNIMP	0,1= unimproved land	0.11	0.31	0	1
LVAREA	Total living area (square feet)	886.42	480.79	0	1920
HEAT	0,1 = central heating system	0.78	0.41	0	1
FULLBATH	0,1 = presence of a full bathroom	0.88	0.33	0	1
LAKEWATER	0,1 = primary source of drinking water is from lake	0.48	0.50	0	1
LOT	Lot size (acres)	0.66	1.55	0.08	11.91
FF	Total lot frontage on lake (feet)	104.90	68.59	15	410
DIST	Distance to the nearest business district (mile)	18.93	4.04	15.80	28.90
INTWC	ln(water clarity)* surface area of lake (acres)	3012.31	1714.55	105.85	4723.49
LAKEAREA	Surface area of lake (acres)	1619.43	878.65	54	2360
WATERCLARITY	Water clarity (meters)	6.24	0.89	3.8	7.4
EWM	Eurasian watermilfoil percent cover rating	4.09	1.30	1	6
TOTAL	Total aquatic macrophyte percent cover rating	4.78	1.26	1	6
Beebe Pond	Fixed effect dummy for Beebe Pond	0.05	0.21	0	1
Lake Bomoseen	Fixed effect dummy for Lake Bomoseen (base group)	0.57	0.50	0	1
Echo Lake	Fixed effect dummy for Echo Lake	0.02	0.12	0	1
Lake Hortonia	Fixed effect dummy for Lake Hortonia	0.14	0.35	0	1
Lake St. Catherine	Fixed effect dummy for Lake St. Catherine	0.23	0.42	0	1

A common problem with estimated hedonic models is endogeneity. For example, milfoil infestation may be correlated with some unobserved characteristics not accounted for in the explanatory variables and, therefore, failure to include lake-specific effects could lead to bias in estimated coefficients (βs). Horsch and Lewis (2009) argued that milfoil is spread from lake to lake by the movement of boaters and anglers, who are more likely to visit popular lakes with desirable amenities that are usually unobservable to researchers. This potential concern is addressed in the current estimation by including lake fixed-effect variables [$g(lakes)$] that account for lake-specific characteristics that are not represented by the explanatory variables yet may be correlated with the level of milfoil infestations.[11] Lake Bomoseen is the omitted lake in the estimated equations.

2.4.2 Selection of Functional Specification

Four base equations are estimated, quadratic and exponential specifications of milfoil (*EWM*) and total plant coverage (*TOTAL*) variables. This allows investigations of whether milfoil and total plant coverage have differential effects on sale prices. A J-test is used to investigate which functional specification, quadratic or exponential, fits the data best (Davidson and MacKinnon, 1981). This test is applied to choose between two non-nested models. Let:

$$ln(SP) = f(X_{QUA}) + \alpha ln(\widehat{SP}_{EXP}) + \mu \qquad (4a)$$

[11] Some hedonic studies corrected for identification problem caused by the endogeneity of environmental variable of interest (e.g., Poudyal et al., 2009; Irwin and Bockstael, 2001) by adopting IV regression.

$$ln(SP) = f(X_{EXP}) + \delta ln(\widehat{SP}_{QUA}) + v \tag{4b}$$

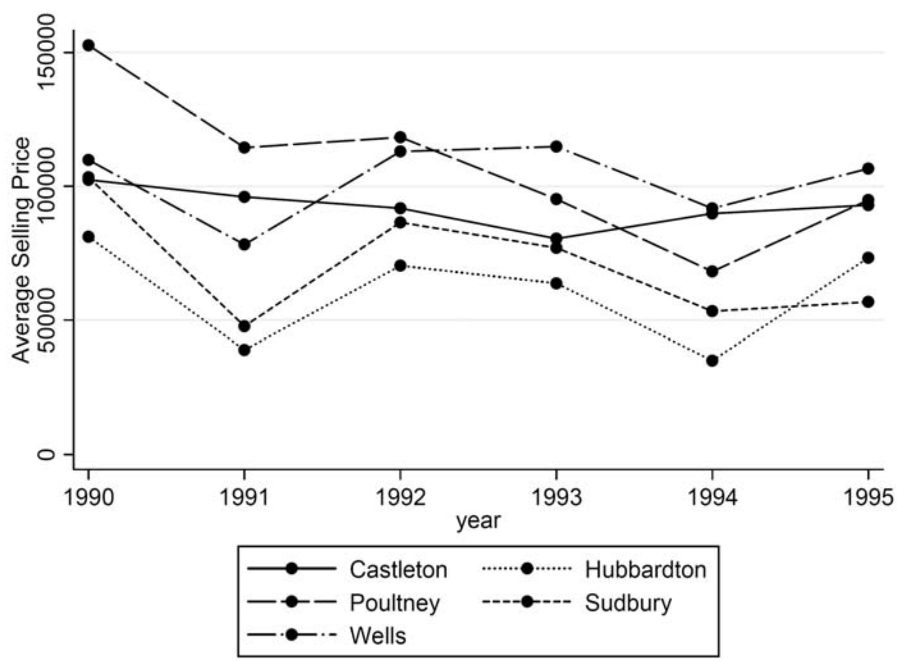
where $f(X_{QUA})$ and $f(X_{EXP})$ are the hedonic equations with the plant coverage variables specified as quadratic and exponential terms, respectively (e.g., equations 3a and 3b). The second terms in each equation, $ln(\widehat{SP}_{EXP})$ and $ln(\widehat{SP}_{QUA})$, are predicted sale prices using the exponential and quadratic specifications, respectively. If α is significant and δ is insignificant, this is evidence that the exponential specification fits the data best. The converse pattern of results would suggest that the squared specification fits the data best. If both α and δ are insignificant, then the test is indeterminate.

2.4.3 One Real Estate Market

An important assumption of hedonic theory is that all the property sales used to estimate a hedonic regression must occur within the same housing market. This is because a hedonic-price function represents an equilibrium envelope of sale points that are arms-length transactions between willing buyers and sellers in a specific market. Markets are deemed as being separated, for example, if consumers in one market do not consider houses in the other market when making their purchase decisions.

There is no uniform theory to distinguish between housing markets and market segmentation usually rests on empirical observation and local market knowledge. Local data strongly supports the assumption that the four lakes and the pond are in the same market. As shown in Figure 2.2, all waters are located within close proximity to each other in the same county and all lakefront properties are within 18 to 30 miles of the only major business district in the county and that region of Vermont, Rutland City. The waters are located in five adjacent towns and the longest distance between any two waters is about 22 miles (Lake Hortonia to Lake St. Catherine).

Figure 2.3 Price Test for One Market Assumption



Another approach, borrowed from the industrial organization literature, uses the "price test" comparison for market segmentation (Stigler and Sherwin, 1985). If prices demonstrate closely parallel movements, then this suggest the loci of prices are in the same market. If significant nonparallel price movements are observed, then the loci of the prices are not in the same market unless the discordance in movements can be traced to differences such as commuting costs to work and shopping. Figure 2.3 shows the trend of mean selling price of vacation houses located on less than 6 acres land from 1990 to 1995 in the five towns where the four lakes and the pond are located. The average prices of these properties increased and decreased together for Hubbardton, Poultney, Sudbury and Wells with one exception; the average price for Wells increased slightly from 1992 to 1993 while the other three averages decreased. This is evidence of a common price trend. The exception is Castleton whose average prices remained relatively stable and were bounded by the average prices from the other towns. Castleton is right in the middle of the five towns and has the most direct access to Rutland city so the price stability might be explained by reduced commuting costs to work and shopping.

Given the physical proximity of the five waters, their adjacency to the major business district in the area and the similarity of price trends we feel that it is reasonable to assume that transactions occurred within a common real estate market.

2. 5 Data

This study uses lakefront property sales from four lakes and one pond in Rutland County, Vermont. These waters were selected because that the Vermont Department of Environmental Conservation could provide data on aquatic plant coverage in front of each sold property.

Only single family residential or vacation homes and unimproved land were used in this study. Information on property sales and sale prices were collected from transfer tax records held in town offices (Table 2.3). Property sales data were collected for all lakefront properties on the selected lakes (pond) that sold during the period January 1, 1990 through December 31, 1995. The sale prices are converted to 1995 dollars. This resulted in 65 usable observations. Property tax records provided data on structural characteristics of any residences on the property and lot characteristics. Seven of the 65 observations were sales of undeveloped lots.

Data on water clarity, lake area, and aquatic macrophyte coverage were provided by the Vermont Department of Environmental Conservation. Water clarity is measured using a secchi disk that is 8 inches in diameter and is alternatively black and white in each quadrant. The disk is lowered into the lake water and the depth at which the disk disappears from sight is the measure of water clarity. The minimum water clarity during the summer months, the period of lowest water quality due to eutrophication, is used as the measure of water clarity.

Aquatic macrophyte growth is measured using a percent coverage rating. The percentage of the water surface covered by aquatic macrophytes is computed for the water surface area in front of each sold property. That is, for a fixed water surface area immediately in front of each shoreline property the percent coverage is computed as the surface area covered by the macrophyte growth divided by the total surface area under consideration. This was done for milfoil and for total plant growth. The Vermont Department of Environmental Conservation assigned categorical ratings to these percent coverages. Each number corresponds to the percent coverage of macrophytes ranging from 1 (less than 1% coverage) to 6 (81 to 100% coverage), e.g., 2 → 1-20%, 3 → 21-40%, 4 → 41-60%, and 5 → 61-80%. Table 2.4 summarizes the milfoil and total aquatic plant coverage data used to estimate the hedonic equations. These data indicate that milfoil as a percentage of total aquatic plant growth ranges from 14% (Beebe Pond) to 100% (Lake Hortonia), and the average across all five waters is 71%.[12] Among the 65 observations, 44 have milfoil ratings equal to the total aquatic plant ratings, indicating that for about two thirds of the properties milfoil is the primary aquatic plant growing in the water immediately in front of the property.

Table 2.4 Lake Macrophyte Percent Cover Rating and Water Clarity

		Milfoil (Percent Cover Rating)			Total Aquatic Plant (Percent Cover Rating)			Water Clarity (Meters)		
	N	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Beebe Pond	3	2.0	2	2	5.0	4	6	7.1	6.7	7.4
Lake Bomoseen	37	4.1	2	6	4.8	2	6	6.6	5.6	7.4
Echo Lake	1	3.0	3	3	6.0	6	6	7.1	7.1	7.1
Lake Hortonia	9	5.8	4	6	5.8	4	6	4.7	3.8	4.9
Lake St. Catherine	15	3.7	1	5	3.9	1	6	6.0	5.3	6.4
Total	65	4.1	1	6	4.8	1	6	6.2	3.8	7.4

2.5.1 Estimation Robustness

These data provide a unique opportunity to examine the effect of an aquatic invasive species on property values because the invasion data are specific to individual properties. This strength is tempered by the limitation of the small number of observations, $n = 65$. To investigate the robustness of the estimation results two supplementary analyses are conducted.

Atkinson and Crocker (1987) found that including a large number of characteristics (explanatory variables) in hedonic price equations can result in unreliable parameter

[12] Percentages are calculated using the midpoint of the percentages for each integer on the rating scale. For example, if the percentage coverage rating is 2, then we view the percent coverage as 10%, the midpoint of 1% to 20%.

estimates, which is more likely to be problematic for a study with a small sample size. A small sample size, such as 65, can easily result a low degree of freedom and a high mean square error. Including correlated independent variables can increase the possibility of multicolinearity, leading to inflated standard errors. Some of the correlations reported in Table 2.5 are larger than 0.6. On the other hand, omitting relevant variables from a hedonic equation, especially those potentially correlated with the variables of interest, can lead to omitted variable biases.

In order to avoid omitted variable biases while reducing the number of explanatory variables, principal-component-analysis (Greene, 1994) and all-possible-regressions procedures (Neter, 1996) are used to investigate the effect of reducing the number of explanatory variables on the estimation results for the coefficient estimates on *EWM* and *TOTAL* variables.

2. 5.1.1 Principal Component Analysis (PCA)

Principal component analysis uses a small number of indicator variables (*L* principal components) constructed from the *K* original independent variables ($L<K$) as new regressors. These *L* principal components are linear combinations of the *K* original variables, and they reduce the number of regressors (increasing degrees of freedom) and reduce colinearity between independent variables.

Table 2.5 Correlation Matrix

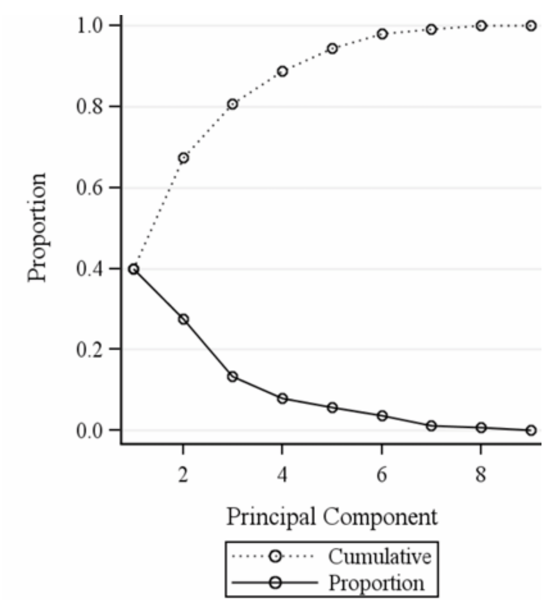
	EWM	*TOTAL*	*UNIMP*	$ln(LVAREA)$	*HEAT*	*FULLBATH*	*R*	*LOT*	*FF*	*DIST*	*INTWC*
EWM	1										
TOTAL	0.63	1									
UNIMP	-0.33	-0.02	1								
$ln(LVAREA)$	0.32	-0.01	-0.99	1							
HEAT	0.33	0.18	-0.66	0.69	1						
FULLBATH	0.28	0.01	-0.93	0.93	0.60	1					
LAKEWATER	0.20	0.04	-0.33	0.29	0.05	0.26	1				
LOT	0.21	0.14	-0.05	0.07	0.10	0.06	0.03	1			
FF	0.12	0.01	-0.03	0.01	-0.04	0.05	-0.03	0.43	1		
DIST	0.13	-0.01	-0.03	0.01	-0.37	0.07	0.24	0.27	0.43	1	
INTWC	-0.04	-0.03	-0.17	0.17	0.41	0.12	-0.23	-0.27	-0.37	-0.87	1

The problem with PCA is that it is unclear how to interpret the coefficient on the L principal component variables. To avoid this problem, the PCA technique is applied to all property characteristics variables in the equation (2) except the *EWM* and *TOTAL* variables and the lake specific variables. Four principal components, which are the linear combination of the 9 omitted characteristic variables, are retained and used in the new hedonic price function:

$$ln(SP) = \beta_0 + \beta_1 PC_1 + \ldots + \beta_4 PC_4 + h(MC) + g(lakes) + u \qquad (5)$$

where PC_i (i =1,2,3,4) denotes the ith principle component. L is set to four principle components because these are sufficient to account for nearly 90% percent of total variance in the original 9 independent variables (Figure 2.4).

Figure 2.4 Variance Explained by the Principal Components



2. 5.1.2 All-Possible-Regressions (APR) Procedure

The all-possible-regressions is a systematic procedure to reduce the number of independent variables to a parsimonious subset. Based on the assumption that the functional specification of equation (2) is correct, the all-possible-regressions considers all possible subsets of the pool of independent variables to identify "good" subsets according to a selected criterion. Given a dependent variable Y and a set of potential independent variables, X (X_1, X_2,…, X_p), the problem is to find and fit the "best" model of the form $Y = \beta^* X^* + u^*$, where X^* is a subset of X. A variety of selection criteria are available to select among the 2^p possible submodels, including MSE, PRESS, and C (Neter, 1996). The MSE (mean square error) criterion seeks the subset of explanatory variables such that MSE is at the minimum or so close to minimum that adding more variables is not worthwhile. The PRESS (prediction sum of squares) criterion is a measure of how well the fitted subset model can predict the observed dependent variable, Y. Models with small PRESS values are considered "good" candidate models. The C criterion is computed based on MSE and SSE (sum of squared errors), and we seek to identify subsets of X variables for which C value is small. Specific to this study, we have p =15 potential independent variables that include the liner and quadratic or exponential plant coverage variables and lake binary variables. The all-possible-regressions procedure estimates 32768 submodels (2^{15}).

Table 2.6 Baseline Hedonic Models

	Milfoil		Total Macrophytes	
	Quadratic	Exponential	Quadratic	Exponential
UNIMP	2.3078[a]	2.3143	1.5990	1.5921
	1.0463[b]	(1.0469)	(1.0580)	(1.0452)
ln (LVAREA)	0.6112	0.6182	0.4703	0.4553
	(0.1538)	(0.1544)	(0.1608)	(0.1599)
HEAT	-0.2730	-0.2712	-0.1603	-0.1666
	(0.1986)	(0.1986)	(0.1940)	(0.1922)
FULLBATH	-0.3553	-0.3978	-0.2424	-0.1621
	(0.3956)	(0.3979)	(0.3894)	(0.3916)
LAKEWATER	-0.0773	-0.0822	-0.1248	-0.1151
	(0.1140)	(0.1136)	(0.1112)	(0.1098)
LOT	0.1007	0.0992	0.1011	0.1023
	(0.0354)	(0.0354)	(0.0345)	(0.0342)
FF	-0.0003	-0.0002	0.0003	0.0003
	(0.0008)	(0.0008)	(0.0009)	(0.0008)
DIST	-0.1181	-0.1134	-0.0981	-0.1091
	(0.0341)	(0.0334)	(0.0328)	(0.0319)
INTWC	-0.0004	-0.0004	-0.0002	-0.0002
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
EWM	**-0.2470**	**-0.0366**	NA	
	(0.2516)	**(0.0868)**		
EWM ^{2}	**0.0378**		NA	
	(0.0315)			
exp (EWM)		**0.0010**	NA	
		(0.0009)		
TOTAL	NA[c]		**0.4475**	**0.1344**
			(0.2487)	**(0.0897)**
TOTAL ^{2}	NA		**-0.0587**	
			(0.0293)	
exp (TOTAL)	NA			**-0.0016**
				(0.0007)
Beebe Pond	-1.0601	-1.0750	-0.3689	-0.2522
	(1.1365)	(1.1396)	(1.0999)	(1.0956)
Echo Lake	-1.4575	-1.5015	-0.4438	-0.2725
	(1.2207)	(1.2269)	(1.2293)	(1.2301)
Lake Hortonia	-0.8785	-0.9436	-0.0137	0.1798
	(1.0403)	(1.0512)	(1.0060)	(1.0115)
Lake St. Catherine	-0.3270	-0.3509	0.1062	0.2231
	(0.7890)	(0.7916)	(0.7749)	(0.7752)
constant	12.0294	11.6582	10.4646	10.9634
	(1.8483)	(1.7515)	(1.8144)	(1.7196)
Adjusted R^2	0.6116	0.6115	0.6303	0.6373
N	65	65	65	65

[a] ***, **, * denotes significance at 0.01, 0.05 and 0.1 levels.

[b] Numbers in parentheses are standard errors.

[c] NA denotes not applicable.

2. 6 Results

Estimation results for base models will all explanatory variables included (equation 2) are presented in Table 2.6.[13] The second and third columns present the results for the milfoil coverage (*EWM*) equations (quadratic and exponential specifications) and the fourth and fifth columns present the respective results for the total aquatic plant coverage (*TOTAL*) equations.[14]

Living area [*ln*(*LVAREA*)], lot size (*LOT*), and distance to the nearest business district (*DIST*) are significant and positive in all four equations, and unimproved land dummy (*UNIMP*) is significant and positive in the two milfoil equations. None of the lake-specific binary variables are significant, which suggests there are not unique aspects of the lakes, which are not controlled for by the variables in the equations.[15]

The results also show that the *EWM* does not significantly affect property values in the neither the quadratic nor the exponential specifications. In contrast, both *TOTAL* and $TOTAL^2$ are significant in the quadratic equation, and *exp*(*TOTAL*) is statistically significant in the exponential equation. Both of these results indicate that total aquatic plant coverage diminishes property values because the coefficients on $TOTAL^2$ and

[13] Two kinds of spatial relationships were investigated, spatial dependence (or spatial autocorrelation) and spatial heterogeneity. These potential problems were investigated by using Lagrange multiplier (LM) tests (Anselin, 2005). The spatial weight matrix takes a dichotomous form where all "neighbors" (all properties abutting the same lake) are assigned a value of 1 and all "non-neighbors" are assigned a value of 0. For each of the four models in Table 2. 6 neither LM-error nor LM-lag is significant, suggesting that neither spatial dependence nor spatial heterogeneity exist in the data.

[14] A variety of other specifications of the hedonic equation were estimated for both the milfoil and total aquatic plants, e.g., [$h(MC) = \beta_{10}MC$] and [$h(MC) = \beta_{10}MC^2$] separately. Other specifications of the macrophyte coverage variables were generally not significant and when significant suggest that total aquatic plant coverage, not milfoil coverage, affects sale prices.

[15] The lack of significance suggests that there is not an endogeneity problem, that the lakes are not in separate markets and there are not other unique aspects of individual lakes that are not controlled by the independent variables.

exp(*TOTAL*) are negative, but the quadratic result is surprising because the sign of the coefficient on *TOTAL* is positive. The quadratic specification suggests that plant coverage on the water surface up to a rating of 3 (21-40% coverage) increases property values and then decreases values for further increases in plant coverage. Given government documentation and media reporting there is no logical reason for this result; all conjectural evidence suggests that sale prices should decrease with increase in plant coverage.

The PCA reduces the number of explanatory variables by 5 and results in the same pattern of results for the plant coverage variables (Table 2.7). *EWM* is not significant in the quadratic or the exponential specifications. Both *TOTAL* and $TOTAL^2$ have significant coefficients with *TOTAL* being positive and $TOTAL^2$ being negative. This pattern of results again indicates that property prices increase with plant coverage ratings up to 3 and then decline thereafter. The coefficient for the exponential term [*exp*(*TOTAL*)] is statistically significant and negative, which indicates that total aquatic plant coverage diminishes property values.

Table 2.8 presents the independent variables that the APR procedure indicated were the "best" models.[16] For the two milfoil models, *EWM* was not significant in the quadratic or exponential specifications and neither of these models is reported. The coefficients on *TOTAL* and $TOTAL^2$, as well as *exp*(*TOTAL*), were significant and parsimonious

[16] The APR procedure estimated 2^{15}=32768 submodels, they were then ranked according to MSE, PRESS and C criteria respectively. By "best", we mean that the submodel selected for each of the 4 specifications (2 for milfoil and 2 for total macrophyte) is top 10 for all the 3 criteria and has the highest total rank.

specifications of these models are reported. Again, the quadratic results indicate that property values increase with plant coverage ratings up to 3 and decline thereafter.

Table 2.7 Estimated Hedonic Models with PCA

	Milfoil		Total Macrophytes	
	Quadratic	Exponential	Quadratic	Exponential
PC1	0.2484[a]	0.2478	0.2231	0.2290
	0.0434[b]	(0.0432)	(0.0394)	(0.0388)
PC2	0.2151	0.2132	0.1502	0.1855
	(0.1201)	(0.1180)	(0.1068)	(0.1036)
PC3	0.1840	0.1837	0.2088	0.2130
	(0.0611)	(0.0611)	(0.0537)	(0.0537)
PC4	-0.1221	-0.1208	-0.0846	-0.1073
	(0.0758)	(0.0744)	(0.0668)	(0.0654)
EWM	**-0.0659**	**-0.0120**	NA	
	(0.2918)	**(0.1011)**		
EWM ^{2}	**0.0105**		NA	
	(0.0363)			
exp(EWM)		**0.0003**	NA	
		(0.0010)		
TOTAL	NA[c]		**0.6462**	**0.1449**
			(0.2559)	**(0.0935)**
TOTAL ^{2}	NA		**-0.0882**	
			(0.0298)	
exp (TOTAL)	NA			**-0.0022**
				(0.0007)
Beebe Pond	0.7907	0.7841	0.5421	0.6176
	(0.4309)	(0.4320)	(0.4094)	(0.4016)
Echo Lake	0.8931	0.8832	1.0216	1.1141
	(0.5360)	(0.5327)	(0.4843)	(0.4835)
Lake Hortonia	0.7914	0.7711	0.8165	0.9650
	(0.5087)	(0.5112)	(0.4310)	(0.4231)
Lake St. Catherine	0.9447	0.9424	0.7092	0.7839
	(0.3572)	(0.3545)	(0.3187)	(0.3127)
constant	11.1449	11.0796	10.1940	10.8375
	(0.5477)	(0.3314)	(0.5017)	(0.3278)
Adjusted R^2	0.4470	0.4474	0.5593	0.5599
N	65	65	65	65

[a] ***, **, * denotes significance at 0.01, 0.05 and 0.1 levels.

[b] Numbers in parentheses are standard errors.

[c] NA denotes not applicable.

Table 2.8 "Best" Models from All-Possible-Regressions Procedure

	Quadratic	Exponential
UNIMP		2.0665
		(0.9429)
ln (LVAREA)	0.1883[a]	0.4789
	0.0280[b]	(0.1357)
HEAT		
FULLBATH		
LAKEWATER	-0.1792	
	(0.1003)	
LOT	0.1076	0.1005
	(0.0312)	(0.0310)
FF		
DIST	-0.0994	-0.0985
	(0.0304)	(0.0293)
INTWC	-0.0001	-0.0002
	(0.0002)	(0.0002)
TOTAL	**0.5118**	
	(0.2271)	
$TOTAL^2$	**-0.0692**	
	(0.0266)	
exp (TOTAL)		**-0.0007**
		(0.0003)
Beebe Pond	-0.0240	-0.2867
	(1.0530)	(1.0417)
Echo Lake	0.1018	-0.5146
	(1.1531)	(1.1450)
Lake Hortonia	0.3257	0.0220
	(0.9659)	(0.9565)
Lake St. Catherine	0.3637	0.1113
	(0.7433)	(0.7331)
constant	11.7022	10.8223
	(1.3885)	(1.4806)
Adjusted R^2	0.6351	0.6442
N	65	65

[a] ***, **, * denotes significance at 0.01, 0.05 and 0.1 levels.

[b] Numbers in parentheses are standard errors.

These results collectively indicate that the estimation results are robust to the inclusion and exclusion of explanatory variables despite the small sample size. Milfoil (*EWM*), by itself, does not after property values, but as the major component of total plant coverage (*TOTAL*) it does diminish property prices.

The theoretical question that remains is does the quadratic or the exponential specification of *TOTAL* fit the data best? The J-test[17] results indicate α (equation 4a – α = 0.826, se_α = 0.486, p_α = 0.095) is significant, but δ (equation 4b – δ = 0.668, se_δ = 0.458, p_δ = 0.151) is not, suggesting the exponential specification fits the data better.

