ABSTRACT

TÍMÁR, LEVENTE. Modeling the Anthropogenic Spread of an Aquatic Invasive Species: The Case of Zebra Mussels and Transient Recreational Boating in Wisconsin. (Under the direction of Daniel J. Phaneuf.)

Transient recreational boating is the principal overland vector of dispersal for several freshwater invasive species. Traditional models of human-mediated aquatic invasions use aggregate data on boating patterns, and are unable to incorporate changes in boater behavior. This limits their practicality because most control policies, by targeting the human vector, are expected to have behavioral consequences, and a complete policy assessment needs to consider the resultant shift in invasion pressure. In this dissertation, I develop a framework that utilizes site choice probabilities from a travel cost recreation demand model to quantify anthropogenic invasion threat to a water body. This integrated behavioral approach enables the researcher to trace the effects of any perturbation that elicits a behavioral response, and to translate them into changes in site-specific invasion probabilities. I employ spatial and temporal data on the distribution of zebra mussels in Wisconsin, and the results of a large revealed preference boater survey in the state to empirically implement the model. Finally, I demonstrate via counterfactuals that a control policy that reduces the desirability of an infested choice alternative is less effective than a traditional model would imply. In some situations, the policy may be not only expensive in terms of lost consumer surplus, but also counterproductive in terms of increasing (instead of decreasing) the probability of zebra mussel spread to the uninfested water bodies it is designed to protect. Similar unintended consequences would be impossible to identify in a traditional, non-behavioral approach.
Modeling the Anthropogenic Spread of an Aquatic Invasive Species: The Case of Zebra Mussels and Transient Recreational Boating in Wisconsin

by
Levente Timár

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Economics

Raleigh, North Carolina 2008

APPROVED BY:

_________________________                           _________________________
Daniel J. Phaneuf            Barry K. Goodwin
Chair of Advisory Committee

_________________________                           _________________________
Roger H. von Haefen          Laura O. Taylor
Dedication

To Mariann whom I hope to spend the rest of my life with

To Mom and Dad who inspired me from an ocean away
Biography

Levente Tímár was born in Székesfehérvár, Hungary in 1978 to parents Béla Tímár and Irma Molnár. After graduating from high school, Levente was lured to the United States by the collegiate athletic system: he dreamed of receiving a university education while pursuing his running career. He accepted an athletic scholarship from the University of Northern Iowa, where he would become a multiple academic All-American in track and field and cross-country, and earn a Bachelor of Arts degree in economics in the spring of 2001. He began his doctorate in economics at North Carolina State University later that year while working as a graduate teaching assistant. Upon passing the Ph.D. written preliminary examinations two years later, Levente moved back to his native Hungary to focus on his running ambitions once again for the upcoming Olympic year. He returned to Raleigh on a research assistantship in the fall of 2004, developed an interest in environmental economics, and decided to conduct his dissertation research in the field. He concluded his graduate studies in August 2008.
Acknowledgements

I appreciate the guidance the chair of my doctoral committee, Daniel Phaneuf, provided throughout my graduate career, and this dissertation in particular. His experience and knowledge have benefited me immensely; his dedication often motivated me to work harder. I have learned a lot from Dan, and feel fortunate to have had him as my advisor. Likewise, I am grateful for the valuable insight Barry Goodwin, Roger von Haefen, Laura Taylor and Ada Wossink offered. Their comments contributed not only to the quality of my research, but to my training as an environmental economist as well.

I thank my parents for all the sacrifice they made for me, their unfailing support and encouragement. I thank my Mom who shed tears every time (twenty) over the last eleven years when I left for the United States, but did not once question my decision to do so, and I thank my Dad whose good spirits and optimism radiated through the phone line whenever we talked.

Finally, I wish to thank Mariann for everything she is. Though an ocean may lie between us, she has always been beside me. Her strength, intelligence and ambition were an inspiration, and I became a better person through her. Mariann, thank you for sharing your life with me.
Table of Contents

List of Tables ........................................................................................................................ vii
List of Figures ........................................................................................................................ viii

1 Introduction ........................................................................................................................ 1
   1.1 The Economics of Biological Invasions .............................................................. 2
   1.2 An Aquatic Invader: the Zebra Mussel ............................................................ 5
   1.3 Modeling Framework ...................................................................................... 7
   1.4 Plan of the Dissertation .................................................................................. 8

2 A Review of the Ecological Problem and the Relevant Literature ......................... 9
   2.1 Zebra Mussel Background .......................................................................... 10
   2.2 Ecological and Economic Damages ............................................................. 12
   2.3 Ecological Literature ................................................................................... 15
   2.4 Economic Literature .................................................................................... 24

3 Data ................................................................................................................................. 28
   3.1 Zebra Mussel Data ....................................................................................... 29
   3.2 Wisconsin Hydrography Data ...................................................................... 30
   3.3 Boating Data ................................................................................................. 31
   3.4 Lake Clarity Data .......................................................................................... 37

4 Analysis ........................................................................................................................... 51
   4.1 The Random Utility Model ............................................................................ 52
      Data Preparation ............................................................................................ 53
      Specification .................................................................................................. 59
      Estimation ...................................................................................................... 64
      Results ............................................................................................................. 69
   4.2 The Discrete Duration Model .......................................................................... 73
      Specification .................................................................................................. 74
      Estimation ...................................................................................................... 75
      Results ............................................................................................................. 77
4.3 The Integrated Economic and Ecological Model ........................................... 77
  The Human Threat Variable ................................................................. 78
  The Integrated Model ............................................................................ 84
  Discussion of Estimation Results ......................................................... 88
  Environmental Threat ........................................................................... 91
  Model Evaluation .................................................................................. 94

5 Policy Simulations and Conclusions ........................................................ 114
  5.1 Clarity Feedback ................................................................................ 115
  5.2 Policy Counterfactuals ...................................................................... 117
  5.3 Concluding Remarks ........................................................................ 126

References ................................................................................................. 140

Appendices ................................................................................................. 145
  Appendix A The Questionnaire of the Wisconsin Recreational Boating
  Study ........................................................................................................ 146
  Appendix B Corrected Second-Step Standard Errors ............................. 150
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 3.1</td>
<td>Boating data summary</td>
<td>42</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>The temporal distribution of boating trips</td>
<td>43</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>Seasonal variation in statewide mean water clarity</td>
<td>44</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Random utility model estimation results</td>
<td>103</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Random utility model specification summaries</td>
<td>104</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>Hazard model estimation results</td>
<td>105</td>
</tr>
<tr>
<td>Table 4.4</td>
<td>Integrated model estimation results</td>
<td>106</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>Policy and focus lakes</td>
<td>130</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>Lake Winnebago policies</td>
<td>131</td>
</tr>
<tr>
<td>Table 5.3</td>
<td>Lake Monona policies</td>
<td>132</td>
</tr>
<tr>
<td>Table 5.4</td>
<td>Welfare analysis</td>
<td>133</td>
</tr>
<tr>
<td>Table B.1</td>
<td>Two-step statistical inference</td>
<td>158</td>
</tr>
</tbody>
</table>
List of Figures

| Figure 3.1 | The distribution of zebra mussels in the Upper Midwest as of 2006 .......... | 45 |
| Figure 3.2 | Boating survey study area ................................................................. | 46 |
| Figure 3.3 | The number of boating days by county ................................................ | 47 |
| Figure 3.4 | Seasonal variability in statewide mean Secchi-depth ............................ | 48 |
| Figure 3.5 | Inland lake clarity prediction surface over the state of Wisconsin .......... | 49 |
| Figure 3.6 | Secchi-disk readings over time in Little Muskego Lake .......................... | 50 |
| Figure 4.1 | The spatial distribution of alternative specific constants ....................... | 107 |
| Figure 4.2 | The distribution of the residences of surveyed boaters by zip code area .... | 108 |
| Figure 4.3 | The human threat variable as a function of site choice probability .......... | 109 |
| Figure 4.4 | Invasion predictions from Model I .......................................................... | 110 |
| Figure 4.5 | Invasion predictions from Model II .......................................................... | 111 |
| Figure 4.6 | Invasion predictions from Model III .......................................................... | 112 |
| Figure 4.7 | Maximum simulated invasion probabilities over time in Model I ............... | 113 |
| Figure 5.1 | Invasion predictions from Model I with a clarity feedback effect .............. | 134 |
| Figure 5.2 | The geographic location of policy lakes and focus lakes ....................... | 135 |
| Figure 5.3 | Absolute changes in predicted choice shares from the Lake Winnebago ... | 136 |
| Figure 5.4 | Percentage changes in invasion probabilities from the Lake Winnebago ... | 137 |
| Figure 5.5 | Absolute changes in predicted choice shares from the Lake Monona ... | 138 |
| Figure 5.6 | Percentage changes in invasion probabilities from the Lake Monona ... | 139 |
Chapter 1

Introduction
Invasive species are the focus of considerable interdisciplinary attention: biological invasions have been addressed by various fields of science using frameworks as diverse as pest-control, predator-prey models, pollution abatement, risk analysis, dynamic control, queuing theory and even the physical theory of nucleation.\(^1\) Most biological invasions are directly caused by human activity. In this dissertation, I examine modeling the human behavior responsible for spreading an aquatic pest, and then model the spread of the invasive organism as a function of that behavior. This chapter lays down the economic foundation of biological invasions in general, and of aquatic invasions in particular. I also introduce zebra mussels, the freshwater invasive specific to my application, briefly describe the modeling framework I employ, and preview some of the findings of my investigation.

1.1 The Economics of Biological Invasions

Species that are moved beyond their natural range are referred to as non-indigenous in their new habitat. The movement of species is sometimes intentional and beneficial: non-indigenous food crops and livestock provide more than ninety-nine percent of the United States agricultural output (Pimentel, Lach, Zuniga & Morrison, 2000). Whether introduced intentionally or unintentionally, non-indigenous species often successfully compete with or prey upon native species. Because they typically lack natural predators, they may become invasive and inflict great ecological and economic harm. Ecosystems are so severely affected that invasive species are the second most important cause of biodiversity loss (Holmes, 1998), and have led to more vertebrate extinctions than direct human exploitation (Olson and

---

\(^1\) For an analysis of exotic species introductions using nucleation theory (originally developed to model the process of crystallization) see Korniss and Caraco (2005).
Roy, 2002). It is estimated that there are around 50,000 non-indigenous species in the United States, and that they cause over $136 billion of damages annually (Pimentel et al., 2000). This figure is most likely an underestimate because it deals with only a subset of species and impacts, and ignores many non-market effects. The agricultural sector (which, as noted, is itself mostly made up of non-indigenous species) loses approximately one fourth of the value of its potential output due to invasive plants and the costs of controlling them (Simberloff, 1996). The damages sustained by other sectors of the economy are also substantial – sectors as diverse as water and energy supply, recreation and even transportation are affected. Moreover, losses are likely to rise even further in the future: the rate of spread of invasive species is increasing, and reaches 1,800 hectares per day in certain parts of the country (Eiswerth & Johnson, 2002).

Human activities, particularly those associated with the movement of people and commodities, are largely responsible for the transportation and introduction of non-indigenous species. Other human activities result in habitat fragmentation and agricultural disturbance, and increase the susceptibility of ecosystems to biological invasions, thereby facilitating the establishment of exotic species after they have been introduced. For both of these reasons, high economic activity in a region tends to be associated with high rates of biological invasions (Perrings, Williamson & Dalmazzone, 2000). Indeed, biological invasions are becoming more frequent (Heywood; Parker et al. in Horan, Perrings, Lupi & Bulte, 2002), and are unlikely to subside because of the intensification of agricultural production and the increasingly open global economy.
The economics of invasive species deals with the interdependence of biological invasions and human behavior. Virtually all research (even non-economic studies) recognize that people are ultimately responsible for transporting non-indigenous species. The other side of the interaction is nonetheless often ignored: the risks associated with the spread of an invasive may directly or indirectly change the very behavior that is responsible for transporting the organism. That is, the actions of people are not completely exogenous; causality may operate in both directions. I attribute the general failure to address the latter point to two well-known facts about biological invasions. First, unintentional non-indigenous species introductions are typically external to the market: prices do not reflect all expected economic losses resulting from the invasion risk imposed by an activity. Several introductions have been ascribed to the failure of people to change their behavior in response to a change in the risks associated with that behavior, because they did not themselves face the resulting costs (Perrings et al., 2002). Second, another type of market failure dominates the post-introduction spread of invasive species: control effort is a public good. If left to the uncoordinated actions of individual decision makers, the level of control will be insufficient from a social perspective. The benefits of an inspection routine or a quarantine policy, for example, are neither rival nor exclusive. Both of these arguments underscore the lack (or sub-optimality) of a direct behavioral response. However, neither imply that the behavioral component can be ignored. To align individual incentives with policy goals, a federal or local authority may implement various policies specifically designed to influence the actions of individual decision makers. Since an effective policy will elicit a behavioral response, it will also affect the risk of introduction or the spread pattern of exotic species. Modeling the
potential behavioral response is especially important when the objective is to evaluate different policy options. Nonetheless, even detailed ecological case studies tend to ignore this human dimension of biological invasions. Perrings et al. (2002) correctly note that "economics is much more than just a method for calculating costs. It is a framework for understanding the complex casual interactions between human behavior and natural processes, and for finding institutional and behavioral solutions to seemingly intractable environmental problems." My objective is to examine the spread of the zebra mussel to inland lakes in Wisconsin in this spirit, and allow for the possibility of adjustments in the actions of individual decision makers.

1.2 An Aquatic Invader: the Zebra Mussel

One of the most serious threats to the world's freshwater habitats is the zebra mussel (Dreissena polymorpha), a small species of mollusk whose common name comes from its dark, zebra-like stripes (Aldridge et al., 2004). The U.S. Fish and Wildlife Service estimates the potential economic losses from zebra mussels within the Great Lakes region alone for the ten-year period from 2000 to 2010 to be around $5 billion. The colonization of North American lakes by the zebra mussel is the subject of my dissertation not only because of the magnitude of the damages caused, but also because a combination of unique characteristics make this invasion especially suitable for economic analysis. In general, “freshwater habitats offer a particularly tractable system for investigating the spread of exotic species” (Johnson, Ricciardi & Carlton, 2001, p. 1789), because they have well-defined borders and connections. Moreover, the relationship between human activities and the spread of aquatic
pests is remarkably strong: many aquatic species, among them zebra mussels, require a vector to move between unconnected water bodies. Ecological evidence indicates that a very specific human activity, recreational boating, is the most potent vector for several exotic plants and animals, and is the main cause of the overwhelming majority of inland lake zebra mussel invasions (Johnson & Carlton, 1996).

Other aquatic invasive species that are spread primarily by recreational boaters include quagga mussels, the Eurasian watermilfoil, the purple loosestrife and hydrilla. In some instances, the spread of these various exotic plants and animals may be related since mussels are most often transported to other water bodies on vegetation entangled in the boat’s rotor or the trailer. An important difference is that zebra mussels can survive several days of aerial exposure (and even weeks under ideal conditions), while most aquatic plants require a relatively short time period between successive boat launches for survival. Host plant survival is of course not required for mussel survival. Furthermore, once they are introduced, zebra mussels are practically impossible to eradicate from a water body, considerably simplifying the economic analysis of their dispersal (and limiting available policy options). A final argument for concentrating my efforts on zebra mussels is that their significant impact on both economic and ecological systems, and their unprecedented rate of spread in the first few years after introduction made them the focus of legislative action and monitoring efforts. Therefore, relatively accurate data on their spatial and temporal distribution exists.
1.3 Modeling Framework

One method of accomplishing my objectives is to explicitly model the behavior of the human vector, and then to model the spread of the invasive in a manner that takes vector behavior as an input. In this framework, anything that elicits a behavioral response will also affect invasion risk at various sites. A similar interdependence has been explored in the literature of human diseases: the probability of infection affects human behavior, which in turn affects the epidemiology of the disease. In the present context, by responding to zebra mussel control policies, people may change the dynamics of mussel spread.

Using a large boater survey, I model recreation decisions via a multinomial logit travel cost model. The model generates site choice probabilities for all boaters and all water bodies in the choice set. Any new invasion event requires that the same boater visits both an infested site and an uninfested site (within a short time period). I construct a time- and site-specific variable that captures the threat of mussel transport to uninfested water bodies, and is based on the set of currently infested sites and the multinomial logit trip probabilities. Finally, this variable is used as an input in a discrete duration model that results in site-specific invasion probabilities. The model is able to explain the currently observed spatial distribution of zebra mussels in the study area fairly well, even when using mussel distribution data from more than a decade ago to establish initial conditions. In this framework, the effect of a hypothetical policy change on mussel spread can also be assessed because the behavioral model enables us to predict post-policy trip probabilities and the corresponding new invasion probabilities.
1.4 Plan of the Dissertation

The rest of this dissertation is divided into four chapters. In the following (second) chapter, I examine the reasons behind the success of zebra mussels as an invasive species, describe the various types of damages they cause, and review the relevant ecological and economic literature. In the third chapter, I summarize the major data sets I employ. They include spatial and temporal data on the distribution of zebra mussels, a large revealed preference boater survey from the Wisconsin Department of Natural Resources that I use to estimate the behavioral component of the model, and data on water clarity measurements in Wisconsin inland lakes. Chapter four contains most of the empirical analysis. I specify the discrete choice and discrete duration models in this chapter, integrate the two via the threat variable constructed from the site choice probabilities, and present estimation results. I also consider various methods to evaluate model performance and compare different specifications. The last chapter presents the results of policy simulations. I demonstrate that a zebra mussel control policy may either increase or decrease the probability of mussel invasion at specific lakes. The direction of the probability change is not always intuitive because of the policy-induced redistribution of boating trips. Thus, to assess the effectiveness of a policy passed in response to the spread of an aquatic invader, it is important to explicitly model recreational boater behavior.
Chapter 2

A Review of the Ecological Problem and the Relevant Literature
The impact of zebra mussels on the economy and the environment has lead to both legislative action to slow their spread, and substantial scientific interest in the success of this aquatic pest. A large volume of ecological literature has emerged to determine lake characteristics needed for mussel survival and reproduction, to identify their primary vectors of transmission, and to analyze spatial and temporal aspects of their spread. In this chapter, I recount the initial introduction into North America and the rapid post-introduction invasion, and identify the characteristics of the zebra mussel that make it an especially successful freshwater invader. Mussel invasion may lead to environmental degradation and substantial economic losses. I discuss a wide range of actual and potential damages in detail, and existing control mechanisms. Finally, I review the relevant literature, and consider how my research supplements the existing body of knowledge on the subject.

2.1 Zebra Mussel Background

Zebra mussels grow to a maximum length of approximately fifty millimeters, but most adults are about the size of a fingernail. They inhabit freshwater ecosystems at depths of two to seven meters, and live four to five years. Native to Eastern and Central Europe, they were first detected in North America in Lake St. Clair, a small water body connecting Lake Huron and Lake Erie, in 1988. It is generally believed that larval zebra mussels were introduced into Lake St. Clair when a ship sailing from a European freshwater port discharged its ballast
water into the lake.\(^2\) Within only two years, mussels were detected in all five of the Great Lakes. They soon entered the Mississippi watershed via shipping routes, and were spotted in Louisiana, thousands of kilometers downstream from the original point of introduction, in 1992. The Nonindigenous Aquatic Nuisance Prevention and Control Act of 1990 was passed due to the concern caused by the explosive spread of the zebra mussel. In fact, their rate of spread was so alarming that Ludyanski et al. predicted that “by the year 2000, the zebra mussel can be expected to have colonized all North American rivers, lakes and reservoirs that fit its broad ecological requirements” (Kraft & Johnson, 2000, p. 993). Dispersal between unconnected watersheds, however, requires a vector, and has been slower than initially anticipated. Wisconsin is one of the most severely affected states: about fifty of its inland lakes have been confirmed to contain zebra mussels – a small fraction of the total number of lakes in the state. The first invasions west of the Rocky Mountains did not occur until January 2008, when mussels were found in the San Justo Reservoir in California, and in the Pueblo Reservoir in Colorado.

Zebra mussels possess characteristics that are unusual for freshwater benthic species. They are active filter feeders meaning that they are capable of actively pumping water for filtration. They do this so effectively that a single mussel can filter a liter of water each day. According to information found on the U.S. Geological Survey’s Great Lakes Science Center webpage, mussel densities of more than one million per square meter have been recorded in some parts of Lake Erie. When mussels are so abundant, their removal of suspended

---

\(^2\) Cargo ships use ballast water to ensure stability and maneuverability when not carrying a full load. Ships take on ballast water in bays, estuaries and inland waters, and discharge it in ports-of-call before cargo is loaded. The global movement of ballast water now appears to be the single largest mechanism for marine non-indigenous species introductions, and it is estimated that over 4000 different species are being moved around the globe in ship ballast tanks every day (Batabyal & Beladi, 2006).
particulate matter from the water may dramatically increase clarity: mean Secchi-disk\textsuperscript{3} transparencies almost doubled during the first years of invasion in western Lake Erie (Leach in Neary & Leach, 1992). As the sudden increase in water clarity may indicate, zebra mussels are highly productive: a single female produces up to a million eggs in a spawning season. Their larvae, called veligers, are free swimming, and spend weeks of development in this stage, ensuring the widespread dissemination of offspring. Adult zebra mussels, on the other hand, attach themselves securely to hard surfaces by byssal threads – another characteristic not generally found in freshwater bivalves. Remarkably, they may also move and reattach themselves. But the most important characteristic that contributes to their success is their hardiness: zebra mussels can colonize lakes with a wide range of physical and chemical characteristics (Padilla, Chotkowski & Buchan, 1996), and even survive prolonged periods of aerial exposure. With this unique combination of traits, they found a previously unoccupied niche in North America, and became a particularly successful invader.