Based on the coefficient estimates for the exponential specification, the total aquatic plant coverage variables in Table 2.8 are used to compute the marginal effect of aquatic plants on property values. Marginal values are computed for each of the 5 increments on the six-point, macrophyte-coverage scale (e.g., 1→2, 2→3, … , 5→6). This is done for increases (infestation) and decreases (remediation) of total plant coverage. If current aquatic plant coverage is "*k*" then the increment is "*k*+1" for an increase and "*k*-1" for a decrease. The marginal effects for increases and decreases in aquatic plant coverage, using the exponential specification, are computed as follows:

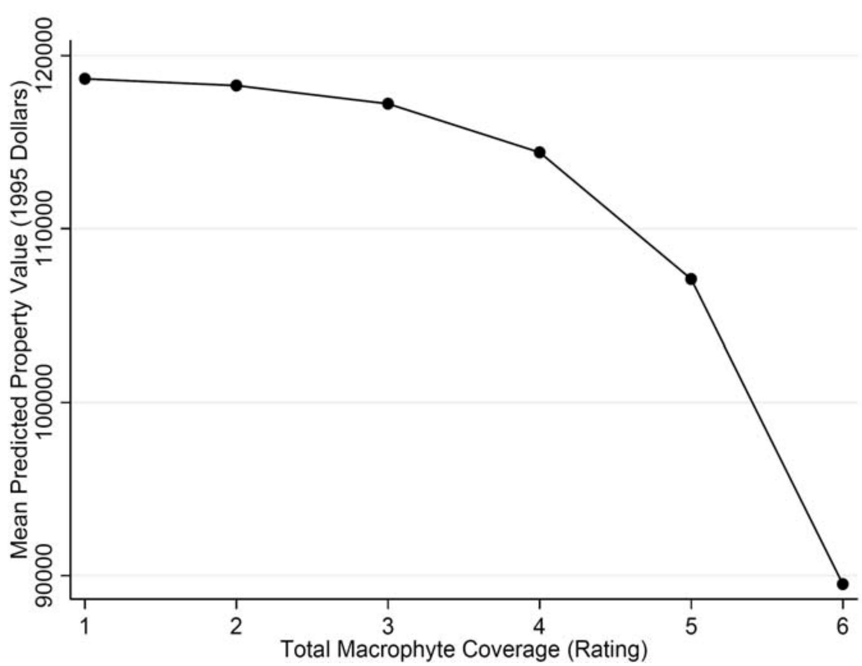
$$MP_{EXP}^{+} = \frac{(SP_{k+1} - SP_k)/SP_k}{(k+1)-k} = \frac{SP_{k+1} - SP_k}{SP_k} = e^{\beta_{10}(e^{k+1}-e^{k})} - 1\text{, and} \tag{6a}$$

$$MP_{EXP}^{-} = \frac{(SP_{k-1} - SP_k)/SP_k}{k-(k-1)} = \frac{SP_{k-1} - SP_k}{SP_k} = e^{\beta_{10}(e^{k+1}-e^{k})} - 1\,. \tag{6b}$$

[17] The J-test results reported here are for quadratic and exponential specifications in Table 2. 8. The J-test is also applied to specifications in Table 2. 6 (quadratic and exponential specification for total macrophyte coverage) and Table 2. 7 (quadratic and exponential specification for total macrophyte coverage). The results show that both α and δ are insignificant, indicating the test is indeterminate.

Marginal prices for increases in macrophyte coverage range from $355 to $17,764, which correspond to percentage reductions in property values ranging from 0.3% to 16.4% (Table 2.9). If a lake has heavy aquatic plant coverage, removing the milfoil such that the rating drops from 6 to 5 would increase property values by $21,356 (19.65%). This, for example, would be the projected property average value improvement for Lake Hortonia that has an average coverage rating of 5.8 that is entirely composed of milfoil. The price diminutions for incremental increases in infestations according to the 6-point plant coverage scale are shown in Figure 2.5.

Figure 2.5 Mean Predicted Lakefront Property Value (dollars) [a]



[a] Predicted property values are calculated for each property at aquatic plant coverage level k (k=1,2,3,4,5,6) based on exponential sepcification and the mean value is then computed over all n observations at each level.

Table 2.9 Marginal Effect of Changes in Total Macrophyte Coverage[a]

Increasing Aquatic Plant Coverage (Invasion)		
k =1 to k =2	-0.33%	-\$354.69
k =2 to k =3	-0.88%	-\$961.45
k =3 to k =4	-2.39%	-\$2,593.66
k =4 to k =5	-6.36%	-\$6,906.54
k =5 to k =6	-16.35%	-\$17,764.39
Decreasing Aquatic Plant Coverage (Remediation)		
k =6 to k =5	19.65%	\$21,355.50
k =5 to k =4	6.82%	\$7,414.51
k =4 to k =3	2.46%	\$2,670.91
k =3 to k =2	0.90%	\$975.04
k =2 to k =1	0.33%	\$357.68

[a] Marginal effects of total macrophyte are calculated based on exponential form.

2.7 Conclusions and Implications

This study shows that Eurasian watermilfoil significantly and substantially affects lakefront property values as the primary component of total aquatic macrophyte growth in a lake. As milfoil infests a lake, adding to total macrophyte growth, property values can diminish by <1% to 16% for incremental increases in the infestation. Four of the five percentages are 7% or less for increasing or decreasing milfoil invasions.

It is difficult to compare the marginal effects for different changes in aquatic macrophyte coverage to the all-or-nothing and before-or-after milfoil effects reported by Halstead et al. and Horsch and Lewis. To place these percentages in perspective, Boyle and Keil (2001) reviewed seven hedonic studies of water quality and reported marginal price effects for three studies as percentages of sales prices. The marginal effects are 2% for a 100 unit change in fecal coliform counts, 6% for a one-unit change in PH and 20% for location inside versus outside of a lake bay with eutrophication. The later study is another all-or-nothing application, like Halstead et al., which is not comparable to the current study. Boyle et al. (1998) report that a one meter change in water clarity, from either an improvement or worsening of eutrophication, can have a 4 to 16% effect on property values. Our results, accompanied by the results from other hedonic studies of water quality issues, suggest the binary-modeling approaches of Halstead et al. and Horsch and Lewis may overestimate the property-price impacts of milfoil. The Halstead et al. study may be capturing other attributes that vary between lakes and the Horsch and Lewis study may be capturing other lake attributes that changed at the same time as the milfoil invasions.

The findings from the study reported here have important policy implications. First, milfoil is the primary component of total aquatic plant growth, which means that milfoil significantly reduces property prices even though the milfoil variable, by itself, was not significant. There are a number of reasons why the milfoil variable may not have been significant. Given that milfoil looks similar to some native aquatic plants property owners may not be able to distinguish between milfoil and these native plants. Property owners may find aquatic plant growth in total problematic, not just the milfoil. It is also possible that milfoil might be found to be significant if more data were available; a larger sample size and observations from more lakes. These interpretations are all observationally equivalent with the current data and we cannot comment on the relative credibility of these potential inferences.

Second, once milfoil is introduced into a lake it will grow rapidly and spread and is impossible to eradicate. Hence, management efforts have focused on protecting lakes from Eurasian watermilfoil. The results reported here indicate that policies that successfully prevent infestations have significant economic benefits to owners of lakefront properties and local communities. As shown in Table 2.4, the percentage of macrophyte growth attributable to milfoil ranges from 14% to 100%, and the average milfoil coverage rating across all five lakes is about 4. If milfoil infestation level increases from the average value, 4 (41%-60% coverage), to 5 (61% to 80% coverage), the marginal change can have a 6.4% reduction in property values. Consider a simple example, if the average value of lakefront properties was $100,000 (close to the $109 thousand reported in Table 2.3) and there were 1,000 lakefront properties, then a 6.4% reduction in property values from further milfoil invasion would result in an aggregate property value loss of about $6.4 million. If the property tax rate were 1.5%, then the $6.4 million lost in property value would result in an annual loss in property tax revenue of nearly $100,000. While this is an example for a stylized lake, the intuition applies to all lakes in Vermont. Even if a lake is free of milfoil currently, it is under threat from this invasive aquatic species.

This presents a dilemma for land owners, community leaders and resource managers as those that have the most to lose from milfoil infestation of lakes, property owners and local communities may not be the perpetrators of the spread of milfoil. Milfoil is spread from lake to lake by transient boaters, migratory waterfowl and other sources. Protection efforts rely substantially on programs to educate people to check and clean their aquatic gear of milfoil before and after entering a lake. While there are programs in place to educate boaters to check and clean their boats when they remove them from a lake, cooperation is voluntary and detection of milfoil on all parts of boats, motors and trailers is difficult.[18]

Thus, to enhance the voluntary control of milfoil, lake associations and local communities may want to pay trained professionals to education property owners, and monitor boat launches to educate and help boaters check their boats, motors and trailers for milfoil when launching and removing their boats from lakes. For lakes with milfoil infestations, the property value and property tax impacts can be used to justify efforts to control and reduce the extent of the invasions. Finally, the economic welfare losses estimated here are lower bounds of the total losses because they only count the losses to lakefront property owners and do not count the losses to people who use lakes to recreate but do not own lakefront property.

[18] http://www.anr.state.vt.us/dec/waterq/lakes/docs/ans/lp_milfoilbookletcombined08.pdf, last accessed on April 21, 2010.

References

Anselin, L., 2005. Exploring Spatial Data with GeoDa: A Workbook. Spatial Analysis Laboratory, Department of Geography, University of Illinois, Urbana-Champaign.

Atkinson, S.E., Crocker, T.D., 1987. A Baysian Approach to Assessing the Robustness of Hedonic Property Value Studies. Journal of Applied Econometrics 2(1):27-45.

Boyle, K.J., Lawson, S.R., Michael, H.J., Bouchard R., 1998. Lakefront Property Owners' Economics Demand for Water Clarity in Maine Lakes.Maine Agricultural and Forest Experiment Syation, Miscellaneosu report, 410.

Boyle, M.A., Kiel, K.A., 2001. A Survey of House Price Hedonic Studies of the Impact of Environmental Externalities. Journal of Real Estate Literature 9(2), 117-144.

Boylen, C.W., Eichler, L.W., Sutherland, J.W., 1996. Physical Control of Eurasian Watermilfoil in an Oligotrophic Lake. Hydrobiologia 340, 213-218.

Cropper, M.L., Deck, L.B., McConnell, K.E., 1988. On the Choice of Functional Form for Hedonic Price Functions. Review of Economics & Statistics 70(4), 668-675.

Davidson, R., MacKinnon, J.G., 1981. Several Tests for Model Specification in the Presence of Alternative Hypotheses. Econometrica 49(3), 781-793.

Eiswerth, M.E., Darden, T.D., Johnson, W.S., Agapoff, J., Harris, T.R., 2005. Input–Output Modeling, Outdoor Recreation, and the Economic Impacts of Weeds. Weed Science 53(1), 130-137.

Eiswerth, M.E., Donaldson, S.G., Johnson, W.S., 2000. Potential Environmental Impacts and Economic Damages of Eurasian Watermilfoil (*Myrophyllum spicatum*) in

Western Nevada and Northeastern California. Weed Technology 14 (3), 511-518.

Freeman, A. M., III, 2003. Measurement of the Environment and Resource Values. Resources for the Future.

Gibbs, J.P., Halstead, J.M., Boyle, K.J., Huang, Ju-Chin, 2002. An Hedonic Analysis of the Effects of Lake Water Clarity on New Hampshire Lakefront Properties. Agricultural and Resource Economics Review 31(1), 39-46.

Greene, W.H., 1994. Econometric Analysis. Upper Saddle River, New Jersey, 258.

Halstead, J.M., Michaud, J., Hallas-Burt, S., Gibbs, J.P., 2003. Hedonic Analysis of Effects of a Nonnative Invader (*Myriophyllum heterophyllum*) on New Hampshire Lakefront Properties. Environmental Management 32(3), 391-398.

Holmes, T.P., Murphy, E.A., Bell, K.P., 2006. Exotic Forest Insects and Residential Property Values. Agricultural and Resource Economics Review 35(1), 155-166.

Horsch, E.J., Lewis D.J., 2009. The Effects of Aquatic Invasive Species on Property Values: Evidence from a Quasi-Experiment. Land Economics 85(3), 391-409.

Irwin, E.G., Bockstael, N.E., 2001. The Problem of Identifying Land Use Spillovers: Measuring the Effects of Open Space on Residential Property Values. American Journal of Agricultural Economics 83(3), 698-704

Kaiser, B.A., Burnett, K., 2006. Economic Impacts of E. Coqui Frogs in Hawaii. Interdisciplinary Environmental Review 8, 1-11.

Lancaster, K.J., 1966. A New Approach to Consumer Theoty. The Journal of Political Economy 74(2), 132-157

Leggett, C.G., Bockstael, N.E., 2000. Evidence of the Effects of Water Quality on Residential Land Prices. Journal of Environmental Economics and Management 39(2), 121-144.

Lovell, S.J., Stone, S.F., Fernandez, L., 2006. The Economic Impacts of Aquatic Invasive Species: A Review of the Literature. Agricultural and Resource Economics Review 35 (1), 195-209.

Madsen, J.D., Sutherland, J.W., Bloomfield, J.A., Eichler, L.W., Boylen, C.W., 1991. The Decline of native Vegetation Under Dense Eurasian Watermilfoil Canopies. Journal of Aquatic Plant Management 29, 94-99.

Mendelsohn, R., Hellerstein, D., Huguenin, M., Unsworth, R., Brazee, R., 1992. Measuring Hazardous Waste Damages with Panel Models. Journal of Environmental Economics and Management 22(3), 259-271.

Michael, H. J., Boyle, K. J., Bouchard, R., 2000. Does the Measurement of Environmental Quality Affect Implicit Prices Estimated from Hedonic Models. Land Economics 76 (2): 283-298.

Neter, J., Kutner, M.H., Nachtsheim, C.J., Wasserman, W., 1996. Applied Linear Statistical Models, fourth ed.

Palmquist, R.B., 2003. Property Value Models, in: Mäler, K.G. and Vincent, J. (eds), Handbook of Environmental Economics-volume 2., (Amsterdam: North Holland/ Elsevier Science).

Pimentel, D., McNair, S., Janecka, J., Wightman, J., Simmonds, C., O' Connell, C., Wong

E., Russel, L., Zern, J., Aquino, T., et al, 2001. Economic and Environmental Threats of Alien Plant, Animal and Microbe Invasions. Agriculture, Ecosystems and Environment 84(1), 1-20.

Pimentel, D., Zuniga, R., Morrison, D., 2005. Update on the Environmental and Economic Costs Associated with Alien-Invasive Species in the United States. Ecological Economics 52(3), 273-288.

Poor, P. J., Pessagno, K.L., Paul, R.W., 2007. Exploring the Hedonic Value of Ambient Water Quality: A Local Watershed-Based Study. Ecological Economics 60(4), 797-806.

Poudyal, N.C., Hodges, D.G., Tonn, B., Cho, S., 2009. Valuing Diversity and Spatical Pattern of Open Space Plots in Urban Neighborhoods. Forest Policy and Economics 11(3), 194-201.

Rosen, S., 1974. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. Journal of Political Economy 82(1), 34-55.

Sheldon, S.P., Creed, R.P., Jr., 1995. Use of a Native Insect as a Biological Control for an Introduced Weed. Ecological Applications 5(4), 1122-1132.

Smeltzer, E., Heiskary, S.A., 1990. Analysis and Applications of Lake User Survey Data. Lake and Resevior Management 6(1), 109-118.

Smith, C.S., Barko, J. W., 1990. Ecology of Eurasian Watermilfoil. Journal of Aquatic Plant Management 28(2), 55-64.

Stigler, G.J., Sherwin, R.A., 1985. The Extent of the Market. Journal of Law and

Economics 28(3), 555-585

Taylor, L.O., 2003. The Hedonic Method. In P.A. Champ, K.J. Boyle and T.C. Brown (eds.), A Primer on Nonmarket Valuation. Kluwer Academic Publishers.

Unmuth, J.M.L., Hansen, M.J., Pellett, T.D., 1999. Effects of Mechanical Harvesting of Eurasian Watermilfoil on Largemouth Bass and Bluegill Populations in Fish Lake, Wisconsin. North American Journal of Fisheries Management 19, 1089-1098.

Vermont Agency of Natural Resources, 2010. Report on Aquatic Nuisance Control Activities in Vermont. (http://www.leg.state.vt.us/reports/2010ExternalReports/252250.pdf, accessed May 31, 2010).

Wagner, K.I., Hauxwell, J., Rasmussen, P.W., Koshere, F., Toshner, P., Aron, K., Helsel, D.R., Toshner, S., Provost, S., Gansberg, M., Masterson, J., Warwick S., 2007. Whole-lake Herbicide Treatments for Eurasian Watermilfoil in Four Wisconsin Lakes: Effects on Vegetation and Water Clarity. Lake and Reservoir Management 23 (1), 83-94.

Wilcove, D. S., Rothstein, D., Dubow, J., Phillips, A., Losos, E., 1998. Quantifying Threats to Imperiled Species in the United States. Bioscience 48 (8), 607-615.

Chapter 3

The Impact of 9/11 on Housing Prices near Mosques

ABSTRACT

The terrorist attacks by Islamic extremists that occurred on September 11, 2001 (9/11) had a tremendous influence on American society. An unexplored potential impact of 9/11 is the influence it may have had on housing prices near mosques. If after 9/11 households were less willing to live near mosques because of a perceived increase in the risk of being exposed to terrorist activities, increased surveillance by the government, or because of prejudice towards Muslims or the Islamic religion, this could cause housing prices to decrease. Using a unique dataset that provided the locations of Muslim mosques in the U.S. around the time of 9/11 combined with micro-level housing data in many of these areas, we find that housing prices decreased by approximately 7% ($10,559 for the average home) in areas near mosques along the east coast of the U.S. on average in the two years following the attacks. However, on the west coast we find no evidence that 9/11 caused a systematic decrease in housing prices near mosques.

KEYWORDS: 9/11 Terrorism Attack, Quasi-experiment, Hedonic, Property Values

3.1 Introduction

The terrorist attacks by Islamic extremists that occurred on September 11, 2001 (9/11) had a profound impact on American society. For example, since 9/11 citizens of the United States have been much more concerned about the risks from terrorism. Anecdotal evidence in the media suggests there has also been an increase in anti-Islam and anti-Muslim sentiment. For example, there have been reports of a rise in hate crimes towards Muslims and also strong opposition to the building of new mosques.[1] Muslims have expressed frustration from being the targets of increased surveillance by the government.[2] Besides these broad societal changes, 9/11 also caused a wide array of economic impacts. Research by economists have shown that 9/11 had tremendous impacts on the insurance industry (i.e. Ericson and Doyle, 2004; and Lakdawalla and Zanjani, 2006), the aviation industry (i.e. Carter and Simkins, 2004; and Ito and Lee, 2005), the government's investment in homeland security and national defense (i.e. Hobijn, 2002; and Makinen, 2002), and on the tourism industry (i.e. Goodrich, 2002; and Blake and Sinclair, 2003).

This paper is focused on an economic impact that has not yet been considered in the literature — the impact that 9/11 had on housing prices near mosques.[3] There are several

[1] For an example of the media's documentation of a hate crime see the New York Time's report on the recent stabbing of the taxicab driver in New York City at: http://www.nytimes.com/2010/08/26/nyregion/26cabby.html . For a discussion of the increase in opposition to building mosques since 9/11 see USA Today's write-up at: http://www.usatoday.com/news/religion/2004-03-08-mosque-opposition_x.htm.

[2] "More than half of Muslims in this country say government anti-terrorism policies single them out for increased surveillance and monitoring, and many report increased cases of name-calling, threats and harassment by airport security, law enforcement officers and others," For more reports see http://www.nationofchange.org/paying-high-price-9-11-paranoia-1316103047, http://theglobalconsciousness.wordpress.com/2012/02/25/the-threat-of-terrorism-is-at-an-all-time-low-so-why-the-increases-government-surveillance-on-muslim-communities/, last accessed on 04/21/2012.

[3] While this is the first paper to look at housing prices and mosques, there are two related papers that have looked at the impact that terrorist activities can have on property values. Gautier et al. (2009) looked at a

reasons why 9/11 may have influenced housing prices near mosques. First, it is possible that after 9/11 households were less willing to live near mosques because many households now had stronger preferences for not living near Muslim households or their places of worship. According to a 2006 *USA Today* - Gallup Poll focusing on U.S. attitudes toward Muslims living in the United States,[4] a significant number of American are willing to admit they harbor at least some feelings of prejudice against Muslims. Nearly one quarter of Americans, 22%, say they would not like to have a Muslim as a neighbor. Another possibility is that households perceived an increase risk in terrorist and criminal activities near mosques following 9/11. Following 9/11 there was heightened coverage of any criminal or terrorist acts that took place near mosques. For example, the burning or desecration of mosques or the accusations that Muslim leaders at mosques posed a threat to U.S. security were widely covered in the media. _Finally, it may be that Muslims were less likely to move near a mosque or Muslims living near mosques were more likely to leave because of increased government surveillance and investigation.[5] This paper is not focused on trying to determine an exact social mechanism that leads to a possible housing price impact, rather it attempts to document whether or not a housing price impact occurred as a result of 9/11.

---

terrorism event in the Netherlands and the impact it had on house prices and Abadie and Dermisi (2008) who looked at the impact of 9/11 on the office real estate market in downtown Chicago. In Section 2 we discuss these papers in more detail.

[4] http://www.gallup.com/poll/24073/antimuslim-sentiments-fairly-commonplace.aspx, last accessed on August 2nd, 2010. Survey methods: "These results are based on telephone interviews with a randomly selected national sample of 1,007 adults, aged 18 and older, conducted July 28-30, 2006. For results based on this sample, one can say with 95% confidence that the maximum error attributable to sampling and other random effects is ±3 percentage points."

[5] "In general, mosques and other houses of worship do not have special protection from surveillance under U.S. law … Since the Sept. 11, 2001, terrorist attacks, several federal investigations have used informants, surveillance and electronic eavesdropping to gather information about mosques." For more information see http://muslimsforasafeamerica.org/?p=11, last accessed on 04/21/2012.

Two unique datasets make our analysis possible. The first dataset comes from the 2000 Mosque Study Project which was the largest, most comprehensive survey of mosques conducted in the United States prior to 9/11.[6] This project collected information on 1,209 mosques built from 1925 to 2000 in the United States, including the physical addresses of the mosques. This allowed us to geocode the exact location of mosques that existed in the United States at approximately the time of 9/11. The second critical dataset for our analysis is micro-level data on individual housing prices and characteristics for many populated areas along the east and west coasts in the United States. This dataset was purchased from a commercial vendor of assessor data. The housing data was also geocoded so that we could calculate the distance from individual homes to the nearest mosque. Thus these two unique datasets allow us to conduct our analysis at the micro-level and also allows us to look at differences in 9/11 impacts between the eastern and western United States.

Our identification strategy exploits 9/11 as an unanticipated and exogenous event. Specifically, we compare prices of houses that sold within a quarter of a mile of a mosque (the treatment group) and houses that sold between a quarter of a mile and a half mile (the control group) before and after 9/11 to houses located between a half mile and a mile away from the nearest mosque. Results from this basic differences-in-differences strategy suggest that housing prices within a quarter of a mile of a mosque in areas along the east coast of the United States, decreased by approximately 8% in the two years following 9/11. Houses located between a quarter mile and one half mile from a mosque saw no change in their housing prices relative to houses between one half and one mile

[6] This project was primarily sponsored by the Council on American-Islamic Relations.

from a mosque. Furthermore, the same analysis applied to areas along the west coast of the United States finds no evidence that 9/11 decreased property values near mosques.

We also conduct a meta-analysis to examine what factors contribute to a decrease in property values. We find that mosque areas (within one mile of a mosque) with higher population density, higher percentage of white population, and located on eastern coast are more likely to have a negative impact. Furthermore, the meta-analysis suggests that there is substantial heterogeneity in where the price decreases occurred. The average price decrease we document in the eastern part of the U.S. is driven primarily by only about 16-18% of the mosques that show a statistically significant price decrease when a generalized difference-in-differences analysis is run solely on that individual mosque area instead of jointly with all mosque areas in the east. While on the west coast only about 2-4% of the mosques have a statistically significant impact when analyzed alone; or about what would be expected due purely to randomness.

We think our results provide two important contributions to the literature. First, it helps to fill in a gap in the 9/11 literature by analyzing the impact that 9/11 had on housing prices near mosques. The magnitude of the capitalization of the 9/11 event into property values, as shown in our results, is approximately 7% ($10,559 for the average home), which suggests that the impact is a non-negligible part of the economic costs of the 9/11 attacks. Second, our results add to the literature on how people react and adapt to the threat of terrorism. The economic effect of terrorism such as 9/11 and the behavioral responses they elicit are important to society and policymakers. Empirical assessments of these behavioral responses are crucial to understand the appropriate remedies and economic policy responses to terrorist activities. Our results indicates that many households preferred not to live near mosques along the east coast of the U.S. after 9/11 (these households may have been Muslim OR non-Muslim).

The remainder of the study will proceed as follows. Section 2 provides background on the issue at hand and provides a review of some of the relevant literature. Section 3 describes in more detail the unique datasets we have acquired that make our analysis possible. Section 4 outlines the identification strategy for the hedonic price regressions used to determine the impact of 9/11 on housing prices near mosques. Section 5 discusses the baseline results and robustness checks. Section 6 introduces meta-analysis and applies the techniques to all the mosque areas in our dataset. The last section concludes the study and discusses the policy implications.

3.2 Further Background and Literature Review

In this section we provide additional background on some of the issues covered in this study. First we describe the existing evidence regarding potential mechanisms that could cause housing prices to decrease near mosques after 9/11 in the U.S. Next we review the relevant papers in the literature on how people react and adapt to terrorist threats. Finally we review the growing literature that combines quasi-experiments with the hedonic method that provides the basis for the identification strategy outlined later on in this paper.

3.2.1 Survey and Media Evidence on Potential Mechanisms

It is possible that after 9/11 households were less willing to live near mosques because many non-Muslim households had stronger prejudice towards Muslim households or their places of worship. There are several surveys that have asked questions related to this potential for an increase in anti-Islam and anti-Muslim sentiment post 9/11. According to a 2006 *USA Today* - Gallup Poll focusing on U.S. attitudes toward Muslims living in the United States, a significant number of American are willing to admit they harbor at least some feelings of prejudice against Muslims. Nearly four in ten Americans (39%) say they do feel some prejudice. Survey questions that ask about Americans' reaction to being near Muslims in different situations are apt to reflect American's personal discomfort with Muslims: nearly one quarter of Americans, 22%, say they would not like to have a Muslim as a neighbor. Slightly fewer, 18%, say they would feel nervous if they noticed a Muslim woman flying on the same airplane as themselves, while significantly more – 31% – say they would feel nervous if they noticed a Muslim man on their flight.

Meanwhile, a poll released on March 9, 2006[7] by the Council of American-Islamic Relations (CAIR), a leading American Muslim civil rights group, suggests that nearly one in five Americans maintain a strong anti-Muslim attitude and showed that approximately one in four Americans believes that Islam is a religion of hatred and violence. Only 6 percent of Americans have a positive first impression of Islam and Muslims.

A similar poll conducted more recently in 2009 (a *Washington Post – ABC News* poll[8]) confirmed the results of *USA Today* - Gallup Poll and CAIR poll. It found that a

---

[7] http://www.cair.com/Portals/0/pdf/american_public_opinion_on_muslims_islam_2006.pdf, last accessed on August 2nd, 2010. Survey methods: "Samples for the 2005 polls are random digit samples of telephone numbers selected from telephone exchanges in the United States. Interviews were conducted by telephone. The interview took place during November 2005 and consisted of a total of 1001 interviews. The response rate was 23.2 percent and the cooperation rate was 41.5 percent. The margin of error for this survey, with 95 percent confidence, is ±3.1 percent.

[8] http://csis.org/files/media/csis/pubs/090421_islampollreport.pdf, last accessed on August 2nd, 2010. This ABC News/Washington Post poll was conducted by telephone March 26-29, 2009, among a random national sample of 1,000 adults including both landline and cell-phone-only respondents. Results for the full sample have a 3-point error margin.

growing proportion of Americans are expressing unfavorable views of Islam, and a majority of Americans say that Muslims are prone to violence. Nearly half of Americans – 48 percent – hold an unfavorable opinion of Islam and 29 percent express the belief that mainstream Islam encourages violence against non-Muslims. The poll also found that one in three Americans has heard prejudiced comments about Muslim lately. One in four Americans admitted to harboring prejudices towards Muslims.

Another possibility for a decrease in housing demand near mosques post 9/11 is that households perceived an increase risk for terrorist or criminal activities near mosques. The mainstreaming of anti-Muslim rhetoric has contributed to a rash of attacks on American mosques. The American Civil Liberties Union (ACLU) has so far chronicled mosque attacks in 21 states.[9] For example, Mosques in New York have frequently faced hate rhetoric by opponents after 9/11, as did several others in Tennessee. A mosque in Jacksonville, Florida was recently the target of a bomb attack. A Houston radio host made a controversial on-air call for the bombing of a proposed mosque and community center in New York City.

A final possibility for why housing prices may have decreased near mosques that we will briefly discuss is that the perceived increase in government surveillance and investigations after 9/11 and the general suspicion of one's neighbors made it less nice for Muslims to live in these neighborhoods and therefore many of them moved out. Furthermore, it could be that social interaction at work and other public places became less pleasant for Muslims after 9/11. As a result, some Muslims may have wanted to

---

[9] For the map see http://www.aclu.org/maps/map-nationwide-anti-mosque-activity, last accessed on 04/21/2012.

integrate more by moving to non-Muslim neighborhoods and changing the way they dress and act in public. Both of these explanations are based on the behavior of Muslims as opposed to non-Muslims.

3.2.2 Terrorism Adaptation Literature

There have been two other studies that have evaluated the impact of terrorism on real estate activities that are related to this paper. The first study by Gautier et al. (2009) used the 2004 murder of film maker and journalist Theo van Gogh by a recent convert to radical Islam as an event study in the Netherlands. Their paper attempted to address the question of whether or not fundamentalist-Islamic terrorism has an effect on the public's attitudes towards mainstream Muslims. They used the hedonic price method to show that after the murder, house prices in Amsterdam neighborhoods with more than 25% Muslims decreased by about 3% in the 10 months following the murder. The second study by Abadie and Dermisi (2008) investigated the impact of 9/11 on the office real estate market in downtown Chicago, IL using building-level panel data on vacancy rates. Their study was interested in understanding whether or not there was an increased perception of terrorist risk after 9/11 in a large U.S. city like Chicago. They used a fixed effects identification strategy and found that vacancy rates increased in the three most distinctive Chicago landmark buildings and buildings nearby that would be considered the most likely targets for a terrorist attack.

Our paper is also related to an existing literature on the economic cost of terrorism. Saxton (2002) categorized and briefly summarized both the short-term and long-term economic costs of the 9/11 attacks. As Krugman (2004) pointed out, the economic costs of terrorism can be divided into three categories: first, immediate loss done by terrorist attacks such as buildings and infrastructure destroyed; second, increased government's budget on national defense and homeland security in an attempt to fight terrorism; and last, economic costs of behavioral response to terrorism.

To illustrate this last category, Krugman used vacation plans as an example of how terrorism distorts individual choices. In Krugman's words, "A Midwesterner decides to forgo a theater trip to New York and make a musical trip to Branson instead; a European traveler chooses a Mediterranean location rather than one in the Basque region; a transatlantic traveler goes to Europe rather than Israel." House purchase plans in some ways are like vacation plans: the 9/11 attacks changed people's attitude toward Muslims and distort their choices on where to live. The distorted individual choices can depress the housing prices in a neighborhood with mosques and higher Muslim population density and thereby decrease the property tax revenue. However, this indirect cost caused by terrorism is largely overlooked in previous studies.

An exception is Frey et al. (2007), who discussed the pros and cons of using contingent valuation surveys and the hedonic price method to assess the indirect costs of terrorism such as people's utility (welfare) loss. The author proposed a method called the "life satisfaction" approach and revealed that terrorism leads to a considerable reduction in life satisfaction and that individuals would need to receive substantial increases in income in order to compensate for the harm caused by terrorism. In some ways our paper can be thought of as focusing on this last category of economic costs of terrorism and using a "revealed preference" method to further understand whether or not it appears that living near a mosque imposes life satisfaction costs on households.

3.2.3 Quasi-Experimental Hedonic Literature

The hedonic price method has been widely applied to housing markets to evaluate household's willingness to pay for changes in a wide range of housing characteristics including environmental and urban amenities and disamenities. Rosen (1974) provided the theoretical underpinnings for the hedonic price model by demonstrating that the hedonic price function is simply an envelope of the equilibrium of transactions between all buyers and sellers of a differentiated good such as a house. Under the assumptions of his model, the marginal value consumers place on individual amenities and disamenities can be recovered by regressing sale prices on the attributes of the house.

This "traditional" approach, however, may often lead to biased parameter estimates because of omitted variables that confound a cross-sectional identification strategy. For this reason, a number of studies have begun to combine the "traditional" hedonic model with quasi-experiments that have occurred from nature or man-made policies to identify a causal impact of the change in a (dis)amenity on housing prices. Recent examples of (dis)amenities that have been analyzed using quasi-experimental hedonic applications include: air quality (Chay and Greenstone, 2005), cancer risk (Davis, 2004), airport noise (Pope, 2008a), crime risk from sex offenders (Linden and Rockoff, 2006; Pope, 2008b), school quality (Black, 1999), and water quality (Horsch and Lewis, 2009). This increased ability to identify a causal impact is not without some cost as was recently demonstrated by Kuminoff and Pope (2012). By identifying a causal impact through the use of panel housing data before and after a natural experiment, the price function may in fact change or shift over time making it unclear if the estimated "treatment effect" is the MWTP for the amenity or simply a measure of capitalization. Kuminoff and Pope present theoretical and empirical evidence that this "wedge" between a capitalization estimate and the true welfare impact is likely in most quasi-experimental hedonic applications. The results from this paper are clearly capitalization effects and may or may not line up well with the true welfare impact.

Another concern with the quasi-experimental hedonic method is that it is difficult to establish whether the event or policy change is truly exogenous. If the event or policy change was expected by households, then the expectations will blur the discrete timing of when the event or policy change occurs and it becomes more difficult to interpret the differences in housing prices before and after the change causally. However, our application makes an ideal quasi-experiment since the 9/11 attacks can clearly be viewed as an unexpected and exogenous event.

3.3 Data

3.3.1 Mosque Data

To conduct our analysis, it is crucial to have information on the location of mosques at approximately the time that 9/11 occurred. We were fortunate that this type of information existed. The mosque location data comes from the 2000 Mosque Study Project, the largest, most comprehensive survey of mosques conducted in the United States prior to 9/11.[10] The project compiled a list of 1,209 mosques that were built between the years 1925 and 2000 in the United States. The existence of each mosque was verified by telephone contact. A mosque is defined to be an organization that holds

[10] http://sun.cair.com/Portals/0/pdf/The_Mosque_in_America_A_National_Portrait.pdf, last accessed on August 5th, 2010. We thank Dr. Ihsan Bagby for providing us with this information.

Jum`ah Prayers (Friday Prayers) and other Islamic activities. Jam`ah Prayers held in hospitals and businesses are not considered mosques but student associations that hold Jum`ah and other Islamic activities are considered a mosque although they might not have a permanent facility (i.e. they may use a university classroom). The original mosques list contains information on mosques' name, physical or P.O. Box address, state, city and zip code.

We checked the information for each mosque in the survey by using "Google maps" or "Bing maps" and assigned a categorical rating from 0 to 4 for each of them based on their building size and appearance. A rating of 0 means the mosque is a Muslim student association or its physical address is not available or there is no related information on the internet. A rating of 1 means its physical address and other information is available on the internet but its image cannot be found on either "Google maps" or "Bing maps". A rating of 2 means the mosque has its own stand-alone building, but it does not have a mosque-like appearance and it does not have a website. A rating of 3 means it is a stand-alone building and is not mosque-like in appearance, but it does have a website or it has a mosque-like building but does not have a website. Finally, the highest rating of 4 means the building is very mosque-like in appearance and it has a website. Before conducting our analysis, we dropped all of the mosques that were rated "0" since they are unlikely to have a physical building and are therefore much less noticeable to households in the area. The addresses provided in our list of mosques were geocoded using GIS software and this process allows us to measure the distance between houses and mosque locations. We were able to match eighty-eight percent of mosques to a physical street address, while the remaining unmatched mosques were dropped from our analysis.

3.3.2 Housing Data

The second key dataset for our analysis is housing data near mosques in the United States. Our micro-level housing data was purchased from a real estate data vendor that collects data from local assessors and then provides this data in a consistent format. The data we purchased contains housing sales and characteristics information from 1998 to 2005 for single-family residential properties in many parts of the United States. Since more populous areas are more likely to digitally archive and distribute their assessor data, our dataset tends to be derived from counties in the U.S. with high population. This means that we have good coverage for populous areas along the east and west coast of the U.S. but much less consistent coverage in the interior of the U.S. To balance the need for geographic parsimony and to maintain a reasonable number of housing sale observations, *a priori* to the analysis we broke our available housing data into 2 regions: east coast states and west coast states.[11] In what follows, we focus on the east coast sample. However, we will highlight corresponding results for the west coast in the results section of the paper later on.

In addition to the sales price of a home that was sold during the time frame of our data, this micro-level dataset also provides a consistent set of housing characteristics such as the square feet of the house, number of bathrooms, number of bedrooms, year built, and the size of the lot. Some neighborhood demographics at the block-group level have also been attached to the data from the 2000 decennial census such as the percentage of non-

[11] All other housing data not on one of the coasts was combined to form a third de facto region. However, since this area combines housing in areas as distant as Denver Colorado and Cleveland Ohio, it is more difficult to think that they constitute a reasonable "housing market" and that one should estimate these observations jointly.

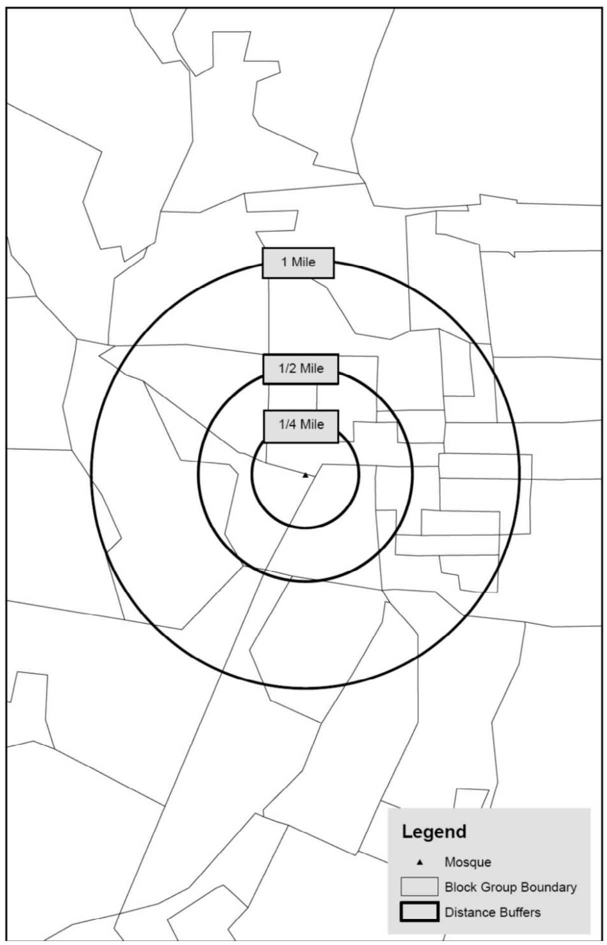
white population, percentage of population under 18 years old, and population density. The addresses of the houses have been geocoded to the street level which means the houses can be spatially related to any level of geography (i.e. block group, census tract, municipality, or county).

3.3.3 Combining the Mosque and Housing Data

The distances from each house transaction to each mosque within 1 mile were calculated using the mosques location information and the geocoded street-level housing addresses. Housing transactions that occurred further than 1 mile from a mosque were dropped from the analysis. Most houses only have one mosque within 1 mile while a smaller subset of houses (14%) have multiple mosques within 1 mile. As is common in the literature, only the information on the nearest mosque is used to conduct the analysis.

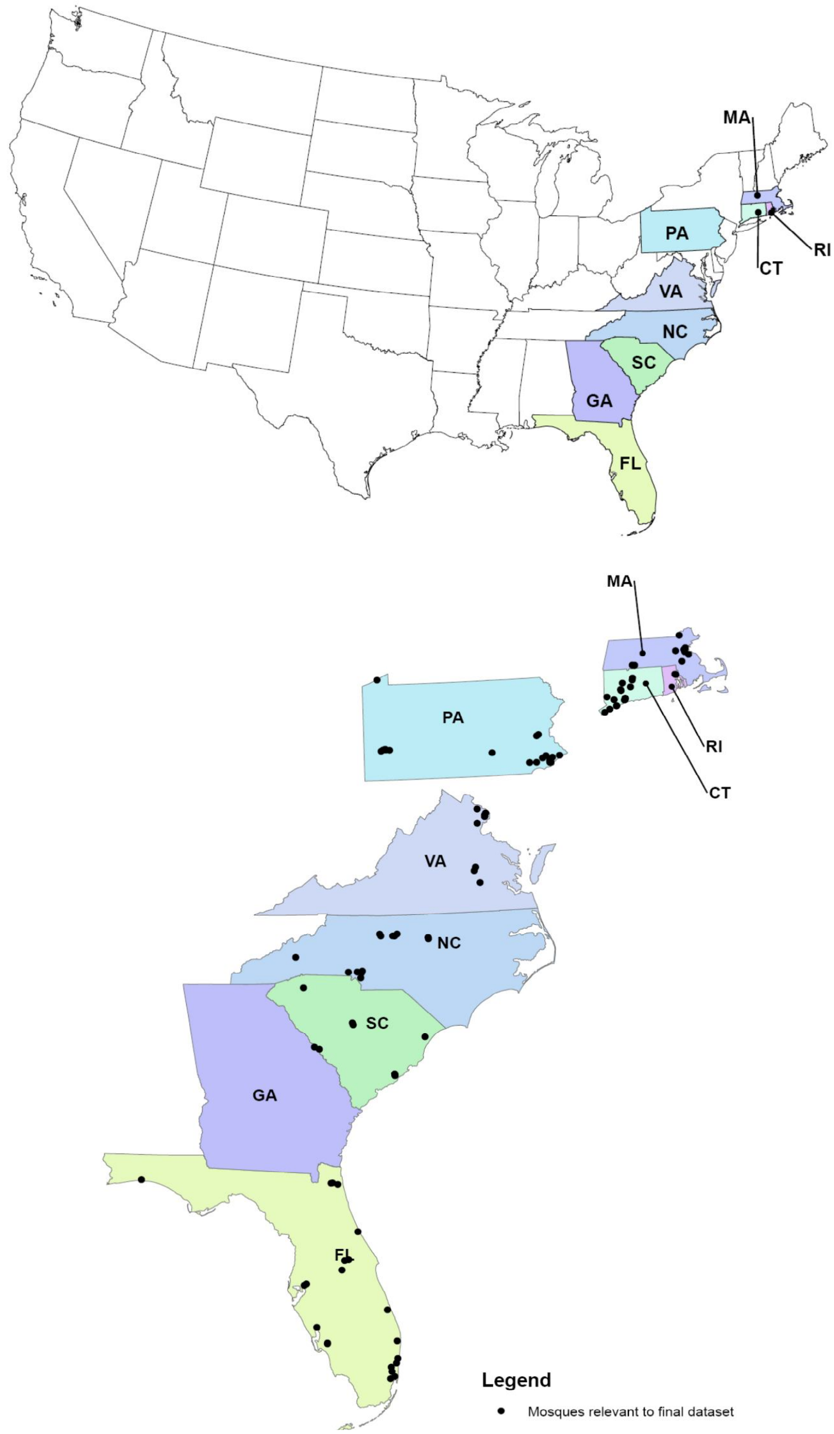
Using the timing of the housing sale in relation to 9/11 as well as the linear distance between mosques and houses, four key variables were created: (i) $D^{1/4}$ is a dummy variable indicating whether or not the property sale occurs within a quarter mile of a mosque; (ii) $D^{1/2}$ is an indicator variable for property sold within a half mile of a (iii) $D^{1/4}\_post$ is a dummy variable equal to 1 for houses that sold within a quarter mile of a mosque *after* the 9/11 event, and (iv) $D^{1/2}\_post$ is a dummy variable equal to 1 for houses that sold within a half mile of a mosque *after* the 9/11 event, and 0 otherwise. The reasoning behind the creation of these indicators and their usefulness in identifying the impact of the 9/11 event on housing prices will be discussed in the subsequent Figure 3.1 shows an example of a mosque in relation to block groups within a quarter, a half, and one mile buffers.[12] We define a "mosque area" as those homes within a one mile radius – the area within the largest circle in Figure 3.1. We will use these areas in our identification strategy to help mitigate omitted variable bias through the use of mosque-specific fixed effects.

Figure 3.1: Example of Parcels within 1/4, 1/2, and 1 Mile of a Mosque



[12] A census block group is smaller than a census tract. Whereas the typical census tract has approximately 4,000 people within its boundaries, a block group typically has a population of around 1,500. Census block groups never cross state, county or census tract boundaries.

Figure 3.2: Map of East Coast Study Area



To better ensure the stability of the hedonic price function over time, we limit our analysis to house sales that occur within a four-year window surrounding 9/11 (i.e., two years prior and two years after 9/11. Specifically, the time period is from 09/11/1999 to 09/11/2003). Limiting the sample to two years before and after, helps us to better mitigate potential temporal confounders in our identification strategy. As discussed in Kuminoff and Pope (2012), this narrowing of our temporal window around the exogenous event makes it more likely that the hedonic price function is stable over the time frame of the analysis, and means that our capitalization estimates are more likely to have a welfare interpretation.

Table 3.1 Summary Statistics of Single Family Transactions in the 4-Year Sample (East)

Variable	Description	Mean	S.E.	Min	Max
SP	sale price	150848.10	110691.50	17000.00	802000.00
SQFT	size of living area	1650.76	721.27	326.00	8597.00
LOT	lot size (acre)	0.22	0.24	0.01	4.75
AGE	age	35.10	29.65	0.50	103.00
BR	# of bedrooms	3.10	0.83	1.00	10.00
BATH	# of bathrooms	1.81	0.81	0.50	6.00
WHITE	white percentage	0.61	0.32	0.00	0.99
POPD	population density	0.05	0.04	0.00	0.43
N			10385		

Table 3.2 Summary Statistics of Mosques Relevant to the 4-Year Sample (East)

State	No. of mosques	Rating			
		Mean	S.E.	Min	Max
Georgia	1	1.0	.	1	1
Rhode Island	1	1.0	.	1	1
South Carolina	3	1.0	0.0	1	1
Virginia	4	3.5	0.6	3	4
Massachusetts	8	2.9	1.0	1	4
Connecticut	8	2.0	1.2	1	4
North Carolina	10	2.7	1.1	1	4
Pennsylvania	13	2.4	1.3	1	4
Florida	14	2.6	1.1	1	4
Sum	62	2.5	1.2	1	4

To provide balance across our pooled cross-sections in a given mosque area, we limit the analysis to areas where we have housing sales before and after 9/11 for each buffer band around a mosque (see Figure 3.1, each mosque area includes three buffer bands: 0 to a quarter mile, a quarter to a half mile, and a half to one mile). After this restriction on the data, we end up with 10,385 housing sales in our "4-year sample" on the east coast that occurred within one mile of 62 mosques. Table 3.1 provides summary statistics of the housing transactions in the 4-year sample for the east coast. Figure 3.2 shows the geographic distribution of the 62 mosques that are relevant to our analysis along the east coast. Table 3.2 provides some summary information for these mosques and the states in which they are located.

3.4 Identification Strategy for the Hedonic Analysis

The major problem in identifying the implicit price of local (dis)amenities is the potential for omitted variable bias. The problem is that variation in the local (dis)amenities may be correlated with factors that are unobserved to economists and therefore the effect of the local (dis)amenities may be (under)overestimated. In our case, the covariance of a mosque location with both observable and unobservable neighborhood characteristics makes it difficult to identify the effect of the 9/11 attacks on property values by merely comparing areas with mosques to areas without them. If for example, mosques tend to be built in poorer communities then variation in the sale prices of properties around the mosque may reflect the distaste for the location where the mosque was built rather than distaste for living near Muslims or the mosque.

3.4.1 Treatment Groups and Control Groups

In Table 3.3 we compare the summary statistics of houses near a mosque (within a half mile radius of a mosque) to houses farther away (between half and one mile of a mosque) in our east coast sample of homes. Although they are quite similar, the main differences on average are that houses near a mosque tend to have lower prices, are slightly older, and have more non-white residences in the block group where the house is located. Instead of comparing aggregated areas, however, our micro-level data allows us to know the specific mosque and housing locations. This allows us to compare the value of property sales within very small areas where the housing stock is more homogenous.

Table 3.3 Comparison of House Sales Near to a Mosque and Farther Away

	Within 1/2 miles of mosque	Between 1/2 and 1 mile of mosque
	Mean (S.E.)	Mean (S.E.)
SP	133511.90	157061.50
	(100466.10)	(113496.90)
SQFT	1599.48	1669.14
	(691.87)	(730.69)
LOT	0.21	0.23
	(0.21)	(0.25)
AGE	38.62	33.84
	(31.54)	(28.85)
BR	3.10	3.10
	(0.83)	(0.83)
BATH	1.73	1.84
	(0.79)	(0.82)
WHITE	0.51	0.64
	(0.33)	(0.31)
POPD	0.05	0.05
	(0.04)	(0.04)
N	2740	7645

As will be illustrated in the results section, our data provides us with the ability to better control for the differences between houses near mosques and slightly further away within a mosque area. It is this control that will allow us to try and identify if prices of homes near mosques fell post 9/11. Table 3.4 also shows that homes within a quarter of a mile and homes between a quarter and a half mile, appear to be very similar to one another before 9/11 occurred. These summary statistics show little tabulated evidence of any preexisting differences in housing characteristics for homes near mosques prior to 9/11.

Table 3.4 Housing Sales Summary for Controls and Treatments before 9/11

	Within 1/4 miles of mosque	Between 1/4 and 1/2 miles of mosque
	Mean	Mean
	(S.E.)	(S.E.)
SP	120891.50	117826.20
	(96481.63)	(87637.09)
SQFT	1643.27	1599.76
	(797.14)	(688.29)
LOT	0.20	0.21
	(0.20)	(0.19)
AGE	36.47	35.39
	(32.68)	(30.36)
BR	3.21	3.08
	(0.87)	(0.75)
BATH	1.71	1.74
	(0.76)	(0.78)
WHITE	0.46	0.52
	(0.33)	(0.33)
POPD	0.05	0.05
	(0.04)	(0.04)
N	374	866

3.4.2. Empirical Models: Cross-Sectional and Generalized Differences-in-differences (GDID) Estimation

We estimate two empirical models: a cross-sectional model and a generalized difference-in-differences model. First, we use the cross-sectional estimator to statistically check that we can provide adequate control for preexisting differences in the prices of properties located within a quarter of a mile of a mosque and those located between a quarter and a half mile of a mosque. Based on the assumption that these geographic areas are similar, we then use a generalized differences-in-differences model where properties that are sold within a quarter mile of a mosque are used as treatments and properties that are sold between a quarter and a half mile and a half and one mile of a mosque are used as controls. A key point of Kuminoff, Parmeter and Pope (2010) is that if the size of the change in the amenity is not marginal, the shape of the hedonic price function can still change even if we look at a very small temporal window. In order to overcome this problem, they suggested a generalized version of difference-in-difference strategy which recognizes that the shape of the hedonic price function may have changed over time and therefore allows both intercept and slope coefficients to change over time.

The cross-sectional specification takes the following form:

$$ln\,(SP) = \alpha + \beta X + \theta D^{1/4} + \gamma Spatial\_fix + \delta Time\_fix + \varepsilon \qquad (1)$$

where *SP* is the actual sale price of a property; *X* is a set of observable property characteristics; $D^{1/4}$ is a dummy variable indicating whether or not the property sale occurs within a quarter mile of a mosque; and *Spatial_fix* represents spatial fixed effects. Also, ε is an error term and α, β, θ, γ are parameters to be estimated.

Starting with a standard DID regression, interact every property characteristic with a time dummy *T* (*T* indicates whether or not the property was sold after 9/11), the GDID takes the following form:

$$ln\,(SP) = \alpha + \beta X + \gamma Spatial\_fix + \delta Time\_fix + \sigma X{*}T$$
$$+ (\theta_0 D^{1/4} + \pi_0 D^{1/2}) + (\theta_1 D^{1/4}\_post + \pi_1 D^{1/2}\_post) + \varepsilon \qquad (2)$$

where $D^{1/2}$ and two interaction terms, $D^{1/4}\_post$ and $D^{1/2}\_post$ were described in the preceding section. This specification indicates that the estimated impact of the 9/11 attacks on property values is given by the term θ_1. Aside from controlling directly for physical and neighborhood characteristics, there are two other components that need to be addressed to make sure that the GDID estimators are adequately capturing causal treatment effects: time and space. To control for time we introduce *Time_fix* which represents a series of year dummies.

Special attention is given to controlling for spatial confounders. If one defines a spatial resolution for the model that is too high, the confounding overlap assumption may be violated.[13] If this happens, too many spatial confounders in the model will reduce estimate precision especially in a scenario where observations are limited. On the other hand, if the spatial resolution is low it might not be able to capture all the spatial amenities that have an impact on housing price. We think a reasonable choice for the *Spatial_fix* variables is a set of "mosque area" dummies. These are defined as the area within a one mile radius of a mosque. Another attractive alternative to controlling for time and space is to include year by mosque specific effects which allows for more flexible and simultaneous control. In our most rigorously controlled models we include these year by mosque controls.

3.5 Results

3.5.1 Graphical Evidence

While the DID framework (and its GDID counterpart) is an attractive specification, it may nevertheless yield biased estimates of the causal impact of 9/11 on housing prices near mosques if there are other uncontrolled temporal confounders or if there are preexisting trends in the treatment and control groups. Thus it is often useful to do a graphical analysis that analyzes these trends and the timing of any impact on housing prices. Figure 3.3A shows the raw quarterly mean prices for properties that sold within a quarter mile and between a quarter, and a half mile in our east coast sample. This figure illustrates that the differences in quarterly mean housing price appear to be small before the 4th quarter of 2001. However, beginning with the 4th quarter of 2001 where we would

[13] This assumption is discussed in Parmeter and Pope (2009). The confounding overlap assumption is imposed to ensure that there are observations for both the treatment and control group for the covariate value of X.

expect any price impacts to first be observed, the price differences increase which is suggestive of a 9/11 impact. However, the change in price differences is somewhat noisy as might be expected given that this figure is simply using the raw averages of housing prices without controlling for other housing characteristics.