2.2 Ecological and Economic Damages

Zebra mussels pose a range of ecological and economic threats. They can fundamentally alter river and lake ecosystems and are known to have caused extinctions of many aquatic species worldwide (Ricciardi, Neves & Rasmussen, 1998). In North America, they have already lead to declines in the richness and abundance of native unionid mussels, an important component of freshwater biodiversity (Drake & Bossenbroek, 2004). Their filtering activity can remove a large fraction of phytoplankton and some small forms of zooplankton from a lake –

\textsuperscript{3}Secchi-disk readings measure the clarity (transparency) of water by submerging a disk in the water and measuring the depth at which it can no longer be seen by naked eye.
microscopic plants and animals that form the base of the food chain. Small fish depend on them for survival and growth, and large fish and other animals often depend on small fish for their survival and growth: the effect of the invasive mussel thus reverberates through the entire food chain.

One might hope that native species of fish and birds will feed on veligers or adult zebra mussels, and that the presence of mussels may perhaps benefit these species by serving as a food source for them. Empirical evidence, however, does not seem to bear out this hope. French and Bur (1996) found that the mean total length of six-year-old female freshwater drum has declined since the zebra mussel invaded Lake Erie, even though the mussels comprised more than two thirds of gut samples in these fish. The total lengths of five-year-old drum changed little. The authors conclude that zebra mussels do not benefit freshwater drum by serving as a staple in their diet. In another study, Jennings (1996) determined that the survival and growth of juvenile fathead minnows is not reduced by low to moderate mussel densities, but fish growth might be adversely affected at higher densities. An indirect channel through which zebra mussels may also affect some fish populations is the filtering-associated increase in water clarity. The more transparent water allows light to penetrate deeper, thereby enhancing aquatic plant growth, and potentially reducing predator effectiveness by providing denser cover for prey. Additionally, the Great Lakes Sea Grant Network reports that because of the large volumes of water they filter and their high body-fat content, zebra mussels can accumulate ten times more contaminants than other, native mussels. These contaminants accumulate in waterfowl and fish that eat zebra mussels.
Another set of concerns relates to the mussels' attachment to hard surfaces. They colonize not only rocks and human-made structures (peers, boats, intake pipes) but other stationary or slow-moving creatures as well. They may interfere with the feeding, growth and reproduction of crayfish, turtles, snails, and native mollusk species. Zebra mussels can, for example, colonize a clam to such an extent that the animal cannot open its shells to eat. The colonization of rocky reefs, some of which are spawning grounds of different (commercially valuable) species of fish, is also disturbing to many ecologists, even though Fitzsimons, Leach, Nepszy and Cairns (1995) found no apparent adverse impact on walleye reproduction despite the almost total coverage of spawning beds.

The tendency of zebra mussels to colonize any hard substrate is what causes the greatest direct economic losses. They may clog the water intake pipes of industrial facilities that rely on fresh water from lakes or rivers; electric power generating stations, drinking water treatment plants, steel plants, golf courses and other facilities have been affected. They may attach to the hulls or motors of boats creating significant additional drag and reducing fuel efficiency. Small zebra mussels can even get into engine cooling systems, causing overheating and other damage. Navigational buoys have been sunk under the weight of zebra mussels, and the clarity-induced plant growth creates additional navigational hazards. Wood, steel, and concrete are all damaged by prolonged attachment of the mussels. Some pathways through which zebra mussels may affect the recreational experience have already been noted. Besides the effects on biodiversity, fish populations and boating, mussels often foul beaches and swimming areas, their sharp shells posing a risk to bare feet. On the other hand, some people, no doubt, value the increased water clarity that sometimes accompanies the
establishment of this exotic mussel. The net effect on recreation behavior is therefore ultimately an empirical question.

For all practical purposes, once zebra mussels become established in a lake, they are impossible to eradicate. Biological control methods do not exist. Some species feed on them, but not heavily enough to keep their population under control. There are chemical means of killing the mussels, but they are so hardy that everything else in the water would also have to be poisoned to completely destroy their population. Therefore, facilities usually apply local control methods: some of them install equipment to treat the water at the point of intake, while others rely on chemical coatings, mechanical scraping, or filtration.\textsuperscript{4} A 1995 survey of infrastructure operators indicated total zebra mussel-related expenses of over $17.7 million at 339 facilities in that year alone. One plant reported total expenditures of over $5.9 million in the six years between 1989 and 1995 (O'Neill, 1997).

2.3 Ecological Literature

Since they first appeared in Lake St. Clair, zebra mussels have been the subject of numerous ecological studies. Early research typically focused on predicting the eventual distribution of mussels based on environmental factors alone. Recall that the Great Lakes and major rivers were colonized so quickly that zebra mussels were expected to invade all water bodies that fit their broad ecological requirements within less than a decade (Kraft & Johnson, 2000). The particular pattern of invasions was of relatively little interest due to the short expected time

\textsuperscript{4} Many European facilities have two sets of intake pipes so that they can operate while the primary pipes are shut down for cleaning. However, most American facilities were designed before zebra mussels appeared in the region, and are not adapted for this procedure.
horizon, and the initial emphasis on the final outcome, the eventual geographical range of mussel distribution, is therefore perhaps not surprising.

Strayer (1991) determined potential geographic extent based on the mussel’s requirements for temperature, frost frequency, and precipitation. Likewise, Neary and Leach (1992) mapped the potential distribution of the zebra mussel in Ontario based on the premise that there are physical and chemical constraints which limit the successful reproduction of the mussel. Lakes which do not meet the ecological requirements for reproduction or survival are less likely to be invaded. The two constraints that are most likely to be binding for zebra mussels are low pH value and low calcium concentration. Neary and Leach used results from the European literature on survival rates under different chemical conditions to determine the limiting pH and calcium values. Not all of the six thousand lakes included in the study had calcium and pH measurements; therefore additional data on conductivity and bedrock geology was used to predict missing values. The authors also acknowledge the role of human-mediated dispersal, and further classify lakes based on their distance from major waterways and highways. When humans act as a vector of dispersal, but data on vector movement is not available, variables that affect lake access and lake attractiveness are often included in the aquatic invasions context. These include proximity to highways, type and number of public boat launches on lakes, the number of residences on or near lakes, lake size and depth, and the relative abundance of game fish species (Buchan & Padilla, 2000). The results of the Neary and Leach (1992) study show that the probability of invasion to inland lakes is highly variable in Ontario. A problem with using chemical covariates is their

5 Not surprisingly, the roots of the zebra mussel literature go back to Europe where invasion started early in the 19th century: mussels were first recorded at Rotterdam in 1826 (Kerney & Morton in Aldridge et al., 2003).
seasonal variability; they are rarely static in a lake environment. Due to differences in climate and other factors, the use of European data on survival and reproductive success is also questioned by some ecologists.

In a more recent and more sophisticated paper, Drake and Bossenbroek (2004) revisit the issue of potential geographic extent. They propose six environmental factors (average annual temperature, frost frequency, annual precipitation, solar radiation, minimum temperature and maximum temperature) and five physical and geological factors (bedrock geology, elevation, flow accumulation, slope and surface geology) to forecast potential distribution in the United States via a machine-learning algorithm for nonparametric prediction of species distributions, but drop two of the temperature variables from the final model. Their results also suggest that the likelihood of invasion is highly variable across different regions, which is not unexpected considering the large geographic extent of the study area. Much of the Western part of the country is probably not habitable for the zebra mussel because of its higher elevation.6

A different strand of the ecological literature implicitly or explicitly recognizes that focusing on the eventual outcome, the maximal geographic range of an invader, misses a rich area of investigation: patterns in spatial and temporal spread before that maximal range is achieved (Johnson & Padilla, 1996). In this context, identification of potential vectors and assessment of their significance is an important task. Before Johnson and Carlton’s 1996 study, little was known about the role of human and natural vectors in spreading zebra mussels. Recreational boating had been implicated in aquatic invasions, and it was known

6 The 2008 invasion of the Pueblo Reservoir in Colorado may serve as evidence to the contrary. It lies at an altitude of nearly five thousand feet.
that aquatic birds were capable of transporting other aquatic organisms, but their relative significance was debated. Johnson and Carlton examined external transport of young zebra mussels by mallard ducks. The ducks were permitted to swim in water containing high concentrations of veligers or juveniles and then directed overland to pools containing tap water. Although zebra mussels were transported in all cases, the usual rate was less than half a mussel per duck per trip. Internal transport by ingestion and defecation was ruled out as a vector because none of the mussels survived passage through the digestive system. Transport by waterfowl is even less likely to occur in nature due to the larger distances and the more extreme conditions that prevail during flight.\(^7\)

The authors listed seven different mechanisms by which recreational boats may contribute to the spread of zebra mussels. These include adult mussels attached to exterior surfaces, adults attached to anchors or material snagged by the anchor, adults attached to aquatic macrophytes (plants) entangled on the boat trailer during boat launch or retrieval, larvae in bilge water, larvae in engine cooling water, larvae in bait buckets, and larvae in live wells – the last two are associated with fishing activities. To assess the role of each of these potential mechanisms, fishing and pleasure boaters were interviewed, and boats were inspected and sampled at public boat ramps on Lake St. Clair (the lake in which zebra mussels were first detected) in 1992. Entangled vegetation and live wells were determined to pose the greatest risk. On average, about eight percent of trailers were observed to have entangled macrophytes with attached adult mussels, but in some situations, nearly a third of all trailers carried adult mussels by this means. Potential mussel densities of over a thousand

\(^7\) The distance between pools was less than three meters in Johnson and Carlton’s controlled experiment.
individuals per meter stem length of vegetation point to the significance of this mode of transport. Live wells contained larvae in eighty-three percent of the cases, and larval densities indicated that a typical well could transport over four thousand of them.\textsuperscript{8} However, because of their sessile nature, the introduction of adults leads to more spatial aggregation, which may be a requirement for establishing a viable population.

Even though it is clear that, on a per trip basis, boats are capable of transporting a considerably higher number of zebra mussels than birds, an assessment of vector significance must also address survival during transit and frequency of vector movement between infested and uninfested lakes. Both of these considerations further implicate recreational boating. Veligers are more likely to stay alive in wet live wells and engine cooling systems than on birds. Adult mussels may survive three to five days under extreme conditions and several weeks in wet fishing nets (Buchan & Padilla, 1999). Likewise, the temporal distribution of vector movements also points to the importance of the human vector. Boats move not only more frequently between bodies of water than birds do, but boating activity is highest in the summer and early fall when veligers and aquatic plants are abundant, while migration of waterfowl does not, in general, coincide so conveniently with the mussel's reproductive cycle.

Having established transient recreational boating as a significant vector in aquatic invasions, Johnson et al. (2001) moved on to examine in more detail the seven boating-related transport mechanisms identified in the previous study, and, employing the same 1992 dataset, evaluate their relative significance in transporting and delivering zebra mussels. The

\textsuperscript{8} In a subsequent paper, Johnson et al. (2001) note that this figure is incorrect due to calculation error: a typical live well contains about 1230 larvae.
more thorough examination lead to the same conclusions: live wells may move up to a hundred times more larvae on a per trip basis than other mechanisms, but it is adult mussels on entangled vegetation that poses the greatest threat, and is most likely to lead to successful new introductions. Using conditional probabilities, the authors estimated that a maximum of twelve out of ten thousand boats deliver a live zebra mussel on aquatic macrophytes to inland waters.\(^9\) Putting the numbers into perspective, this means that just one of the boat launches they examined on Lake St. Clair could have been the source of one hundred seventy invasion events in a single season. Recreational boating, and entangled macrophytes in particular, could be responsible for hundreds or even thousands of dispersal events in the region.

Ecological applications frequently rely on diffusion models to predict the spread of invasive species. In a typical diffusion model, the velocity of the advancing population front depends on the intrinsic rate of population growth and a diffusion coefficient (Buchan & Padilla, 1999). This structure makes diffusion models suitable for modeling range expansion in a homogeneous environment where individuals move randomly – often the case in marine and terrestrial invasions. However, modeling spread in a heterogeneous habitat, such as isolated inland lakes, poses a unique set of challenges. Diffusion models are clearly less applicable in such a scenario because freshwater organisms cannot disperse part of the way to another lake or stream: they require a vector to transport them all the way between the

---

\(^9\) The probability of successful delivery to an inland lake is the product of three terms: the probability that the boat transports live mussels, the probability of boat movement between infested and uninfested waters, and the probability that mussels survive conditional on the movement between lakes: 

\[
P(\text{delivery}) = P(\text{transport}) \cdot P(\text{move}) \cdot P(\text{survival|move}).
\]

\(P(\text{transport})\) was further broken down into 

\[
P(\text{exposure}) \cdot P(\text{transport|exposure}),
\]

where \(P(\text{exposure})\) is the probability of a boat's exposure to the potential mechanism, and \(P(\text{transport|exposure})\) is the conditional probability that mussels were indeed being transported if exposed. For bait buckets and live wells, interview data on usage was also utilized: 

\[
P(\text{exposure}) = P(\text{equipped}) \cdot P(\text{used|equipped}) \cdot P(\text{exposure|used}).
\]

All of these probabilities were estimated to assess the importance of the different mechanisms.
habitats. Even the more advanced stratified diffusion models have difficulty predicting long-range dispersal events (Bossenbroek, Kraft & Nekola, 2001). For these reasons, gravity models are often used in analyses of freshwater invasions. As the name implies, gravity models are derived from the equation of gravitational attraction as described by Isaac Newton's law of gravity. These models contain elements that are analogous to mass and distance in the original equation. Thus, in the present context, dispersal events between two lakes may depend on the size of the lakes (or, in a betterpecified model, on boating activity at each of the lakes), and on the distance that separates them.\footnote{Gravity models are also used in economics and other social sciences. The gravity model of trade, for example, predicts trade flows based on the economic sizes of the trading partners and distances between them. Rosenthal (1987) investigated the role of substitute prices in travel cost models with three different specifications of the travel cost model, including a gravity/logit model specification.} Gravity models acknowledge the importance of the human vector. They do not estimate movement rates by the organism; they estimate the force of “attraction” between two sites, and movement rates are a function of this force (Bossenbroek et al., 2001). As such, they are better able to account for the spatial location of sites and for long-distance dispersal events that result from vector movement than diffusion models.

Studies making use of gravity models attempt to relate different approximations to boating activity to the observed spatial and temporal spread of the zebra mussel. Bossenbroek et al. (2001) utilize aggregate county-level data on the number of registered boats from counties in Michigan, Wisconsin, and parts of Illinois, Indiana and Ohio to forecast zebra mussel dispersal to inland lakes in four of the states. The force of attraction between a source (county) and a destination (lake) depends on the number of registered boaters in the county, the area of the lake, and the distance between them. Only lakes with surface area larger than
twenty-five hectares were included in the analysis, and it was assumed that larger lakes are more attractive than small ones. The nine lakes suspected of having been colonized by water flow from an upstream source were excluded to focus explicitly on the role of recreational boating. Initial zebra mussel sources included the Great Lakes, the Mississippi River and the Illinois River, but newly invaded inland lakes were added as potential sources for subsequent iterations of the model. Uniquely among studies focusing on boating patterns, the authors account for the ecological requirements of zebra mussel colonies by estimating the suitability of a given lake based on the pH and calcium levels established by the literature. Lake chemistry tends to exhibit spatial autocorrelation because it is influenced by bedrock geology, and therefore missing pH and calcium values were estimated using data from the nearest twenty lakes with records. About half of the 3,600 lakes included in the study area were determined to be environmentally inappropriate for zebra mussels, and many of these lakes were concentrated in northern Wisconsin. Regional differences between prediction accuracy were observed, but overall, the authors claim, the model was successful in duplicating actual patterns of spread.

In another gravity-type application, Schneider, Ellis and Cummings (1998) assess relative invasion risks to native mussel communities by the zebra mussel in Illinois lakes and streams. Generally high calcium concentrations across the whole state mean that environmental factors are less likely to play an important role in the Illinois invasion, and the chemical component of the model can be ignored. This study, however, addresses both overland transport between watersheds by recreational boaters, and downstream spread of the free swimming veligers within a watershed. There are two additional structural differences
between the Bossenbroek et al. and the Schneider et al. studies. First, the force of attraction between source and destination in Schneider et al. is based not on lake size and the number of registered boats, but on actual boat-use data taken from creel surveys at sites on Lake Michigan and fifty-five inland lakes.\(^{11}\) That is, the likelihood of boat movement, and zebra mussel transport, between two sites depends on boat use at each site and the distance separating the two sites. Second, both sources and destinations are lakes, unlike in Bossenbroek et al. (2001) where the movement of boats was between counties and lakes. The probability of a successful invasion is proportional to the total number of infested boats arriving at a lake.\(^{12}\) As in other studies, boat landings on Lake Michigan and along the major rivers were considered initially infested. The analysis stresses the large impact of factors that affect overland transport over those that affect downstream spread, and leads to the perhaps counterintuitive conclusion that prevention efforts should not be aimed directly at the important native mussel habitats, but rather at currently uninfested popular boating sites that are likely to become future sources of invasion.

To interpret and predict invasion patterns, one needs to understand the velocity, frequency, distance and direction of movement of non-indigenous organisms and their vectors. Gravity models are useful for modeling aquatic invasions because they are able to estimate the force of attraction (that is, vector movement) between pairs of sites. Padilla et al. (1996), and Buchan and Padilla (1999) use actual data on vector movement instead of estimated figures. They employ a randomized survey of 58,800 boaters conducted by the

---

\(^{11}\) For another fifty-two sites included in the study, no creel information was available, and boat use was estimated by contacting the official responsible for the boat access site.

\(^{12}\) For parameterization, it was assumed that two thousand introductions are required for infestation (Bossenbroek et al. [2001] determined a threshold of eight hundred fifty trips by best-fit parameterization) and that once a mussel population is established, veligers can infest downstream sites by simple diffusion.
Wisconsin Department of Natural Resources in 1989-1990. Some of the questions in the survey allowed them to characterize not only lake use, but also subtler details of spatial interactions – namely, the number of people who actually boated on both a Great Lake (source of invasions) and an inland lake within two weeks, or on different inland lakes within two weeks. The spatial resolution of some parts of the study was limited to counties because the survey did not always provide site-specific detail. All of the six inland lakes that were infested at the time of the study (1995) were among the most frequently visited lakes, and all of them were used by at least one person who had also boated on Lake Michigan within the two-week time frame of the survey. The Mississippi River was determined to be a less significant source because Mississippi boaters tended to stay within the river basin and visit flowages connected to the river. The authors contend that explicitly accounting for spatial patterns in boater movement between the Great Lakes and inland lakes (and between county pairs) is important because overall lake use is not the best indicator of invasion risk: connections to infested lakes is what determines risk.

2.4 Economic Literature

Most of the economic literature on invasive species is concerned with the risk of new invasions and their causes (Costello, Springborn, McAusland & Solow, 2007; Horan & Lupi, 2005), the development of optimal inspection routines to prevent introduction (Batabyal & Nijkamp, 2005; Batabyal & Beladi, 2006), the optimization of post-introduction control effort (Eiswerth & Johnson, 2002), or the choice between preventive and post-introduction control efforts (Wilman, 1996). The ecological literature, though extensive and diversified,
does not typically endogenize behavior when humans are responsible for biological invasions. However, vector behavior is rarely modeled even in economics – the likely reason is that terrestrial (and marine) exotic species do not, in general, require a vector to spread after the initial introduction, and it is precisely terrestrial invasions that cause the most economic harm through their effect on agriculture. Therefore, they tend to be the subject of most economic analyses.

Eiswerth and Johnson (2002) use dynamic optimization to analyze investment in invasive species management. Eiswerth and Johnson tie together two strands of economic literature: pollution control and renewable resource economics. Pollution, not unlike invasive species, has a negative impact on environmental service flows, and costs have to be incurred to reduce its stock. An important difference is that the natural growth rate of the stock of pollutants is negative, while living creatures reproduce and grow over time. The connection to renewable resource economics is even more self-evident: invasive species may represent a public “bad” whose stock dynamics can be modeled with standard lumped-parameter ecological models that also allow for harvesting the stock. The complication is that in the biological invasions context, because the natural population growth rate is positive, the impact of marginal changes in any of the model parameters is ambiguous. This, they claim, underlines the need for specific case-studies and for the collaboration with ecologists to parameterize economic models.

Macpherson, Moore and Provencher (2006) were the first to endogenize boater behavior in the management decisions related to an aquatic invader. Their model is

---

13 To a large extent, their working paper motivated my interest in this topic.
characterized by inter-seasonal management decisions and intra-seasonal boater trip
decisions. The spread of the invasive species depends on the probability that a representative
boater visits an infested lake in the current time period, the probability that the boater picks
up and transports the invasive organism, the probability that the boater visits an uninfested
lake in the next time period, and the probability that the introduction also leads to the
establishment of a colony. The site choice probabilities are from the logit formula in a
random utility model. A forward-looking lake manager chooses from two management
options at the beginning of each season: the implementation of mandatory boat cleaning at
the lake, and the closure of the lake to transient boating traffic. Boater behavior depends on
the current state of the invasion, lake-specific variables and the controls chosen by the
manager, while the dynamics of the invasion depends on stochastic ecological processes, the
trip-taking decisions of boaters and the control efforts chosen by the lake manager. Two
management objectives are considered in a dynamic optimization framework: maximizing
the discounted net benefits of invasion control, and minimizing the rate of spread. The model
is then applied to the case of the Eurasian watermilfoil, an invasive aquatic plant whose
primary vector of spread is recreational boating, with hypothetical data in a four-lake system.
Although the two management objectives have different implications, the authors
demonstrate that a policy that accounts for boater responses to controls is more effective
under both scenarios.