Figure 3.3A: Quarterly Mean Price Trend in the 4-Year Window

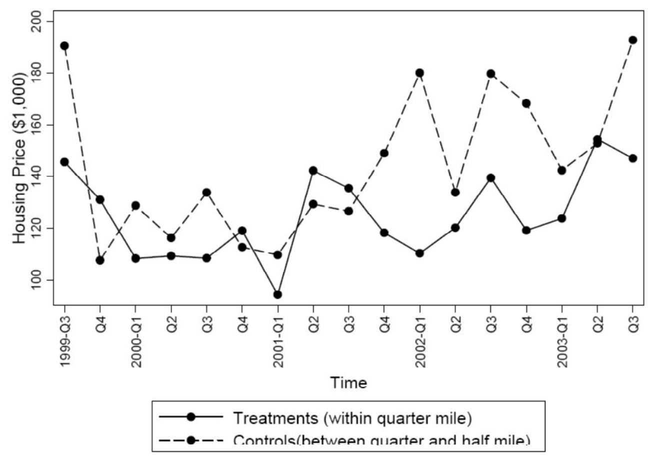


Figure 3.3B: Quarterly Mean Price Trend in the 4-Year Window
(Conditional on housing characteristics)

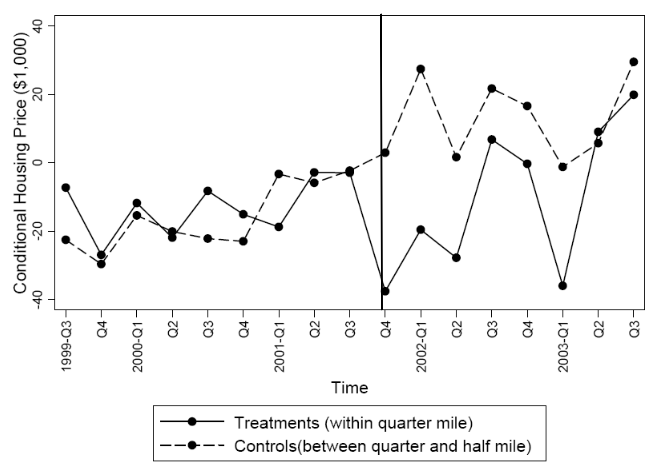


Figure 3.3C: Quarterly Mean Price Trend in the 4-Year Window
(Conditional on housing characteristics and year dummies)

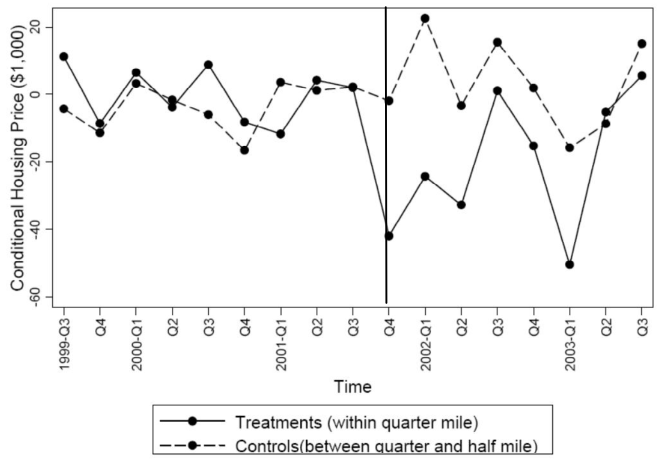


Figure 3.3D: Quarterly Mean Price Trend in the 4-Year Window
(Conditional on housing characteristics, and mosque area dummies)

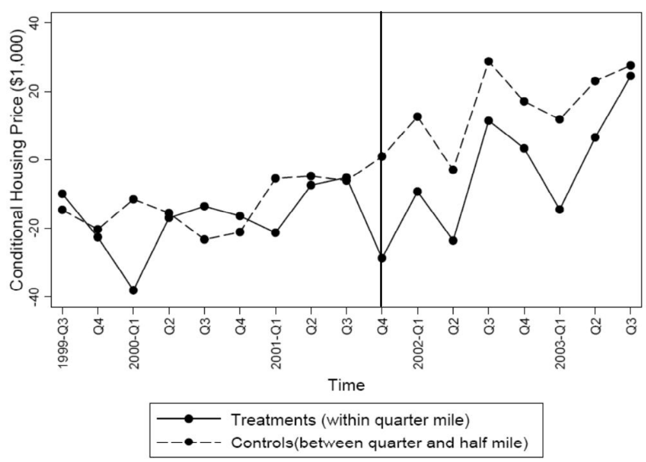
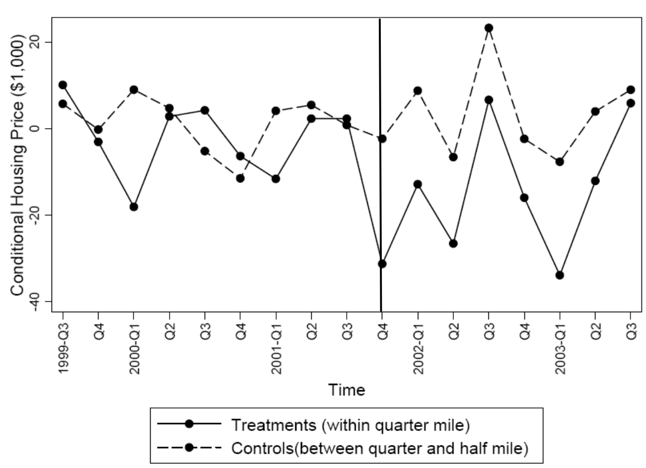


Figure 3.3E: Quarterly Mean Price Trend in the 4-Year Window
(Conditional on housing characteristics, year dummies and mosque area dummies)



In order to better capture the evolution of housing prices before and after 9/11, we create a series of additional figures that graph residuals from a series of regressions that condition out housing characteristics, year dummies, and mosque area dummies. As can be seen in figures 3.3B to 3.3E, the housing price trend for houses within a quarter of a mile of a mosque and those between a quarter and a half mile saw very similar price paths up until quarter 3 of 2001. Then beginning in quarter 4 of 2001 we see a steep drop in housing prices for the homes nearest the mosques. The divergence between the two price paths is quite striking and provides convincing graphical evidence that 9/11 may have caused housing prices very near mosques to decrease. Furthermore, it lends credibility that we have a reasonable control group in that the evolution of prices appears to be very similar prior to 9/11.

3.5.2 Cross-sectional Results

Table 3.5 provides cross-sectional evidence (based on equation 1) that we can provide adequate control for preexisting differences in the prices of properties located within a quarter of a mile of a mosque and those located between a quarter and a half mile of a mosque. These regressions use the same 4-year sample of east coast housing data that was used to create the figures in the previous section. Moving from column [1] to column [4] increases the set of controls that we impose on the data. As can be seen in column [4], once housing characteristics, year dummies, and mosque area dummies are added to the model, there is no longer a statistical difference between homes located within a quarter of a mile of a mosque and those located further away within a mosque area.

Table 3.5 Cross-Sectional Regression Results

	Dependent Variable = $ln\ (SP)$			
	[1]	[2]	[3]	[4]
$D^{1/4}$	-0.1868	-0.0828	-0.0981	-0.0342
	(0.0251)***[a]	(0.0181)***	(0.0304)**	(0.0260)
House characteristics[b]		x	x	x
Year dummies[c]	x	x	x	x
State dummies			x	
Mosque area dummies				x
S.E. clustered by…			state	mosque area
Adjusted R^2	0.0207	0.4975	0.6118[d]	0.7756
N	10385	10385	10385	10385

Note: [a] Standard errors are in parentheses. ***, **, * represent significance at 0.01, 0.05, 0.1 level, respectively. In column [3] and [4], robusted standard error is reported.

[b] House characteristics include square footage and its square term, lot size, age, no.of bedrooms, percentage of white population, and population density.

[c] Year dummies are created based on relative sale year rather than actual sale year. For example, houses sold between 09/11/1999 to 09/11/2000 receive a relative year equal to 1, so on and so forth.

[d] In column [3] and [4], R^2 is reported instead of adjusted R^2 as it is not available in clustered OLS estimation.

Table 3.6 GDID Regression Results (East)

	[1]	[2]	[3]	[4]	[5]
	4-year	4-year	4-year	6-year	2-year
$D^{1/4}$	-0.0291	0.0137	0.0065	0.0092	0.0003
	(0.0261)	(0.0341)	(0.0359)	(0.0350)	(0.0358)
$D^{1/4}$_ *post*	**-0.1053**	**-0.0837**	**-0.0708**	**-0.0707**	-0.0748
	(0.0384)***[a]	(0.0369)**	(0.0386)*	(0.0381)*	(0.0486)
$D^{1/2}$		-0.0193	-0.0144	-0.0159	-0.0225
		(0.0185)	(0.0191)	(0.0166)	(0.0218)
$D^{1/2}$_ *post*		0.0222	0.0161	0.0061	0.0439
		(0.0162)	(0.0161)	(0.0137)	(0.0278)
House characteristics	x	x	x	x	x
$X*T$[b]	x	x	x	x	x
Year dummies	x	x		x	x
Mosque area dummies		x		x	x
Year-by-mosque dummies			x		
S.E. clustered by…		mosque area	mosque area	mosque area	mosque area
R^2	0.4996	0.7768	0.7852	0.7990	0.7762
N	10385	10385	10385	17692	4006

Note: [a] Standard errors are in parentheses. ***, **, * represent significane at 0.01, 0.05, 0.1 level, respectively.

[b] As shown in equation (2), $X*T$ represents interaction terms between every property characteristics with a time dummy T.

3.5.3 GDID Results

Table 3.6 shows our GDID results. Column [1] in Table 3.6 reports estimates from a simple pre-post comparison for the treatment group (equation 2 without the dummy variable for properties sold between a quarter and a half miles from the mosque) using the 4-year sample. The estimates suggest that homes located within a quarter miles of a mosque sold for approximately 11 percent less, on average, than homes further away *after* the 9/11 attacks, but approximately 3 percent less on average *prior* to the 9/11 attacks. This 11 percent decline in price is statistically significant at 0.01 level while the 3 percent difference is not significant.

Columns [2] and [3] in Table 3.6 present the primary GDID results (full specification of equation 2) that exploit a pre-post comparison for both the treatment and control group. The specification in column [2] includes year dummies and mosque area dummies separately and the standard errors are clustered by mosque area. With this specification our estimate of the impact of the 9/11 attacks on properties near mosques is slightly lower at about 8.4 percent, but it is still statistically significant at 0.05 level. The estimated change in value for a home located between a quarter and a half mile of a mosque after the 9/11 attacks is positive (approximately 2 percent) but statistically insignificant. Thus households living just slightly farther away from the mosque (between a quarter and a half mile) experienced no decrease in property values on average. Column [3] presents the results from a similar regression but this time including year-by-mosque fixed effects that provide the most flexible time-space control. These results are consistent with those in column [2] with a slightly lower impact of 7.1 percent, confirming that the price of houses nearest to mosques (within a quarter mile) experienced a price decline after the 9/11 attacks while households living just slightly farther away from the mosque (between a quarter and a half mile) experienced no decrease in property values on average.

Columns [4] and [5] provide a robustness check on the sensitivity of our results to the temporal cutoffs we have imposed on the data before and after 9/11. We create two additional sample cutoffs of 1 year before and 1 year after (2-year window) and 3 years before and 3 years after (6-year window). Both of these regressions also include year dummies and mosque area dummies separately. For our 6-year sample, the price reduction in the treated area is statistically significant and nearly the same in magnitude. The 2-year window estimates the impact of 9/11 at approximately 7.5 percent but it is not significant at conventional levels.

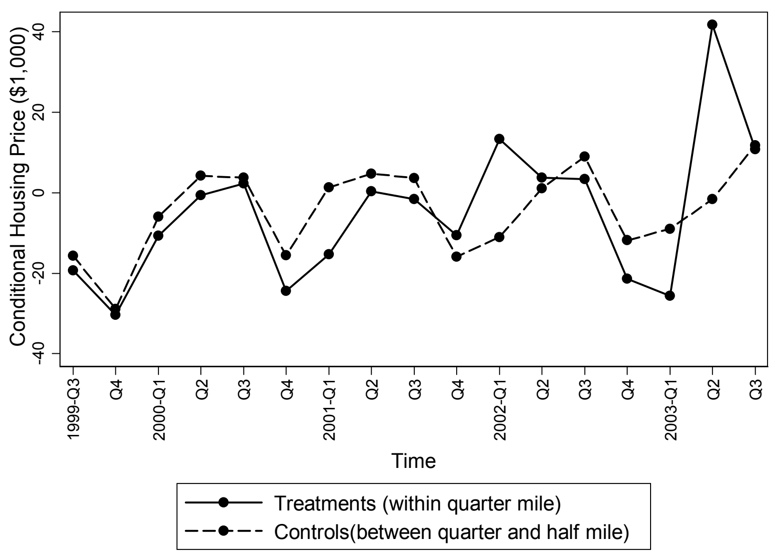
Table 3.7 Generalized DID Regression Results for West

	[1]	[2]	[3]
	West/4-year	West/6-year	West/2-year
$D^{1/4}$	-0.0181	-0.0102	-0.0082
	(0.0086)**	(0.0075)	(0.0123)
$\boldsymbol{D^{1/4}\_post}$	0.0140	0.0073	0.0215
	(0.0099)	(0.0070)	(0.0134)
$D^{1/2}$	-0.0166	-0.0171	-0.0124
	(0.0068)**	(0.0060)***	(0.0081)
$D^{1/2}\_post$	-0.0035	-0.0001	-0.0088
	(0.0060)	(0.0049)	(0.0083)
House characteristics	x	x	x
$X*T$	x	x	x
Year dummies			
Mosque area dummies			
Year-by-mosque dummies	x	x	x
S.E. clustered by…	mosque area	mosque area	mosque area
R^2	0.8598	0.8654	0.8585
N	39948	63247	17948

3.5.4 West Coast results

We now turn our attention to the analysis of the data we have for the housing market on the west coast of the United States (California, Oregon, Nevada and Washington). An analogous GDID analysis was performed using mosque and housing data for the west coast and the results are shown in Table 3.7. In column [1] it can be seen that the regression results suggest that 9/11 *increased* housing prices near mosques by approximately 1.4% but the increase is not statistically significant. Looking at the corresponding graphical evidence in Figure 3.4, it is clear that there is not a sharp break after 9/11 like we saw for the east coast sample, which supports our GDID regression result. Furthermore, when we enlarge or narrow the temporal window for our analysis, the regression result is still positive and not statistically significant. Thus the evidence suggests that 9/11 had no impact on housing prices near mosques on the west coast of the U.S.[14] The implications of these differences between the west and east coasts is discussed more fully in the conclusions below.

Figure 3.4: Quarterly Mean Price Trend in the 4-Year Window (West Coast)
(Conditional on housing characteristics, year dummies and mosque area dummies)



3.6 Internal Meta-Analysis

Our estimates in the preceding sections reflect our maintained assumptions about the market extent as well as additional modeling choices. First, by estimating a single

[14] We also analyzed the remaining data between the east coast and west coast states for which we had more limited coverage, and also found no effect of 9/11 on housing prices near mosques.

hedonic price function for these eastern states we implicitly treat the entire eastern area as one housing market. It may be the case, however, that different metro areas on the east coast have very different hedonic price functions. Another modeling choice is that we chose one quarter of a –mile as our cutoff distance to differentiate the treatment group (within a quarter mile) from the control group (within a quarter mile to a half mile). The reason for this choice is that the housing stock is relatively homogenous for these two groups prior to 9/11. However, we could have used one third mile instead. Table 3.8 shows that the housing stock is also relatively homogenous across treatment and control groups before 9/11 for this cutoff distance. Therefore, the choice of cutoff distance seems arbitrary and could use further investigation. Third, another subjective choice is the temporal window. We use a 4-year sample for our baseline results and use 6-year and 2-year samples as a type of robustness check.

Table 3.8 Housing Sales Summary for Controls and Treatments before 9/11

	Within 1/3 miles of mosque	Between 1/3 and 2/3 miles of mosque
	Mean (S.E.)	Mean (S.E.)
SP	119853.40	136786.30
	(95778.95)	(107326.50)
SQFT	1605.17	1636.28
	(728.46)	(728.84)
LOT	0.22	0.24
	(0.22)	(0.27)
AGE	34.30	33.51
	(31.36)	(28.72)
BR	3.15	3.11
	(0.83)	(0.81)
BATH	1.75	1.81
	(0.76)	(0.84)
WHITE	0.48	0.58
	(0.32)	(0.31)
POPD	0.05	0.05
	(0.04)	(0.04)
N	629	1637

Table 3.9 Summary Statistics of Mosque Areas Used in Meta-Analysis

	Total No. of Mosque Areas[b]	No. of Mosque Areas with Price Effect[a]	Price Effect			
			Mean	(S.E.)	Min	Max
East						
4-year sample and 1/4mile cutoff	17	3	-0.35	0.06	-0.41	-0.30
4-year sample and 1/3mile cutoff	19	3	-0.26	0.10	-0.35	-0.16
6-year sample and 1/4mile cutoff	29	5	-0.36	0.15	-0.57	-0.22
6-year sample and 1/3mile cutoff	30	5	-0.22	0.10	-0.34	-0.11
2-year sample and 1/4mile cutoff	2	0	N/A	N/A	N/A	N/A
2-year sample and 1/3mile cutoff	2	0	N/A	N/A	N/A	N/A
West						
4-year sample and 1/4mile cutoff	78	3	-0.24	0.08	-0.32	-0.15
4-year sample and 1/3mile cutoff	80	1	-0.09	N/A	-0.09	-0.09
6-year sample and 1/4mile cutoff	100	2	-0.21	0.03	-0.23	-0.19
6-year sample and 1/3mile cutoff	102	1	-0.08	N/A	-0.08	-0.08
2-year sample and 1/4mile cutoff	34	0	N/A	N/A	N/A	N/A
2-year sample and 1/3mile cutoff	35	1	-0.17	N/A	-0.17	-0.17
Middle						
4-year sample and 1/4mile cutoff	14	2	-0.46	0.50	-0.82	-0.11
4-year sample and 1/3mile cutoff	15	0	N/A	N/A	N/A	N/A
6-year sample and 1/4mile cutoff	25	2	-0.51	0.22	-0.67	-0.35
6-year sample and 1/3mile cutoff	26	0	N/A	N/A	N/A	N/A
2-year sample and 1/4mile cutoff	5	0	N/A	N/A	N/A	N/A
2-year sample and 1/3mile cutoff	5	2	-0.18	0.01	-0.19	-0.17
N of obs. used in meta-analysis	618	30	-0.28	0.17	-0.82	-0.08

Note: [a] "price effect" means the coefficient of $D^{1/4}\_post$ is significant and negative.

[b] Mosque areas with very low degree of freedom are dropped in meta-analysis according to the rule: "OLS linear multiple regression with only 1 dependent and 1 independent normally requires a minimum of 30 observations. A good rule of thumb is to add at least an additional 10 observations for each independent variable added to the equation."

In order to further investigate these modeling choices, we conduct an "internal meta-analysis" following the work of Banzhaf and Smith (2007). The internal meta-analysis proceeds as follows. We have three large study areas (east coastal states, west coastal states, and middle states) and two modeling choices: the temporal window (2-year, 4-year, and 6-year), and the cutoff distance (1/3 mile and 1/4 miles). Therefore, there are 3*3*2=18 groups in total. We run a GDID regression for each mosque area (houses within 1 mile of a mosque) for each of the 18 groups to obtain 618 estimates in total on the key coefficient $D^{1/4}$*_post* (or $D^{1/3}$*_post*). These 618 estimates will be the observations in our meta-analysis. In Table 3.9 we summarize the total number of mosques areas in each group, the number of mosque areas that have significant negative price effects for their individual GDID regression and some summary statistics for these price effects. For example, in our 4-year sample with one quarter mile as the cutoff (first row under EAST category), there are 17 mosque areas with only 3 out of the 17 mosque areas having a significant negative price effect. The mean price effect is approximately 35%. Thus it is clear that our average results that we reported for the combined analysis are highly influenced by a smaller subset of the mosques.

We then use a Probit model to see if differences in the mosque areas and our modeling choices can help explain differences in our estimates.[15] The Probit model is used to examine what factors affect the likelihood that we detect a significant price decrease in treated property values in a given mosque area. There are four groupings of independent variables that are included in the Probit model: (1) mosque ratings and average demographics of the mosque area such as rating of a mosque, average percentage of

[15] We also used a Heckman 2-stage to examine what factors affect the magnitude of a significant decrease in treated property values, but found nothing of significance.

white population, average percentage of owner-occupied houses, average population density; (2) our modeling choices reflected in dummies: cutoff choice (1/4 or 1/3 mile), temporal window choice (2year, 4year, or 6year sample) (3) spatial dummies such as whether or not the mosque area is in east coastal states, west coastal states or middle states. State dummies are also included; and (4) some statistics from the hedonic regression such as the number of observations and R^2. We include R^2 into the Probit model in order to exam how the fit of the GDID regression influences the likelihood that 9/11 had a price effect. The description of these variables and their summary statistics are provided in Table 3.10 where it can be seen that there is substantial variation across mosque areas: the average percentage of white population ranges from 0.01 to 0.96; the average population density ranges from 0.01 to 0.2; and the number of observations ranges from 224 to 1258. Some mosque areas do not have enough data for some of the GDID regressions ($n<k$) and therefore are dropped from the meta-analysis.

As shown in Table 3.11, the Probit results reveal that the average percentage of white population is positively related to obtaining a negative and statistically significant coefficient on $D^{1/4}$*_post* (or $D^{1/3}$*_post*). Average population density is also positively related, i.e., houses located in more highly populated areas tend to experience price depression after 9/11. The east coast dummy is also positively related, confirming our baseline results that there is a negative price effect for states along the east coast, but not for west coast states. While these results may be suggestive of anti-Islam or anti-Muslim attitudes potentially of whites in populated areas in the east, it is far from conclusive about whether or not this is the social mechanism that is leading to the price impacts we find.

Table 3.10 Meta-Analysis Variables and Summary Statistics

Name	Description	Mean	S.E.	Min	Max
Dependnet Variable					
POSSIBILITY	= 1 if estimated coefficient on $D^{1/4}\_post$ (or $D^{1/3}\_post$) is significant and negative	0.05	0.22	0	1
MAGNITUDE	*POSSIBILITY* * estimated coefficient on $D^{1/4}\_post$ (or $D^{1/3}\_post$)	-0.01	0.07	-0.82	0
Independent Variables					
RATING 2	= 1 if the rating of the mosque is 2	0.29	0.46	0	1
RATING 3	= 2 if the rating of the mosque is 3	0.24	0.43	0	1
RATING 4	= 3 if the rating of the mosque is 4	0.13	0.34	0	1
AVGWHITE	Average percentage of white population	0.57	0.24	0.01	0.96
AVGOWNER	Average percentage of owner-occupied houses	0.59	0.14	0.23	0.94
AVGPOPD	Average population density	0.06	0.03	0.01	0.19
NUMOBS	Number of observations in the generalized DID regression	464	201	**224**	**1258**
R2	R^2 from generalized DID regression	0.64	0.16	0.18	0.90
2YEAR	= 1 if 2-year sample are used	0.13	0.34	0	1
6YEAR	= 1 if 6-year sample are used	0.50	0.50	0	1
THIRDCUT	= 1 if cutoff point is 1/3 mile	0.51	0.50	0	1
EAST	= 1 if mosque area is in east coastal states	0.16	0.37	0	1
WEST	= 1 if mosque area is in middle states	0.69	0.46	0	1
STATEFIX	State dummies		N/A		

Table 3.11 Probit Results (Dependent Variable : *POSSIBILITY*)

Variable	Coefficient
RATING 2	-0.268
	(0.174)
RATING 3	-0.198
	(0.158)
RATING 4	-0.108
	(0.183)
AVGWHITE	**0.889*****
	(0.302)
AVGOWNER	0.077
	(0.559)
AVGPOPD	**9.001*****
	(2.341)
NUMOBS	-0.000
	(0.000)
R2	-0.366
	(0.408)
2YEAR	0.138
	(0.152)
6YEAR	-0.125
	(0.153)
THIRDCUT	-0.142
	(0.117)
EAST	**1.674*****
	(0.560)
MID	0.360
	(0.615)
STATEFIX	Yes
Constant	-2.845***
	(0.643)
N	618
Log Likelihood	-276.469
LR statistic	98.010
Pseudo R^2	0.151

note: *** p<0.01, ** p<0.05, * p<0.1

3.7 Conclusion

The terrorist attacks by Islamic extremists that occurred on September 11, 2001 (9/11) had a variety of societal and economic impacts on the United States. In this paper we have focused on a previously unexplored potential impact of 9/11—the impact it may have had on housing prices near mosques. Using a unique dataset that combines the locations of functioning mosques with housing transactions near the time of 9/11, combined with a generalized difference-in-differences framework, we are able to investigate the impact of 9/11 on housing prices near mosques in the United States.

We think there are several interesting results from our analysis. First, we find that housing prices decreased by approximately 7% ($10,559 for the average home) in areas near mosques along the east coast of the U.S. in the two years following 9/11 on average. However, on the west coast we find no evidence that 9/11 caused decreases in housing prices near mosques. Furthermore, the price effects we find in the east are largely driven by a relatively small percentage of the mosques. Thus the overall evidence from our analysis suggests that although it appears there was a causal impact of 9/11 on housing prices near some mosques, especially in the east, it clearly did not have a significant impact for many areas near mosques, and this is especially true in the west. The internal meta-analysis that we conducted also suggests that significant housing price impacts were more likely after 9/11 in more densely populated areas with a higher concentration of white households.

We think our results provide two important contributions to the literature. First, it provides the first evidence on the impact that 9/11 had on housing prices near mosques. Second, our results add to the literature on how people react and adapt to the threat of terrorism. The economic effect of terrorism such as 9/11 and the behavioral responses they elicit are important to society and policymakers. Empirical assessments of these behavioral responses are crucial to understanding the appropriate remedies and economic policy responses to terrorist activities. Our results indicate that many households preferred not to live near mosques along the east coast of the U.S. after 9/11. Nonetheless, our analysis is unable to determine what the mechanism was that drove the decreased demand for houses near mosques in some areas. Acquiring additional data that would enable one to distinguish between the hypothesized mechanisms would be an interesting avenue for future research.

References

Abadie, A., Dermisi, S., 2008. Is terrorism eroding agglomeration economies in Central Business District? Lessons from the office real estate market in downtown Chicago. Journal of Urban Economics 64(2), 451-463.

Banzhaf, H.S., Smith, V.K., 2007. Meta-analysis in model implementation: choice sets and the valuation of air quality improvements. Journal of Applied Econometrics.22(6), 1013-1031.

Blake, A., Sinclair, M.T., 2002. Tourism Crisis Management: US Response to September 11. 30(4), 813-832.

Black, S.E., 1999. Do better schools matter? Parental valuation of elementary education. Quarterly Journal of Economics. 114(2), 577-599.

Carter, D.A., Simkins, B.J. 2004. The market's reaction to unexpected, catastrophic events: the case of airline stock returns and the September 11th attacks. The Quarterly Review of Economics and finance. 44(4), 539-558.

Chay, K.Y., Greenstone, M., 2005. Does air quality matter? Evidence from the housing market. Journal of Political Economy 113(2), 376-424.

Davis, D.W., 2004. The effect of health risk on housing values: evidence from a cancer cluster. The American Economic Review 94(5), 1693-1704.

Ericson, R., Doyle, A., 2004. Catastrophe risk, insurance and terrorism. Economy and Society 33(2), 135-173.

Frey, B.S., Luechinger, S., Stutzer, A., 2007. Calculating tragedy: assessing the costs of terrorism. Journal of Economic Surveys 21(1), 1-24.

Gautier, P.A., Siegmann, A., Van Vuuren, A., 2009. Terrorism and attitudes towards minorities: the effect of the Theo van Gogh murder on house prices in Amsterdam. Journal of Urban Economics 65(2), 113-126.