Although human activities are responsible for most non-indigenous species
introductions, the relationship between post-introduction spread and human behavior is rarely
as strong as in the freshwater invasions context. While it is often acknowledged that “range
expansion can be analyzed by explicitly modeling the behavior and movement of the human vectors” (Schneider et al., 1998, p. 789), this has never been done empirically. Ecological studies generally treat human behavior as exogenous – they employ data on boating patterns without explicitly modeling behavior. Given the substantial economic and ecological damages caused by aquatic invasives, and the significant role recreational boaters play in their distribution, the failure of environmental economists to investigate the subject is surprising. Macpherson et al. (2006) recognized the importance of modeling boater preferences to assess the effect of various policy scenarios. They base their analysis on simulated data in a small, four-lake system to deal with the challenges their dynamic optimization stage imposes. In the following chapters, I outline a model that can be used in large-scale applications of human-mediated freshwater invasions, and estimate it using actual data on the distribution of zebra mussels and recreational boating trips, representing the first empirical effort in the economic literature.
Chapter 3

Data
My objective is to analyze the spread of zebra mussels by modeling boater behavior via a recreation demand model. Ecological models of human-mediated freshwater invasions are generally based on aggregate share data (either actual or estimated) such as the number of boaters visiting a particular site. My objectives, however, require using individual-level information on boater preferences. In this chapter, I describe the various data sources I will rely on to perform my analysis. I start by introducing data on the spatial distribution of zebra mussels and on the hydrography of Wisconsin, and then summarize the results of a boater survey that serves as the basis of my behavioral model. Finally, I describe auxiliary data on lake clarity that will also play a role in the analysis.

3.1 Zebra Mussel Data

I acquired geospatial data on the distribution of zebra mussels from the United States Geological Survey. The data features the spatial locations of confirmed zebra mussel sightings from 1988 to spring 2006, and was compiled using reports from a variety of federal, state and municipal agencies, public utilities, universities, engineering and private consultant firms. The U.S. Geological Survey claims that the data are an essential part of the Department of the Interior's Nonindigenous Aquatic Species Program, and are intended to make resource managers aware of zebra mussel occurrences so that they can help to control or prevent the spread of the invasive.

Information for each mussel observation includes the year the sighting occurred, and a short description of the locality where mussels were found. More than two hundred sightings have been recorded in Wisconsin, but several of those are repeat observations.
covering the same site; over half of the data points are Great Lakes observations, and about forty are from the Mississippi River. The first invasions in Wisconsin affected Lake Michigan, Lake Superior and the Mississippi River in 1989, 1990 and 1991, respectively. Inland lake mussel populations were first discovered in 1994. Three lakes in the southeastern part of the state (Elkhart Lake, Lake Okauchee and Silver Lake) were invaded that year – each is located within fifty kilometers of Lake Michigan. Mussels have gradually extended their geographic range westward, and have since then invaded inland lakes in central Wisconsin. However, northern Wisconsin still remains relatively unaffected by the mussel infestation. As of early 2006, forty-eight inland lakes in Wisconsin have been invaded by zebra mussels, and the overwhelming majority of them are located in the southern-southeastern part of the state. Figure 3.1 shows the geographic distribution of sightings in Wisconsin and the neighboring states.

3.2 Wisconsin Hydrography Data

With more than fifteen thousand inland lakes, six hundred fifty miles of Great Lakes shoreline, and over forty thousand miles of rivers and streams, Wisconsin is one of the richest states in hydrological features. The Wisconsin Department of Natural Resources maintains a detailed (1:24,000-scale) statewide hydrography geographic database that also contains various descriptive attributes for surface water features, including names and water body identification codes. I use this data set to identify and locate boating destinations, and to calculate travel distances between residential locations and water features.
3.3 Boating Data

The fundamental component of a study analyzing zebra mussel spread by modeling boater behavior is of course micro-level data on boaters’ recreation site choices. I employ the 1989-1990 Wisconsin Recreational Boating Study, a major revealed preference survey initiated by the Bureau of Law Enforcement of the Wisconsin Department of Natural Resources.\textsuperscript{14} Some of the main objectives of the boating study included

- determining, by county, the boating pressure on Wisconsin’s inland lakes and streams, and on Wisconsin’s Great Lakes coastal waters
- determining the temporal distribution of boating pressure
- describing boating activities
- determining the types and sizes of boats used
- determining the number of people in boating parties, and their spending habits associated with boating trips
- determining the residence of boat operators
- identifying boaters’ perceptions of the quality of their boating experience, especially those related to crowding, and
- identifying potential conflicting recreational uses of Wisconsin’s water resources (Penaloza, 1991).

The study area consisted of all Wisconsin counties and selected counties from the neighboring states of Illinois, Iowa and Minnesota, and is depicted in Figure 3.2. A random

\textsuperscript{14} The same data set was used by Padilla et al. (1996) and Buchan and Padilla (1999) as noted in the previous chapter. Note that the timing of the boater survey coincides with the period mussels were first discovered in Lake Michigan. The preferences recorded in the data are therefore not influenced by mussel presence in different water bodies.
sample of 42,000 people was drawn from the population of the more than 482,000 licensed boat owners in Wisconsin in 1989.\textsuperscript{15} The overwhelming majority of licensed boats were motorboats with a ninety-six percent share in the licensing records; the share of sailboats was only three percent, and the share of canoes a mere one percent (Penaloza, 1991).\textsuperscript{16} These three different types of boats were sampled separately to ensure that each type was adequately represented: 28,000 motorboaters, 7,000 sailboaters and 7,000 canoeists were randomly selected. In addition, 5,600 names from the licensing records of each of the three neighboring states were drawn and included in the study for a combined (resident and non-resident) sample size of 58,800.

Due to the stratified nature of the sampling procedure just described, analysts at the Wisconsin department of Natural Resources applied different weighting factors to the survey responses from each state, and for each type of boat when estimating boating pressure. Motorboaters are by far the largest group in the population of boaters, and they represent the greatest threat in terms of transporting zebra mussels. Therefore, I will base the behavioral analysis only on the responses of motorboaters who lived within the study area, and were licensed in Wisconsin – note that this also includes some motorboaters from the neighboring states. This decision is also convenient in that it relieves one from dealing with the stratified sampling procedure and the varying response rates by the sample groups.

The survey was designed with the objectives of obtaining precise answers and eliciting a high response rate. The questionnaire itself includes several questions that are very

\textsuperscript{15} Ten percent of the state’s population (or about forty percent of households) are registered boaters (Buchan & Padilla, 1999).

\textsuperscript{16} Trailered boats are the biggest concern in spreading non-indigenous species. Survey questions do not enable me to distinguish between resident boats (boats continually moored in the water) and trailered boats, but Padilla et al. assert that trailered boats represented 98.6 percent of registered boats in the survey (1996).
specific, and are therefore subject to recall bias. To minimize errors in reporting, the boating season (April-October) was broken up into fourteen two-week-long periods, and both resident and non-resident samples were accordingly divided into fourteen equally sized groups. Questionnaires were mailed to members of a different group in each period, and respondents were only asked about their boating experiences relating to the two weeks prior to receiving the survey (Penaloza, 1991). A copy of the entire questionnaire is reproduced in Appendix A.

Sampled boaters were contacted as many as three times through personalized letters. Each boat owner received an advance letter to inform them about the study and the questionnaire they would be receiving shortly. The questionnaire itself came with a pre-printed return envelope and the appropriate postage affixed – first class postage stamps were used on all outgoing and return envelopes to signal the importance of the materials. Finally, a third letter with another copy of the questionnaire was sent out to all those who had not returned the original questionnaire after a week. Penaloza reports that over twenty percent of all responses were the replacement questionnaires from the latter mailing. This procedure resulted in a seventy-four percent overall response rate, and, significantly, a response rate as high as eighty-one percent among Wisconsin motorboaters. Penaloza contends that due to the large sample size and the high response rate, the margin of error for the estimated number of boaters on the water during the boating season is remarkably low at around one percent (Penaloza, 1991).

In this section, I review the survey questions most important for my purposes, and summarize what we do and do not know about respondents and the trips they took. We
observe what zip code area respondents live in (question 19) if they actually boated during the two-week survey period; those who did not boat were only asked about the type of boat they owned. Residential location is the first piece of information required for implementing a recreation demand model because it is needed to compute travel distances (and costs) to various sites for each boater. We know the exact days on which a respondent used the boat (question 5), and the type of the boat involved (questions 1-3). We know the Wisconsin counties in which the boat was used, and the number of days it was used in each county on inland lakes and streams (question 8) and on the Great Lakes (question 10). Since no Wisconsin county neighbors more than one of Lake Michigan and Lake Superior, the latter question unambiguously identifies the targeted Great Lake. The name of the inland lake, river or stream on which respondents did most of their boating during the survey period is also identified in question 8.

Questions 8 and 10 thus provide the second key piece of information needed for a site-specific behavioral model: the name of the chosen alternative. Note that although the primary inland boating destination is listed by name, it is not necessarily known how many trips were made to the site, or whether multiple-day trips were involved. This makes it impossible to analyze the frequency of visits to particular sites. We also do not have any information about other inland sites that may have been boating destinations. From question 11 we know what activities each boater was involved in on the water, but unless a single boating trip was made during the relevant time frame, it is impossible to decipher which of these activities the boater engaged in on the water body listed in questions 8 or 10. The
primary activity is also identified in this section. Other questions relating to the boating experience pertain exclusively to the last day that respondents boated (questions 14-18).

Unless a single trip was made during the survey period, there is no reason the site visited on the last trip should be the same as the most heavily used site. Hence, in general, we cannot ascertain what site the perceived crowding and satisfaction measures, specific trip and spending information correspond to, significantly reducing the informational value of this part of the questionnaire. Also missing from the boating data are socioeconomic variables such as income and family status. Only with a great deal of measurement error could one use boat size and power, or spending on the last trip as indicators of financial well-being or income.

Nearly 40,000 surveys were completed and returned to the Department of Natural Resources, and are available for analysis. It is estimated that each response represents about two hundred forty boating events in the population of Wisconsin motorboaters (Penaloza, 1991). The most important findings from the survey are displayed in Tables 3.1 and 3.2, and are summarized below. Slightly more than a quarter of respondents boated on Wisconsin waters during the two-week period pertaining to the questionnaire. A total of 35,005 boating days are represented in the survey, 24,123 of which are associated with motorboats registered in Wisconsin. Figure 3.3 shows the distribution of boating days by county. Each boat owner who used the boat in Wisconsin boated on average on more than three days. Approximately ninety percent of overall boating activity took place on inland lakes and streams, and about ten percent on the Great Lakes. Twelve percent of those who boated on inland lakes, rivers or streams used the boat in more than one county, and two percent used the boat in more than
two counties. Fifteen percent of Great Lakes boaters boated in multiple counties, and the share of those boating in at least three counties was over four percent. As for the temporal distribution of boating trips, the summer months were the busiest, with almost a quarter of all boating trips of the season occurring in July.\textsuperscript{17} Respondents generally boated on the weekends for an average of over five hours at a time. The most popular activity on the water was fishing, followed by cruising or sailing, swimming and water skiing. As could be expected, the popularity of fishing relative to other activities was highest during the spring and fall months; the fraction of motorboaters whose primary activity was fishing from the boat is about sixty-five percent, and varies between a high of nearly ninety percent in early April and a low of slightly below sixty percent in July.

In all, more than eight hundred different inland water bodies were reported as primary boating destinations; many of them were visited by a single boater only. The mean (one-way) distance traveled was slightly under forty-five miles.\textsuperscript{18} Perhaps not surprisingly, the majority of the most popular sites have been identified to contain zebra mussels. The most often referenced inland water body among Wisconsin motorboaters was the Mississippi River with 253 reported visits, closely followed by Lake Winnebago (251), the Wisconsin River (189), Lake Mendota (119), Lake Geneva and the Wolf River (104 each). Relatively few (112) boaters said they boated on both inland waters and a Great Lake during the specified time period. Although inland water bodies with several boater connections to a Great Lake tend to

\textsuperscript{17} The study period overlapped with two boating seasons, and lasted May 6, 1989 – October 20, 1989 and April 7, 1990 – May 4, 1990.
\textsuperscript{18} About half of Wisconsin’s population lives within this distance from one of the Great Lakes.
be popular boating destinations in general, this is not always the case, as noted by Padilla et al. (1996).

Lastly, note that boaters may freely cross (and transport mussels across) state boundaries. The boating survey contains information only on trips that targeted lakes or streams in Wisconsin, and does not enable me to model inter-state mussel spread. However, this is unlikely to become a significant problem: as shown in Figure 3.1, most states neighboring Wisconsin are relatively unaffected by zebra mussels. Iowa, Minnesota and the Upper Peninsula of Michigan have only a few zebra mussel sightings each (in water bodies other than the Mississippi River and the Great Lakes). All of those invasions took place fairly recently. Northeastern Illinois has a higher incidence of invasions, but most occurred only after the majority of southeastern Wisconsin lakes were already invaded. This suggests that boating destinations in neighboring states did not play a significant role in Wisconsin invasions, and that the most important mussel sources (the Great Lakes, the Mississippi River, and inland lakes in Wisconsin) will in fact be included in the analysis.

3.4 Lake Clarity Data

In the freshwater recreation demand literature, water clarity is often used as an observed site attribute. It is an easily gaugeable aspect of water quality (both to recreationists and to individuals collecting data on it), and is therefore not subject to some of the criticism regarding the use of other physical and chemical measures of water quality as site attributes in behavioral models. For example, dissolved oxygen, total phosphorus or impaired status based on total maximum daily loads may be good indicators of water quality, but are not
always readily observable to recreationists, and may have little effect on behavior. Another reason to use water clarity in the context of my research is the often-reported effect zebra mussels exert on it through their filtering activity.

Wisconsin’s Citizen Lake Monitoring Network and other organizations performed about 90,000 Secchi-disk measurements at over 1,200 inland lakes statewide between 1987 and 2005. Access to this extensive data set was provided by officials at the Wisconsin Department of Natural Resources. Each observation identifies the water body assessed by name and a unique water body identification code, and the precise location where the measurement was carried out by coordinates of latitude and longitude. The exact date and time of each measurement were recorded along with the observed Secchi-depth, and comments about the circumstances obtaining at the time of the survey. It was also noted whether the disk hit bottom at the recorded depth.

Lake clarity is subject to large seasonal variation due to changes in climate, aquatic plant abundance, algal levels and other factors. Changes in the mean Secchi-depth of all Wisconsin lakes over nearly two decades is depicted in Figure 3.4. Clarity is typically highest early in the spring, and lowest at the end of the summer. Since not all lakes are monitored with the same intensity year-round, seasonal variation needs to be addressed when comparing measurements across specific water bodies. The most common method to control for changes in clarity over the course of the year is to use readings from certain months only when computing mean Secchi-depth. The Wisconsin Department of Natural Resources routinely uses data from the summer period only in such situations (J. Filbert, personal communication, January 19, 2007). I decided to use Secchi-disk observations performed in
July, August and September based on the similarity of the means and standard deviations of measurements from those months. Note that this period largely coincides with the highest level of boating activity, so the clarity values recorded are those that are relevant for most boaters. Table 3.3 summarizes the seasonal variability in observations.

Aside from seasonal variation, statewide mean Secchi-depth has stayed rather stable over time. This, however, is not necessarily true for observations from individual lakes. Site-level clarity data occasionally display a temporal trend over the expected seasonal variability – this may be due to either natural processes or human activity near the lake. Since boaters’ recreation decisions were observed in 1989-1990, water clarity values relevant to that era should be used in the behavioral model. The drawback of restricting attention to a short time interval would of course be the loss of potentially useful information from other years. Secchi-readings are influenced by weather conditions, for example, and an accurate characterization of water clarity requires averaging several observations to account for such idiosyncrasies. I therefore use July, August and September readings from the years 1987-1994 when calculating site-specific mean Secchi-depth. This time span is short enough for Secchi-measurements based on it to accurately reflect the conditions pertaining at the time of the boater survey, but long enough to provide sufficient amount of data for analysis. Note that the first inland lake mussel invasion (in Wisconsin) took place in 1994. Consequently, potential mussel-induced clarity changes are not reflected in the lake-level averages.

There are nearly 16,000 relevant observations, and more than five hundred lakes with at least five Secchi-readings. Average Secchi-depth ranges from less than a foot in Sinissippi Lake (central Wisconsin) to over twenty-five feet in Lake Lucerne (northern Wisconsin).
Although there are occasionally large differences in water clarity within nearby lakes, it is nevertheless possible to make general statements about the spatial distribution of clarity values. The clearest lakes are predominantly concentrated in the northeastern and northwestern regions of the state; lakes in south and central Wisconsin tend to be less clear. Since Secchi-depth is not observed for all inland lakes designated as primary boating sites in the survey, this spatial regularity can be exploited to interpolate clarity values for those lakes. Figure 3.5 illustrates the prediction surface over the state based on the spline interpolation technique. Sites with missing clarity observations are assigned the predicted value that coincides with the centroid of the water body.

Recall that large improvements in the water clarity of Lake Erie were attributed to zebra mussels. Changes in inland lake water clarity are also important since they may influence boaters’ recreation decisions. It is theoretically possible to identify the effect zebra mussels have on clarity using the more than 3,800 Secchi-disk measurements from lakes that are currently infested. Twenty-five of the infested lakes have at least five pre-invasion and five post-invasion observations from the relevant months, and on average, mean post-invasion water clarity (nine feet) is about a half foot higher than mean pre-invasion water clarity. This change is probably not significant in the practical sense, and it is certainly not as dramatic as the increase documented in some cases – the Secchi-disk data does not support the hypothesized large clarity increase in infested inland lakes. Indeed, most lakes show

---

19 A spline is a technique that uses a mathematical function minimizing overall curvature, and results in a smooth surface that passes exactly through the input points. I used a high tension weight parameter to predict clarity values more closely constrained by the sample data range. Another spatial interpolation method, the inverse distance weighted, results in similar estimates, but a less smooth surface.

20 Mean pre-invasion and post-invasion Secchi-disk values are calculated using a symmetric one-year lag (in each direction) from the year of mussel discovery in a specific water body. This is intended to account for the fact that lags in discovery may exist, and that mussels are not expected to change water clarity instantaneously.
virtually no change in mean water clarity. A notable exception is Little Muskego Lake whose
clarity readings are displayed in Figure 3.6: mean Secchi-depth more than doubled following
its invasion by zebra mussels in 1999. It is, however, impossible to assert whether the
observed improvement is in fact due to zebra mussels. The apparent conflict between the
results of the Secchi-disk data analysis and claims made in the ecological literature highlights
the fact that lakes are different. Changes in clarity have in some cases been observed
following colonization, but general statements about the outcome are hard to make. The
effect mussels have on clarity is idiosyncratic to the lake, and may depend on bedrock
geology, the source of water supply, drainage type, the degree of colonization and other
factors. For example, seepage lakes are landlocked bodies of water and are not influenced by
streams. Drainage lakes, on the other hand, have both surface water inflows and outflows;
since their water generally circulates faster, clarity in drainage lakes may be expected to
change less following an invasion by zebra mussels.
Table 3.1. Boating data summary

<table>
<thead>
<tr>
<th></th>
<th>All boat owners</th>
<th>Wisconsin motorboaters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of respondents</td>
<td>39,508</td>
<td>19,047</td>
</tr>
<tr>
<td>Used boat in Wisconsin</td>
<td>10,508</td>
<td>6,983</td>
</tr>
<tr>
<td>Boated in at least one inland county</td>
<td>9,100</td>
<td>6,184</td>
</tr>
<tr>
<td>Number of boating days in county one</td>
<td>29,338</td>
<td>20,490</td>
</tr>
<tr>
<td>Boated in at least two inland counties</td>
<td>1,122</td>
<td>812</td>
</tr>
<tr>
<td>Number of boating days in county two</td>
<td>2,009</td>
<td>1,483</td>
</tr>
<tr>
<td>Boated in at least three inland counties</td>
<td>159</td>
<td>112</td>
</tr>
<tr>
<td>Number of boating days in county three</td>
<td>250</td>
<td>172</td>
</tr>
<tr>
<td>Boated in at least one county on Great Lakes</td>
<td>1,046</td>
<td>630</td>
</tr>
<tr>
<td>Number of boating days in county one</td>
<td>3,057</td>
<td>1,788</td>
</tr>
<tr>
<td>Boated in at least two counties on Great Lakes</td>
<td>161</td>
<td>93</td>
</tr>
<tr>
<td>Number of boating days in county two</td>
<td>272</td>
<td>145</td>
</tr>
<tr>
<td>Boated in at least three counties on Great Lakes</td>
<td>47</td>
<td>26</td>
</tr>
<tr>
<td>Number of boating days in county three</td>
<td>79</td>
<td>45</td>
</tr>
<tr>
<td>Engaged in fishing</td>
<td>7,027</td>
<td>5,177</td>
</tr>
<tr>
<td>Cruising, sailing</td>
<td>3,894</td>
<td>2,517</td>
</tr>
<tr>
<td>Water skiing</td>
<td>1,465</td>
<td>1,218</td>
</tr>
<tr>
<td>Swimming</td>
<td>1,492</td>
<td>1,043</td>
</tr>
<tr>
<td>Other enjoyment boating</td>
<td>1,743</td>
<td>1,004</td>
</tr>
</tbody>
</table>
Table 3.2. The temporal distribution of boating trips

<table>
<thead>
<tr>
<th>Period</th>
<th>Date</th>
<th>Percentage of total boating days</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6 May – 19 May 1989</td>
<td>5.13</td>
</tr>
<tr>
<td>2</td>
<td>20 May – 2 Jun 1989</td>
<td>8.06</td>
</tr>
<tr>
<td>3</td>
<td>3 Jun – 16 Jun 1989</td>
<td>7.09</td>
</tr>
<tr>
<td>4</td>
<td>17 Jun – 30 Jun 1989</td>
<td>10.73</td>
</tr>
<tr>
<td>5</td>
<td>1 Jul – 14 Jul 1989</td>
<td>13.61</td>
</tr>
<tr>
<td>6</td>
<td>15 Jul – 28 Jul 1989</td>
<td>10.73</td>
</tr>
<tr>
<td>7</td>
<td>29 Jul – 11 Aug 1989</td>
<td>11.00</td>
</tr>
<tr>
<td>8</td>
<td>12 Aug – 25 Aug 1989</td>
<td>8.68</td>
</tr>
<tr>
<td>10</td>
<td>9 Sep – 22 Sep 1989</td>
<td>5.34</td>
</tr>
<tr>
<td>11</td>
<td>23 Sep – 6 Oct 1989</td>
<td>4.56</td>
</tr>
<tr>
<td>12</td>
<td>7 Oct – 20 Oct 1989</td>
<td>3.48</td>
</tr>
<tr>
<td>13</td>
<td>7 Apr – 20 Apr 1990</td>
<td>1.10</td>
</tr>
<tr>
<td>14</td>
<td>21 Apr – 4 May 1990</td>
<td>2.37</td>
</tr>
</tbody>
</table>
### Table 3.3. Seasonal variation in statewide mean water clarity