Goodrich, J.N., 2002. September 11, 2001 attack on America: a record of the immediate impacts and reactions in the USA travel and tourism industry. Tourism Management 23(6), 573-580.

Hobijn, B., 2002. What Will Homeland Security Cost? Economic Policy Review 8(2), 21-23.

Horsch, E.J., Lewis D.J., 2009. The effects of aquatic invasive species on property values: evidence from a quasi-experiment. Land Economics 85(3), 391-409.

Ito, H., Lee D., 2005. Assessing the impact of the September 11 terrorist attacks on U.S. airline demand. Journal of Economics and Business 57(1), 75-95.

Krugmen, P., 2004. The Nexus of terrorism & WMDs: developing a consensus. Breiefing Note. Princeton University.

Kuminoff, N.V., Parmeter, C.F., Pope, J.C., 2010. Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities. Journal of Environmental Economics and Management 60 (3): 145-160.

Kuminoff, N.V., Pope, J.C., 2012. Do "capitalization effects" for public goods reveal the public's willingness to pay? Working Paper.

Linden, L., Rockoff, J.E., 2008. Estimates of the impact of crime risk on property values from Megan's law. American Economic Review 98(3), 1103-1127.

Makinen, G., 2002. The economic effects of 9/11: a retrospective assessment. Congressional Research Service. pp. CRS–5. http://www.fas.org/irp/crs/RL31617.pdf.

Parmeter, C.F., Pope, J.C., 2009. Quasi-experiments and hedonic property value methods. Handbook on Experimental Economics and the Environment, Edward Elgar Publishers.

Pope, J.C., 2008a. Buyer information and the hedonic: the impact of a seller disclosure on the implicit price for airport noise 63(2), 498-516.

Pope, J.C., 2008b. Fear of crime and housing prices: household reactions to sex offender registries. Journal of Urban Economics 64(3), 601-614.

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

Saxton, J., 2002. The economic costs of terrorism. Joint Economic Committee, United States Congress. (http://www.house.gov/jec/, last accessed on August 11th,2010)

Sheridan, L.P., 2006. Islamophobia pre- and post- September 11th, 2001. Journal of Interpersonal Violence 21(3), 317-336.

Chapter 4

Partial Identification of Hedonic Demand Functions

ABSTRACT

This paper proposes a partial identification strategy identifying consistent bounds on demand function parameters based on the concept of "imperfect instrument (IIV)" developed by Nevo and Rosen (2010). Furthermore, I extend their analysis to demonstrate that partially identified demand functions can be used to bound welfare measures for non-marginal changes. I use the new methodology to measure the benefits from improvements in lakewater quality.

This paper contributes to the hedonic literature in three ways. First, it provides an identification strategy for hedonic demand estimation without the need for assumptions on instrument validity that are difficult to justify; Second, it demonstrates how partial identification of parameters leads to partial identification of welfare measures. Third, it provides new evidence on the WTP for non-marginal improvements in lake water quality.

KEYWORDS: Hedonic Demand Estimation; Endogeneity Problem; Partial Identification; Imperfect Instrument; Consumer Surplus; WTP for Lake-water Quality

4.1 Introduction

Hedonic theory has been applied to differentiated goods markets as varied as housing markets, labor markets, child care services, agricultural products, and computers. Most often, these applications are "first-stage" analyses, estimating the hedonic price function, but not underlying demands for characteristics of goods. There are two limitations of this approach: One is that the welfare measurement is limited to marginal changes. The changes in which policy makers are interested, however, are unlikely to be marginal. The second limitation is that measures of marginal willingness to pay are limited to a particular market. They cannot be transferred across markets. Therefore, developing a "second-stage" approach to estimate the demand function is especially important to policy evaluation.

The purpose of this essay is to propose a solution to the endogeneity problem that arises in hedonic demand estimation. Bartik(1987) and Epple(1987) demonstrate that the nonlinearity of the hedonic price function invalidates instruments based on the usual within-market supply shifters. Instead, Bartik(1987) suggests using data from multiple markets with market dummies as instruments. This essay and some other studies (Bishop and Timmins, 2008, 2011) point out that sorting behavior across markets invalidates market dummies. Over of the past decade, researchers have sought to address this endogeneity problem by adding more information to the model. One strategy is to add assumptions about the shape of utility functions (e.g., Driscoll et al., 1994; Chattopadhyay, 1999; Bajari and Benkard, 2005; Bajari and Kahn, 2005; Sieg et al., 2002; Sieg et al., 2004; Smith et al., 2004; Klaiber and Phaneuf, 2010). Another strategy is to add more data. For example, Kuminoff and Pope (2012) demonstrate that demand curves can be identified using repeated cross-section data on the same geographic market before and after an unexpected shock. A third strategy is to combine the first two: provide more structure and more data. Bishop and Timmins (2008) track the same individuals over time and space. They also write down a parametric specification for an individual's indirect utility function.

Instead of providing more structure or data, I propose a fundamentally different way to extract useful information about demand, returning to the simple reduced-from approach, using only data that are widely available at a low cost. The key observation I make is that what we call "the endogeneity problem" in hedonic demand estimation is only a problem if we limit ourselves to the extremes of point identification, in which identification is viewed as an all-or-nothing concept. If we take a broader perspective on identification, we can partially identify demand functions and welfare measures.

Specifically, I propose a partial identification strategy identifying consistent bounds on Marshallian consumer surplus based on the concept of "imperfect instrument (IIV)" developed by Nevo and Rosen (2010) (NR thereafter). Furthermore, I extend their analysis to demonstrate that partially identified demand curves can be used to bound welfare measures for non-marginal changes. Marshallian Consumer Surplus (*MCS*) can be single bounded if the key parameter is single bounded and *MCS* can be either double bounded or point identified if the key parameter is double bounded.

Overall, this study contributes to the hedonic literature in three ways. First, it provides an identification strategy for hedonic demand estimation without the need for assumptions on instrument validity that are difficult to justify. Second, it demonstrates how partial identification of parameters leads to partial identification of welfare measures.

Third, it provides an empirical demonstration to measure the benefits from improvements in lake water quality.

The essay proceeds as follows: Section 4.2 illustrates the endogeneity problem in hedonic demand estimation and explains why market dummies may not be valid instruments. Section 4.3 derives a linear demand function and applies NR's partial identification strategy to define sufficient conditions for the identification of single-sided or double-sided bounds on the key parameter. Section 4.4 then discusses bounds on *MCS*. Section 4.5 provides a demonstration to water quality and section 4.6 concludes.

4.2 Motivation

4.2.1 The endogeneity problem with hedonic demand estimation

In order to illustrate the endogeneity problem with estimating the demand function for a particular attribute of a differentiated good, we begin from the hedonic price function and the consumer's utility maximization problem.

Let $\mathbf{x}$ represent a differentiated good with K characteristics $\mathbf{x} = [x_1, x_2, ..., x_K]$. The differentiated good is assumed to be sold in a perfectly competitive market m and the consumer/supplier interactions determine an equilibrium price schedule for the good in that market, $p^m(\mathbf{x})$. As such, any single consumer or supplier takes $p^m(\mathbf{x})$ as exogenous[1].

Consumers obtain utility from two goods: the differentiated good, $\mathbf{x}$, and a composite numeraire good, b (representing income left over after purchasing $\mathbf{x}$). A consumer with income y and a vector of preference parameters α faces the utility maximization problem:

[1] Marginal prices may not be constant resulting from the fact that consumers are unable to "repackage" the differentiated goods.

$$
\begin{aligned}
&\max_{\mathbf{x},b} \quad U(\mathbf{x},b;\alpha) \\
&s.t. \quad y = p^m(\mathbf{x}) + b
\end{aligned}
\tag{2.1}
$$

The consumer is assumed to purchase only one unit of the differentiated good and he seeks to maximize utility by choosing $\mathbf{x}$ and b subject to his budget constraint. The first order condition to this utility maximization problem is expressed as:

$$
\frac{\partial p^m(\mathbf{x})}{\partial x_k} = \frac{\partial U(\mathbf{x},b;\alpha)/\partial x_k}{\partial U(\mathbf{x},b;\alpha)/\partial b}
\tag{2.2}
$$

The consumer chooses a level of x_k to set its implicit price equal to his marginal willingness to pay (MWTP). If we further assume that the marginal utility of income is constant, then the RHS of equation (2.2) traces out the consumer's MWTP function (or compensated demand function) for x_k as x_k varies, holding $\mathbf{x}_{-k}$ constant.

When one wants to estimate the compensated demand function, there are two equations to consider:

$$
p^m_{x_k} = \frac{\partial p^m(\mathbf{x})}{\partial x_k}
\tag{2.3}
$$

$$
\frac{\partial U(\mathbf{x},b;\alpha)/\partial x_k}{\partial U(\mathbf{x},b;\alpha)/\partial b} = D(x_k;\mathbf{x}_{-k},\alpha)
\tag{2.4}
$$

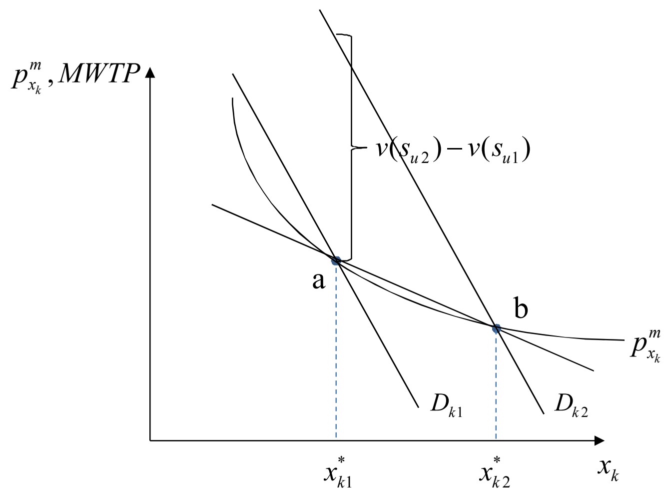
Equation (2.3) states that the marginal price function of x_k is a function of itself and other characteristics of the differentiated good. Equation (2.4) states that the consumer's MWTP for x_k is a function of itself, other characteristics of the differentiated good (x_k's substitutions and complements) and the consumer's attributes (demand shifters). An econometric approximation to the demand function in (2.4) can be expressed as:

$$\frac{\partial p^m(\mathbf{x})}{\partial x_k} = \hat{D}(x_k; \mathbf{x}_{-k}, \mathbf{s}_o, \delta) + v(\mathbf{s}_u) \tag{2.5}$$

where δ is a vector of parameters; the consumer's preference α is decomposed into observed and unobserved parts $(\boldsymbol{s}_o, \boldsymbol{s}_u)$. For example, $\mathbf{s}_o$ may be a small set of observable demographic characteristics such as age, income, gender, and race; and $\mathbf{s}_u$ may represent unobserved elements of genetics and past experiences that influence current preferences. The econometric error term depends on the unobserved part: $v(\mathbf{s}_u)$. The difficulty with estimation is that, in general, the (observed) marginal price function (2.3) only intersects the (unobserved) demand function (2.4) at a single point. Thus, $\hat{D}(x_k; \mathbf{x}_{-k}, \mathbf{s}_o, \delta)$ cannot be identified from OLS estimation of the hedonic price function for a single geographic market.

The source of bias can be seen from Figure 4.1 which is essentially reproduced from Bartik (1987). Consider two households identical in all observed variables ($\mathbf{s}_o$ and their choices for $\mathbf{x}_{-k}$). Consumer 2 has a greater unobserved taste for x_k ($v(\mathbf{s}_{u2})$) and thus a demand function for x_k that is greater by $v(\mathbf{s}_{u2}) - v(\mathbf{s}_{u1})$. Because consumer 2 has a greater taste for x_k, he chooses a higher level of x_k ($x_{k2}^* > x_{k1}^*$). In the figure, points *a* and *b* are the only points observed on consumer 1 and 2's demand curves. Both of them are on the implicit price function in market m. An OLS regression of marginal prices $\hat{p}_{x_k}^m$ on x_k will result in a straight line through a and b, which has a flatter slope than the true demand functions D_{k1} and D_{k2}. Hence, the positive correlation between unobserved tastes $v(\mathbf{s}_u)$ and x_k leads to a positive bias in OLS estimates of the slope of the demand function[2].

Figure 4.1 The Endogeneity Problem



4.2.2 Market dummies as imperfect instruments

Developing instruments to solve the endogeneity problem is a difficult task. In order to estimate the MWTP function in equation (2.4), the exclusion restriction for IV estimation requires that the instrument shifts the implicit price function (2.3) without shifting the demand function (2.4). Perhaps the most widely used strategy is to develop instruments based on indicators for distinct geographic markets (Palmquist, 1984; Bartik, 1987).

Recall that the m superscript in (2.3) indicates that the implicit price function is specific to geographic market m . Spatial variation in the conditions of supply and demand may lead to equilibrium price functions that vary from market to market. If we

[2] See Epple(1987) for a characterization of the identification problem in a more general simultaneous equations framework.

can observe the same "type" of consumer in multiple markets, then variation in the shape of the hedonic price function will trace out multiple points on the demand curve for that consumer type. If the market dummies are valid instruments, then Bartik(1987) observes that interactions between market dummies and demand shifters will also be valid instruments.

Unfortunately, the prospect of using market dummies as instruments is complicated by the possibility that consumers may sort themselves across markets. In an extreme case we would never observe the same "type" of consumer in two different markets because they would be fully stratified across markets according to their preferences. This is the basic logic behind empirical models of Tiebout sorting (Tiebout, 1956)[3].

The geographic extent of "the market" remains an open question, with different researchers maintaining different assumptions about the extent of free mobility. For example, in the literature on using hedonic property value models to estimate the marginal willingness to pay for non-market goods and service, the definitions for markets vary from as small as a single city to nation-wide. Hedonic demand studies often define markets as distinct cities: Zabel and Kiel (2000) define 4 cities, Chicago, Denver, Philadelphia, and Washington D.C., as different markets; Palmquist (1984) uses 7 cities, Atlanta, Denver, Houston, Louisville, Miami, Oklahoma City, and Seattle, as separate markets; Cheshire and Sheppard (1998) uses two cities in England, Reading and Darlington, as separate markets; and Witte, Sumka, and Erekson (1979) uses four cities in North Carolina: Greenville, Kinston, Lexington, and Statesville as separate markets.

[3] This sorting process says that consumers with different preference for some attributes will sort themselves to markets satisfying their preference. The Tiebout sorting models of Epple and a variety of co-authors exploited such relationship existing in the hedonic equilibrium. For example, Epple (1987), Epple, Filimon, and Romer (1984), Epple and Romano (1988), Epple and Platt (1988) and Epple and Sieg (1999).

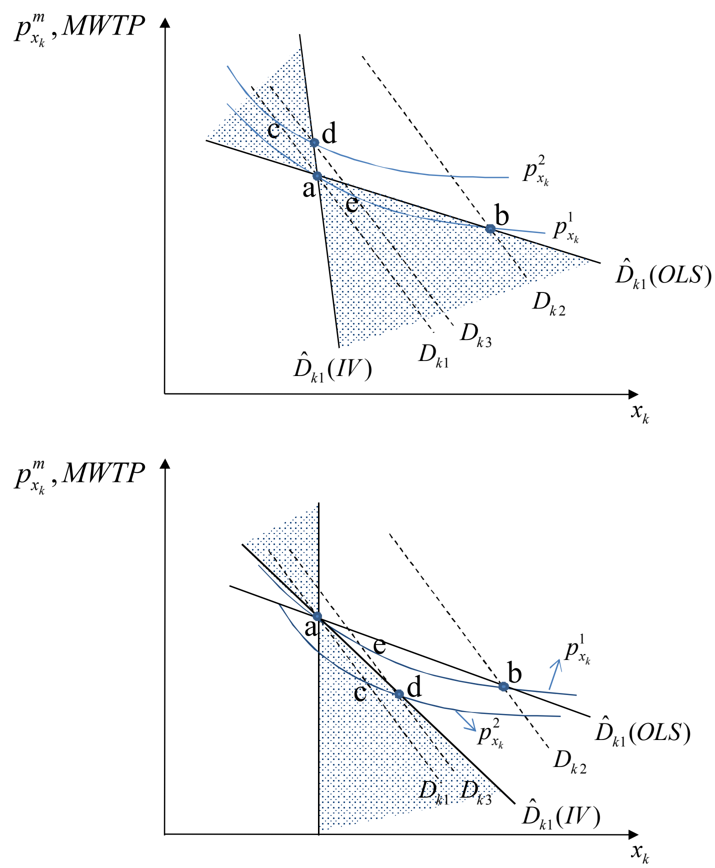
Although all of these papers treat a single city as a market, they make different implicit assumptions about moving costs. The minimum driving distance between any two cities used in these studies ranges from 50 minutes in Witte, Sumka, and Erekson (1979) (Greenville to Kinston in NC) to about 7 hours in Palmquist (1984) (Oklahoma City to Houston)! Other authors make the seemingly heroic assumptions that the entire U.S. is a single market and there is zero cost of moving anywhere in the country. The possibility of assuming a national housing market has been advocated by Linneman (1980) and maintained by Linneman (1981), Cobb (1984), Smith and Deyak (1975), Deyak and Smith (1974) and Chay and Greenstone (2005).

If the markets are not truly separated, market dummies will be invalid instruments and therefore IV regression will be biased. Nevertheless, it may still be possible to identify bounds on the true demand curve. Figure 4.2 shows the consequence with using market dummies as imperfect instruments. Without loss of generality, we assume there are two markets with different marginal price functions. Some consumers are assumed to be free to move between markets. A market dummy is defined as equal to 0 when the consumer is in market 1 and equal to 1 when the consumer is in market 2.

In case 1 (upper panel), consumer 1's optimal choice is a in market 1 and c in market 2. Assuming he is free to move between markets, he can choose either a or c, but we observe that he chooses a. If we can also observe the same type of consumer locating in market 2, then IV regression with a market dummy as a "perfect" instrument will connect point a and c and correctly estimate consumer 1's true demand curve. However, we cannot observe c due to the Tibout sorting process where consumers would be stratified according to their preferences. Instead, we observe a "similar" type of consumer,

consumer 3, who has a higher unobserved preference than consumer 1 (D_{k3} is higher than D_{k1}). Consumer 3 can also choose either e in market 1 or d in market 2, but we observe that he chooses d located in market 2. Due to Tibout sorting process, therefore, the market dummy becomes an "imperfect" instrument and IV regression will connect a and d, resulting in a negatively biased slope of consumer 1's demand curve. Therefore, the true slope of consumers' demand function is "double bounded" by OLS and IV estimates. The shaded area in the figure illustrates these bounds.

Figure 4.2 Partial Identification for Demand Function



In case 2 (lower panel), the marginal price function in market 2 is below the marginal price function in market 1. But the sorting process is the same: consumer 1 chooses a in market 1 and consumer 3 chooses d in market 2. IV regression with market dummy will connect a and d and result a positively biased demand curve in this case. But the IV estimate is “less” biased than OLS. Hence IV and OLS both provide upper bounds. The true slope of the consumers' demand function is “single bounded”. Of course, as long as we believe the demand curve is downward sloping, we have a trivial lower bound defined by a perfectly inelastic demand curve.

Whether the slope of the demand curve is single or double bounded depends on patterns of correlation between variables. There are two key correlations to highlight. First, notice that the market dummy and x_k share the same (positive) direction of correlation with unobserved preferences. Second, notice that the market dummy and x_k are negatively correlated in case 1 where we get double-sided bounds and positively correlated in case 2 where we get single-sided bounds. Knowledge of these correlations can be used to predict whether OLS and IV estimates will collectively define single sided or double sided bounds on demand curves. NR(2010) formalized this logic in the more general context of IV estimation for linear functions. We will utilize their strategies in the context of hedonic demand estimation and then extend the results to welfare analysis.

4.3 Partial Identification of Linear Demand Function

Most previous applications of hedonic demand estimation utilize linear, log-linear, and log-log models for their demand functions (see Zabel and Kiel, 2000; Ohsfeldt, 1988; Palmquist, 1984; Cheshire and Sheppard, 1998; Witte, Sumka, and Erekson, 1979; Bishop and Timmins, 2010; Epple, 1987). Recent results from NR(2010) allow us to partially identify parameters of these demand functions. In this section, we also assume a linear demand function and tailor their assumptions and propositions to the hedonic demand estimation context.

Let the empirical counterpart of the demand function in (2.5) be denoted:

$$p_x = x\beta + W\alpha + u, \tag{3.1}$$

where $x \equiv x_k$ represents the attribute of interest and p_x denotes the implicit price of x. W is a vector of empirical counterparts of $\mathbf{x}_{-k}$ and $\mathbf{s}_o$ (a constant may also be included in W). u represents the unobserved preference for x and we assume that $E(u) = 0$. Note that the only endogenous variable is x. The variables included in W are all assumed to be exogenous[4].

Suppose we have multiple instruments for x which denoted as $z_j (j = 1, 2, ..., J)$. Some important assumptions have to be made in order to partially identify the parameters in (3.1). Following NR(2010), we use σ_{xy} and ρ_{xy} to denote the covariance and correlation coefficient between x and y, and σ_x to denote the standard deviation of x:

Assumption NR1 (sampling process): The n observations of (p_x , x , W , z_j) are stationary and weakly dependent.

Assumption NR2 (W exogenous): $E(W'u) = \mathbf{0}$.

Assumption NR3 (same direction of correlation): $\rho_{xu}\rho_{z_j u} \geq 0, j = 1, 2, ..., J$.

Assumption NR1 is used to ensure the consistency of estimated bounds. Stationarity requires stability of the model over time. In the hedonic context, most studies use cross-

[4] In this paper, we focus on the case where demand function contains only one endogenous regressor. NR(2010) also discussed the multiple endogenous regressors case, it would be interesting to use their results to extend our arguments.

sectional data or pooled cross-sectional data in a relatively short time period. In the housing market, for example, stationarity will hold if there are no large fluctuations in housing prices, structures, neighborhood composition, and public goods. Weakly dependent indicates that as the variables get farther apart in time or location, the correlation between them becomes smaller. This assumption is likely to hold in housing market data as the correlation in housing prices and characteristics should be dissipating as time elapses.

Assumption NR2 says that the variables included in W are all exogenous. There are two components of W: x_{-k} represents all attributes other than the variable of interest. $\mathbf{s}_o$ represents consumers' observable demographic characteristics, which are believed to be correlated with preferences.

Assumption NR3 considerably weakens the maintained assumptions of usual hedonic IV studies which would require that $\rho_{z_j u} = 0$. Instead, it nests the usual condition as a special case and requires that the endogenous variable x and the instrument z have the same direction of correlation with the error term. NR label instruments satisfying this assumption "imperfect instrumental variables (IIV)"[5].

Because our main interest is to estimate the slope parameter, β, we transform the demand function (3.1) to a simple bivariate model. We regress both the independent variable p_x and the endogenous regressor x on the covariates, W, and obtain the

[5] In NR (2010), they add an additional assumption NR4 asserting that the IIV is less correlated with the error term than the endogenous variable: $|\rho_{xu}| \geq |\rho_{z_j u}|$. We drop this assumption due to two reasons: first, the partial identification can work only with assumption NR3, and NR4 only helps to narrow the bounds; second, in the demand function estimation context, it is difficult to justify whether or not our instruments meet NR4.

residuals from these regressions. Let $\tilde{p}_x$, $\tilde{x}$ denote the residuals. Using these residuals, the original linear estimation for β can be rewritten as:

$$\tilde{p}_x = \tilde{x}\beta + u. \tag{3.2}$$

Define β^{OLS} and $\beta^{IV}_{z_j}$ to be the probability limits of the standard OLS and IV estimators for β. They can be calculated as follows:

$$\beta^{OLS} = \frac{\sigma_{\tilde{x}\tilde{p}_x}}{\sigma^2_{\tilde{x}}} \tag{3.3}$$

$$\beta^{IV}_{z_j} = \frac{\sigma_{z_j\tilde{p}_x}}{\sigma_{z_j\tilde{x}}} \tag{3.4}$$

It can be shown that under assumptions NR1-NR3, the true value of the parameter, β, can be bounded by β^{OLS} and $\beta^{IV}_{z_j}$.

Proposition1. Let NR1-NR3 hold.

Case 1: If $\sigma_{\tilde{x}z_j} < 0$, then $\min\{\beta^{OLS}, \beta^{IV}_{z_j}\} \le \beta \le \max\{\beta^{OLS}, \beta^{IV}_{z_j}\}$.

Case 2: If $\sigma_{\tilde{x}z_j} > 0$ and $\sigma_{xu}, \sigma_{z_ju} \ge 0$, then $\beta \le \min\{\beta^{OLS}, \beta^{IV}_{z_j}\}$; if $\sigma_{\tilde{x}z_j} > 0$ and σ_{xu}, $\sigma_{z_ju} \le 0$, then $\beta \ge \max\{\beta^{OLS}, \beta^{IV}_{z_j}\}$.

Proof: See Appendix A.

Proposition 1 is modified slightly from Lemma 1 in NR(2010). They invoke the same assumptions (NR1-NR3) for the simple bivariate linear model. In contrast, we start from a multivariate model (3.1) and then apply the Frisch-Waugh-Lovell theorem to net out the effect of exogenous variables W, obtaining a modified bivariate model (3.2). Proposition 1 stated above is based on this modified bivariate model.

As we can see from Proposition 1, a potential drawback is that when $\sigma_{\tilde{x}z_j} > 0$ for each IIV, β is only single bounded. Double-sided bounds would be more informative. Fortunately, if we have two IIVs, z_1 and z_2, we can attempt to construct a new IIV that combines z_1 and z_2 in a way that provides double-sided bounds. The new IIV is constructed as a weighted difference of z_1 and z_2:

$$w(\gamma) = \gamma z_2 - (1-\gamma) z_1 \tag{3.5}$$

NR(2010) show that if we can find a $\gamma \in (0,1)$ such that $w(\gamma)$ satisfies the conditions in the first case in Proposition 1, then $w(\gamma)$ can be used as an instrument to obtain a double-sided bounds:

Proposition 2. Let NR1-NR3 hold.

If $\sigma_{\tilde{x}w(\gamma)} < 0$, and $\sigma_{xu}, \sigma_{w(\gamma)u} \geq 0$, then we have double-sided bounds for β:

$\beta^{IV}_{w(\gamma)} \leq \beta \leq \min\{\beta^{IV}_{z_1}, \beta^{IV}_{z_2}, \beta^{OLS}\}$.

where $\beta^{IV}_{w(\gamma)}$ is defined by using $w(\gamma)$ as an instrument in the definition given in equation (3.4).

Proof: See proof of Proposition 5 in NR(2010).

NR(2010) also provide a testable condition that guarantees the existence of a value for γ satisfying the first case in Proposition 1[6]:

$$\sigma_{z_1\tilde{p}_x}\sigma_{z_2\tilde{x}} < \sigma_{z_2\tilde{p}_x}\sigma_{z_1\tilde{x}} \tag{3.6}$$

[6] See NR (2010) Lemma 2.

This test is very helpful because all of the variables are observable. While we can test if there exists some value of γ that will yield double-sided bounds, we do not know what value γ should take. However, if we are willing to assert that one instrument is better than another in the sense that it is both more relevant ($\rho_{\tilde{x}z_1} > \rho_{\tilde{x}z_2}$) and more valid ($\rho_{z_1u} < \rho_{z_2u}$), then γ can take the value of $\frac{\sigma_{z_1}}{\sigma_{z_1} + \sigma_{z_2}}$ to ensure that $w(\gamma)$ satisfies the conditions in Proposition 2[7].