<table>
<thead>
<tr>
<th>Month</th>
<th>Number of observations</th>
<th>Mean clarity (feet)</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>19</td>
<td>7.45</td>
<td>3.65</td>
</tr>
<tr>
<td>February</td>
<td>43</td>
<td>14.28</td>
<td>8.69</td>
</tr>
<tr>
<td>March</td>
<td>147</td>
<td>11.42</td>
<td>6.67</td>
</tr>
<tr>
<td>April</td>
<td>2,264</td>
<td>10.42</td>
<td>6.36</td>
</tr>
<tr>
<td>May</td>
<td>10,011</td>
<td>11.39</td>
<td>6.45</td>
</tr>
<tr>
<td>June</td>
<td>16,332</td>
<td>11.00</td>
<td>6.12</td>
</tr>
<tr>
<td>July</td>
<td>19,587</td>
<td>9.73</td>
<td>5.37</td>
</tr>
<tr>
<td>August</td>
<td>18,710</td>
<td>9.16</td>
<td>5.24</td>
</tr>
<tr>
<td>September</td>
<td>12,812</td>
<td>9.07</td>
<td>5.11</td>
</tr>
<tr>
<td>October</td>
<td>7,343</td>
<td>9.59</td>
<td>4.91</td>
</tr>
<tr>
<td>November</td>
<td>777</td>
<td>9.90</td>
<td>4.88</td>
</tr>
<tr>
<td>December</td>
<td>39</td>
<td>9.49</td>
<td>5.79</td>
</tr>
<tr>
<td>Total</td>
<td>88,084</td>
<td>9.95</td>
<td>5.64</td>
</tr>
</tbody>
</table>
Figure 3.1. The distribution of zebra mussels in the Upper Midwest as of 2006
Figure 3.2. Boating survey study area
Figure 3.3. The number of boating days by county
Figure 3.4. Seasonal variability in statewide mean Secchi-depth
Average depth is shown in feet. Vertical reference lines mark September of each year – generally the month with the lowest value of mean water clarity.
Figure 3.5. Inland lake clarity prediction surface over the state of Wisconsin
Deeper shades correspond to higher clarity.
Figure 3.6. Secchi-disk readings over time in Little Muskego Lake
    Depth is measured in feet; the reference line marks the year of zebra mussel invasion (1999).
Chapter 4

Analysis
I employ a framework consisting of two distinct models to assess the threat of mussel invasion to individual water bodies while explicitly accounting for human behavior. The first is a site-specific discrete choice model that characterizes the recreation decisions boaters make (the economic model). The importance of a large boater survey becomes clear when the goal is to estimate a travel cost model containing hundreds of choice alternatives. The second component is a discrete duration model that predicts invasion probabilities based on observed characteristics of the water bodies (the ecological model). The main innovation of my approach lies in how the two models are integrated. Invasion threat is assessed in the discrete duration model using a time- and site-specific variable constructed from site choice probabilities and current invasion status. I first describe the random utility model, and then a simple discrete duration model. Finally, I integrate the two, present estimation results, and discuss various methods of evaluating model performance.

4.1 The Random Utility Model

In this section, I introduce a site choice model to characterize boater behavior using the survey data from the Wisconsin Department of Natural Resources. First, I explain in detail the steps I needed to take to identify and geo-locate boating destinations precisely. After the boating data is prepared for site-specific analysis, I specify the utility index of decision makers, describe the explanatory variables, identify challenges in estimation and strategies for dealing with them, and finally I present estimation results. The site choice model I use is a multinomial logit commonly used in random utility models of recreation demand, and in discrete choice applications in general.
Data Preparation

The boater survey was designed primarily with a county-level investigation in mind, and, according to its stated objectives, was probably not intended to serve as a data source for micro-economic analysis. However, one of the survey questions presents an opportunity for site-specific discrete choice modeling: the inland lake, river or stream on which boaters did most of their boating during the relevant time period (that is, the primary inland boating destination) is identified by name. Using this information requires carefully matching reported site names with actual geographic locations. Accordingly, the most challenging aspect of preparing the survey data for input into a site choice model was identifying the inland lakes visited.

There are several common lake names that repeat themselves in Wisconsin. For example, there are at least sixty-eight Bass Lakes and fifty-nine Long Lakes in the state. Many counties contain more than just one of the fifty-nine Long Lakes in Wisconsin: four are located in Oneida county, four in Polk county, three in Chippewa county, two in Waushara county, two in Florence county and another two in Price county. Likewise, there are nine Bass Lakes in Burnett county, six Clear Lakes in Oneida county, and four Deer Lakes in Polk county. Of the four Horseshoe Lakes (one of which is also officially called Garfield Lake) and three Little Horseshoe Lakes in Polk and Barron counties, the largest is shared by the two counties. Greater Bass Lake, Little Bass Lake and Lower Bass Lake are all official lake names in Langlade county, but some boaters only said they visited Bass Lake in Langlade county. Chequamegon Waters Flowage in Taylor county is also known as Miller Dam Flowage, but is apparently more often referred to as simply Miller Lake in everyday
language. Pike Lake in Bayfield county (one of nine Pike Lakes in Wisconsin) had a surprisingly large share of boater traffic considering its small surface area and non-central location – investigation revealed that four more prominent lakes (Eagle Lake, Lake Millicent, Hart Lake and Twin Bear Lake) in that county are collectively referred to as the Pike Chain of Lakes. This list is certainly far from exhaustive, and it is only intended to give a flavor of the types of issues that may arise, besides the more obvious problems of imperfect geographic knowledge on the part of survey respondents and misspelled site names, when one faces the task of identifying specific boating destinations. However, the examples clearly demonstrate that a site name and a county of location are not sufficient to identify a site uniquely – and in some cases, a choice has to be made between as many as three counties, each of which may form a feasible combination with the a water body named.

The most promising means to advance with a site-specific model is to adopt the identification system employed by the Wisconsin Department of Natural Resources that uses Water Body Identification Codes (WBIC) to catalog lakes, streams and rivers. The response of each Wisconsin motorboater to survey question 8 was examined to locate the water body specified, and to assign the appropriate WBIC to it (using the Wisconsin hydrography data). If the destination was ambiguous due to any of the potential reasons referred to above, a more thorough investigation was undertaken that started with a review of the relevant GIS data in ArcMap, and occasionally included searching for and studying other lake data sets, fishing maps, satellite photos, and various other documents. In the case of infeasible site-county combinations it was generally assumed, for example, that the boater was mistaken about the exact county of location, but was correct about the larger geographic area: that is,
neighboring counties were given priority over more distant counties when searching for a specific site. The majority of such cases did not represent local (within-county) boating, suggesting the potential for unfamiliarity with the area, and lending some credibility to my assumption. In other cases, the source of uncertainty was the abundance rather than the lack of feasible site-county combinations. While no foolproof method for the assignment exists, following a few general guidelines can aid in establishing with some confidence the exact geographic locations visited in many such cases: lakes with larger surface area, and occasionally lakes located closer to the residence of the boater were favored over smaller, more distant sites. Lake area and proximity to population centers have of course been widely used in the literature to approximate boater traffic in the absence of micro-level data. Moreover, the choice was frequently quite obvious: of the four Horseshoe Lakes and three Little Horseshoe Lakes in Polk and Barron counties, for example, one is larger than the six others combined leading me to believe that assigning all Horseshoe Lake boaters in those counties to the largest lake (the one shared by the two counties) is probably justifiable. In the end, relatively few observations needed to be dropped because of an inability to identify the boating location (and assign a WBIC) with reasonable confidence.

Some lakes in Wisconsin are actually parts of a chain of connected lakes. Each chain of lakes was treated as a single site for both practical and theoretical reasons. Lake chains tend to be comprised of small contiguous lakes, and the lake visited on a boating trip was frequently only referred to by the name of the system it is a member of. Thus, in most cases it is not possible to treat all the individual lakes separately. Pooling them into chains will have a negligible effect on distance measurements in the travel cost model since the lakes in each
chain occupy a compact geographic region. Because the lakes are contiguous or they are connected by a short stretch of river or stream, boaters may well visit several of them on a single trip. Also, because of the direct freshwater connections, we may expect that, following an initial invasion, mussels would eventually spread to all parts of a chain within a short time period.

The discussion of determining inland boating locations has thus far focused only on inland lakes. Rivers and the Great Lakes were naturally easier to locate. At the same time, it is not sensible to treat each river or Great Lake as one site given their large geographic extent. Precisely geo-locating boating sites on rivers and the Great Lakes is impossible, but resolution can be improved by resorting to survey information identifying the counties in which respondents boated. We can, in effect, divide each stream, river and Great Lake shoreline into segments by county. The 200-mile stretch of the Mississippi River runs through eight Wisconsin counties, and will constitute eight boating sites, while Lake Michigan borders eleven Wisconsin counties, and will therefore be divided into eleven choice alternatives.21

Occasionally, inland boating took place in multiple counties that are all traversed by the river named in the survey. In such situations, it is impossible to know with certainty which segments of the river were boating destinations. We do know, however, that at least one of the segments was visited, and it is possible to make an intelligent guess as to which

---

21 Defining site choice as the choice of the primary boating site is especially convenient when the feature is a river or a Great Lake: segments of a contiguous water body are not mutually exclusive alternatives – unless, of course, we are considering the choice of a primary boating site.
county was the primary destination: the boater is assigned to the stretch of river in the county that is closest to his or her residence.\textsuperscript{22}

The nature of the problem is again different when it comes to Great Lakes boating. Each Wisconsin county has a shoreline with at most one of the Great Lakes: the name of a county visited on a Great Lakes boating trip (from question 10) unambiguously reveals whether the destination was Lake Michigan or Lake Superior.\textsuperscript{23} An interesting consequence is that in principle we could preserve up to three Great Lakes sites visited per boater (each county named for question 10). However, if we wish to model all boating trips together, then there is a good reason we should avoid doing so: respondents were asked to identify only one inland boating site, and including multiple Great Lakes sites for each boater would inflate the importance of those sites relative to the importance of inland sites. Therefore, the task was to choose the primary boating site among Great Lakes counties if more than one of them were visited. Since the inland water body identified is the one on which respondents did most of their boating during the survey period, and since the number of days for which respondents boated on Great Lakes waters in each of three Wisconsin counties is known, the choice is straightforward: the county with the highest number of Great Lakes boating days is kept for each boater.

I was able to construct a consistent choice set (where mutual exclusivity is satisfied) by invoking the concept of a primary boating site. The limitations this method entail should

\textsuperscript{22} The surface area criterion applied in the case of inland lakes to assess the importance of a site is not readily applicable to river segments. Another possible method would be to assign the boater to the county first listed in the questionnaire. However, there is no reason why the county mentioned first would be the primary boating destination. In fact, an examination of the responses reveals that it is often not. Since, ceteris paribus, larger distance reduces the likelihood of visiting a site, basing the choice of primary boating destination on shortest distance seems to be a better alternative.

\textsuperscript{23} Incidentally, we even know how many days the individual boated in each county.
also be noted. By definition, the primary boating site is the site receiving the largest share of a respondent’s boating activity. Conversely, a site visited on a multiple-day trip is more likely to receive the largest share of a respondent’s boating activity. Thus, the employment of primary sites increases the likelihood that multiple-day trips will be included in the model. To assess the value of access to a site in a travel cost application, the resources given up in order to gain access to the site need to be measured accurately. They also need to be expended for the single purpose of visiting the site of interest (Phaneuf & Smith, 2003). This is a problem with including multiple-day trips. Driving costs to the site are relatively easy to estimate, but a multiple-day trip may entail several other types of expenses as well, and is more likely to be a multiple-objective trip. Estimating a travel cost model with multiple-day trips included, but not appropriately accounted for, has the consequence of reducing the parameter estimate on the cost variable. Substitution patterns in forecasting, and welfare measures will therefore be affected.

The boating survey does not allow us to identify single-day trips with certainty, nor can we measure expenses other than driving costs (recall that questions about various expenses pertain only to the last trip taken, which may or may not have been a trip to the primary boating site). We could, however, use various rules of thumb to exclude all trips that appear to be multiple day trips – the trade-off, obviously, is that several single-day trips would be eliminated as well. Depending on the criteria used, eliminating all potential multiple-day trips may reduce the number of trips by a factor of eight, and the number of sites in the choice set by a factor of five. Leaving out multiple-day trips would also result in an under-representation of Northern Wisconsin sites that, being farther away from population
centers, are often targeted on such trips. Finally, note that in the present application I am interested in analyzing the spread of an invasive that is not dependent on the length of a visit. Launching a trailered boat with entangled vegetation is not less likely to result in infestation if the boat is used on the water for more than one day; a single and a multiple-day trip both increase the probability of invasion by the same amount. This argument combined with the large reduction in the amount of data associated with excluding multiple-day trips suggests that it may be worthwhile to explore ignoring the conventional wisdom, and handling the set of trips we regard as acceptable more flexibly.

Specification

Travel-cost recreation demand models are based on the premise that by observing the costs incurred by a sample of decision makers to visit recreation sites, we are able to assess the relative significance of the attributes that characterize those sites. In the present context, we observe the choice of a primary boating site made by recreational boaters in Wisconsin. The spatial heterogeneity of residences and recreation locations ensures the variability in travel costs needed to identify the rate at which boaters are willing to trade additional travel expenses for specific site attributes. This section describes how the model is specified and implemented.

We observe a sample of $N$ boaters choosing among $J$ primary boating sites. Each boater is assumed to choose the alternative whose combination of attributes and associated costs yields the highest level of utility to the boater. In the context of my application, the
utility boater $n$ obtains from visiting site $j$ is a linear combination of travel costs, site attributes and boater characteristics:

$$U_{nj} = \gamma C_{nj} + \theta g_j + \phi S_j d_n^a + \beta B_n d_n^l + \alpha^a d_j^a + \alpha^l d_j^l + \alpha^g d_j^g + \xi_j + \varepsilon_{nj},$$  \hspace{1cm} (4.1)

where $C_{nj}$ captures the travel cost to boater $n$ of visiting site $j$, $S_j$ is the water clarity of inland lake $j$ measured in feet of Secchi-depth, and $B_n$ represents the size of boater $n$’s boat expressed as boat length in feet. There are three categorical variables in the above specification: $d_n^a$, $d_j^l$ and $d_j^g$. They enter utility either as part of an interaction term or as individual variables. The first, $d_n^a$, varies over individuals, and is equal to one if boater $n$ is an angler, and zero otherwise. The remaining two categorical variables vary over sites: $d_j^l$ and $d_j^g$ equal one if site $j$ is an inland lake or a Great Lake, respectively, and zero otherwise.

Implicit in this specification is that the coefficient of the third site type (river) is normalized to zero, as required for identification. The parameter $\xi_j$ represents unobserved site characteristics, or in other words, the characteristics of site $j$ that are not explicitly modeled by explanatory variables in the utility specification. As usual, the last term in the utility index, $\varepsilon_{nj}$, accounts for the fact that the utility boaters derive from visiting a water body is not completely observable: $\varepsilon_{nj}$ captures the idiosyncratic tastes of boater $n$ for recreation site $j$ that are only known to the boater. Therefore, from the perspective of the researcher, $\varepsilon_{nj}$ is a random term. Finally, $\gamma$, $\theta$, $\phi$, $\beta$, $\alpha^a$ and $\alpha^g$ are parameters to be estimated. In the paragraphs that follow, I review some of the terms entering utility in more detail.

---

24 More specifically, $d_n^a$ equals one if the activity the boater spent the most time on during the survey period was fishing from the boat, and zero otherwise. Note that we know the primary boating site and the primary boating activity, but in some cases, the primary activity may not be the one the boater engaged in at the primary boating site.
The common feature of all travel-cost recreation demand models is a variable capturing the costs associated with visiting an alternative. I assume that travel costs are directly related to travel distance. Residential locations are identified by the zip code of each boater’s home, and straight-line travel distances between zip code area centroids and boating site centroids are straightforward to determine in GIS software. Travel costs, \( C_{nj} \), are then calculated by multiplying the two-way distance between the residence of boater \( n \) and alternative \( j \) by the per mile cost of travel. The geographic distribution of boater residences and boating destinations gives rise to travel costs that vary over both decision makers and choice alternatives. It is easy to verify from the specification of utility that the coefficient of the travel cost variable is the negative of the marginal utility of income, and therefore we expect its estimate to be negative.

Water clarity, \( S_j \), is an observed site attribute that enters the utility function via two terms, and in a fashion that allows modeling heterogeneous preferences for it. The multinomial logit model can only accommodate non-random taste variation, that is, preferences that vary with observed individual characteristics. One such characteristic over

---

25 I had to abandon my original intention of using highway distances due to inconsistencies in the PC Miler results. Straight-line distances, while not always correct, provide a reasonable approximation for distances driven, and were judged to be significantly more accurate than the distances reported by PC Miler.

26 Per mile travel costs are assumed to be 40 cents for all boaters. In a fishing-related travel cost application, Feather uses the American Automotive Association’s (AAA) 1989 travel cost estimate of 30.6 cents per mile for midsize cars (Feather, 1994). Per mile travel costs in the transient recreational boating context will tend to be higher for two reasons: first, towing a boat reduces the gas mileage of any vehicle, and second, towing a boat may require driving a more powerful vehicle that has a lower base gas mileage. The latter effect was quantified by comparing the AAA’s current per mile cost estimates for midsize sedans and sport utility vehicles.

27 As noted in section 3.4, Secchi-disk measurements are available for only a subset of inland lakes in the choice set. Water clarity was spatially interpolated for inland lakes with missing clarity values, and is restricted to zero for rivers and the Great Lakes. Therefore, clarity affects the utility of inland lake boaters only – hence the \( l \) superscript in \( S_j \). In effect, \( S_j^l \) is an interaction term between water clarity and the dummy variable for inland lakes. In the rest of this chapter, I use \( l, r \) and \( g \) superscripts to differentiate between variables that apply to inland lakes, rivers or Great Lakes respectively.
which preferences for water clarity may be expected to differ is $d_n^a$, the variable indicating whether the boater is an angler. The rationale for such a specification is simple: while many boaters can be expected to value water clarity, anglers are more concerned with fish catch rates, and less concerned with clarity per se. Catch rates are in fact potentially negatively related to water clarity: nutrient-poor oligotrophic lakes tend to be very clear and have low biological productivity; nutrient-rich eutrophic lakes, on the other hand, generally display lower levels of visibility and support large amounts of biomass. Heterogeneous preferences for water clarity are enabled by the interaction of $S_j^l$ and $d_n^a$. The interaction term allows a boater who is primarily interested in fishing to have different preferences for water clarity than a boater who is mainly interested in other activities: a one-foot increase in Secchi-depth increases the utility of a non-angler by $\theta$, and it increases the utility of an angler by $\theta+\phi$. From the discussion above it follows that $\theta$ is expected to be positive, while $\phi$ is expected to be negative.

Inland lakes, rivers and the Great Lakes are assumed to be distinct alternatives. Ceteris paribus, the utility of visiting an inland lake is different from the utility of visiting a river or a Great Lake. The site-type dummy variables $d_j^r$ and $d_j^g$ operationalize this assumption. It is not immediately clear whether the average tastes of boaters are such that they prefer to visit inland lakes, rivers or Great Lakes, and therefore we may not necessarily have a priori expectations for the sign of the site-type coefficients. However, since boating on Lake Michigan or Lake Superior requires a larger craft than boating on inland water bodies, we do expect boat size to affect site choice in a certain way: owners of large boats are hypothesized to be more likely to visit one of the Great Lakes. The interaction term between
the individual-specific variable $B_n$ and the site-specific variable $d_j^x$ is intended to capture this effect. It allows the utility of visiting a Great Lake to vary depending on the length of the decision maker’s boat. The coefficient of the interaction term, $\beta$, is thus expected to be positive.

Finally, unobserved site characteristics $\xi_j$ are also represented in the utility function. Murdock (2006) demonstrated that ignoring unobserved site attributes in recreation demand models may lead to biased travel cost parameter estimates if there is correlation between travel costs and unobserved attributes, and proposed an approach for dealing with the problem. Traditional travel cost models either ignore unobserved site characteristics, or deal with them by including alternative specific constants for a subset of sites. This strategy reduces variation useful for estimating the parameter on any observed site characteristic, or prevents simultaneous estimation of those parameters altogether. Including unobserved site attributes $\xi_j$ for all choice alternatives addresses the correlation and facilitates the unbiased estimation of the travel cost parameter, while also allowing for the estimation of parameters on other site characteristics in a second stage regression.

Note that implicit in the utility specification is that zebra mussels do not directly affect the boating experience. Mussel presence per se does not make boaters more or less likely to visit a choice alternative. For the most part, this assumption is probably realistic: aside from a few extreme cases where navigational buoys and engine cooling systems are damaged, mussels do not directly affect transient recreational boaters.28 While the assumption may be realistic, it is also necessary. It would be impossible to measure the direct

---

28 Hulls of resident boats, on the other hand, are easily damaged by prolonged exposure to mussels.
Recall that the boater survey identifies the primary inland boating site separately from the primary Great Lake boating site. Less than four percent of all boaters visited both an inland water body and one of the Great Lakes within the survey period. That is, less than four percent of boaters are observed to visit more than one choice alternative. My aim is to model boaters’ choices conditional on taking a trip, as opposed to modeling their decision on whether to engage in recreational boating on a given choice occasion. Since I estimate a site choice model, not a participation model, I treat each choice situation by each decision maker as a separate observation: a boater visiting primary inland site $j_l$ and primary Great Lakes site $j_g$ is indistinguishable from two boaters – one visiting $j_l$, and the other, identical boater, visiting $j_g$. Although for simplicity I will continue referring to $N$ as the number of boaters, it should be clear that it is in fact the number of choice occasions represented in the survey, and not the actual number of boaters.