4.4 Partial Identification of Marshallian Consumer Surplus

In this section, we extend the analysis to welfare measurement. Specifically, we still consider the linear demand function and suppose a policy leads to a non-marginal change in x . Marshallian consumer surplus (*MCS*) is shown to be either single or double bounded depending on whether β is single or double bounded.

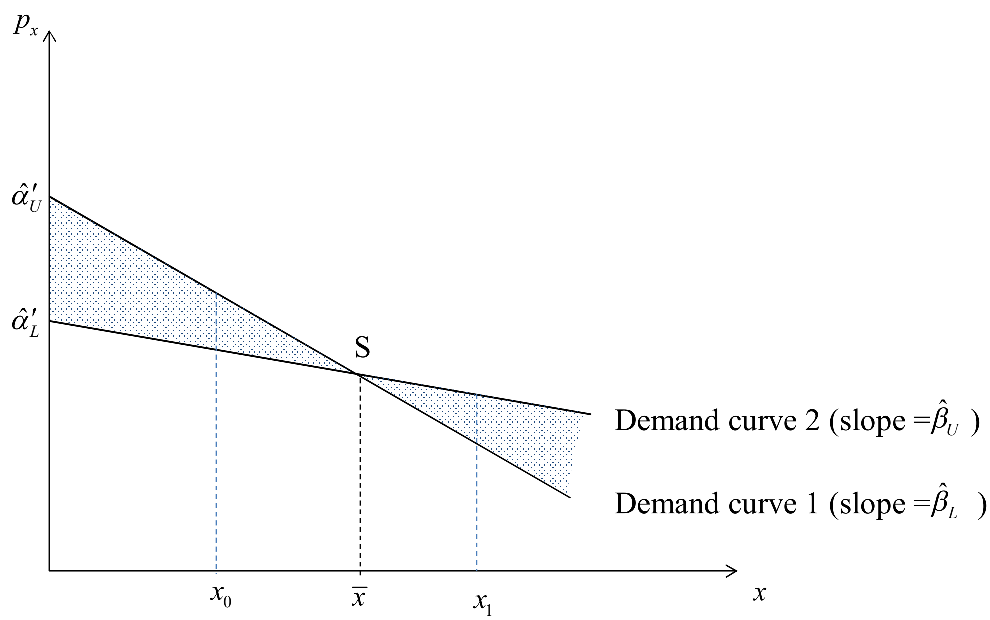
Denote the bounds on β as $\hat{\beta}_L \le \beta \le \hat{\beta}_U$. The corresponding bounds on $\alpha' \equiv W\alpha$ are given by $\overline{p}^x - \hat{\beta}_U \overline{x} \le \alpha' \le \overline{p}^x - \hat{\beta}_L \overline{x}$, where $\overline{p}^x$, $\overline{x}$ represent the average value of p^x , x respectively[8]. Note that $\hat{\alpha}'$ is a decreasing function of $\hat{\beta}$ since $\hat{\alpha}' = \overline{p}^x - \hat{\beta}\overline{x}$. Therefore, the upper and lower bounds on β give us two demand curves with different slopes and intercepts. Figure 4.3 provides an example. When β increases from $\hat{\beta}_L$ to $\hat{\beta}_U$, the demand curve rotates counterclockwise from demand curve 1 to demand curve 2. All of

[7] This statement is proved in Appendix A.

[8] An alternative method to derive bounds on α' is: first obtain the identification region for each parameter in vector α using Proposition 3 in NR (2010), then calculate the identification region for α' using bounds on α and mean value of each variable in W .

the potential demand curves are bounded by them and share a common intersection $S=(\overline{x},\overline{p}^x)=(\sum_{i=1}^{N}x_i,\sum_{i=1}^{N}p_i^x)$ (where i indexes observations in a sample)[9].

Figure 4.3 Bounds on Welfare Measure



Consider a public policy designed to produce a non-marginal change in x : x increases from x_0 to x_1. Facing demand curve 1, this non-marginal change yields a welfare gain equal to:

---

[9] To see this, pick two arbitrary demand curves with slope denoted by $\hat{\beta}'$ and $\hat{\beta}''$ satisfying $\hat{\beta}_L \le \hat{\beta}' \le \hat{\beta}'' \le \hat{\beta}_U$,

$$p^x = \hat{\alpha}' + \hat{\beta}' x$$
$$p^x = \hat{\alpha}'' + \hat{\beta}'' x$$

Solve the above equation system we get: $x=\dfrac{\hat{\alpha}''-\hat{\alpha}'}{\hat{\beta}'-\hat{\beta}''}$. Plug $\hat{\alpha}'=\overline{p}^x-\hat{\beta}'\overline{x}$ and $\hat{\alpha}''=\overline{p}^x-\hat{\beta}''\overline{x}$ in to get $x=\overline{x}$ and $p^x=\overline{p}^x$, which are independent of the value of $\hat{\alpha}$ and $\hat{\beta}$. Therefore, all possible demand curves are bounded by demand curve 1 and 2 and they intersect at a common point $S=(\overline{x},\overline{p}^x)$.

$$MCS_1 = \int_{x_0}^{x_1} (\hat{\alpha}_U + x\hat{\beta}_L)dx = \hat{\alpha}_U(x_1 - x_0) + \frac{\hat{\beta}_L}{2}(x_1^{\ 2} - x_0^{\ 2}). \tag{4.1}$$

Similarly, demand curve 2 yields a *MCS* equal to

$$MCS_2 = \int_{x_0}^{x_1} (\hat{\alpha}_L + x\hat{\beta}_U)dx = \hat{\alpha}_L(x_1 - x_0) + \frac{\hat{\beta}_U}{2}(x_1^{\ 2} - x_0^{\ 2}). \tag{4.2}$$

Intuitively, if we have double-sided bounds on β, we should also have double-sided bounds on *MCS*. It is straightforward to demonstrate this. Proposition 3 provides a formal statement.

Proposition 3. If NR1-NR3 hold, $\min\{MCS_1, MCS_2\} \le MCS \le \max\{MCS_1, MCS_2\}$.

Proof: See Appendix A.

Proposition 3 says if β is double bounded, then *MCS* is also double bounded by the corresponding value of *MCS* calculated from lower and upper bounds of β. Notice that the lower (upper) bound of β does not necessarily give the lower(upper) bound for *MCS*. The opposite might occur. As shown in Figure 4.3, when $|x_0 - \bar{x}| > |x_1 - \bar{x}|$, $MCS_2 \le MCS \le MCS_1$; when $|x_0 - \bar{x}| < |x_1 - \bar{x}|$, $MCS_1 \le MCS \le MCS_2$.

In the special case where the original and new levels of x are equidistant from $\bar{x}$, *MCS* can be point identified! Proposition 4 states this special case formally.

Proposition 4. If NR1-NR3 hold and $|x_0 - \bar{x}| = |x_1 - \bar{x}|$, then *MCS* is point identified: $MCS = MCS_1 = MCS_2$.

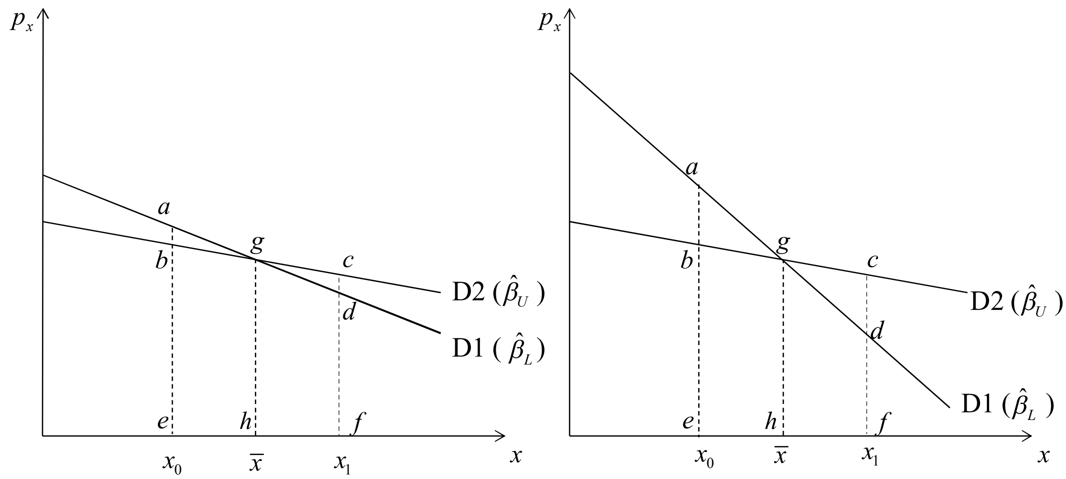
Proof: See Appendix A.

Proposition 4 may seem counterintuitive at first, and is worth repeating. Even if β is only partially identified, *MCS* is point identified as long as the original level and the new level of x are the same distance to the mean value of x. It also implies that, given x_0 and x_1, narrower bounds on β imply narrower bounds on *MCS*. These results can be demonstrated via two figures.

The demand curves in the left panel in Figure 4.4 provide narrower bounds on β than the demand curves in the right panel. We consider three sets of change in the level of x. First, suppose x increases from x_0 to $\bar{x}$. Then MCS_1 equals the area *aehg* and MCS_2 equals the area *behg*, which is smaller than MCS_1. Therefore, we have $MCS_2 \leq MCS \leq MCS_1$ and the width of these bounds is *abg*. Clearly, the size of area *abg* is smaller in the left panel where we have narrower bounds. Now suppose x moves from $\bar{x}$ to x_1. MCS_1 (represented by *ghfd*) is now smaller than MCS_2 (represented by *ghfc*). The bounds for *MCS* are $MCS_1 \leq MCS \leq MCS_2$ and their width is *gdc*. As in case 1, the width is smaller in the left panel than in the right panel. Lastly, we consider the case stated in Proposition 4: a change in x that is symmetric with respect to the mean. That is, x moves from x_0 to x_1 and $|x_0 - \bar{x}| = |x_1 - \bar{x}|$. We can observe that regardless of the width of the bounds on β, $MCS_1 = MCS_2$ so that MCS is point identified!

In summary, there are two situations that can give us narrow, informative bounds on *MCS*. First, the bounds on β, the slope of the demand curve, are narrow; and second, the current and prospective levels of x are symmetric or "near-symmetric" to the average level of x in the population. Even though it is unlikely that any policy will produce a quality change that is exactly centered at the mean, it may often be the case that the change spans the mean, so that the bounds on *MCS* are narrow enough to be informative even though the difference between IV and OLS estimates are large.

Figure 4.4 Bounds on Welfare Measure



Change	MCS_1	MCS_2	$MCS_2 - MCS_1$
x_0 to $\bar{x}$	*aehg*	*behg*	*-abg*
$\bar{x}$ to x_1	*ghfd*	*ghfc*	*gdc*
x_0 to x_1	*aefd*	*befc*	*0*

For example, consider the Clean Air Act, or the Clean Water Act. While they both impose minimum standards, actions taken to meet these standards may tend to improve environmental quality everywhere. Areas that were initially the most polluted may observe large improvements over a range that lies below $\bar{x}$, while areas that were initially the least polluted may observe small improvements that lie over a range above $\bar{x}$. However, at least some areas are likely to be close to the mean in the original population. They will experience improvements that cross the $\bar{x}$ threshold. Therefore we would expect that near-symmetry result to be useful for identifying the benefits of improvements for at least some locations.

Propositions 3 and 4 assume that double-sided bounds for \$\beta\$ can be obtained. What if β is only single bounded? Denote the single-sided bound as $-\infty < \beta \le \hat{\beta}_U$.

Graphically, one would rotate *D1* clockwise in Figure 4.4 until it is vertical to the x axis. Returning to the three scenarios, when x changes from x_0 to $\bar{x}$, the true MCS is bounded by MCS_2 and $+\infty$. When x changes from $\bar{x}$ to x_1, MCS is bounded by 0 and MCS_2. In the third case, we can view MCS_1 as the sum of area *aehg* and *ghfd* which approach $+\infty$ and 0, respectively. Therefore, MCS_1 approach $+\infty$. Thus, single-sided bounds on β produce single-sided bounds on Marshallian consumer surplus.

4.5 A Demonstration: Water Quality in Markets for Lakefront Properties

In this section we demonstrate the methodology, using a hedonic model of housing markets. The primary variable of interest is the lake water quality measured by water transparency. Our objective is to partially identify a demand function for water quality and then derive bounds for Marshallian consumer surplus for prospective improvements.

We first describe three sets of data which allow us to estimate demand: lakefront property transactions data; water quality data; and demographic data describing property purchasers. We will discuss why the data make a suitable application for partial identification. Then we derive bounds on demand parameters and estimate consumer surplus under several different scenarios.

4.5.1 Data Description

The data we use here cover three states: Vermont, New Hampshire, and Maine. The data for each state have been used individually in previous papers or theses (Hsu, 2000; Lawson, 1997; Boyle et al., 1999; Gibbs et al., 2002). This paper is the first to pool the data.

Table 4.1 Description of 1st Stage Variables

Variable	Description
P	sale price of the property (1995 dollars)
BARE	0,1 = unimproved land
SQFT	total living area (square feet)
LOT	lot size (acres)
HEAT	0,1 = central heating system
FULLBATH	0,1 = a full bathroom
FF	total lot frontage abutting the water (feet)
LAKESIZE	surface area of lake (acres)
AGE	age of house (years)
PLUMB	0,1 = full plumbing
DIST	distance to the nearest large town (miles)
DENSITY	housing density (lots/1,000 feet of lake frontage)
TAX	property tax rate in the year of purchase
WT	water transparency (meters)

Property transaction data

We use sales of residential properties located on selected freshwater lakes and ponds in Vermont, New Hampshire, and Maine over the period 1990 to 1995[10]. Single family residential, vacation homes, and unimproved land are all used in this study. The sale prices were collected from transfer tax records and the property characteristics (structural and lot) were obtained from assessment records, both of which are maintained at the individual town offices. This yields a sample of 230 transactions in Vermont, 518 transactions in New Hampshire, and 851 transactions in Maine. Table 4.1 defines variables and Table 4.2 reports the summary statistics of property sales data for each state. Note that the available variables are not entirely consistent across states, especially for

[10] The period 1990 and 1995 was selected because the real estate market for lakefront properties in the study area was stable 4. during this period. Poor (2001) tested this stability statistically using the same source of data for Maine.

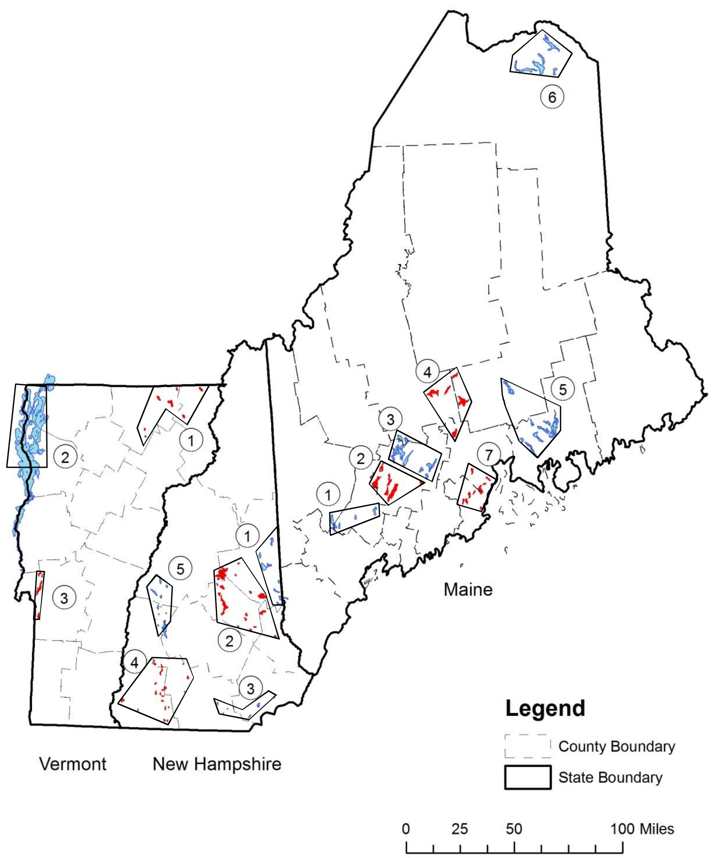
New Hampshire. Therefore, different sets of variables will enter into the hedonic price functions for each state.

Table 4.2 Summary Statistics of 1st Stage Variables

Variable	Mean	Std. Dev.	Min.	Max.
Panel A: Vermont				
P	99033.76	61313.78	4000	340000
BARE	0.18	0.39	0	1
SQFT	810.48	578.08	0	3560
LOT	1.01	1.84	0.07	11.91
HEAT	0.53	0.50	0	1
FULLBATH	0.68	0.47	0	1
FF	132.94	117.34	15	1201
LAKESIZE	1575.06	840.90	54	2795
WT	5.26	1.87	1.9	9.5
N		230		
Panel B: Maine				
P	71535.76	57155.35	400	500000
BARE	0.25	0.43	0	1
SQFT	715.33	614.12	0	4128
LOT	1.37	2.57	0.04	20
HEAT	0.46	0.50	0	1
FULLBATH	0.59	0.49	0	1
FF	154.27	140.31	10	1800
LAKESIZE	3515.09	2428.37	171	8239
WT	4.15	1.98	0.3	9.4
N		851		
Panel C: New Hampshire				
P	159298.50	109832.50	12500	815253.8
AGE	39.47	25.84	0	181
SQFT	1126.53	723.84	107	6532
PLUMB	3.89	0.50	0	4
FF	135.80	187.69	5	3395
DIST	12.36	7.90	0	33
DENSITY	8.51	3.10	1	20
TAX	22.53	7.80	8.2	41
LAKESIZE	1240.51	1507.80	31	9091
WT	4.79	2.05	0.7	11.5
N		518		

For each state, property sales are divided into several market groups which were defined by the proximity of the lakes to each other. Each group represents a different multiple listing service region which real estate agents define to be a distinct market. There are 20 lakes within 3 market groups in Vermont, 53 lakes within 5 market groups in New Hampshire, and 37 lakes within 7 market groups in Maine. Figure 4.5 shows the map of the three states, the locations of the lakes, and the delineation of market groups.

Figure 4.5 Lakes and Market Group Delineations



Note: Lakes and surrounding properties within each polygon are in one market.

Several features of these data illustrate the empirical relevance of the theoretical issues raised in the previous sections. First, each state includes data from multiple markets. This enables us to use market dummies as instruments to identify demand functions.

Second, the spatial extent of a market is difficult to determine with certainty. While this problem complicates point identification with market dummies as instruments, it motivates using market dummies as “imperfect” instruments because if MLS regions are not truly distinct from each other then unobserved preferences may be correlated with market dummies, making them “imperfect”. There are two reasons why the markets in our data may not be truly separated. First, as we can see from Figure 4.5, even if the different housing markets are in different MLS areas, many of them are in close proximity(market 2 and 3 in Maine, for example). Second, commuting patterns provide some clues about market segmentation. A working household would probably consider moving closer to their jobs. Table 4.3 shows the county-to-county worker flow in the three states. We can see that almost no workers commute across states. Within each state, most of working population chooses to work in the market where they live. But there are still sizable fractions of workers who commute to a different market. This is especially true in New Hampshire where 21.8 percent of working population in market 1 and 17.4 percent in market 2 chooses to work in market 3.

Third, the property transaction data mix sales of “summer houses” with “permanent houses”. Purchasers of summer homes may be more likely to search across all three states, whereas purchasers of permanent homes may be more likely to search in a local area, close to their job, or family and friends. Therefore, market dummies may be correlated with consumers' unobserved preferences for at least a subset of the sample.

One limitation of these data is the small number of observations compared to recent studies (see Kuminoff, Parmeter, and Pope, 2010, for examples). We use a parsimonious set of explanatory variables in the regressions in order to avoid low degrees of freedom and high MSEs.

Table 4.3 2000 VT/ME/NH County-to-County Worker Flow (Percentage)

Home County	Work County														
	VT 1	VT 2	VT 3	HN 1	NH 2	NH 3	NH4	NH5	ME1	ME2	ME3	ME4	ME5	ME6	ME7
VT 1 (Orleans/Essex)[a]	**77.3**[b]	1.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VT 2 (Franklin/Grand Isle/Chittenden)	0.1	**94.9**	0.1	0.0	0.0	0.1	0.0	2.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
VT 3 (Rutland)	0.0	0.7	**86.4**	0.0	0.0	0.0	0.1	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NH 1 (Carroll)	0.0	0.0	0.0	**65.3**	*[c]	21.8	1.6	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HN 2 (Carroll/Belknap/Strafford)	0.0	0.0	0.0	*	**68.1**	17.4	2.0	1.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HN 3 (Hillsborough/Rockingham)	0.0	0.0	0.0	1.3	1.4	**74.7**	*	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
HN 4 (Cheshire/Hillsborough)	0.0	0.0	0.0	0.2	0.3	*	**74.0**	0.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
NH 5 (Sullivan/Grafton)	0.0	0.0	0.0	0.4	3.1	1.2	1.0	**81.9**	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ME 1 (Androscoggin)	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.0	**74.0**	2.7	2.7	0.2	0.1	0.0	0.0
ME 2 (Kennebec)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.1	**80.4**	*	6.1	0.8	0.0	1.5
ME 3 (Kennebec)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	4.1	*	**80.4**	6.1	0.8	0.0	1.5
ME 4 (Penobscot/Somerset/Waldo)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	5.3	5.3	**86.8**	*	0.3	*
ME 5 (Hancock/Penobscot)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.6	*	**93.9**	0.4	1.1
ME 6 (Aroostook)	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	2.0	2.0	**96.1**	0.0
ME 7 (Waldo/Knox)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	4.6	4.6	*	8.5	0.0	**80.5**

Source: County-by-county work flow files, U.S. Census Bureau, 2000.

Note: [a] Counties in the parenthesis are the ones that housing market is included in. For example, VT1 is included in Orleans and Essex County.

[b] Each element in this Table reads as: x percent of Home County working population works in Work County. For example, 77.3% of Orleans/Essex's working population works in the Orleans/Essex County. 1.8% of Orleans/Essex's working population works in the Franklin/Grand Isle/Chittenden County.

[c] Missing value means Home County and Work County has overlapping. For example, NH1 is included in Caroll County, part of NH2 is also included in Caroll County.

Water quality data

The lakes in the study areas are known for having clear, high-quality water, but some are threatened by "cultural eutrophication" which originates from residential development, silviculture, and agricultural activities within the watershed. One of the results from eutrophication is reduced water transparency. We use transparency as the indicator of water quality for two reasons. It is highly correlated with overall water quality, and it is the only measure that is consistently maintained at a specific interval over the summer months. Other water chemistry tests are only conducted for specific "hot spots" where eutrophication is known to be particularly problematic (Poor et al., 2001).

Data on lake water transparency were provided by the Vermont Department of Environmental Conservation (VTDEC), the Maine Department of Environmental Protection (MEDEP), and the New Hampshire Department of Environmental Services (NHDES). It is measured using a secchi disk that is 8 inches in diameter and alternatively black and white in each quadrant. The disk is lowered into the lake water and the depth (in meters) at which the disk disappears from sight is the measure of water transparency. The minimum water transparency during the summer months is used as the measure of water transparency because in spring and fall, transparency is subject to fluctuations of water flows and silt disturbance, and in winter the ice prohibits measurement. Since summer months are the time when the eutrophication levels are most affected by long exposure of the water to sunlight, which stimulates algae growth, it is the most appropriate time to measure the water's trophic status.

Water transparency varies over the lakes within each market area and over time for each individual lake. As eutrophication is a long-term process, the temporal variation for each lake is very small. The major source of variation in water transparency is between lakes. The summary statistics for water transparency are reported in Table 4.2.

Survey Data

Data for demographic variables in the demand function come from a mail survey of the purchasers of the lakefront property sales used in the first stage. Respondents were asked to provide information about the socioeconomic characteristics of their households, such as the age and employment status of the household head, the total household income and the number of kids in the household. They were also asked questions about their familiarity with the lake. Responses to these questions were used to create variables that served as demand shifters in the demand function estimation. Since some purchasers failed to answer the survey and some surveys we received were not valid, the survey data is only a subset of the property transaction data. Table 4.4 lists the variables created from the mail survey responses and Table 4.5 summarizes these variables for each state.

Table 4.4 Description of 2nd Stage Variables

Variable	Description
RESAGE	age of the mail survey respondent
INC	total after-tax household income
RETIRED	0,1 = respondent is fully retired
KIDS	total number of people under 18-year old in the household
VISIT	0,1 = visited the lake before purchasing the property
FRIEND	friends or relatives own property on the lake

Table 4.5 Summary Statistics of 2nd Stage Variables

Variable	Mean	Std. Dev.	Min.	Max.
Panel A: Vermont				
RESAGE	47	11	25	72
INC	89579	69848	17500	350000
RETIRED	0.29	0.46	0	1
KIDS	3.01	1.45	1	9
VISIT	0.97	0.18	0	1
FRIEND	0.45	0.50	0	1
N		95		
Panel B: Maine				
RESAGE	45	11	22	74
INC	77979	50810	7500	212500
RETIRED	0.10	0.29	0	1
KIDS	0.77	1.04	0	4
VISIT	0.14	0.35	0	1
FRIEND	0.50	0.50	0	1
N		240		
Panel C: New Hampshire				
RESAGE	46	11	22	79
INC[a]	15	5	5	26
RETIRED	0.27	0.45	0	1
KIDS	0.85	1.07	0	4
VISIT	0.89	0.32	0	1
FRIEND	0.54	0.50	0	1
N		203		

Note: [a] After-tax household income in New Hampshire is recored as rankings: minimum income 5 represents \$20,000 to \$ 24,999; maximum income 26 represents \$350,000 and above; mean income 15 represents \$90,000 to \$99,999. The specific rankings are provided in Hsu, 2000, Appendix D.

4.5.2 Hedonic model and partial identification

The hedonic price function (stage-one) and demand function (stage-two) will be estimated for each state separately and then comparisons will be made across states. In general, there are two potential sources of error associated with this two-step procedure. In the first stage, we calculate consumers' MWTP for environmental quality by estimating the hedonic price function. If there is an omitted variable problem with the estimation, then estimated MWTP will be biased. This bias will be carried over to the demand function estimation in the second stage. Second, suppose we use an ideal quasi-experiment to recover the "true" MWTP in the first stage. If we follow the standard practice of approximating WTP by multiplying MWTP by the size of the change then our measure of *MCS* will be biased. It is difficult to predict which source of error is more important (Kuminoff, Smith, and Timmins, 2010).

There is a large literature dealing with the first-stage error using quasi-experimental designs (See Parmeter and Pope (2009) for citations to the literature). Since we do not have any source of quasi-experimental variation in water quality and the focus of this paper is demand estimation, we abstract from omitted-variable issues in the first stage and focus on the second-stage identification. But any quasi-experimental approach can be integrated into this phase of our analysis without influencing second-stage identification.

First-stage: hedonic price function

The first stage model regresses the price of house i in market m on its structural attributes, neighborhood attributes, and environmental amenities. There are many potential functional specifications, but we opt for a standard liner-in-parameter specification[11]. For each market we estimate a separate hedonic price function[12]:

[11] I also estimated a semi-log model, which produced very similar results. The results are reported in the Appendix B.
[12] Due to the data availability, the set of independent variables used for NH is different.

$$P_{im} = \theta_{0m} + \theta_{1m} BARE_{im} + \theta_{2m} SQFT_{im} + \theta_{3m} LOT_{im} + \theta_{4m} HEAT_{im} + \theta_{5m} FULLBATH_{im} + \theta_{6m} FF_{im} + \theta_{7m} WQ_{im} + \epsilon_{im}, \quad (5.1)$$

where $WQ = LAKESIZE \times \ln(WT)$ is a interaction between lake size and the natural log of water transparency, which reflects the nonlinear relationship between housing price and water quality. It is expected that at lower levels of water transparency property owners are willing to pay more for a one meter improvement in transparency. In fact, changes in transparency occurring above four meters are not as visibly noticeable as are changes in transparency below this threshold (Smeltzer and Heiskary, 1990), supporting the assumption that the relationship between property prices and water transparency is nonlinear.

In the first stage, there is a potential "omitted variable" problem. Everything else equal, a more developed watershed means less transparent lake water[13], but properties located around more developed lakes may enjoy more facilities and hence be more expensive. It is unclear whether this source of bias will cause us to overstate or understate the effect of lake water clarity. Usually, one way to solve this problem is to add lake fixed effects. In our case, however, there is too little variation in water transparency within a lake to add lake fixed effects.

Table 4.6 reports the results for all markets in Vermont and Maine; Table 4.7 reports results for New Hampshire because it employs a different set of independent variables.

[13] This statement can be supported by information from some governmental websites. "A watershed is the land area surrounding a lake that drains into the lake. Any activities occurring in the watershed affect lake water quality when the pollution resulting from an activity that enters streams or groundwater that eventually feed into the lake." (\url{http://www.vtwaterquality.org/cfm/lakerep/lakerep_details.cfm?id=RICHVILLE}, last accessed on 10/02/2011.). "Common sources of pollution to streams include agricultural activities such as crop production, cattle grazing; municipal dischargers such as sewage treatment plants; and urban runoff from city streets, parking lots, sidewalks, storm sewers, lawns, golf courses, and building sites."(\url{http://water.epa.gov/type/rsl/monitoring/vms21.cfm}, last accessed on 10/02/2011.)

Indicated by the adjusted R^2, the models fit very well with Vermont and Maine data. As expected, the total living area (*SQFT*) is positive and significant in all the VT/ME markets. The existence of central heating system (*HEAT*) significantly increase the property price in one VT market and four ME markets, by $27378 on average. Since the winter in these states is long and cold, it is reasonable that the value of a central heating system is high. A full bathroom (*FULLBATH*) also significantly increases the property price in four ME markets. Total lot frontage abutting the lake (*FF*) is significant in one VT market and seven ME markets. The environmental variable *WQ* is significant and positive in two VT markets and four ME markets. Based on equation (5.2), the implicit price for water transparency (*WT*) can be calculated.

Table 4.7 also suggests that the models fit well with New Hampshire data. *SQFT* and *WQ* are significant and positive in all five of the markets as expected, implying that living area and water transparency does significantly affect property prices.

An important feature to note is the heterogeneity in hedonic gradients across markets. This will provide us with an important source of variation to identify the demand function. Also note that the coefficients for *WQ* are not significant for all markets. In the second stage, we only use markets where there is enough data to significantly distinguish the water quality coefficient from zero[14].

The implicit price for water transparency for use in stage-two of the estimation procedure is:

$$P_{im}^{WT} = \theta_{7m} \frac{LAKESIZE_{im}}{WT_{im}} \tag{5.2}$$

[14] As shown in Table 3.6, market VT3 has a coefficient for WQ that is substantially lower than the other 2 markets in Vermont; Market ME2, ME5, and ME6 have coefficients for WQ that have negative sign and are also substantially lower than the other 4 markets in Maine. Therefore, it is reasonable not including these markets in the second stage.

Table 4.6 Hedonic Price Function Estimates for VT/ME

	VT1	VT2	VT3	ME1	ME2	ME3	ME4	ME5	ME6	ME7
BARE	-11,678.8	36,537.5	36,442.1	-4,167.8	17,346.2	11,208.1	-10,613.8	11,156.4	11,038.3***	27,806.3
	(9,894.3)	(22,544.0)	(32,707.5)	(22,254.6)	(14,162.6)	(8,692.5)	(11,828.5)	(7,369.5)	(4,213.6)	(30,125.4)
SQFT	33.9***	60.2***	71.4***	22.0**	31.7***	48.1***	51.3***	48.5***	35.1***	59.7***
	(9.1)	(10.7)	(16.5)	(10.5)	(8.1)	(5.0)	(13.4)	(6.0)	(4.3)	(19.3)
LOT	-264.7	-214.8	10,448.0***	-3,830.0	-2,607.2	311.0	15,509.8***	416.1	-1,268.4	-6,322.7***
	(1,599.1)	(3,488.9)	(3,995.3)	(3,529.4)	(1,830.4)	(1,008.9)	(5,576.3)	(902.6)	(1,360.2)	(2,288.6)
HEAT	14,839.6*	45,282.2***	6,761.5	30,800.8**	17,674.9**	23,995.8***	2,972.4	19,169.9***	1,348.6	-25,475.5
	(7,673.5)	(13,366.2)	(19,654.3)	(12,399.2)	(8,976.8)	(5,420.3)	(8,525.7)	(5,644.3)	(4,020.2)	(31,573.7)
FULLBATH	4,606.4	-4,959.3	24,483.8	19,006.4	24,360.0**	15,523.0**	2,233.8	15,556.3**	15,141.1***	59,507.4
	(7,172.6)	(20,106.3)	(27,951.9)	(20,409.5)	(11,136.9)	(7,511.8)	(8,787.0)	(6,315.2)	(3,562.2)	(41,584.2)
FF	127.5***	-4.6	41.1	256.3***	115.7**	57.7**	-46.4***	83.7***	65.3***	126.8***
	(39.8)	(45.1)	(95.2)	(67.3)	(47.2)	(23.5)	(17.9)	(19.6)	(17.6)	(35.4)
WQ	**11.4*** **	**10.6*** **	**2.0**	**7.5*** **	**-0.9**	**1.6*** **	**11.4*** **	**-0.1**	**-0.4**	**35.0*** **
	(2.1)	(4.0)	(3.7)	(1.4)	(2.6)	(0.4)	(3.4)	(1.2)	(0.5)	(13.5)
_cons	10,294.3	3,869.0	-3,683.3	-19,755.4	10,179.4	2,763.3	2,431.3	2,216.8	-4,602.5	-14,566.2
	(8,980.9)	(23,612.9)	(30,590.2)	(21,691.0)	(13,738.2)	(7,746.1)	(8,993.7)	(7,757.4)	(4,812.5)	(26,190.7)
N	60	99	71	83	105	254	39	164	150	56
Adjusted R^2	0.63	0.51	0.42	0.54	0.44	0.59	0.80	0.66	0.67	0.62

Note: *** p<0.01, ** p<0.05, * p<0.1

Table 4.7 Hedonic Price Function Estimates for NH

	NH1	NH2	NH3	NH4	NH5
AGE	-105.9	-453.0**	-1,746.0**	-1,817.1**	2,115.3**
	(687.2)	(221.4)	(847.1)	(858.0)	(1,075.1)
AGE2	-0.4	-0.0	10.6	12.2*	-15.3**
	(8.3)	(0.0)	(8.7)	(7.1)	(7.8)
SQFT	41.1***	50.5***	43.0***	41.0***	109.6***
	(11.6)	(6.9)	(9.0)	(13.5)	(20.5)
PLUMB	23,074.6	6,112.2	2,025.8	-2,371.5	9,646.8
	(17,504.2)	(8,222.3)	(14,863.5)	(13,126.9)	(18,497.1)
FF	21.1	107.0*	165.2**	130.3	211.7*
	(15.1)	(57.6)	(82.7)	(88.1)	(119.6)
DIST	-2,422.7***	-601.3	-714.3	201.8	-216.0
	(740.6)	(877.3)	(1,330.4)	(1,480.8)	(3,525.8)
DENSTY	-1,855.2	-3,648.2*	943.6	-7,663.4**	-5,648.3
	(1,599.4)	(2,005.2)	(2,113.9)	(3,078.2)	(4,220.4)
TAXRT	-3,214.0***	-1,418.7*	304.1	368.1	-4,634.2*
	(816.6)	(726.2)	(1,077.7)	(1,227.4)	(2,451.4)
WQ	**4.6****	**5.9*****	**193.3****	**124.7*****	**23.8*****
	(2.0)	(2.0)	(83.4)	(21.1)	(6.6)
_cons	110,645.4	125,738.5**	59,451.5	136,745.5**	54,573.0
	(76,556.8)	(49,175.7)	(67,418.5)	(68,729.7)	(120,199.9)
N	110	170	68	73	97
Adjusted R^2	0.43	0.50	0.62	0.67	0.68

Note: *** p<0.01, ** p<0.05, * p<0.1

Second-stage: bounds on demand function parameter

Next, we specify the demand function. The implicit price of water transparency calculated using equation (5.2) is a function of itself, its substitutes and complements such as square feet of living area and lake frontage, household's characteristics including age, income, retirement status, and the number of children in the family. Several other factors from survey that may influence a household's MWTP for water quality are also included, such as whether they have visited the lake before, and whether they have a friend or relative owning a lakefront house. In particular, we estimate the following specification:

$$P_i^{WT} = \beta WT_i + (\alpha_0 + \alpha_1 SQFT_i + \alpha_2 FF_i + \alpha_3 RESAGE_i + \alpha_4 INC_i + \alpha_5 RETIRED_i + \alpha_6 KIDS_i + \alpha_7 VISIT_i + \alpha_8 FRIEND_i) + u_i \tag{5.3}$$

We consider using market dummies and their interactions with demand shifters as IIVs. However, we first need to make some adjustments to them. NR's method requires using only one IIV to estimate each bound for β. This requirement complicates the idea of using dummies as IIVs. When the number of markets exceeds two, it makes little sense to use only one market dummy as an instrument because this implicitly treats all other markets the same. One way to overcome this problem is to create a categorical variable M to capture the variation across markets. We generate M based on the average level of WT in each market. With K markets, M takes the value of 1 for the market with lowest average water transparency and a value of K for the market with highest average water transparency. This categorical instrument is based on the same idea as the rank-based instruments that were introduced by Epple and Sieg (1999) and used in the subsequent sorting applications (Sieg et al., 2002; Sieg et al., 2004; Smith et al., 2004)[15]. According to the definition of categorical instrument M, the interaction terms between market dummies and demand shifters are also replaced by interactions between M and demand shifters. Here we consider using one demand shifter, household income (INC), and generate another IIV represented as $M * INC$.

The concern in traditional demand estimation is that the water transparency, the endogenous variable, is positively correlated with unobserved preferences, u. In order to satisfy assumption NR3 we require that M and $M * INC$ are positively correlated with

[15] Epple and Sieg (1999) used the income rank of communities as their instruments because they assume that if preferences were homogeneous then communities are stratified by income alone. In our case, we use the water transparency rank of markets as our instruments because it facilitates the analysis of correlation direction between instrument and error term.

the error term. The sorting models of Epple and co-authors suggest that in equilibrium there should be a relationship between the quantity of the amenity being consumed and the unobserved preferences of the individuals choosing the consumption. This sorting process implies that, all else constant, households with stronger preference for water quality will sort themselves into markets with higher average water quality. Furthermore, it is natural to assume that households with higher income have stronger preferences for water quality. This may be because that people with lower income place more emphasis on house conditions that fulfill their basic needs rather than the environmental amenities. Furthermore, higher income people are more likely to own boats, jet-skis, and other expensive equipment that would allow them to spend more time on the water and therefore are more likely to care about water quality. Some other studies also provide evidence that wealthier people have stronger preferences for water quality. For example, Kosenius (2010) investigated heterogeneous preference for water quality and found that higher income people are willing to pay more for increased water transparency and decreased occurrences of algae bloom. Hence, we expect that both M and $M*INC$ are positively correlated with error, satisfying assumption NR3.

By construction, however, it is also the case that WT is positively correlated with M and $M*INC$. Thus according to proposition 1, these two IIVs will only yield single-sided bounds. If one instrument is better than the other in that it is both more relevant and more valid, then a new IIV can be generated that yields double-sided bounds by constructing a weighted difference of the original IIVs. To investigate this possibility, we first define instrument M as z_1 and $M*INC$ as z_2 and check if the inequality (3.6) holds. If the inequality holds, then we set $\gamma = \frac{\sigma_{z_1}}{\sigma_{z_1} + \sigma_{z_2}}$ which implies $\rho_{z_1 x} > \rho_{z_2 x}$ and $\rho_{z_1 u} < \rho_{z_2 u}$. Last, a weighted difference between z_1 and z_2 is constructed using equation (3.5) and the new IIV $w(\gamma)$ is used to repeat the estimation[16].

4.5.3 Results

The results are presented in Table 4.8. The dependent variable in all columns is the implicit price for water transparency. Column (1) presents results from OLS regression, the estimated *WT* coefficient is negative as expected. Column (2) reports IV regression results using the market categorical variable M as an instrument. Column (3) reports IV regression results using the interaction variable $M * INC$ as an instrument. Column (4) reports the bounds if we view M and $M * INC$ as IIVs and apply case 2 in Proposition 1 to them. For all three states, we can only get single-sided bounds because none of them passed NR's test for the existence of a valid weight γ.

The bounds for the demand curve slopes vary greatly across markets. The upper bound in Maine is only -293, while in Vermont it is nearly ten times larger, -2921. In New Hampshire, the upper bound is even larger, -5055. The 95% confidence interval for the single-sided bounds is reported in column (5).

We also calculate bounds on *MCS* for non-marginal changes in water transparency. Three reference points are used for the *MCS* calculation. Taking Maine for example, the

[16] In this procedure note that z_1 and z_2 cannot be reversed because we need to invoke the assumption that z_1 is more relevant and valid than z_2 to justify the value of γ. The relevance condition $\rho_{z_1 x} > \rho_{z_2 x}$ can be checked using data; The validity condition $\rho_{z_1 u} < \rho_{z_2 u}$ is believed to hold because the unobserved preference is assumed to correlated with both market dummy and income distribution.

Maine Department of Environmental Protection has determined that 3.0 meters of lakewater transparency is the threshold below which lakes have significantly compromised water quality. That is, it is nearly impossible from a management perspective to take any actions that will improve water transparency (Boyle et al., 1999)[17]. In our sample, the average visibility for all lakes in Maine is 4.6 meters, the average visibility for lakes that do not have compromised water clarity (>3 meters) is 5.2, and the average visibility for lakes with compromised water clarity ($\leq$ 3 meters) is 1.7 meters. We then consider three scenarios of water transparency change: an increase from 4.6 to 5.2 meters; a decrease from 4.6 to 1.7 meters; and an increase from 1.7 to 5.2 meters[18]. The *MCS* regions, with different sets of original level and new level are also reported in column (4) in Table 4.8 and 95% confidence intervals on the *MCS* bounds reported in column (5). Since the estimated demand curve represents demand for water transparency for an average person in the sample, and the property price has not been annualized, the *MCS* here is interpreted as an average person's gain/loss associated with water transparency change.

Column (5) reports the confidence intervals for bounds on key parameter and welfare measurement, the point is to distinguish between the economic and statistical uncertainty in the demand estimation. Economic uncertainty comes from the uncertainty on the correlation between instrument and error term and hence we get bound identification instead of point identification. Confidence intervals account for the statistical uncertainty of the upper bound or the lower bound itself.

---

[17] This threshold must be different in Vermont and New Hampshire, but we use this standard for all three states here.

[18] With "low demographic" and "pooled data" case, we use different sets of reference points as increased samples are used in these cases.

Table 4.8 Demand Function Estimation for VT\ME\HN

	(1) OLS	(2) IV1	(3) IV2	(4) IIV	(5) 95% CI
Panel A: Vermont					
WT	-878.12***	-2368.14***	-2921.49***	$(-\infty, -2921.49]$	$(-\infty, -909.52)$
	(173.70)	(529.48)	(1001.70)		
x_0=2.6, $\bar{x}$=4.6, x_1=5.5					
MCS ($\bar{x} \uparrow x_1$)				[0, 2702.20]	(0, 3571.38)
MCS ($\bar{x} \downarrow x_0$)				$(-\infty, -14477.23]$	$(-\infty, -10574.00)$
MCS ($x_0 \uparrow x_1$)				$[17179.44, +\infty)$	$(14145.38, +\infty)$
Panel B: Maine					
WT	-239.25	718.66	903.32	$(-\infty, -293.24]$	$(-\infty, 89.17)$
	(193.69)	(471.49)	(580.69)		
x_0=1.7, $\bar{x}$=4.6, x_1=5.2					
MCS ($\bar{x} \uparrow x_1$)				[0, 1939.67]	N/A[a]
MCS ($\bar{x} \downarrow x_0$)				$(-\infty, -10863.32]$	N/A
MCS ($x_0 \uparrow x_1$)				$[12802.99, +\infty)$	N/A
Panel C: New Hampshire					
WT	-1111.52***	-4673.07***	-5055.63***	$(-\infty, -5055.63]$	$(-\infty, -3351.40)$
	(247.15)	(797.63)	(864.07)		
x_0=2.3, $\bar{x}$=4.9, x_1=5.3					
MCS ($\bar{x} \uparrow x_1$)				[0, 2068.23]	(0, 2219.80)
MCS ($\bar{x} \downarrow x_0$)				$(-\infty, -33160.51]$	$(-\infty, -27499.15)$
MCS ($x_0 \uparrow x_1$)				$[35228.75, +\infty)$	$(29718.95, +\infty)$

Note: [a]As long as we believe the demand curve for WT is downward sloping, the coefficient of WT should be negative. Since the upper bound of 95% confidence interval is positive, we did not report the corresponding welfare measure.

Boyle et al. (1999) attempted to address the endogeneity problem with point identification of demand function for water quality. They employed the same data in Maine and specified the same functional form for the hedonic price function and demand function. But we employ a slightly different set of explanatory variables which we think are more reasonable in both functions[19]. The instruments they used are based on the same idea as market dummies: they are variables which are likely to cause variation in water quality across markets but not within a single market, such as total population of all towns within each market group divided by the total feet of lake shoreline in the market group region; total number of lakes in each market; the minimum distance between the center of each market to the nearest business center; and the unemployment rate for each market in the year the property was purchased. Their point estimate for β is -16,287, falling within our bounds: $(-\infty,-293.24]$. Based on their estimated demand function, there will be a \$1270.36 consumer surplus for an increase from 4.6 to 5.2 meters. The figure falls within our bounds (0 to \$1939.67). A -\$88794.78 consumer surplus for a decrease from 4.6 to 1.7 meters also falls within our bounds ($-\infty$ to $-\$10863.32$).

We also report the results with pooled data from the three states in Table 4.9[20]. They follow the same general pattern as the state-specific results.

[19] In hedonic price function, we dropped two variables which are originally used in Boyle et al. (1999), distance to the nearest city and a dummy indicating whether or not the property's primary source of water is the lake. The reason is that the distant variable is highly correlated with the key variable WQ and the dummy is not significant in all markets and the adjusted R^2 will not be influenced after dropping it. We also added a new dummy variable indicting whether or not the property is bare land. In demand function, we dropped variables indicating whether or not the purchaser expected an improvement, decline, or no change in the water clarity at the time the property was purchased but added some demographic variables such as the purchaser's age, whether or not he is retired, and how many children under 18 in the family. These demographic variables are more widely available and used in the second stage hedonic estimation.

[20] As noted in Table 4. 5, the after-tax household income is measured in dollar for Vermont and Maine, but in rank for New Hampshire. This measurement difference does not influence the results for state-specific demand estimation, but will influence the demand estimation with pooled data. Therefore, I converted the dollars in Vermont and Maine to rank in order for the consistency.

Table 4.9 Demand Function Estimation with Pooled Data

	(1) OLS	(2) IV1	(3) IV2	(4) IIV	(5) 95% CI
Pooled Data					
WT	-709.45***	-2157.21***	-2631.94***	$(-\infty, -2631.94]$	$(-\infty, -1829.67)$
	(144.73)	(364.31)	(408.17)		
$x_0=2.1, \bar{x}=4.7, x_1=5.4$					
MCS ($\bar{x} \uparrow x_1$)				[0, 2807.95]	(0, 2984.97)
MCS ($\bar{x} \downarrow x_0$)				$(-\infty, -21720.51]$	$(-\infty, -18936.31)$
MCS ($x_0 \uparrow x_1$)				$[24528.46, +\infty)$	$(21921.28, +\infty)$

4.5.4 Estimation with "low" consumer characteristics data

In the preceding sections, a variety of household's characteristics other than income was added into the hedonic demand function such as age, retirement status, number of children in the family. In most of the hedonic dataset, however, these micro demographic variables are not available to match to property sales. In fact, difficulty in obtaining consumer demographic data also historically limited demand function estimation. One of the publicly available source of such data is the data set on mortgage applications published by Federal Financial Institutions Examination Council (FFIEC) in accordance with the Home Mortgage Disclosure Act (HMDA). The HMDA data provide three relevant types of information on mortgage applications: (1) Property location including census tract number or identifier; (2) Loan information including loan amount, loan date, loan type, property type, loan purpose, and owner-occupancy; and (3) Applicant (buyer) information including ethnicity, race, gender, and gross annual income. Bayer et al. (2008) and Bishop and Timmins (2010) demonstrate that HMDA data can provide house owners' demographics information in a way that can be linked to micro data on property transactions[21].

However, the HMDA does not track all the variables we might be interested in. The only useful variables provided by HMDA are race and annual income. While sorting models have focused on income as the most important demand shifter, it is not clear what would be gained from including more variables like the ones in the preceding sections. Therefore, it would be interesting and useful to perform a test on the importance of income relative to other micro demographic variables.

So now suppose we only have data on income, but no other demographics data ("low demographics"). This would mimic a realistic situation where we have access to HMDA data. Comparing this to the results from the "full demographics" scenario would indicate what is gained when we go from HMDA data to a fuller set of demographic characteristics.

Specifically, the "low demographics" case only enables us to utilize the following demand function, which is a reduced version of a "full demographics" case specified in equation (5.3):

$$P_i^{WT} = \beta WT_i + (\alpha_0 + \alpha_1 SQFT_i + \alpha_2 FF_i + \alpha_3 INC_i) + u_i \tag{5.4}$$

Table 4.10 reports the "low demographics" results for the three states separately and Table 4.11 reports the results for the pooled data. We get quite similar results in "low demographics" case compared to "full demographics" case. This result suggests that the parsimonious HMDA demographic data might capture the most important variables for measuring household's demand function.

[21] Specifically, Bayer et al. (2008) merged the property data with HMDA data on the basis of census tract, loan amount, lender's name, and date.

Table 4.10 Demand Function Estimation ("Low Demographics") for VT\ME\NH

	(1) OLS	(2) IV1	(3) IV2	(4) IIV	(5) 95% CI
Panel A: Vermont					
WT	-738.32***	-2328.46***	-2595.77***	$(-\infty, -2595.77]$	$(-\infty, -956.16)$
	(152.74)	(535.16)	(815.46)		
$x_0=2.6, \bar{x}=4.7, x_1=5.6$					
MCS ($\bar{x}\uparrow x_1$)				$[0, 2832.39]$	$(0, 3530.59)$
MCS ($\bar{x}\downarrow x_0$)				$(-\infty, -14785.60]$	$(-\infty, -11278.29)$
MCS ($x_0\uparrow x_1$)				$[17617.99, +\infty)$	$(14808.88, +\infty)$
Panel B: Maine					
WT	-318.57*	787.48	869.30	$(-\infty, -318.57]$	$(-\infty, 51.37)$
	(187.41)	(476.36)	(540.70)		
$x_0=1.7, \bar{x}=4.6, x_1=5.5$					
MCS ($\bar{x}\uparrow x_1$)				$[0, 2860.26]$	N/A[a]
MCS ($\bar{x}\downarrow x_0$)				$(-\infty, -10971.73]$	N/A
MCS ($x_0\uparrow x_1$)				$[13831.99, +\infty)$	N/A
Panel C: New Hampshire					
WT	-1064.48***	-4497.16***	-4935.70***	$(-\infty, -4935.70]$	$(-\infty, -3357.54)$
	(220.90)	(728.04)	(800.69)		
$x_0=2.4, \bar{x}=4.9, x_1=5.4$					
MCS ($\bar{x}\uparrow x_1$)				$[0, 2495.00]$	$(0, 2666.68)$
MCS ($\bar{x}\downarrow x_0$)				$(-\infty, -30983.86]$	$(-\infty, -25924.17)$
MCS ($x_0\uparrow x_1$)				$[33478.86, +\infty)$	$(28590.84, +\infty)$

Note: [a]As long as we believe the demand curve for WT is downward sloping, the coefficient of WT should be negative. Since the upper bound of 95% confidence interval is positive, we did not report the corresponding welfare measure.

Table 4.11 Demand Function Estimation ("Low Demographics") with Pooled Data

	(1) OLS	(2) IV1	(3) IV2	(4) IIV	(5) 95% CI
Pooled Data					
WT	-708.17***	-1967.72***	-2425.99***	$(-\infty, -2425.99]$	$(-\infty, -1704.19)$
	(134.30)	(336.55)	(367.30)		
$x_0=2.1, \bar{x}=4.8, x_1=5.5$					
MCS ($\bar{x}\uparrow x_1$)				[0, 2715.23]	(0, 2902.59)
MCS ($\bar{x}\downarrow x_0$)				$(-\infty, -21608.32]$	$(-\infty, -19017.92)$
MCS ($x_0\uparrow x_1$)				$[24323.55, +\infty)$	$(21920.50, +\infty)$

4.6 Conclusions and Policy Implications

In this paper we investigate partial identification of the parameters of the hedonic demand function with endogenous regressors and instruments that fail to satisfy the validity condition. Based on Nevo and Rosen(2010)'s methodology, we use market dummy and its interaction with demand shifter as "IIVs" to partially identify parameters in the hedonic demand function. We argue that due to the sorting process and the uncertainty of the geographic extent of the market the market dummy and hence the interactions can be justified as "IIVs".

Since the goal of estimating demand function is to calculate welfare measure associated with a non-marginal change in non-market good, we extend the partial identification on demand function parameters to Marshallian consumer surplus (*MCS*). We show that *MCS* can be either double bounded or point identified when the parameters are double bounded; it can be single bounded when the parameters are single bounded.

Relative to conventional hedonic demand estimation, imposing weaker assumptions on instruments leads to a range for parameters rather than point identification. The advantage, however, is that this method avoids the need for highly restrictive assumption on instrument validity that is usually difficult to justify. And in some circumstances, we can still obtain narrow bounds on welfare measures.

We can standardize the steps for estimating hedonic demand function and associated welfare measure by adding partial identification strategy to the conventional point estimation. Step 1, pool the data from multiple markets and obtain OLS estimates for demand parameters. Step 2, obtain IV estimates using rank-based market instrument (IV1). Step 3, obtain IV estimates using the interaction of rank-based market and income as instrument (IV2). The partial identification based on the OLS, IV1, and IV2 would give us single-sided bounds on demand parameters. Therefore Step 4 is to perform the test given in equation (3.6) to see if double-sided bounds could be obtained. The last step would be to calculate bounds on welfare measures based on bounds on parameters. This procedure is easy to follow and can be widely applied to non-market goods valuation other than water quality.