**Estimation**

I estimate the unknown parameters following the two-stage approach proposed by Murdock (2006). A two stage approach is necessary because both observed site characteristics and unobserved site characteristics are represented in the utility function, and a standard (single-stage) discrete choice model is unable to identify the two types of parameters simultaneously. Therefore, the utility function is rewritten in the following manner:

29 However, my model can potentially accommodate an indirect link between mussel presence and utility through the water clarity variable, which I discuss below.
The two equations correspond to the first stage and the second stage, respectively, and the new terms $\delta_j$ are alternative specific constants: they include all characteristics (observed or unobserved) that vary only across recreation sites. Since the full set of alternative specific constants in the first stage soak up all site-specific variation, any variable that varies only over the $j$ dimension (over alternatives) needs to be estimated in the second stage. In particular, water clarity, the site-type categorical variables and unobserved site attributes only enter in the second-stage regression. In my application, the first stage involves the usual maximum likelihood estimation for a discrete choice model, while second stage estimation is accomplished via ordinary least squares. In this section, I describe both stages in detail.

To proceed with the first stage, note that the utility individual $n$ obtains from visiting choice alternative $j$ can be decomposed into a deterministic part and a random part: $U_{nj} = V_{nj} + \varepsilon_{nj}$. The deterministic component, $V_{nj}$, is called representative utility, and includes all non-random components of utility:

$$V_{nj} = \delta_j + \gamma C_n + \phi S^l_j n + \beta B_n d_n^x + \varepsilon_{nj}. \quad (4.4)$$

Representative utility is a function of the observed explanatory variables of the first-stage model, and is therefore known with certainty given the values of parameters $\gamma$, $\varphi$, $\beta$, $\delta_1, \ldots, \delta_J$. Conversely, $\varepsilon_{nj}$ is a term that captures all factors that affect utility, but are not included in the model (Train, 2002). In Murdock's approach, the second stage estimating equation makes it clear what those omitted (from the first stage) factors are. Note that as in a standard linear regression model, dummy variables cannot be introduced for all choice alternatives because of identification issues. Hence, including a full set of alternative specific constants for $N$ alternatives entails estimating $N-1$ parameters: I normalize $\delta_j$ to zero.

---

30 Alternative specific constants capture the average effect on utility of all factors that are not included in the model (Train, 2002). In Murdock's approach, the second stage estimating equation makes it clear what those omitted (from the first stage) factors are. Note that as in a standard linear regression model, dummy variables cannot be introduced for all choice alternatives because of identification issues. Hence, including a full set of alternative specific constants for $N$ alternatives entails estimating $N-1$ parameters: I normalize $\delta_j$ to zero.
specification of representative utility. Boater $n$ knows the value of $\varepsilon_{nj}$, but it is unknown to an outside observer, and is therefore treated as a random variable. Because utility includes a random component, the boater’s choice cannot be predicted exactly; it can only be described in a probabilistic manner (Train, 2002).

Different distributional assumptions about $\varepsilon_{nj}$ lead to different discrete choice models. I assume that each $\varepsilon_{nj}$ is independently and identically distributed type I extreme value – this assumption gives rise to the most popular discrete choice model, the multinomial logit. One of the appeals of this model is that logit choice probabilities have a simple, closed-form solution. The probability that boater $n$ visits choice alternative $j$ can be mathematically derived from the representative utility equation, and is expressed as

$$P_{nj} = \frac{e^{V_{nj}}}{\sum_{j=1}^{J} e^{V_{nj}}} ,$$

(4.5)

where $J$ is the total number of choice alternatives. Note that the choice probability is a function of representative utility only, and therefore, estimating the multinomial logit model does not require simulation. Using the individual choice probabilities expressed in this manner, it is straightforward to derive the likelihood function for the observed sample of boaters. The likelihood function can then be used to find the values of the unknown parameters that maximize its value, that is the first-stage maximum likelihood parameter estimates. I turn my attention to these tasks next.

To write the likelihood function, we need to introduce some additional notation. Let $y_{nj}$ denote an observed choice outcome for boater $n$: $y_{nj}$ equals one if boater $n$ chose site $j$; it
equals zero otherwise. Consequently, the model-implied probability that boater \( n \) chooses the alternative that he or she was actually observed to choose can be expressed as

\[
L_n(\lambda) = \prod_{j=1}^{J} (P_{nj})^{y_{nj}},
\]  

(4.6)

where \( \lambda \) represents the parameters of the utility function in equation 4.2: \( \lambda = \{\gamma, \varphi, \beta, \delta_1, \ldots, \delta_J\} \). Utility function parameters affect this probability because \( P_{nj} \) is a function of representative utility \( V_{nj} \). Note that since only a single choice situation is modeled for each boater, the probability expression above will simply equal the probability of the alternative actually chosen: \( L_n(\lambda) = P_{nj} \). This is individual \( n \)’s contribution to the likelihood function. As the site choice of each decision maker is assumed to be independent, the probability of all boaters in the sample making the particular choice they were observed to make is the product of all the individual likelihood contributions. The likelihood function is therefore

\[
L(\lambda) = \prod_{n=1}^{N} \prod_{j=1}^{J} (P_{nj})^{y_{nj}},
\]  

(4.7)

and its logarithm, the log-likelihood function, becomes

\[
LL(\lambda) = \sum_{n=1}^{N} \sum_{j=1}^{J} y_{nj} \ln P_{nj} .
\]  

(4.8)

The first stage parameter estimates are the values of \( \gamma \), the travel cost parameter, \( \varphi \), the water clarity-angler interaction term parameter, \( \beta \), the boat length-Great Lakes interaction term parameter, and \( \delta_1, \ldots, \delta_J \), the full set of alternative specific constants that maximize the value of the log likelihood function.

Estimating a full set of alternative specific constants in the first stage can be computationally challenging if the number of choice alternatives is large. Therefore, I
augment the gradient-based numerical search algorithm used in maximizing the likelihood function with a contraction mapping designed to increase the speed of recovering parameter estimates. This estimation strategy was originally developed by Berry, Levinsohn and Pakes (1995), and was slightly modified by Murdock (2006) to fit in the framework of recreation demand models. A contraction mapping can work in this situation because of the role alternative specific constants play in a random utility model: including an alternative specific constant for a choice alternative ensures that the predicted choice share of that alternative will equal its observed choice share. This property is exploited by the contraction mapping in the following manner. The maximum likelihood routine searches over the parameter vector except the alternative specific constants, which instead are calculated by the contraction mapping during each iteration:

\[
\hat{\delta}_j^{\text{new}} = \hat{\delta}_j^{\text{old}} + \ln(s_j) - \ln\left[\sum_{n=1}^{N} P_{nj}(\hat{\delta}_j^{\text{old}})\right],
\]

(4.9)

where \(s_j\) denotes the observed share of alternative \(j\). The contraction mapping calculates the values \(\delta_j\) based on the previous estimates and the difference between logged actual and predicted shares (the term inside the brackets). If the actual choice share of alternative \(j\) is larger than the predicted choice share (which is a function of the current value of the alternative specific constant), then the alternative is more desirable than implied by the model, and its alternative specific constant is increased as shown by the contraction mapping. This approach yields the same estimates as the computationally more challenging complete maximum likelihood routine would yield (Murdock, 2006).
The alternative specific constants estimated in the first stage enter as left-hand-side variables in the second stage. Here, they allow us to measure the effect of observed site characteristics whose parameters would otherwise be impossible to estimate in the first stage. Second stage estimation is via ordinary least squares according to equation 4.3, and yields parameter estimates for water clarity $\theta$, the site-type dummy variables $\alpha^l$ and $\alpha^g$, and an ordinary least squares constant, $\delta$. Note that $\delta$ has no effect on behavior. To see this, imagine substituting the second stage equation (4.3) back into the utility function (4.2): the same constant is added to the utility of every choice alternative. As utility is merely an ordinal ranking of preferences, the absolute level of utility is irrelevant. The addition of $\delta$ only affects the level of utility, and therefore it has no effect on the ranking of choice alternatives and behavior.

**Results**

Estimation was implemented in MATLAB using data from the set of motorboaters who were licensed in Wisconsin. I estimated several specifications that differ either in the composition of the choice set, or the set of explanatory variables included in the model. I discuss estimation results and the various model structures below. In addition, Table 4.1 summarizes estimation results, and Table 4.2 provides an overview of the different model specifications.

---

31 I borrowed from computer code written by my colleague, Allen Klaiber, for the workshop “Revealed Preference Outside Markets: Micro-Econometrics in Environmental Economics” (held at North Carolina State University in 2006) to estimate the random utility model.
Models 1-8 are based on the same explanatory variables; they include all variables except the clarity-angler interaction term from equation 4.1, which I consider later due to concerns that it may be endogenous. Models 1, 2 and 3 feature progressively smaller choice sets: a site enters the choice set of decision makers if it was visited by at least five, ten or twenty boaters, respectively, in the three models. Boaters who visited a recreation site not in the choice set are not represented in the model. Consequently, both the number of choice alternatives and the number of individuals (or trips) decrease as we move from Model 1 toward Model 3. All statistically significant parameter estimates have the expected sign in all three specifications (the estimated alternative specific constants are not reported in the table). The travel cost parameter is negative and highly statistically significant in all three cases, as required for any recreation demand model. The boat length-Great Lake interaction term has a positive and significant parameter estimate in all three models, once again confirming our expectations. Recall that the least squares intercept term has no economic meaning in the present context, and therefore we are not concerned with the sign or significance of its estimate. The parameter on inland lake clarity is positive indicating that the average boater values water clarity. Its significance decreases as the choice set shrinks. This pattern is general for all variables since fewer boaters and fewer sites mean less variation is available for estimation. However, the effect is especially pronounced for site attributes such as lake clarity because sites are less numerous than boaters. The only parameter (other than the OLS constant) that is consistently not significant is the coefficient of the inland lake dummy variable. This variable captures the utility difference that exists between boating on inland lakes and boating on inland rivers. However, to correctly interpret its coefficient, note that
the inland lake dummy variable also affects utility through the inland lake clarity variable. Thus, what the term actually captures is the utility difference between inland lakes of zero Secchi-depth and inland rivers. I had no prior expectation for the sign of this parameter, and it seems plausible that it in fact equals zero. Finally, the Great Lake dummy parameter is significant and negative over all three models. This, by an argument analogous to the interpretation of the inland lake dummy parameter, means that a boater who owns a boat of zero length (if such a thing was possible) would prefer an inland site to a Great Lakes site. A larger boat makes the Great Lakes more desirable relative to inland sites, and the parameter estimates imply that the owner of a twenty-foot long boat will be more likely to visit a Great Lake than an inland lake (of near-zero Secchi-depth) under all three scenarios.

Models 4 and 5 include lakes with at least five or ten (respectively) actual Secchi-disk measurements in the choice set. Lake clarity is not interpolated in these two models, and lakes that do not have Secchi-disk observations are not represented. Model 6, on the other hand, uses a different spatial interpolation technique than the first three models. Model 7 is based on summer boating trips only, and Model 8 includes only trips of less than two hundred kilometers, or about one hundred twenty-five miles (one-way, straight-line). In all cases, parameter estimates are similar to those that were estimated in the first three models.

Models 9 and 10 introduce the clarity-angler interaction term. Model 9 is based on the largest possible choice set, while Model 10 includes only summer trips of less than two hundred kilometers (one-way, straight-line) in distance. Once again, estimation results are
very similar to those before.\textsuperscript{32} The new interaction term enters with a significant negative coefficient, as expected. While all boaters value lake clarity, anglers value it less than others do. Since the negative of the travel cost parameter is the marginal utility of money, we can use it to assess willingness to pay for the different site attributes. The estimates from Model 9 suggest that, ceteris paribus, a non-angler would be willing to face a $1.86 increase in travel costs for a one-foot increase in Secchi-depth, compared to $1.30 for an angler.\textsuperscript{33} Also, each one-foot increase in boat length implies a willingness to incur extra travel expenses of $1.90 in order to visit one of the Great Lakes. Since Model 9 maintains the largest choice set and the largest number of explanatory variables, and since the parameter estimates are robust to model specification, Model 9 is my preferred model. All subsequent analyses will be based on this version of the random utility model.

By mapping the estimated alternative specific constants we observe an interesting pattern. Their spatial distribution (from Model 9) is shown in Figure 4.1. The farther north and the farther west a site is located, the larger its alternative specific constant – the relationship is indeed so strong that differences in water clarity and water body classification (which, as shown by equation 4.3, are also included in the alternative specific constants) do not have a perceptible effect on the pattern. That is, the differences are mainly due to unobserved site attributes. The most likely explanation for this striking distribution can be derived from the spatial distribution of boater residences, which is depicted in Figure 4.2.

Boater residences are heavily concentrated in the southeast, precisely in the area where

\textsuperscript{32} Note that parameter estimates from discrete choice models based on different choice sets are technically not directly comparable, but the ratios of marginal utilities are. Since estimates are fairly stable over models, so are marginal utility ratios.

\textsuperscript{33} Figures reported are in 1989 dollars. According to the Consumer Price Index, $1 in 1989 had the same purchasing power as $1.74 has in 2008.
negative alternative specific constants are predicted. To account for the observed frequency of visits to northwestern sites, and for the fact that trips to northwestern sites tend to be long-distance trips by default, those sites are predicted to exhibit desirable unobserved characteristics resulting in large alternative specific constants. This, of course, does not contradict our expectations: northern territories are richer in water features, and are in general less crowded and more scenic. These attributes are unobserved, and presumably valued by most boaters causing them to undertake longer trips in order to reach northern destinations. Additionally, note that I did not exclude potential multiple day trips from the estimation of Model 9. Inasmuch as those trips tend to target water bodies farther away from the residence of boaters, this will also inflate the alternative specific constants of northwestern sites relative to those of southeastern sites.

4.2 The Discrete Duration Model

I now introduce a very simple discrete duration (or hazard) model to analyze the spread of zebra mussels. It will serve subsequently as the basis of a more sophisticated model that incorporates results from the random utility analysis. The discrete duration model is used to assess invasion risk to water bodies in the choice set of recreational boaters. The main purpose of this section is to illustrate the conceptual tool, a binary logit model, and to provide some insight into the way it captures the spatial dynamics of mussel invasion. It is intended to serve as a motivation for subsequent analysis. Empirical results from the model in this section, however, are not necessarily meaningful, and should be interpreted while keeping
this in mind. I first introduce the hazard model and the objective function, and then describe estimation and present maximum likelihood results.

**Specification**

My aim is to characterize the probability that a site will be invaded using explanatory variables that may affect this probability. Let $H_{jt}$, the hazard index for site $j$ at time period $t$, depend on site $j$’s water body type. Inland lakes, rivers and the Great Lakes may not be equally habitable by zebra mussels, or the main mechanism of mussel spread may be different for the three types of aquatic ecosystems. For example, spread by commercial shipping or natural dispersion may dominate the Great Lakes invasion, downstream dispersion may dominate the spread within rivers and streams, and transport by recreational boaters may dominate the spread between hydrologically isolated inland lakes. This specification does not, of course, account for the complex pattern of spatial interactions that characterizes the actual invasion mechanism – the goal here is merely to demonstrate some of the properties of the simple binary model. I write the hazard index as a linear combination of the site-type dummy variables:

$$H_{jt} = \mu + \tau^r d^r_j + \tau^g d^g_j,$$  \hspace{1cm} (4.10)

where $\mu$ is a constant analogous to an intercept term, and the dummy variables denote rivers and Great Lakes, respectively. If the dummy variables were not modeled, then the estimate of $\mu$ would simply reflect the average hazard for any site. Note that none of the explanatory variables change over time, and therefore the predicted site hazard will not change over time.
either (its time subscript is retained nevertheless because it will become important in a subsequent model).

A linear predictor does not constrain values to be in any interval, so the hazard index may take any value. Probabilities, on the other hand, are necessarily between zero and one. Therefore, to characterize the probability of mussel invasion with the hazard index, we need to transform the index in a manner that produces proper probabilities. One way to accomplish this is to use the logit transformation:

\[ \pi_{jt} = \frac{e^{H_{jt}}}{1 + e^{H_{jt}}}, \quad (4.11) \]

where \( \pi_{jt} \) denotes the probability that site \( j \) becomes infested in time period \( t \).\(^{34}\) Invasion probabilities will remain constant over time in this specification since the hazard index is time constant. Rearranging the expression for invasion probability, it becomes clear that it is actually a non-linear function of the probability that depends linearly on the covariates:

\[ \ln \left[ \frac{\pi_{jt}}{1 - \pi_{jt}} \right] = \mu + \tau^r d_j^r + \tau^e d_j^e. \quad (4.12) \]

Thus, a simple linear hazard index can be used to generate probabilities that are necessarily between zero and one.

**Estimation**

We can describe the zebra mussel invasion status of a site by a binary variable \( z_{jt} \) that, in each time period \( t \), takes a value of one if site \( j \) is infested in that period and a value of zero if the site is not infested. Invasion status is described by \( \pi_{jt} \) ex ante, and by \( z_{jt} \) ex post. Models for

\(^{34}\) Thus, the probability that site \( j \) does not become infested during time period \( t \) is \( 1 - \pi_{jt} \).
binary data generally consider the realizations of a variable corresponding to $z_{jt}$ as random. Note, however, that mussel invasion status cannot be viewed as completely random. If site $j$ is mussel-free in time period $t$, then $z_{jt+1}$ can in fact be considered random, because its value is unknown in time period $t$. Once a site becomes infested, however, future values of $z_{jt}$ are known with certainty. The site will never return to an uninfested state since, as I have remarked, once introduced, zebra mussels are unlikely to disappear from a water body. That is, I treat $z_{jt}$ as a binary random variable only until time period $t_j$, the period in which site $j$ is invaded. This time period is site-specific. The risk set, that is, the set of sites that can potentially be invaded in the current period, shrinks over time as new invasions take place.35 Newly invaded sites fall out of the risk set of future periods, and subsequent realizations of the binary variable for those sites should not affect the likelihood function since its values are entirely deterministic.

The likelihood of observing site $j$’s actual invasion outcomes over time can be expressed as

$$ L_j(\kappa) = (1 - \pi_{jt}) \sum_{t=1}^{T} (1 - z_{jt}) \prod_{t'=t}^{T} \pi_{jt'}^{\pi_{jt'}} , \quad (4.13) $$

where the sum in the first exponent is simply the number of time periods site $j$ was not infested, $z_{jT}$ is the invasion indicator in the final time period of observation, and $\kappa$ denotes the parameters of the hazard index. If site $j$ becomes invaded at some point, then the last term ensures that $\pi_{jt}$, the probability of invasion is represented in the likelihood function. The likelihood and log-likelihood for the whole sample of sites become

35 Thus, the choice set of boaters can be partitioned into the risk set (uninfested sites), and the complement of the risk set (infested sites).
\begin{equation}
L(\kappa) = \prod_{j=1}^{J} (1 - \pi_j) \sum_{t=1}^{T} (1 - z_{jt}) (\pi_j)^{z_{jt}}, \tag{4.14}
\end{equation}

\begin{equation}
LL(\kappa) = \sum_{j=1}^{J} \sum_{t=1}^{T} (1 - z_{jt}) \ln(1 - \pi_j) + z_{jt} \ln(\pi_j). \tag{4.15}
\end{equation}

I use the geospatial data on the distribution of zebra mussels from the U.S. Geological Survey to estimate the binary logit model. Recall that the mussel data features the locations of confirmed zebra mussel sightings from 1989 to 2006; each year corresponds to one time period in my model.

**Results**

The estimation results of the hazard model are shown in Table 4.3. All parameter estimates are statistically significant. The large negative constant is due to the fact that invasions, on average, are unlikely events: mussels spread to only about two percent of the risk set in each period. According to the results, Great Lake and river status both increase invasion risk. This occurs because a larger proportion of Great Lakes sites and river sites are invaded than inland lakes (mussels have been found in coastal waters of nearly all counties bordering one of the Great Lakes, and in all counties along the Mississippi River basin). The parameter estimates can be translated into probabilities via equation 4.11. Table 4.3 also shows the probability of invasion for each type of aquatic ecosystem separately.

4.3 The Integrated Economic and Ecological Model

I have specified a site choice model to describe recreational boater behavior, and introduced a discrete duration framework to model the spread of zebra mussels. My ultimate goal is to
make a connection between the two models, and to analyze the mussel invasion while explicitly acknowledging that recreational boaters are at least partially responsible for transporting mussels between water bodies. In this section, I propose a technique to quantify human-induced mussel threat to a water body that entails using trip probabilities from the behavioral model in a novel fashion. I provide a detailed description of what I call the human threat variable, and of using it in a discrete duration framework to model mussel spread. I compare several specifications and evaluate them both objectively (based on the log-likelihood value at the maximum likelihood parameter estimates, and the accuracy of their in-sample predictions), and subjectively (based on the similarity of the actual and predicted patterns of spread in in-sample and out-of-sample predictions).

The Human Threat Variable

Gravity models prevalent in the ecological literature on freshwater invasions characterize invasion threat by the force of attraction between a source and a destination. The definition of sources and destinations, and the measurement of attraction between them are application-specific, though in all models the force of attraction is a function of a quantity analogous to mass in Newtonian equations, and distance. Thus, in a simple gravity model, attraction between two lakes may depend, for example, on their surface area. Lake size may be correlated with boating pressure, but it is more appropriate to use an actual boater survey when available to measure boating traffic at various sites, and to use the number of boaters at each site to quantify the attraction between them. When survey data is not available, figures on the number of registered boaters and lake size have been used to approximate the
attraction between a source county and a destination lake. Note that neither of these specifications for attraction replicates the two-step process by which zebra mussels actually disperse. Mussels are not moved to another water body directly; they are transported to another water body by the boater. Thus, the interaction is not between water bodies directly, and it is also not merely between population centers and water bodies as gravity models implicitly suggest. I propose replacing the gravity-type force of attraction with a human threat variable calculated from the multinomial logit trip probabilities that better represents the complexity of vector movement, and, crucially, is also able to accommodate changes in recreational boater behavior.