The partial identification strategy can be also applied to the first stage hedonic price function estimation where there is potential omitted variable bias. As long as one can predict the directions of correlation between instruments, endogenous variable, and the error term, either single-sided or double-sided bounds can be obtained for the parameter of interest and its implicit price.

A limitation of our method is that the demand function is specified to be linear with a single endogenous regressor. It would be of interest to consider a more complex hedonic demand model in future research.

Appendix A: Proofs

Proposition 1

Proof: The proof of Proposition 1 closely follows the proofs of Propositions 1 and 2 in NR (2010). NR3 includes two cases: (1) $\rho_{xu} \geq 0$ and $\rho_{z_j u} \geq 0$; (2) $\rho_{xu} \leq 0$ and $\rho_{z_j u} \leq 0$.

Under case (1):

$$\begin{aligned}
&\rho_{xu} \geq 0 \\
&\Leftrightarrow \sigma_{xu} \geq 0 \\
&\Leftrightarrow \sigma_{x(\tilde{p}_x - \beta\tilde{x})} \geq 0 \\
&\Leftrightarrow \sigma_{x\tilde{p}_x} - \beta\sigma_{x\tilde{x}} \geq 0 \\
&\Leftrightarrow \beta \leq \frac{\sigma_{x\tilde{p}_x}}{\sigma_{x\tilde{x}}} = \beta^{OLS}
\end{aligned}$$

where the third line uses $\tilde{p}_x = \beta\tilde{x} + u$, which is Lemma 3 in NR's paper. And $\frac{\sigma_{x\tilde{p}_x}}{\sigma_{x\tilde{x}}} = \beta^{OLS}$ because:

$$\begin{aligned}
&\sigma_{xu} = 0 \\
&\Leftrightarrow \sigma_{x(\tilde{p}_x - \beta^{OLS}\tilde{x})} = 0 \\
&\Leftrightarrow \frac{\sigma_{x\tilde{p}_x}}{\sigma_{x\tilde{x}}} = \beta^{OLS}
\end{aligned}$$

Similarly, condition $\rho_{z_j u} \geq 0$ indicates:

$$\begin{aligned}
&\rho_{z_j u} \geq 0 \\
&\Leftrightarrow \sigma_{z_j u} \geq 0 \\
&\Leftrightarrow \sigma_{z_j(\tilde{p}_x - \beta\tilde{x})} \geq 0 \\
&\Leftrightarrow \beta\sigma_{z_j\tilde{x}} \leq \sigma_{z_j\tilde{p}_x}
\end{aligned}$$

if $\sigma_{z_j\tilde{x}} > 0$, then $\beta \leq \frac{\sigma_{z_j\tilde{p}_x}}{\sigma_{z_j\tilde{x}}} \equiv \beta^{IV}_{z_j}$.

if $\sigma_{z_j\tilde{x}} < 0$, then $\beta \geq \frac{\sigma_{z_j\tilde{p}_x}}{\sigma_{z_j\tilde{x}}} \equiv \beta^{IV}_{z_j}$.

Combining these inequalities and a symmetric reasoning for the case (2) gives the conclusion of the proposition.

Proposition 3

Proof: It must be true that $|x_0 - \bar{x}| > |x_1 - \bar{x}|$ or $|x_0 - \bar{x}| < |x_1 - \bar{x}|$ or $|x_0 - \bar{x}| = |x_1 - \bar{x}|$.

(1) Consider the first case: $|x_0 - \bar{x}| > |x_1 - \bar{x}|$. This condition implies that $x_0 + x_1 < 2\bar{x}$. Using (4.1) and (4.2), it follows that:

$$\begin{aligned} MCS_2 - MCS &= (x_1 - x_0)[\hat{\alpha}_L - \alpha + \frac{\hat{\beta}_U - \beta}{2}(x_1 + x_0)] \\ &\leq (x_1 - x_0)[\hat{\alpha}_L - \alpha + \frac{\hat{\beta}_U - \beta}{2} \times 2\bar{x}] \\ &= (x_1 - x_0)[\hat{\alpha}_L - \alpha + \frac{\hat{\beta}_U - \beta}{2} \times 2\frac{\alpha - \hat{\alpha}_L}{\hat{\beta}_U - \beta}] \\ &= 0 \end{aligned}$$

Therefore $MCS_2 \leq MCS$.

$$\begin{aligned} MCS_1 - MCS &= (x_1 - x_0)[\hat{\alpha}_U - \alpha + \frac{\hat{\beta}_L - \beta}{2}(x_1 + x_0)] \\ &\geq (x_1 - x_0)[\hat{\alpha}_U - \alpha + \frac{\hat{\beta}_L - \beta}{2} \times 2\bar{x}] \\ &= (x_1 - x_0)[\hat{\alpha}_U - \alpha + \frac{\hat{\beta}_L - \beta}{2} \times 2\frac{\alpha - \hat{\alpha}_U}{\hat{\beta}_L - \beta}] \\ &= 0 \end{aligned}$$

Therefore, $MCS_2 \leq MCS \leq MCS_1$.

(2) Applying symmetric reasoning to the case where $|x_0 - \bar{x}| < |x_1 - \bar{x}|$ leads to $MCS_1 \leq MCS \leq MCS_2$.

(3) $|x_0 - \bar{x}| = |x_1 - \bar{x}|$ and $x_0 < \bar{x} < x_1$ jointly implies $x_0 + x_1 = 2\bar{x}$, in which case it follows from above that $MCS = MCS_1 = MCS_2$, MCS is point identified.

Proposition 4

Proof: If $|x_0 - \bar{x}| = |x_1 - \bar{x}|$ and $x_0 < \bar{x} < x_1$, then $x_0 + x_1 = 2\bar{x}$. Following from the third case of Proposition 3, we have $MCS = MCS_1 = MCS_2$. Thus MCS is point identified.

Footnote 7

Proof: Suppose we are facing the case 2 in Proposition 1 where β is only single bounded. Without loss of generality, we assume $\sigma_{\tilde{x}z_j} > 0$ and $\sigma_{xu}, \sigma_{z_j u} \geq 0$, $(j = 1, 2)$.

We need to demonstrate that if $\rho_{\tilde{x}z_1} > \rho_{\tilde{x}z_2}$ and $\rho_{z_1 u} < \rho_{z_2 u}$, then $\gamma = \dfrac{\sigma_{z_1}}{\sigma_{z_1} + \sigma_{z_2}}$ ensures that $\sigma_{\tilde{x}w(\gamma)} < 0$ and $\sigma_{w(\gamma)u} > 0$, and hence that Proposition 2 can be applied to obtain double-sided bounds.

The demonstration includes two parts: (1) if $\rho_{\tilde{x}z_1} > \rho_{\tilde{x}z_2}$, then $\sigma_{\tilde{x}w(\gamma)} < 0$ when $\gamma = \dfrac{\sigma_{z_1}}{\sigma_{z_1} + \sigma_{z_2}}$; and (2) if $\rho_{z_1 u} < \rho_{z_2 u}$, then $\sigma_{w(\gamma)} u > 0$ when $\gamma = \dfrac{\sigma_{z_1}}{\sigma_{z_1} + \sigma_{z_2}}$.

$$
\begin{aligned}
&(1) \quad \rho_{\tilde{x}z_2} < \rho_{\tilde{x}z_1} \\
&\Leftrightarrow \rho_{\tilde{x}z_2}\sigma_{z_2} < \rho_{\tilde{x}z_1}\sigma_{z_2} \\
&\Leftrightarrow \rho_{\tilde{x}z_1}\sigma_{z_1} + \rho_{\tilde{x}z_2}\sigma_{z_2} < \rho_{\tilde{x}z_1}\sigma_{z_1} + \rho_{\tilde{x}z_1}\sigma_{z_2} \\
&\Leftrightarrow \sigma_{z_1}(\rho_{\tilde{x}z_1}\sigma_{z_1} + \rho_{\tilde{x}z_2}\sigma_{z_2}) < \rho_{\tilde{x}z_1}\sigma_{z_1}(\sigma_{z_1} + \sigma_{z_2}) \\
&\Leftrightarrow \frac{\sigma_{z_1}}{\sigma_{z_1} + \sigma_{z_2}} < \frac{\rho_{\tilde{x}z_1}\sigma_{z_1}}{\rho_{\tilde{x}z_1}\sigma_{z_1} + \rho_{\tilde{x}z_2}\sigma_{z_2}} \\
&\Leftrightarrow \frac{\sigma_{z_1}}{\sigma_{z_1} + \sigma_{z_2}} < \frac{\rho_{\tilde{x}z_1}\sigma_{\tilde{x}}\sigma_{z_1}}{\rho_{\tilde{x}z_1}\sigma_{\tilde{x}}\sigma_{z_1} + \rho_{\tilde{x}z_2}\sigma_{\tilde{x}}\sigma_{z_2}} \\
&\overset{(i)}{\Leftrightarrow} \gamma < \frac{\sigma_{\tilde{x}z_1}}{\sigma_{\tilde{x}z_1} + \sigma_{\tilde{x}z_2}} \\
&\Leftrightarrow \gamma(\sigma_{\tilde{x}z_1} + \sigma_{\tilde{x}z_2}) < \sigma_{\tilde{x}z_1} \\
&\Leftrightarrow \gamma\sigma_{\tilde{x}z_1} + \gamma\sigma_{\tilde{x}z_2} - \sigma_{\tilde{x}z_1} < 0 \\
&\Leftrightarrow \gamma\sigma_{\tilde{x}z_2} - (1-\gamma)\sigma_{\tilde{x}z_1} < 0 \\
&\overset{(ii)}{\Leftrightarrow} \sigma_{\tilde{x}(\gamma z_2 - (1-\gamma)z_1)} < 0 \\
&\Leftrightarrow \sigma_{\tilde{x}w(\gamma)} < 0
\end{aligned}
$$

Note: Step (*i*) holds because the definition of correlation coefficient is $\rho_{\tilde{x}z_1} = \dfrac{\sigma_{\tilde{x}z_1}}{\sigma_{\tilde{x}}\sigma_{z_1}}$. Rearrange the equation we have $\rho_{\tilde{x}z_1}\sigma_{\tilde{x}}\sigma_{z_1} = \sigma_{\tilde{x}z_1}$. Similarly, $\rho_{\tilde{x}z_1}\sigma_{\tilde{x}}\sigma_{z_1} = \sigma_{\tilde{x}z_1}$ and $\rho_{\tilde{x}z_2}\sigma_{\tilde{x}}\sigma_{z_2} = \sigma_{\tilde{x}z_2}$. Step (*ii*) holds because:

$$
\begin{aligned}
&\sigma_{\tilde{x}(\gamma z_2 - (1-\gamma)z_1)} \\
&\equiv \operatorname{cov}(\tilde{x}, \gamma z_2 - (1-\gamma)z_1) \\
&= \operatorname{cov}(\tilde{x}, \gamma z_2) - \operatorname{cov}(\tilde{x}, (1-\gamma)z_1) \\
&= \gamma\operatorname{cov}(\tilde{x}, z_2) - (1-\gamma)\operatorname{cov}(\tilde{x}, z_1) \\
&\equiv \gamma\sigma_{\tilde{x}z_2} - (1-\gamma)\sigma_{\tilde{x}z_1}
\end{aligned}
$$

(2) $\rho_{z_2u} > \rho_{z_1u}$

$$\Leftrightarrow \rho_{z_2u}\sigma_{z_2} > \rho_{z_1u}\sigma_{z_2}$$

$$\Leftrightarrow \rho_{z_1u}\sigma_{z_1} + \rho_{z_2u}\sigma_{z_2} > \rho_{z_1u}\sigma_{z_1} + \rho_{z_1u}\sigma_{z_2}$$

$$\Leftrightarrow \sigma_{z_1}(\rho_{z_1u}\sigma_{z_1} + \rho_{z_2u}\sigma_{z_2}) > \rho_{z_1u}\sigma_{z_1}(\sigma_{z_1} + \sigma_{z_2})$$

$$\Leftrightarrow \frac{\sigma_{z_1}}{\sigma_{z_1} + \sigma_{z_2}} > \frac{\rho_{z_1u}\sigma_{z_1}}{\rho_{z_1u}\sigma_{z_1} + \rho_{z_2u}\sigma_{z_2}}$$

$$\Leftrightarrow \frac{\sigma_{z_1}}{\sigma_{z_1} + \sigma_{z_2}} > \frac{\rho_{z_1u}\sigma_{z_1}\sigma_u}{\rho_{z_1u}\sigma_{z_1}\sigma_u + \rho_{z_2u}\sigma_{z_2}\sigma_u}$$

$$\Leftrightarrow \gamma > \frac{\sigma_{z_1u}}{\sigma_{z_1u} + \sigma_{z_2u}}$$

$$\Leftrightarrow \gamma(\sigma_{z_1u} + \sigma_{z_2u}) > \sigma_{z_1u}$$

$$\Leftrightarrow \gamma\sigma_{z_1u} + \gamma\sigma_{z_2u} - \sigma_{z_1u} > 0$$

$$\Leftrightarrow \gamma\sigma_{z_2u} - (1-\gamma)\sigma_{z_1u} > 0$$

$$\Leftrightarrow \sigma_{(\gamma z_2 - (1-\gamma) z_1)u} > 0$$

$$\Leftrightarrow \sigma_{w(\gamma)u} > 0$$

Appendix B: Robustness Check Results

Table 4.B1. Demand Function Estimation based on Semi-Log Hedonic Price Function (Individual State)

	(1) OLS	(2) IV1	(3) IV2	(4) IIV	(5) 95% CI
Panel A: Vermont					
WT	-412.65	-2529.75***	-3331.19**	(-∞, -3331.19]	(-∞, -215.48)
	(284.08)	(844.83)	(1551.97)		
$x_0=2.6$, $\bar{x}=4.6$, $x_1=5.5$					
MCS ($\bar{x} \uparrow x_1$)				[0, 3245.36]	(0, 4576.17)
MCS ($\bar{x} \downarrow x_0$)				(-∞, -16872.34]	(-∞, -10794.15)
MCS ($x_0 \uparrow x_1$)				[20117.69, +∞)	(15370.32, +∞)
Panel B: Maine					
WT	102.38	-854.62	-880.32	(-∞, -880.32]	(-∞, 655.83)
	(279.66)	(620.65)	(778.08)		
$x_0=1.7$, $\bar{x}=4.6$, $x_1=5.2$					
MCS ($\bar{x} \uparrow x_1$)				[0, 2335.39]	N/A[a]
MCS ($\bar{x} \downarrow x_0$)				(-∞, -15755.36]	N/A
MCS ($x_0 \uparrow x_1$)				[18090.75, +∞)	N/A
Panel C: New Hampshire					
WT	-2614.82***	-8555.07***	-10639.71***	(-∞, -10639.71]	(-∞, -6389.20)
	(705.37)	(1879.79)	(2155.21)		
$x_0=2.3$, $\bar{x}=4.9$, $x_1=5.3$					
MCS ($\bar{x} \uparrow x_1$)				[0, 4371.81]	(0, 4689.73)
MCS ($\bar{x} \downarrow x_0$)				(-∞,-69911.67]	(-∞, -55401.15)
MCS ($x_0 \uparrow x_1$)				[74283.48, +∞)	(60090.89, +∞)

Note: [a]As long as we believe the demand curve for WT is downward sloping, the coefficient of WT should be negative. Since the upper bound of 95% confidence interval is positive, we did not report the corresponding welfare measure.

Table 4.B2. Demand Function Estimation based on Semi-Log Hedonic Price Function (Pooled Data)

	(1) OLS	(2) IV1	(3) IV2	(4) IIV	(5) 95% CI
Pooled Data					
WT	-1172.43***	-4697.53***	-6687.28***	$(-\infty, -6687.28]$	$(-\infty, -4490.32)$
	(381.65)	(942.78)	(1117.78)		
$x_0=2.1$, $\bar{x}=4.7$, $x_1=5.4$					
MCS ($\bar{x} \uparrow x_1$)				[0, 4460.79]	(0, 4913.41)
MCS ($\bar{x} \downarrow x_0$)				$(-\infty, -45257.04]$	$(-\infty, -37513.27)$
MCS ($x_0 \uparrow x_1$)				$[49717.83, +\infty)$	$(42426.68, +\infty)$

Reference

Bajari, P., and C. L. Benkard (2005). Demand estimation with heterogeneous consumers and unobserved product characteristics: a hedonic approach, *Journal of Political Economy*, 113(6), 1239-1276.

Bajari, P., and M. E. Kahn (2005). Estimating housing demand with an application to explaining racial segregation in cities, *Journal of Business and Economic Statistics*, 23(1), 20-33.

Bartik, T. J. (1987). The estimation of demand parameters in hedonic price models, *Journal of Political Economy*, 95(1), 81-88.

Bayer, P., R. McMillan, A. Murphy, and C. Timmins (2008). A dynamic model of demand for houses and neighborhood, *Working Paper*.

Bishop, K., and C. Timmins (2008). Simple, consistent estimation of the marginal willingness to pay function: recovering Rosen's second stage without instrumental variables, *working paper*.

Bishop, K., and C. Timmins (2011). Hedonic prices and implicit markets: consistent estimation of marginal willingness to pay for differentiated products without exclusion restrictions, *working paper*.

Boyle, K. J., P. J. Poor, and L. O. Taylor (1999). Estimating the demand for protecting freshwater lakes from eutrophication, *American Journal of Agricultural Economics*, 81(5), 1118-1122.

Chattopadhyay, S. (1999). Estimating the demand for air quality: new evidence based on the Chicago housing market, *Land Economics*, 75(1), 22-38.

Chay, K. Y., and M. Greenstone (2005). Does air quality matter? evidence from the housing market, *Journal of Political Economy*, 113(2), 376-424.

Cheshire, P., and S. Sheppard (1998). Estimating the demand for housing, land, and neighborhood characteristics, *Oxford Bulletin of Economics and Statistics*, 60(3), 357-382.

Cobb, S. (1984). The impact of site characteristics on housing cost estimates, *Journal of Urban Economics*, 15(1), 26-45.

Deyak, T. A., and V. K. Smith (1974). Residential property values and air pollution: some new evidence, *Quarterly Review of Economics and Business*, 14(9), 93-100.

Driscoll, P., B. Dietz, and J. Alwang (1994). Welfare analysis when budget constraints are nonlinear: The case of flood hazard protection, *Journal of Environmental Economics and Management*, 26(2), 181-199.

Epple, D. (1987). Hedonic prices and implicit markets: estimating demand and supply functions for differentiated products, *Journal of Political Economy*, 95(1), 59-80.

Epple, D., and G. J. Platt (1998). Equilibrium and local redistribution in an urban economy when households differ in both preferences and incomes, *Journal of Urban Economics*, 43(1), 23-51.

Epple, D., and H. Sieg (1999). Estimating equilibrium models of local jurisdictions, *Journal of Political Economy*, 107(4), 645-681.

Epple, D., and R. E. Romano (1998). Competition between private and public schools, vouchers, and peer-group effects, *American Economic Review*, 88(1), 33-62.

Epple, D., R. Filimon, and T. Romer (1984). Equilibrium among local jurisdictions: toward an integrated treatment of voting and residential choice, *Journal of Public Economics*, 24(3), 281-308.

Gibbs, J. P., Halstead, J. M., Boyle, K. J., and Huang J. C. (2002). An hedonic analysis of the effects of lake water clarity on New Hampshire lakefront properties, *Agricultural and Resource Economics Review*, 31(1), 39-46.

Hsu, T. I. (2000). Hedonic study of the effects of lake-water clarity and aquatic plants on lakefront property prices in Vermont, *University of Maine, Master Thesis*.

Klaiber, H. A., and D. J. Phaneuf (2010). Valuing open space in a residential sorting model of the Twin Cities. *Journal of Environmental Economics and Management*, 60(2), 57-77.

Kosenius, A. K. (2010). Heterogeneous preferences for water quality attributes: The Case of eutrophication in the Gulf of Finland, the Baltic Sea. *Journal of Ecological Economics*, 69(3), 528-538.

Kuminoff, N. V., and J. C. Pope (2012). A novel approach to identifying hedonic demand parameters. *Economics Letters,* forthcoming.

Kuminoff, N. V., C. F. Parmeter, and J. C. Pope (2010). Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities? *Journal of Environmental Economics and Management*, 60(3), 145-160.

Kuminoff, N. V., V. K. Smith, and C. Timmins (2010). The new economics of equilibrium sorting and its transformational role for policy evaluation. *National Bureau of Economic Research*.

Lawson, S. R. (1997). Estimating the benefits of water quality in Maine's lakes: a hedonic property value model, *Master thesis, University of Maine*.

Linneman, P. (1980). Some empirical results on the nature of the hedonic price function for the urban housing market, *Journal of Urban Economics*, 8(1), 47-68.

Linneman, P. (1981). The demand for residence site characteristics, *Journal of Urban Economics*, 9(2), 129-148.

Nevo, A., and A. M. Rosen (2010). Identification with imperfect instruments, *The Review of Economics and Statistics*, forthcoming.

Ohsfeldt, R. L. (1988). Implicit markets and the demand for housing characteristics, *Regional Science and Urban Economics*, 18(3), 321-343.

Palmquist, R. B. (1984). Estimating the demand for the characteristics of housing, *The Review of Economics and Statistics*, 66(3), 394-404.

Parmeter, C. F., and J. C. Pope (2009). Quasi-experiments and hedonic property value methods, *prepared for the Handbook on Experimental Economics and the Environment*.

Poor, P. J., K. J. Boyle, L. O. Taylor, and R. Bouchard (2001). Objective versus subjective measures of water clarity in hedonic property value models, *Land Economics*, 77(4), 482-493.

Sieg, H., V. K. Smith, H. S. Banzhaf, and R. P. Walsh (2002). Interjurisdictional housing prices in locational equilibrium, *Journal of Urban Economics*, 52(1), 131-153.

Sieg, H., V. K. Smith, H. S. Banzhaf, and R. P. Walsh (2004). Estimating the general equilibrium benefits of large changes in spatially delineated public goods, *International Economic Review*, 45(4), 1047-1077.

Smeltzer, E., and S. Heiskary (1990). Analysis and application of lake survey data, *Lake and Reservoir Management*, 6(1), 109-118.

Smith, V. K., and T. A. Deyak (1975). Measuring the impact of air pollution on property values, *Journal of Regional Science*, 15(3), 277-288.

Smith, V. K., H. Sieg, H. S. Banzhaf, and R. P. Walsh (2004). General equilibrium benefits for environmental improvements: projected ozone reductions under EPA's prospective analysis for the Los Angeles air basin, *Journal of Environmental Economics and Management*, 47(3), 559-584.

Tiebout, C. M. (1956). A pure theory of local expenditures, *Journal of Political Economy*, 64(5), 416-424.

Witte, A. D., H. J. Sumka, and H. Erekson (1979). An estimate of a structural hedonic price model of the housing market: an application of Rosen's theory of implicit markets, *Econometrica*, 47(5), 1151-1173.

Zabel, J. E., and K. A. Kiel (2000). Estimating the demand for air quality in four U.S. cities, *Land Economics*, 76(2), 174-194.

Chapter 5

Conclusion

This dissertation focuses on the methodology and applications of the hedonic price model, which is widely used to estimate implicit values for changes in environmental quality. The dissertation consists of three novel essays. The first essay is a "traditional" hedonic-price model where the 1st stage price function is estimated with a cross-sectional dataset. The novel aspect of this research is the property-specific data that is used to investigate price impacts on lakefront properties. The second essay exploits a quasi-experimental with a "traditional" hedonic model, and implements a novel internal meta analysis to summarize statistical findings. The third essay develops a new method to estimating 2nd stage hedonic demand estimation using partial identification to identify welfare for nonmarginal changes in environmental quality.

Chapter 2 uses a "traditional" hedonic model to estimate the effect of an invasive aquatic plant, Eurasian watermilfoil, on lakefront property values at selected Vermont lakes. The unique contribution of this essay is the use of a qualitative scale that indicates the percent the lake surface covered by Eurasian watermilfoil immediately in front of each property, and the data differentiates between native plant growth and milfoil. Thus, the environmental variable in the hedonic price function is unique to each property, and it is possible to see if property prices are responsive to a specific plant species versus all plant growth. This data refinement is better able to capture price effects of Eurasian watermilfoil than more commonly used environmental variables that are crude indicators of the presence or absence of amenities or disamenities used in previous studies. The results indicate that total plant growth, not solely milfoil, affects property values. Thus, milfoil, as a component of total plant growth, decreases property values by 1-16%. This suggests that policies to prevent the spread of the invasive Eurasian watermilfoil will protect property values and will consequently protect local property tax revenues; thereby providing a benefit to local citizens.

The limitations of this study is that there were only a small number of observations (65) and the presence of Eurasian milfoil may be correlated with variables that there was no data to include in the model, e.g., frequency of boats accessing the lake from boat launches that may have milfoil on the boats or trailers and that this boat use might be undesirable to property owners. This suggest that future research is needed where more extensive data are available on the presence of Eurasian watermilfoil in many lakes and there is an opportunity to apply a quasi-experimental design as discussed in the second model.

Chapter 3 uses a quasi-experimental hedonic model to explore the impact that the 9/11 terrorist attacks had on property prices near mosques. A “traditional” hedonic model without the quasi-experimental design may lead to biased parameter estimates because of omitted variables that confound identification of an effect when using cross-sectional data. One concern with the quasi-experimental hedonic model is that it is difficult to establish whether the event or policy change is truly exogenous. If the event or policy change was expected by households, then the expectations will blur the discrete timing of when the event or policy change occurs and it becomes more difficult to interpret the differences in property prices before and after the change causally. However, our application makes an ideal quasi-experiment since the 9/11 attacks can clearly be viewed as an unexpected and exogenous event. The results indicate that the presence of selected mosques in the eastern U.S. results in property price diminutions. A novel extension of this research is the use of an internal meta analysis to compare results across mosques, and the results of this analysis reveals that there is more likely to be an effect of mosques in areas with high population density and a high proportion of the population that is white.

A key limitation of both of the first two applications is that 1^{st} stage hedonic models only reveal marginal changes in value that do not reveal values for nonmarginal changes in quality that are often the policy change that economic values are needed to evaluate the feasibility of changes. This issue is addressed in the third essay that presents a new approach to estimating the 2^{nd} stage hedonic demand functions.

The key contribution of the 2^{nd} stage approach described in Chapter 3 is that what is called "the endogeneity problem" can be addressed if we do not limit ourselves to the extreme of point identification of demand parameters. If we take a broader perspective on identification, we can partially identify demand functions and welfare measures. After developing the conceptual and empirical frameworks for partial identification of 2^{nd} stage hedonic demand functions, an application to lake water quality is presented. While it is potentially possible to identify lower and upper bounds on demand parameters and welfare estimates, I find for this application it is only possible to identify a lower bound, which benefits will not exceed.

The limitation of partial identification is that it is only possible to provide policy makers on bounds for welfare measures related to policy changes; It is not possible to say a policy change has a point welfare estimate of $x. The advantage, however, is that the point estimate is likely biased and the bounds provide evidence of bounds where welfare will not fall below or will not exceed. These types of bounds provide a way to explicitly consider uncertainty related to benefit estimates in policy analyses.

To sum up, this dissertation investigates several important issues in the hedonic price model:

1st. linking property specific environmental quality to property sales to estimate marginal values;

2nd. using a quasi-experimental design to identify marginal values and an internal meta-analysis to identify reasons for differences in implicit prices across study locations; and

3rd. developing a novel partial identification approach to measure welfare associated with non-marginal change in the quality of public goods.

These three essays provide insights that can enhance future applications of hedonic models and provide important value information for current policy issues at the national, state and local levels.