For a new invasion to occur, it is necessary that two events take place: first, boater \( n \) visits infested site \( i \) and picks up mussels, and second, within a time span short enough for zebra mussels to survive on land, boater \( n \) also visits uninfested site \( j \) (and the mussels thereby colonize site \( j \) via site \( i \)). Both of these events involve the boater’s site choice decision, a decision that I have already described via the behavioral model. Given the specification of representative utility and the maximum likelihood parameter estimates, the site choice probabilities can be calculated using equation 4.5.

The movement of boaters is between residential locations and recreation sites, but we are ultimately interested in characterizing the risk of mussel spread between two water bodies. Consider for the moment a situation in which a single boater faces a choice between only two boating sites: site \( i \) is infested, and site \( j \) is not. Let \( P_{ni} \) be the probability that the boater visits site \( i \), and \( P_{nj} \) be the probability that the boater visits site \( j \). Then, we may
estimate the threat of mussel invasion to site \( j \) by simply taking the product of the site choice probabilities:

\[
R_j = P_{ni} \cdot P_{nj}.
\]  

(4.16)

The value of \( R_j \) is largest if the two sites are equally likely to be chosen by boater \( n \), as illustrated in Figure 4.3. If either of the sites is far superior and hence significantly more likely to be visited, then there will be little boater traffic between the two sites, and the value of the threat variable will be correspondingly small.\(^{36}\) Also note that the boater survey does not permit modeling trip frequency or choices over time because a single choice occasion is recorded for each boater. However, an invasion event requires at least two trips from the boater. By using the product of the site choice probabilities to quantify threat, I implicitly assume that the choice probabilities represent behavior accurately over repeated choice occasions. That is, boater preferences are assumed to be stable over time.

To generalize to multiple decision makers, assume that there are several \( (N) \) boaters making a choice between the two sites. The value of the threat variable now has to reflect the threat contribution from each boater. The easiest way to accomplish this is by summing all the individual contributions:

\[
R_j = \sum_{n=1}^{N} P_{ni} P_{nj}.
\]  

(4.17)

\(^{36}\) I consider \( R_i \) to be zero. In a different context, the relationship could be symmetric: for example, if my intention was to merely characterize the level of boater traffic (attraction) between sites \( i \) and \( j \), then \( R_j \) would equal \( R_i \). However, by focusing on \( R_j \), I acknowledge the fact that in the present context the relationship is in fact asymmetric since a mussel-free site does not pose any threat to an already invaded site.
It is important to emphasize that $R_j$ in equation 4.17 is not a probability. To confirm this, note that $R_j$ is not restricted to be in any interval, and its value can be greater than one. At the same time, it does reflect the threat of zebra mussel transport to the uninfested site: the larger $R_j$, the more likely it is that mussels will be transported by recreational boaters from site $i$ to site $j$. It is useful at this point to take a more rigorous look at the distinction between probability of transport and threat. Recall that the probability of the union of two events is calculated by summing their marginal probabilities, and then subtracting the probability of the intersection of the events from the sum to avoid double counting. If the events are defined as “boater $n$ visits both site $i$ and site $j$”, then $R_j$ differs from the probability of “any boater visits both sites” because the intersection probabilities are not subtracted from $R_j$. I argue that $R_j$ is indeed a better metric for threat than the statistical probability would be because the relevant event for invasions is not “any boater visits both sites”: the number of boaters visiting both sites is actually of interest, and therefore there is no need to subtract intersection probabilities. Intuitively, the threat two (identical) boaters pose is exactly double the threat a single boater would pose.

To generalize to multiple choice alternatives, assume that there are a total of $J$ recreation sites in the choice set of the $N$ decision makers, and that only one of the $J$ sites is currently invaded by mussels. The threat to each non-invaded site is calculated in the same manner as in the two-site scenario. The reason I consider this (not fully general) situation separately is to highlight the spatial properties of the human threat variable. Because $R_j$ is

37 In the single-boater-case, it would be possible to argue that, if the two choice occasions are independent, $R_j$ captures the probability that both sites will be visited by the boater (although a time frame for the site choice decisions is not specified). However, I intentionally avoided referring to $R_j$ as a probability, because in the more general case, it clearly is not.
constructed from site choice probabilities derived from a spatially explicit travel cost model, 
\( R_j \) will also reflect spatial heterogeneity. Imagine a simple example with three recreation 
sites, in which the two uninfested sites are equidistant from the only invaded site. The choice 
alternatives are otherwise identical. If the distribution of boaters is spatially heterogeneous, 
then the threat to the two uninfested sites may differ: the lake located closer to the population 
center will be at a higher risk of invasion because boaters will be more likely to visit it (since 
the sites are assumed to be identical, choice probabilities will be determined by travel 
distance alone). In this example, \( R_j \) will be site-varying due to its dependence on human 
behavior, whereas a simple gravity model may predict equal attraction between the sites. 
Threat is a function of the location of both sites and individuals. Moreover, it is also a 
function of the attributes of the alternatives since site choice probabilities are derived from 
representative utility which includes site attributes. Thus, spatial interactions will change if a 
site attribute changes (or if the composition of the choice set changes,) and \( R_j \) will reflect any 
such change.

Finally, consider the fully general case of \( N \) boaters and \( J \) sites, \( I \) of which contain 
mussels. The total risk of spread to site \( j \) will now comprise the aggregate of threats imposed 
by each invaded site separately. The threat variable becomes

\[
R_j = \sum_{i \in I} \sum_{n=1}^{N} P_i P_{nj},
\]

where the first sum is over the set of invaded sites. Of course, the number of invaded sites 
will increase over time as mussels colonize new water bodies. Incorporating the temporal 
dimension, the threat variable can be written as
\[ R_{jt} = \sum_{i \in I_j} \sum_{n=1}^{N} P_{ni} P_{nj}. \] (4.19)

Consequently, the threat to site \( j \) is time-specific, and depends on the set of currently invaded sites. It will never decrease over time because the set of invaded sites (the complement of the risk set) never shrinks. The threat variable thus defined is able to capture fairly complex spatial and temporal relationships: it reflects both distance and quality aspects of recreation sites, and it reflects changes in both the choice set and the risk set. Moreover, \( R_{jt} \) has a cardinal interpretation: its value will double if either the number of (identical) decision makers doubles, or if all decision makers’ probability of choosing site \( j \) doubles, or if all decision makers’ probability of choosing all invaded sites doubles.\(^{38}\)

Recall that Padilla et al. determined, using the 1989-1990 Wisconsin Recreational Boating Study, that the number of boater connections to the Great Lakes (zebra mussel sources) are better predictors of mussel invasion for a specific site than overall boater traffic at the site (Padilla et al., 1996). Although the authors do not articulate it, this is presumably due to the distances involved: the greater the distance between two water bodies, the less likely that they will be visited by the same boater. The threat variable formalizes this concept. Trip probabilities generally decrease with distance, and threat to sites far from current sources will thus tend to be small. Furthermore, my specification allows for more flexibility by being able to reflect potential changes in behavior, and by allowing for interactions between two inland water bodies as well.

\(^{38}\) I am abstracting from the proportional substitution characteristic of the logit model here. In fact, doubling \( P_{nj} \) for all \( n \) would not exactly double \( R_{jt} \) because other trip probabilities, including all \( P_{ni} \), would decrease by the same proportion to ensure that choice probabilities sum to one.
The Integrated Model

The discrete duration model I employed to predict invasion probabilities in the previous section was spatially and temporally invariant; the hazard for site \( j \) depended only on the type of site \( j \). The human threat variable is a conceptually attractive construct to quantify the invasion pressure a site faces due to recreational boating. To actually model the role recreational boaters play in the mussel invasion, we may directly use the threat variable in the hazard index of the discrete duration framework:

\[
H_{jt} = \mu + \rho R_{jt}.
\]  

(4.20)

The value of the hazard index in equation 4.20 is site and time-specific. Invasion probabilities \( \pi_{jt} \) calculated via equation 4.11 (at the maximum likelihood estimates of \( \mu \) and \( \rho \)) will also vary over alternatives and over time, and they will incorporate the threat to each site posed by recreational boating.

The most evident difference between the hazard index in equation 4.20 and the one in equation 4.10 is that the threat-variable-based hazard index does not include site-specific categorical variables. The rationale for classifying inland lakes, rivers and the Great Lakes separately was twofold. First, they may not provide equally suitable habitats for zebra mussels, and second, different mechanisms may account for mussel spread to the different types of water bodies. Historically, mussels first colonized the Great Lakes, and shortly thereafter the Mississippi River. The first Wisconsin inland lake invasion did not take place until 1994 – about five years after the initial introduction to North America. Since then, forty-seven inland lakes (in the choice set), eight river alternatives (as defined in the site choice model), and two Great Lakes alternatives (again, as defined in the site choice model)
have been invaded.\textsuperscript{39} Mussel spread has clearly been dominated by the invasion of inland lakes for more than a decade. This recent dominance and the compelling ecological evidence pointing to the role of transient recreational boating in inland lake invasions provide a case for concentrating modeling effort on inland lakes. The first year for which I model the spread of zebra mussels is therefore 1994, the year of the first inland lake invasion; invasion status in 1993 is taken as the initial condition, and no attempt is made to explain the ecological process in years prior to 1994.\textsuperscript{40} A majority of the Great Lakes alternatives and river alternatives are thereby excluded from the risk set, leaving little variation to estimate coefficients on potential site-type categorical variables. This is the reason the integrated model specification does not include the dummy variables seen in the simple discrete duration model.

Before I formalize the likelihood function, I digress slightly and discuss a subtle but important point on the timing of events. Recall that in the simple discrete duration model, the likelihood contribution of site $j$ was $1 - \pi_{jt}$ in each year the site remained mussel-free, and $\pi_{jt}$ in the year the site became invaded. Since $\pi_{jt}$ was actually time-invariant, the only important task was to keep track of the number of $1 - \pi_{jt}$ terms in the product (and the multiplication by $\pi_{jt}$ if the site became invaded at some point). I did not have to worry about precisely defining a timeline because invasion probabilities were constant over time. However, the human threat variable introduces a time dimension to the model, and it becomes critical to use the right invasion probabilities in the likelihood function. Specifically, I assume the following sequence of events. Overall invasion status in the beginning of period $t$ determines the

\textsuperscript{39} As of early 2006.
\textsuperscript{40} It is in fact believed that recreational boating had no role in the Great Lakes and Mississippi River invasions.
magnitude of the threat variable (and the value of the hazard index) for the set of sites that are not infested at the beginning of period \( t \) (that is, the period \( t \) risk set). Some of these sites will be subject to a large hazard, and become colonized by mussels at the end of period \( t \).

New invasions are only observed at the beginning of period \( t + 1 \). The newly infested sites will act as mussel sources in period \( t + 1 \), thereby increasing threat to each site in the (now smaller) risk set, some of which become sources themselves by period \( t + 2 \), and so forth. Significantly, it is the period \( t \) hazard that determines period \( t + 1 \) invasion outcomes. When writing the likelihood function, this one-period lag between cause (hazard) and effect (invasion) needs to be addressed explicitly. The likelihood contribution of site \( j \), and the likelihood and log-likelihood functions for the whole sample of \( J \) sites now become

\[
L_j(\kappa) = \left[ \prod_{t < t_j} (1 - \pi_{j,t-1}) \right] \left( \pi_{j,t-1} \right)^{z_{jt}},
\]

(4.21)

\[
L(\kappa) = \prod_{j=1}^{J} \left[ \prod_{t < t_j} (1 - \pi_{j,t-1}) \right] \left( \pi_{j,t-1} \right)^{z_{jt}},
\]

(4.22)

\[
LL(\kappa) = \sum_{j=1}^{J} \left\{ \sum_{t < t_j} \ln(1 - \pi_{j,t-1}) \right\} + z_{jt} \ln(\pi_{j,t-1}),
\]

(4.23)

where time period 0 refers to the initial condition, and I use \( t_j \) to denote the time period in which site \( j \) is first observed to contain zebra mussels. If site \( j \) does not become colonized by mussels, then I set \( t_j = T + 1 \), and \( z_{jt} = 0 \).

As in the simple discrete duration specification, a numerical search algorithm is used to locate the values of the unknown parameters (\( \mu \) and \( \rho \) in this case) that maximize the value of the log-likelihood function. Estimation results for this model are shown in Table 4.4 in the
column labeled Model I (other columns refer to specifications I will introduce later). Both estimates are statistically significant, and of the expected sign. However, as Murphy and Topel (1985) demonstrated, the standard errors computed in this step overestimate precision because the human threat variable is an estimated regressor: it is based on parameters estimated in the random utility model, and is therefore measured with a sampling error. In Appendix B, I discuss the possibility of deriving corrected standard errors for the parameters estimated in this step. The large negative constant arises because invasions are unlikely events – it ensures that predicted invasion probabilities will also be small. The interpretation of the threat variable coefficient, on the other hand, is less obvious. As shown by equation 4.12, the relationship between model parameters and invasion probabilities is highly non-linear: in the integrated economic and ecological model, $\rho$ reflects the change in a non-linear transformation of the invasion probability caused by a unit change in human threat. However, it is difficult to intuitively assess what a one-unit change in the threat variable means. Site choice probabilities, the number of decision makers, and the number of infested and uninfested sites in the choice set all interact to determine $R_{jt}$. The value of the threat variable ranges from near zero to 51.06 with an average of 2.32, and using equation 4.11, we can translate these numbers into invasion probabilities. For a site with an average value of the threat variable, the predicted invasion probability will be about 0.016; for a site with a threat variable value of ten, it will be 0.035; and for the site with the highest threat variable value, invasion probability will be approximately 0.7. In general, the large negative constant dominates the effect of the threat variable, and the overwhelming majority of predicted invasion probabilities are small.
Discussion of Estimation Results

The parameter estimates are statistically significant, but I am ultimately interested in an assessment of their economic significance: how well can the model explain the actual sequence of events? In other words, is the model able to reproduce the observed pattern of zebra mussel spread from a set of initial conditions? Parameter estimates and statistical significance are not necessarily informative in answering these questions, and models are often evaluated on the basis of the accuracy of their in-sample predictions. Note, however, that in this case it is challenging to define what is meant by the accuracy of predictions because invasion events are both spatially and temporally differentiated. For example, Lake Geneva in southeastern Wisconsin was one of the first inland lakes to be colonized by zebra mussels, and a good model would therefore predict an early invasion for Lake Geneva. At the same time, the degree to which a model accurately describes the observed pattern also depends on spatial properties. Inland lake invasions are heavily concentrated in southeastern Wisconsin. Even if its forecasts are not exactly correct, it could be argued that a good model would predict invasions in the right geographic neighborhood. Because they need to consider both spatial and temporal effects, objective criteria to evaluate model performance are difficult to design. In the remainder of this section, I consider these issues in more detail, and I also return to them in the next section.

One strategy to perform in-sample predictions is to define a critical invasion probability. Threat variables and invasion probabilities resulting from the initial conditions (invasion status in 1993) are calculated, and each site with a predicted probability over the critical value is assigned an infested status. In the next time period, threat variable values and
invasion probabilities are updated to reflect the change in predicted mussel distribution, and a new set of sites with probabilities over the critical value is assigned an infested status. A simple method to evaluate model performance is to look at the fraction of correct positives—that is, the fraction of lakes that are correctly predicted to be infested within a specific number of iterations. The disadvantage of this approach is its obvious dependence on an arbitrary quantity: the critical probability; theory does not offer any guidance in choosing its value. A reasonable objective when choosing a critical value may be that the model reproduces approximately the correct number of invasions over a certain time horizon (where each iteration may be thought of as one year). A too large critical probability will result in too few invasions, and a too small critical probability will produce too many. The empirical implementation of this best-fit parameterization approach is, however, difficult. Consider the two extreme critical values of zero and one. If the critical probability is zero, all sites in the risk set will be invaded during the first iteration according to the model forecasts. If the critical probability is one, no site will be invaded during the first period, or in any subsequent periods. In general, both low and high critical probabilities result in short model horizon—either because the risk set becomes empty, or because no site reaches the critical probability, in which case the invasion process stops as there will be no change in the risk set and in the associated future values of the threat variable. The problem arises because the predicted spread process is deterministic, and because there is generally little variation in predicted invasion probabilities.⁴¹

⁴¹ For some model specifications, it is in fact impossible to calibrate critical probability in a manner that does not result in a termination of model dynamics within a few iterations.
Due to these issues, I decided to abandon the idea of specifying a critical probability to perform in-sample predictions. Instead, during each predictive iteration of the model, only the site with the highest predicted hazard is added to the set of invaded sites. The threat variable and the hazard index are updated, as before, during each iteration. There is no predetermined cut-off value, and the process can run as long as there are sites in the risk set. The price of this higher flexibility is the loss of the (potential) correspondence between iterations and years. With this strategy, it becomes impossible to make forecasts relating to a specific time horizon; but it is possible to rank sites according to their predicted position in the invasion sequence. Since 1993, fifty-seven alternatives (among them forty-seven inland lakes) have been invaded. If left to run for fifty-seven iterations, Model I correctly predicts thirty invasions. A visual inspection of the invasion forecast map reveals that all fifty-seven predicted invasion occur in approximately the correct geographic area, as shown in the second panel of Figure 4.4. The first data column of Table 4.4 also displays the first ten predicted invasions using Model I. Eight of the sites listed have actually been discovered to contain mussels, giving the model an eighty percent accuracy over ten iterations.

A comparison of the geographic and temporal distribution of actual and predicted invasion events reveals that spread predictions exhibit larger spatial jumps, and slightly less clustering than the actual data. Clustering is a distinct feature of the Wisconsin mussel invasion: several nearby lakes are occasionally colonized within a short time period. There is often also evidence for the existence of freshwater connections between those lakes, suggesting that some of the invasions may be due to natural processes rather than to the influence of transient recreational boating. Since the hazard only depends on a threat variable
constructed from components of a behavioral model, this feature of observed mussel spread is not captured by the integrated economic and ecological model. In the following section, I discuss a potential method to extend the current model to better characterize the observed pattern of invasion.

Environmental Threat

Even when working with initial conditions that precede the first inland lake invasion, the discrete duration model with the human threat variable yields a reasonable approximation to the actual course of the zebra mussel invasion. Recall that the aim of the human threat variable is to characterize the risk of mussel transport between pairs of recreation sites that is due to transient recreational boating. It seems plausible, however, that other factors also play a role in inland invasions – especially in the short-range spread to hydrologically connected water bodies. Consider specifying an environmental (or ecological) threat variable to account for such situations. A simple characterization of this short-range threat would be

\[
E_{ij} = \sum_{i=1}^{j} D_{ij}^{-2} z_{it},
\]

(4.24)

where \(D_{ij}\) is the straight-line distance between water bodies \(i\) and \(j\), and \(z_{it}\) is the status of \(i\) in time period \(t\). Accordingly, the environmental threat to uninfested site \(j\) in period \(t\) depends inversely on the squared distance between site \(j\) and each currently infested site. Environmental threat dies out faster than human threat since it is based on squared inverse distance, and is therefore appropriate to represent short-range threat. Like human threat, environmental threat also changes as the set of invaded sites grows over time. Note that
although I refer to $E_{jt}$ as environmental threat, it does not necessarily reflect natural processes only: it captures invasion threat from all factors other than transient recreational boating. Proximity to a source may increase the likelihood of mussel spread due to natural population growth (larval dispersion within connected water bodies), transport of adult mussels by other animals (some of which, for example turtles, may be able to transport mussels over land as well for short distances), and transport of mussels by human activities other than recreational boating. The major difference between the two threat variables is that $E_{jt}$ is not derived from behavioral data. It is merely a reduced form representation of direct attraction between different water bodies – much in the spirit of gravity models. Together the human and ecological threat variables account for the primarily human-mediated long-distance dispersal as well as the partly natural short-distance spread of zebra mussels.

Although I appealed to the similarity of ecological threat and the force of attraction in gravity models, recall that the latter depends on variables analogous to both mass and distance. Mass is not represented in the threat specification above. As noted in Chapter 2, lake surface area is often used for mass in the biological invasions literature, and it is certainly possible that a large mussel-infested lake poses a larger threat to its neighboring water body than a small lake would (even excluding boater threat). Conversely, it is also possible that a large uninfested lake (perhaps by attracting more wildlife and more human activity) is at greater risk of being invaded than a small uninfested lake would be in the same location. Therefore, the direct interaction between two water bodies, or the potential for vector traffic between them, depends on the surface area of both lakes. It is straightforward to incorporate lake size into my specification of environmental threat for inland lakes:
\[ E_{ij}^l = s_j \sum_{i=1}^{j \prime} D_{ij} z_{ij} , \quad (4.25) \]

where \( s_j \) and \( s_i \) are the surface areas of inland lakes \( j \) and \( i \), respectively, and \( J^l \) is the set of inland lakes in the choice set (that is, the risk set and its complement). However, recall that the choice set also contains sites that are not inland lakes. It is not immediately obvious how to measure the area of a river segment or of a Great Lakes shoreline (though perimeter could potentially be used to approximate area), and I chose not to differentiate between the infectiousness of different river segments, and of various Great Lakes sites. The ecological threat emerging from infested rivers and Great Lakes is therefore specified similarly to the simple threat variable in equation 4.24, without accounting for surface area:

\[ E_{ij}^r = \sum_{i=1}^{j \prime} D_{ij} z_{ij} , \quad (4.26) \]

\[ E_{ij}^g = \sum_{i=1}^{j \prime} D_{ij} z_{ij} , \quad (4.27) \]

where, as before, \( r \) and \( g \) superscripts denote rivers and Great Lakes. The site-type dummy variables \( d_j^r \) and \( d_j^g \) are implicitly included in these equations since the summations are over the set of rivers and the set of Great Lakes, respectively. However, the concerns that arose about not being able to estimate coefficients on these variables are not relevant here, because they are now (in effect) interacted with distance which varies over both the \( i \) and \( j \) dimensions. Thus, in theory, there is no reason to suppose the parameters on these threat variables are not recoverable from the data.

Note the direction of the interactions resulting from the various threat variables. Human threat (or more appropriately, transient recreational boater threat) is based on trip
probabilities from a multinomial logit model in which the choice set of decision makers includes all three types of sites. Therefore, $R_{jt}$ reflects the threat of mussel spread from any site to any other site. On the other hand, $E_{jt}^l$ captures short-range interactions only among inland lakes. Finally, $E_{jt}^r$ and $E_{jt}^g$ reflect the ecological threat to any water body that is due to invaded rivers and Great Lakes, respectively. Inland lakes, which are the focus of my research, are affected by all four types of threat, while rivers and Great Lakes are not threatened by invaded inland lakes. I do not consider this a major shortcoming since the majority of river and Great Lakes invasions are not modeled: they are included in the initial conditions. The environmental threat variables for the three types of hydrological features are used together with the human threat variable in the discrete duration framework.

**Model Evaluation**

I considered hazard indices based on both the simple and the disaggregate environmental threat variables:

$$H_{jt} = \mu + \rho R_{jt} + \eta E_{jt}^l,$$  \hspace{1cm} (4.28)

$$H_{jt} = \mu + \rho R_{jt} + \eta^l E_{jt}^l + \eta^r E_{jt}^r + \eta^g E_{jt}^g.$$

(4.29)

Maximum likelihood parameter estimates and asymptotic t-statistics based on robust standard errors for the different model specifications are presented in the second and third data columns of Table 4.4. With the exception of the river threat coefficient, all estimates are statistically significant at the ten percent level, and are of the expected sign. Log-likelihood values at the maximum likelihood parameter estimates are also shown for each model in...
Table 4.4. Since Model III has the highest log-likelihood value, it is the best specification by
this criterion. Both models with environmental threat outperform the model that employs
human threat only.

I also evaluated models by the accuracy of their in-sample predictions, and present
three statistics based on in-sample predictions in Table 4.4. The simplest, $PC$ (proportion
correct), is the proportion of correct positives among the first fifty-seven predictions (recall
that fifty-seven is the number of invasions since the initial condition): Model II is the most
accurate, followed by Model III and Model I. The order in which sites are predicted to
become invaded and the spatial distribution of predictions do not influence this statistic; it
merely offers a quick numerical summary at a specific moment in the model. The remaining
two statistics ($SATD$ and $SASD$) are slightly more comprehensive, and are intended to better
characterize the temporal and spatial accuracy of invasion predictions. I turn my attention to
describing these statistics next.

The sum of absolute temporal deviations, $SATD$, is based on the idea that a good
model would approximately replicate the actual sequence of invasion events. The sequence
in which sites were invaded is examined, and each site is assigned an invasion rank according
to its place in the observed sequence. Forecasted ranks are also recorded. In a model that
perfectly predicts all invasions, each site’s predicted rank would be the same as its actual
rank. That is, the deviation between the two sequences would be zero. The statistic is based
on such deviations between the actual and predicted invasion sequences: it is essentially a
temporal evaluation method. $SATD$ is calculated as

$$ SATD = \sum_{j \in I} |\hat{I}_j - \tilde{I}_j|, $$

(4.30)
where \( t_j \) is site \( j \)'s rank in the sequence of predicted invasions, \( \bar{t}_j \) is its observed rank, and \( I^I \) is the set of all invaded inland lakes. The sums are only over inland lakes because modeling inland lake invasions is my primary goal, and, as discussed, I expect predictions for inland lakes to be more accurate than for rivers or Great Lakes.

In practice, the exact order in which sites become infested is not observed because, generally, multiple invasions occur in a year. Therefore, I use a range of values for \( t_j \): the interval the site’s rank is known to fall into. Hence the difference between \( t_j \) and \( \bar{t}_j \) is that the former represents the time period, or year of invasion, and the latter is the range of values corresponding to site \( j \)'s observed rank. For example, three inland lakes were invaded in each 1994 and 1995 (the first and second years modeled). The rank of each lake invaded in 1994 is therefore assigned the interval \([1, 3]\), and the rank of each lake invaded in 1995 is assigned the interval \([4, 6]\). When calculating SATD scores, I use the smallest possible deviation between the two sequences for each site.

The best model using this criterion is the one that minimizes the sum of absolute deviations between the two sequences, that is, the model whose invasion predictions for the entire choice set most closely match reality. Using the temporal deviations statistic to compare models is sensible as long as the different specifications are based on the same initial conditions and contain the same number of sites in the choice set. According to the statistic, models including environmental threat perform better than the model based only on human threat. Also, the richer specification for environmental threat outperforms the simple, aggregated specification.
Although this method of evaluating models is based on the entire sequence of invasions, and is more informative than the proportion of correct positive predictions at a certain point in time, note that it is still an incomplete characterization of model performance. Specifically, it is invariant to the geographic location of false positive forecasts. Since temporal deviations are derived from observed invasion rank, sites that are not infested have no direct effect on $SATD$. False positives affect the sums only through their effect on the predicted ranks of correct positives, but are otherwise indistinguishable. A false positive prediction increases the total score by the same amount irrespective of the geographic location of the site. However, it may not be desirable to treat all false positives as equivalent since they differ in their proximity to existing invasions. A model that predicts early invasions (whether correct or false) in approximately the right geographic neighborhood may be considered a more accurate specification than a model with the same temporal deviations score predicting false positives that are geographically far from actual invasions. Essentially, $SATD$ is only a temporal evaluation technique, and does not accommodate penalties for spatial deviations.

It is of course feasible to devise spatial evaluation methods as well. Consider penalizing false negative predictions based on their geographic distance from the closest actually invaded site, and write

$$\text{SASD} = \sum_{j \in \hat{I}'} \min(D_{i,j}, D_{2,j}, \ldots, D_{g,j}), \text{ where } \{1,2,\ldots,i\} \in I',$$

(4.31)

and $\text{SASD}$ stands for sum of absolute spatial deviations. $D_{ij}$, recall, is the straight line distance between sites $i$ and $j$. $I'$, as before, is the set of all infested inland lakes, and $\hat{I}'$ is the set of inland lakes predicted to become infested, given that it contains the same number of elements
as set $I^t$. Since model predictions can run as long as there are sites in the risk set, it is necessary to choose a time period for evaluation with this method. By requiring that the predicted and actual invasion sets contain the same number of sites, I implicitly chose a period (or iteration) for evaluation. Thus, this method offers a spatial assessment of predictive accuracy, but lacks a temporal component. While $SATD$ is based on temporal errors in correct positives, $SASD$ is based on spatial errors in false negatives. An evaluation technique that incorporates both spatial and temporal effects would entail defining the relative significance of spatial and temporal prediction errors, and would therefore be necessarily arbitrary.

Based on spatial evaluation, the best model is the one that minimizes $SASD$. As shown in Table 4.4, models with environmental threat perform better than the model including human threat only. As opposed to the temporal statistic, it also suggest that the simple environmental threat variable is a slightly better specification than the more complex, disaggregated environmental threat.\(^{42}\) Although both the log-likelihood value and the statistics based on in-sample predictions imply that environmental threat improves model performance, I propose a few caveats that may cause one to question this conclusion.

All of the methods discussed thus far capture only one aspect of model performance, and they are sensitive to outliers: an otherwise well-performing model may not appear to be a good specification if it produces a few large errors. This is a concern especially for the model that only includes the human threat variable. For example, there is a known instance of a lake being invaded via recreational divers in Wisconsin (Lake Racine). This lake is not in the

\(^{42}\) Statistics based on squared (as opposed to absolute) temporal and spatial deviations lead to the same conclusions.
choice set, but its case highlights the fact that in a few cases, other activities or processes may also be responsible for spreading zebra mussels. If these are not correlated with the threat variables, then the model may produce large errors suggesting poor performance. Therefore, I argue that a more comprehensive (but subjective) assessment of model accuracy that also includes out-of-sample predictions is also important. I mapped the actual invasions and predictions over time for the various model specifications, and displayed them in Figures 4.4 to 4.6. Note that model predictions start from 1993 initial conditions, that is, before the first inland lake mussel discovery. By the fifty-seventh invasion prediction each model approximately captures the geographic area of the current distribution of zebra mussels, and at the same time also produces a large number of correct positives. However, in each case, the predicted westward spread is less gradual than it actually was: lakes (mostly large lakes) farther inland are predicted to be invaded faster in all models. Given the construction of the human threat variable and the specification of the random utility model, this is perhaps not surprising. Boaters may be less likely to take multiple trips within a short period of time if one of the trips is a long-distance trip. Since choices over time are not modeled, this is not reflected in site choice probabilities and invasion probabilities. Likewise, mussel survival rate during transport probably also depends on travel distance, and this is not accounted for either. Both of these factors would cause us to overestimate threat to sites far from current sources.

A notable difference between the specifications is that in the second and third models, environmental threat has a recognizable effect on spread dynamics. This is especially apparent for out-of-sample predictions: in those models, once mussels reach northeast
Wisconsin, nearly all lakes in the region are invaded before mussels spread to northwestern areas. There is less emphasis on short-range spread in the first model, which predicts an earlier (in terms of model-time) colonization of some northwestern sites. This observation raises another important point. The type of one-invasion-per-iteration prediction I employ makes it impossible to compare predictions across time. It is based on the relative level of hazard to various lakes as opposed to absolute hazard. Though some northwestern sites are invaded in fewer iterations in Model I, we cannot conclude that Model I implies mussels will spread to those sites sooner than in the other models. It is of course also impossible to establish a timeline within a model: each iteration may correspond to different length of time. In fact, maximum forecasted invasion probabilities in each model tend to decrease over iterations indicating that they may become increasingly less likely, and therefore that the rate of invasions may decrease in the future. For a graphical representation of this phenomenon, see Figure 4.7.

Notwithstanding the qualifications above, the models seem to be able to appraise relative invasion threat in the short term as testified by the high accuracy of their early predictions. Ultimately, predictions far into the future are rarely reliable in any model, and an assessment of policy-induced changes in relative invasion risk for the set of most threatened sites may be a valuable tool in analysis. Also note that, unlike absolute risk, relative invasion risk is invariant to changes in certain aspects of boater behavior. For example, if boater education programs about invasive species have succeeded in inducing boaters to take preventive measures (as suggested by recent studies by the Wisconsin Department of Natural Resources), the probability of future invasions will decrease. However, the order of the most
threatened sites will not change if there are no systematic differences among boaters in the behavioral change.

Which model approximates reality better is an empirical question. The ecological literature suggests that mussels are predominantly transported by recreational boaters, and that invasions by other means are less frequent. Model I is based on this premise, and the higher spatial variability of its invasion forecasts is in closer accordance with actual observations. Although achieving more spatial clustering was the original motivation for the other specifications, the degree of clustering exhibited by Model II (and to a lesser extent, by Model III) predictions is not confirmed by the data. A closer inspection of the prediction maps reveals the reason these models nevertheless perform better than Model I according to the evaluative statistics. Predictions in all models start out in the geographic area where invasions are concentrated (close to population centers). By construction, a model in which environmental threat dominates will progress in small spatial jumps, and will therefore result in the invasion of most sites in a given region before spreading to other areas in the state. Therefore, Models II and III will by default tend to result in correct positives sooner than Model I – even if their dynamics do not accurately reflect reality. Model I, on the other hand, by missing a few sites early in the prediction sequence, and by fluctuating more geographically may accumulate large temporal and spatial errors even if it is a reasonable specification.

Environmental threat in the hazard index resembles the force of attraction encountered in gravity models. It is important to note, however, that the water bodies in the choice set are not located in a vacuum: the hydrology of Wisconsin is more complex than
suggested by this simple specification. A more elaborate environmental threat variable would not ignore the presence or absence of direct and indirect hydrological connections between water bodies, for example. But even accounting for such idiosyncrasies may not improve model performance if the relationship between the two threat variables is not in fact linear, as written in the hazard index. This is of course entirely plausible. Mussels can colonize a lake either via recreational boaters, or via another means, but there is no reason why the two types of threat should be additive. Because of these concerns I consider both Model III, the best specification by most measures considered, and Model I, the model based only on recreational boater threat, for policy simulations.

The fact that hydrological studies suggest that lakes in northern Wisconsin may be less habitable to zebra mussels than southern lakes also needs to be addressed. My models do not account for any differences in chemical composition and physical characteristics, and may therefore overestimate the threat to lakes located in northern Wisconsin. At the same time, none of the model specifications predict invasions outside their current geographic range in the first fifty-seven iterations. Conceptually, the fact that northern lakes have remained mussel-free does not mean that chemical differences are indeed significant, or that those lakes are not threatened by future invasions. The current heterogeneous mussel-distribution can be explained even if there were no differences between lakes as indicated by the in-sample model forecasts.
Table 4.1. Random utility model estimation results

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
<th>Model 3</th>
<th></th>
<th>Model 4</th>
<th></th>
<th>Model 5</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>t-statistic</td>
<td>Estimate</td>
<td>t-statistic</td>
<td>Estimate</td>
<td>t-statistic</td>
<td>Estimate</td>
<td>t-statistic</td>
<td>Estimate</td>
<td>t-statistic</td>
</tr>
<tr>
<td>Travel cost</td>
<td>-0.0872</td>
<td>-155.1995</td>
<td>-0.0852</td>
<td>-110.6036</td>
<td>-0.0842</td>
<td>-39.9281</td>
<td>-0.0847</td>
<td>-125.0762</td>
<td>-0.0848</td>
<td>-119.5308</td>
</tr>
<tr>
<td>Boat length × Great Lake</td>
<td>0.1539</td>
<td>53.9196</td>
<td>0.1534</td>
<td>51.7336</td>
<td>0.1441</td>
<td>20.7904</td>
<td>0.1605</td>
<td>51.8880</td>
<td>0.1624</td>
<td>51.6226</td>
</tr>
<tr>
<td>Clarity × Angler</td>
<td>-0.0484</td>
<td>-10.8687</td>
<td>-0.1449</td>
<td>-0.2531</td>
<td>0.2819</td>
<td>0.5159</td>
<td>-0.2738</td>
<td>-0.5064</td>
<td>-0.0423</td>
<td>-0.0677</td>
</tr>
<tr>
<td>Intercept (OLS)</td>
<td>0.0363</td>
<td>0.0674</td>
<td>0.0810</td>
<td>-0.128</td>
<td>0.1269</td>
<td>-91.7384</td>
<td>-0.0872</td>
<td>-154.0260</td>
<td>0.1658</td>
<td>55.2596</td>
</tr>
<tr>
<td>Clarity × Inland lake</td>
<td>0.1317</td>
<td>2.4878</td>
<td>0.1061</td>
<td>1.3446</td>
<td>0.0979</td>
<td>0.8123</td>
<td>0.1236</td>
<td>1.8320</td>
<td>0.1177</td>
<td>1.5702</td>
</tr>
<tr>
<td>Inland lake</td>
<td>-0.5315</td>
<td>-0.6968</td>
<td>-0.2728</td>
<td>-0.3435</td>
<td>-0.2158</td>
<td>-0.2784</td>
<td>-0.1713</td>
<td>-0.2327</td>
<td>-0.5528</td>
<td>-0.6184</td>
</tr>
<tr>
<td>Great Lake</td>
<td>-1.8898</td>
<td>-1.7850</td>
<td>-1.7292</td>
<td>-1.6227</td>
<td>-2.3393</td>
<td>-2.1966</td>
<td>-1.9974</td>
<td>-1.8784</td>
<td>-2.1765</td>
<td>-1.9100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Model 6</th>
<th></th>
<th>Model 7</th>
<th></th>
<th>Model 8</th>
<th></th>
<th>Model 9</th>
<th></th>
<th>Model 10</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>t-statistic</td>
<td>Estimate</td>
<td>t-statistic</td>
<td>Estimate</td>
<td>t-statistic</td>
<td>Estimate</td>
<td>t-statistic</td>
<td>Estimate</td>
<td>t-statistic</td>
</tr>
<tr>
<td>Travel cost</td>
<td>-0.0873</td>
<td>-154.7696</td>
<td>-0.0842</td>
<td>-112.1878</td>
<td>-0.1269</td>
<td>-91.7384</td>
<td>-0.0872</td>
<td>-154.0260</td>
<td>-0.1258</td>
<td>-78.5741</td>
</tr>
<tr>
<td>Boat length × Great Lake</td>
<td>0.1589</td>
<td>55.6375</td>
<td>0.1533</td>
<td>47.4149</td>
<td>0.1807</td>
<td>30.0011</td>
<td>0.1658</td>
<td>55.2596</td>
<td>0.1760</td>
<td>22.2905</td>
</tr>
<tr>
<td>Clarity × Angler</td>
<td>0.0363</td>
<td>0.0674</td>
<td>-0.1449</td>
<td>-0.2531</td>
<td>0.2819</td>
<td>0.5159</td>
<td>-0.2738</td>
<td>-0.5064</td>
<td>-0.0423</td>
<td>-0.0677</td>
</tr>
<tr>
<td>Intercept (OLS)</td>
<td>0.0363</td>
<td>0.0674</td>
<td>-0.0854</td>
<td>-128.1878</td>
<td>-0.1269</td>
<td>-91.7384</td>
<td>-0.0872</td>
<td>-154.0260</td>
<td>0.1658</td>
<td>55.2596</td>
</tr>
<tr>
<td>Clarity × Inland lake</td>
<td>0.1317</td>
<td>2.9872</td>
<td>0.1395</td>
<td>2.4061</td>
<td>0.0797</td>
<td>1.2681</td>
<td>0.1619</td>
<td>3.0555</td>
<td>0.1522</td>
<td>2.1524</td>
</tr>
<tr>
<td>Inland lake</td>
<td>-0.5315</td>
<td>-0.6968</td>
<td>-0.2728</td>
<td>-0.3435</td>
<td>-0.2158</td>
<td>-0.2784</td>
<td>-0.1713</td>
<td>-0.2327</td>
<td>-0.5528</td>
<td>-0.6184</td>
</tr>
<tr>
<td>Great Lake</td>
<td>-1.8898</td>
<td>-1.7850</td>
<td>-1.7292</td>
<td>-1.6227</td>
<td>-2.3393</td>
<td>-2.1966</td>
<td>-1.9974</td>
<td>-1.8784</td>
<td>-2.1765</td>
<td>-1.9100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Model 6</th>
<th></th>
<th>Model 7</th>
<th></th>
<th>Model 8</th>
<th></th>
<th>Model 9</th>
<th></th>
<th>Model 10</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>t-statistic</td>
<td>Estimate</td>
<td>t-statistic</td>
<td>Estimate</td>
<td>t-statistic</td>
<td>Estimate</td>
<td>t-statistic</td>
<td>Estimate</td>
<td>t-statistic</td>
</tr>
<tr>
<td>Travel cost</td>
<td>-0.0873</td>
<td>-154.7696</td>
<td>-0.0842</td>
<td>-112.1878</td>
<td>-0.1269</td>
<td>-91.7384</td>
<td>-0.0872</td>
<td>-154.0260</td>
<td>-0.1258</td>
<td>-78.5741</td>
</tr>
<tr>
<td>Boat length × Great Lake</td>
<td>0.1589</td>
<td>55.6375</td>
<td>0.1533</td>
<td>47.4149</td>
<td>0.1807</td>
<td>30.0011</td>
<td>0.1658</td>
<td>55.2596</td>
<td>0.1760</td>
<td>22.2905</td>
</tr>
<tr>
<td>Clarity × Angler</td>
<td>0.0363</td>
<td>0.0674</td>
<td>-0.1449</td>
<td>-0.2531</td>
<td>0.2819</td>
<td>0.5159</td>
<td>-0.2738</td>
<td>-0.5064</td>
<td>-0.0423</td>
<td>-0.0677</td>
</tr>
<tr>
<td>Intercept (OLS)</td>
<td>0.0363</td>
<td>0.0674</td>
<td>-0.0854</td>
<td>-128.1878</td>
<td>-0.1269</td>
<td>-91.7384</td>
<td>-0.0872</td>
<td>-154.0260</td>
<td>0.1658</td>
<td>55.2596</td>
</tr>
<tr>
<td>Clarity × Inland lake</td>
<td>0.1317</td>
<td>2.9872</td>
<td>0.1395</td>
<td>2.4061</td>
<td>0.0797</td>
<td>1.2681</td>
<td>0.1619</td>
<td>3.0555</td>
<td>0.1522</td>
<td>2.1524</td>
</tr>
<tr>
<td>Inland lake</td>
<td>-0.5315</td>
<td>-0.6968</td>
<td>-0.2728</td>
<td>-0.3435</td>
<td>-0.2158</td>
<td>-0.2784</td>
<td>-0.1713</td>
<td>-0.2327</td>
<td>-0.5528</td>
<td>-0.6184</td>
</tr>
<tr>
<td>Great Lake</td>
<td>-1.8898</td>
<td>-1.7850</td>
<td>-1.7292</td>
<td>-1.6227</td>
<td>-2.3393</td>
<td>-2.1966</td>
<td>-1.9974</td>
<td>-1.8784</td>
<td>-2.1765</td>
<td>-1.9100</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Trips</th>
<th>Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td>4846</td>
<td>278</td>
</tr>
<tr>
<td>Model 2</td>
<td>4075</td>
<td>155</td>
</tr>
<tr>
<td>Model 3</td>
<td>2828</td>
<td>63</td>
</tr>
<tr>
<td>Model 4</td>
<td>3365</td>
<td>185</td>
</tr>
<tr>
<td>Model 5</td>
<td>3212</td>
<td>168</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Trips</th>
<th>Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 6</td>
<td>4846</td>
<td>278</td>
</tr>
<tr>
<td>Model 7</td>
<td>3531</td>
<td>228</td>
</tr>
<tr>
<td>Model 8</td>
<td>4168</td>
<td>237</td>
</tr>
<tr>
<td>Model 9</td>
<td>4846</td>
<td>278</td>
</tr>
<tr>
<td>Model 10</td>
<td>3042</td>
<td>196</td>
</tr>
</tbody>
</table>
### Table 4.2. Random utility model specification summaries

<table>
<thead>
<tr>
<th>Description</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum number of visitors for inclusion in choice set</td>
<td>5</td>
<td>10</td>
<td>20</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Clarity interpolation method</td>
<td>Spline based on lakes with 5 readings</td>
<td>Spline based on lakes with 5 readings</td>
<td>Spline based on lakes with 5 readings</td>
<td>None, only lakes with at least 5 actual Secchi-readings</td>
<td>None, only lakes with at least 10 actual Secchi-readings</td>
</tr>
<tr>
<td>Time period considered</td>
<td>April-October</td>
<td>April-October</td>
<td>April-October</td>
<td>April-October</td>
<td>April-October</td>
</tr>
<tr>
<td>Maximum trip distance</td>
<td>No limit</td>
<td>No limit</td>
<td>No limit</td>
<td>No limit</td>
<td>No limit</td>
</tr>
<tr>
<td>Number of trips</td>
<td>4846</td>
<td>4075</td>
<td>2828</td>
<td>3365</td>
<td>3212</td>
</tr>
<tr>
<td>Sites in choice set</td>
<td>278</td>
<td>155</td>
<td>63</td>
<td>185</td>
<td>168</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Description</th>
<th>Model 6</th>
<th>Model 7</th>
<th>Model 8</th>
<th>Model 9</th>
<th>Model 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum number of visitors for inclusion in choice set</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Clarity interpolation method</td>
<td>Inverse distance weighted based on lakes with 5 readings</td>
<td>Spline based on lakes with 5 readings</td>
<td>Spline based on lakes with 5 readings</td>
<td>Spline based on lakes with 5 readings</td>
<td>Spline based on lakes with 5 readings</td>
</tr>
<tr>
<td>Time period considered</td>
<td>April-October</td>
<td>June-September</td>
<td>April-October</td>
<td>April-October</td>
<td>June-September</td>
</tr>
<tr>
<td>Maximum trip distance</td>
<td>No limit</td>
<td>No limit</td>
<td>200 km (one-way, straight-line)</td>
<td>No limit</td>
<td>200 km (one-way, straight-line)</td>
</tr>
<tr>
<td>Number of trips</td>
<td>4846</td>
<td>3531</td>
<td>4168</td>
<td>4846</td>
<td>3042</td>
</tr>
<tr>
<td>Sites in choice set</td>
<td>278</td>
<td>228</td>
<td>237</td>
<td>278</td>
<td>196</td>
</tr>
</tbody>
</table>
### Table 4.3.  Hazard model estimation results

<table>
<thead>
<tr>
<th>Variables</th>
<th>Estimate</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-4.307505</td>
<td>-29.326465</td>
</tr>
<tr>
<td>Great Lake</td>
<td>2.422964</td>
<td>7.067558</td>
</tr>
<tr>
<td>River</td>
<td>0.803866</td>
<td>2.804676</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Invasion probability</th>
<th>Inland lake</th>
<th>River</th>
<th>Great Lake</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.013288</td>
<td>0.029209</td>
<td>0.131868</td>
</tr>
</tbody>
</table>
Table 4.4. Integrated model estimation results
Actually infested sites are shown in bold in the Predicted invasions section.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Model I</th>
<th></th>
<th>Model II</th>
<th></th>
<th>Model III&lt;sup&gt;43&lt;/sup&gt;</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>t-statistic</td>
<td>Estimate</td>
<td>t-statistic</td>
<td>Estimate</td>
<td>t-statistic</td>
</tr>
<tr>
<td>Human Threat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggregate</td>
<td>0.1037</td>
<td>5.0891</td>
<td>0.1038</td>
<td>5.2412</td>
<td>0.0878</td>
<td>4.9121</td>
</tr>
<tr>
<td>Environmental Threat</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lake</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.0904</td>
<td>1.7442</td>
</tr>
<tr>
<td>River</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.6076</td>
<td>0.1144</td>
</tr>
<tr>
<td>Great Lake</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2.9395</td>
<td>3.7031</td>
</tr>
<tr>
<td>Evaluation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log-likelihood value</td>
<td>-258.1204</td>
<td></td>
<td>-250.3520</td>
<td></td>
<td>-248.4533</td>
<td></td>
</tr>
<tr>
<td>PC</td>
<td>0.5263</td>
<td></td>
<td>0.6140</td>
<td></td>
<td>0.5965</td>
<td></td>
</tr>
<tr>
<td>SATD</td>
<td>1909</td>
<td></td>
<td>1474</td>
<td></td>
<td>1303</td>
<td></td>
</tr>
<tr>
<td>SASD</td>
<td>442</td>
<td></td>
<td>294</td>
<td></td>
<td>319</td>
<td></td>
</tr>
<tr>
<td>Predicted invasions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Lake Winnebago</td>
<td></td>
<td>Lake Winnebago</td>
<td></td>
<td>Lake Winnebago</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Shawano Lake</td>
<td></td>
<td>Shawano Lake</td>
<td></td>
<td>Lake Butte Des Morts</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Pewaukee Lake</td>
<td></td>
<td>Pewaukee Lake</td>
<td></td>
<td>Lake Poygan</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Lake Geneva</td>
<td></td>
<td>Lake Geneva</td>
<td></td>
<td>Clark Lake</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Green Lake</td>
<td></td>
<td>Nagawicka Lake</td>
<td></td>
<td>Lake Winneconne</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Nagawicka Lake</td>
<td></td>
<td>Okauchee Lake</td>
<td></td>
<td>Shawano Lake</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Okauchee Lake</td>
<td></td>
<td>Green Lake</td>
<td></td>
<td>Pewaukee Lake</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Lake Mendota</td>
<td></td>
<td>Lake Mendota</td>
<td></td>
<td>Green Lake</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Lake Monona</td>
<td></td>
<td>Lake Monona</td>
<td></td>
<td>Lake Geneva</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Lake Wisconsin</td>
<td></td>
<td>Lake Wisconsin</td>
<td></td>
<td>Nagawicka Lake</td>
<td></td>
</tr>
</tbody>
</table>

<sup>43</sup> E<sup>e</sup> scaled by 10, E<sup>f</sup> scaled by 100.
Figure 4.1. The spatial distribution of alternative specific constants
Figure 4.2. The distribution of the residences of surveyed boaters by zip code area
Only study-area motorboaters licensed in Wisconsin are represented.
Figure 4.3. **The human threat variable as a function of site choice probability**
In this example, the choice set consists of only two sites (one infested, and one uninfested). Therefore, the choice probability for site 2 is $P_{n2} = 1 - P_{n1}$. 
Figure 4.4. Invasion predictions from Model I
Each panel shows a schematic map of sites in the choice set. Initial invasions include Lake Michigan counties, Mississippi River counties, and a single Lake Superior county.
Figure 4.5. Invasion predictions from Model II
Each panel shows a schematic map of sites in the choice set. Initial invasions include Lake Michigan counties, Mississippi River counties, and a single Lake Superior county.
Figure 4.6. Invasion predictions from Model III
Each panel shows a schematic map of sites in the choice set. Initial invasions include Lake Michigan counties, Mississippi River counties, and a single Lake Superior county.
Figure 4.7.  Maximum simulated invasion probabilities over time in Model I
Chapter 5

Policy Simulations and Conclusions
The predictions presented in the previous chapter are conceptually not different from those of a well-specified gravity model. They are based on boaters’ recreation choices, but the behavior derived from those choices is static. In particular, trip probabilities do not change over time. Therefore, changes in invasion probabilities are exclusively driven by changes in the spatial distribution of zebra mussels. The motivation for modeling vector behavior, however, was to accommodate potential behavioral changes, which I address in this chapter.

I first present a scenario in which motorboaters respond to zebra mussel presence in a lake. Since the data do not allow me to establish whether infested status affects site choices, I achieve the desired effect through an indirect route: presumed changes in water clarity due to the mussels’ removal of particulate matter from the water via filter feeding. Then, I consider a situation in which boaters do not directly react to zebra mussels, but policy makers do, and boaters are allowed to substitute between recreation sites in response to mussel-control policies. I demonstrate that a policy that changes visitation patterns can have unintended consequences, and may in some cases even increase the probability of mussel spread to some uninfested water bodies. Following the discussion of policy counterfactuals, I perform welfare analyses, and offer concluding remarks.

5.1 Clarity Feedback

First, to demonstrate the power of the integrated model, I explore Model I predictions with a hypothetical clarity feedback effect: during each iteration, the site predicted to be invaded by zebra mussels experiences a one-time permanent increase in Secchi-depth by fifty percent. It would be impossible to evaluate a similar change in a site attribute via a gravity model: there
would be no change in the predicted spread of the invasive. However, since water clarity enters the utility function of boaters (via two separate terms), a change in clarity can have an effect on boater behavior. According to my estimates, the clarity increase makes an infested site more attractive relative to other sites; hence, trip probabilities to the site increase (and so does its predicted choice share). Recall that the value of the threat variable depends on trip probabilities to both infested and uninfested water bodies (Figure 4.3). Therefore, changes in the threat variable are now driven not only by the spatial spread of the invasive as before, but also by the behavioral change. The most significant manifestation of this is that $R_{jt}$ no longer must increase over time for each site. It will in fact fall for a few sites during certain time periods, as will their predicted invasion probability. The predicted sequence of invasions, and their spatial pattern, may therefore be different than without the clarity-feedback effect. A comparison of invasion predictions reveals that the relative hazard faced by the top twenty high-risk sites does change due to the clarity effect: some are displaced by as many as five positions in the list. Figure 5.1 maps predictions for this specification; the pattern is unambiguously different from that observed in Figure 4.4, which is based on the same parameter estimates, but assumes that mussel presence has no effect on water clarity. There are in reality several other mechanisms (including actions of policymakers) that may induce a behavioral change, and this simple exercise has merely illustrated that it is important to account for the possibility of such a change. I next turn my attention to an in-depth analysis of various policy effects on zebra mussel spread.
5.2 Policy Counterfactuals

I will consider three types of policy scenarios. The first involves installing boat cleaning equipment at an infested lake, and requiring all boats leaving the water to be inspected and cleaned. This may involve removing entangled vegetation or any attached zebra mussels, using pressure washers with hot water to clean the boat, and draining all water from the boat, motor and trailer to ensure larval mussels are not transported to another lake either. In the first policy option, the necessary equipment is provided free of charge. The second policy type I consider assesses a fee for the boat inspection and cleaning to cover the costs of installing the equipment and of staffing the cleaning stations. If an infested site is suspected to be a significant threat to other water bodies, an extreme policy tool may involve closing the site to transient boating traffic completely to prevent the transmission of zebra mussels to other lakes. Such site closure is the subject of the third policy scenario.

In terms of human threat, free mandatory boat cleaning in effect eliminates a policy lake (a lake at which the policy is implemented) from the set of infested sites because, if the cleaning is performed properly, boats will not transport zebra mussels from it to another lake. When the monetary cost of the procedure is zero, the utility of boating at the policy site does not change.\textsuperscript{44} Thus, there will be no change in choice probabilities under the first policy, and the only influence on invasion probabilities comes from the reduction in the set of infested sites. On the other hand, assessing a fee to cover installation and operating expenses can be expected to reduce the attractiveness of the site because of the higher costs associated with

\textsuperscript{44} In reality, the opportunity cost of time spent at the inspection station would reduce the attractiveness of the policy lake. Here, I am abstracting from this because questions from the boater survey did not enable me to estimate the opportunity cost of time for different individuals.
visiting it. Trip probabilities to the site will decrease, and as a result, trip probabilities to other boating sites will increase. The magnitude of the probability change will be largest for choice alternatives that are good substitutes for the policy site (mostly lakes nearby), and some of the boating traffic is expected to shift to those lakes. Lastly, closing a lake to boaters will lead to a redistribution of all trips that targeted it to remaining sites. Obviously, this policy also prevents zebra mussel transmission from the lake – at least via the recreational boating vector.

Note that there is a subtle difference in performing policy simulations in the two integrated model specifications. The policies I consider operate on recreational boaters, and effectively turn off the boating vector. In Model I (which only includes human threat), all three policy types are successful in terms of completely removing the policy lake from the set of infested sites. The hazard index in Model III, however, also includes environmental threat. This, recall, is intended to capture invasion risk from all factors other than transient recreational boating, and is not influenced by policies targeting recreational boaters: requiring boat inspections and restricting boater access to the site have no effect on environmental threat. Therefore, a policy lake in Model III retains some of its infectiousness. This idea is operationalized in the following manner: human threat is computed using the reduced set of infested sites (reflecting the policy), but environmental threat is a function of the full invaded set.

To assess the effectiveness of the various policy options, I first determine baseline (no-policy) invasion probabilities for a selected set of focus lakes in each model. Post-policy invasion probabilities will be compared to these baseline probabilities. In all cases, I use the
observed invasion status in 2005 (the last year in the data in which Wisconsin invasions were recorded) to calculate threat variable values, hazard indices, and the corresponding invasion probabilities. Focus lakes are important uninfested boating destinations, or are significant in some other respect. They were selected to represent all regions in Wisconsin; most of them have relatively high baseline invasion probabilities which may make them of interest from a policy perspective. Similarly, policy lakes are prominent, but infested, lakes. Figure 5.2 maps the geographic locations of both policy sites and focus sites, and Table 5.1 displays basic information about the lakes.

Baseline invasion probabilities using both models are shown in the rows named ‘invasion probability’ in Table 5.2. Most forecasted invasion probabilities are below ten percent. A notable exception is Lake Mendota, whose probability of invasion is nearly sixty percent according to Model I estimates, and over seventy percent using Model III. With one hundred fifteen reported boating trips, Lake Mendota is the second most heavily used inland lake in Wisconsin (after Lake Winnebago), and is located close to existing zebra mussel populations. In fact, the straight-line distance between the geographic centers of Lake Mendota and Lake Monona, a popular and infested lake, is only about six kilometers. This extreme proximity results in a high level of environmental threat, which explains why Model III baseline probability is higher than Model I baseline probability. As one would expect given the specification of the threat variables, visitation patterns and location both contribute to the exceptionally high invasion probability predicted for Lake Mendota.

Some river segments, for example the Rock River in Rock county and the Wolf River in Winnebago county, are also predicted to face a high invasion threat. However, I am less confident in predictions relating to rivers, especially with the environmental threat specification (Model III), and decided to concentrate on inland lakes only.
Table 5.2 also shows the percentage change in invasion probabilities under various policy scenarios implemented at Lake Winnebago. As noted, Lake Winnebago is the largest and most heavily boated inland lake in the state – policy makers may consider it the most significant zebra mussel source, and conclude that the marginal benefit of “neutralizing” it would be high. As expected, free inspections at Lake Winnebago substantially (in some cases by more than twenty percent) reduce the threat to all other focus lakes. The reduction is in all cases smaller in Model III due to the existence of other spread mechanisms operating through environmental threat: mussels may disperse from Lake Winnebago in Model III even after all boats leaving the lake are cleaned of zebra mussels.

Charging money for mandatory inspections increases the cost of visiting the policy lake for all boaters by the cost of the inspection. The effect can be modeled by increasing travel costs to the site by the appropriate amount, and using the maximum likelihood parameter estimates from the random utility model to recalculate all choice probabilities. The human threat variable is updated to reflect these new choice probabilities as well as the removal of the policy site from the infested set. Changes in the hazard index are due to changes in the human threat variable only, since environmental threat is not influenced by the policy. Invasion probabilities are then determined using the integrated model parameter estimates. As seen in Table 5.2, inspection with a fee is in all cases less effective at reducing the threat to uninfested lakes than free inspection. The fee reduces the utility of visiting Lake Winnebago, and displaces a portion of the trips that targeted it. Boating pressure at other sites increases, which tends to raise invasion probabilities from non-policy, infested sites, counteracting the decrease achieved by neutralizing Lake Winnebago. The larger the fee, the
greater the resulting redistribution of trips, and the larger the reduction in policy effectiveness. In fact, assessing a $20 fee increases (albeit by a very small amount) the threat to Fox Lake and Lake Koshkonong relative to baseline conditions.

Substitution between boating destinations is obviously largest when Lake Winnebago is placed under a quarantine, removing it from the choice set of all boaters. Such a policy would not only be ineffective at protecting uninfested sites, but would actually increase all focus lake invasion probabilities in both models because the effect of behavioral change outweighs the threat reduction achieved by site closure per se. Compare these perhaps counterintuitive predictions to those that would be obtained without modeling vector behavior. It is the possibility of site substitutions that differentiates the various policies I consider; if behavior were not modeled, all policies would be identical with respect to invasion probability forecasts. A model in which behavior is exogenous would therefore imply that a quarantine policy is exactly as effective as an inspection policy (the predicted effectiveness of which would not depend on the price of inspections). Thus, instead of facing a nearly nine percent increase in invasion probability in Model I, Fox Lake would be predicted about twenty-seven percent less likely to be invaded under a quarantine policy implemented at Lake Winnebago. That is, a model ignoring behavioral responses may overestimate the benefits of control policies that elicit a reaction from boaters.

In reality, some boaters who visited Lake Winnebago might prefer to stay at home, and not boat anywhere after the lake is closed. They are, however, not allowed to do so in the random utility model; all boaters are forced to choose a boating destination (that is, site choice probabilities for each individual have to sum to one). The absence of an outside option
(not boating) leads to overestimating trip probabilities to other sites, and thereby inflates predicted invasion probabilities. The distortion introduced by this choice restriction will be relatively small if good substitutes for the policy lake exist, which may be the case for Lake Winnebago – it is located close to other lakes with similar characteristics. Figure 5.3 illustrates the quarantine-policy-induced change in choice shares in Model I, and Figure 5.4 displays the associated increase in invasion probabilities.

The above results are general, and not driven by the choice of policy lake (although they are more pronounced for lake Winnebago than other infested lakes given the large number of boaters visiting it). Consider for example policies implemented at Lake Monona. Changes in predicted focus site invasion probabilities are displayed in Table 5.3 for free inspections and site closure according to both models. Each policy significantly reduces mussel threat at the three focus lakes (Lake Mendota, Lake Waubesa and Lake Wisconsin) closest to Lake Monona, and results in lower invasion probabilities for most other lakes as well. Model I and Model III policy assessments are generally consistent and very similar. However, from the perspective of Lake Mendota invasion risk, there is a large difference in policy effectiveness in the two model specifications. The difference is due to the high level of environmental threat arising from the proximity of the two lakes. Again, a policy that causes behavioral adjustments is less effective in preventing zebra mussel spread to each focus lake. Changes in choice shares and invasion probabilities for the Lake Monona quarantine policy in Model I are illustrated in Figures 5.5 and 5.6, respectively.

---

46 In fact, it would be hard to argue that any lake in Wisconsin does not have close substitutes.
Note that policy outcomes depend on the location and attributes of all lakes in the choice set because of the role substitution plays in determining trip and invasion probabilities. If Lake Mendota was the only good substitute site for Lake Monona, the quarantine policy analyzed in the previous paragraph would be much less effective: most of the boating pressure from Lake Monona would shift to Lake Mendota, significantly increasing its likelihood of invasion. For example, if the four large uninfested lakes closest to Lake Monona and Lake Mendota (Lake Kegonsa, Lake Koshkonong, Lake Waubesa and Lake Wisconsin) did not exist, the quarantine policy would increase Lake Mendota’s invasion probability by 3.61 percent relative to the baseline according to Model I. Compare this with the 23.5 percent reduction predicted with the availability of nearby alternative boating sites.\(^{47}\)

As demonstrated by the counterfactual exercises in this chapter, policy outcomes are often idiosyncratic, and depend on the location and attributes of not only the policy and focus sites, but of all other sites as well. In all cases, a policy that has a large impact on the utility of choosing an infested alternative leads to a (relatively) large change in choice probabilities. Such adaptation to the new policy environment on the part of boaters will render the policy less effective at reducing invasion probabilities due to the resulting redistribution of boating pressure. The policy may, in extreme cases, even increase the threat to some lakes. Similar unintended consequences would be impossible to find in a framework not based on an explicitly integrated behavioral approach.

---

\(^{47}\) I evaluated the effect substitution plays in this example by removing the four uninfested lakes from the choice set, and simultaneously removing all boaters who visited those lakes to ensure their trips are not redistributed to other sites.
The behavioral adjustments are, as noted, a manifestation of utility changes caused by the various policy options, and a complete evaluation of control policies also needs to address the utility costs they impose on boaters. Under the assumptions of the logit model, the change in expected consumer surplus that results from a change in a site attribute, or a change in the choice set can be calculated by

$$\Delta E(CS_n) = \frac{1}{(-\gamma)} \left[ \ln \left( \sum_{j=1}^{J^0} e^{V_{nj}^0} \right) - \ln \left( \sum_{j=1}^{J^1} e^{V_{nj}^1} \right) \right],$$  \hspace{1cm} (5.1)

where $\gamma$ is the coefficient of the travel cost variable, and $J^0, V_{nj}^0, J^1, V_{nj}^1$ reflect the choice set and representative utility before and after the policy change, respectively. The term inside the brackets is the difference between the attainable expected maximum utilities before and after the change, and the negative of the travel cost coefficient is the marginal utility of income – the division by this term translates utilities into dollars. In the case of a mussel-control policy that makes boaters worse off by increasing visit costs or otherwise reducing access, this amount will be negative. That is, the negative of $\Delta E(CS_n)$ measures willingness to accept: it is the minimum amount of money a boater in the original state would require to agree to the imposition of conditions that pertain in the post-policy state.

Table 5.4 displays the results of the welfare analysis for the policy scenarios I considered. The first data column of the table shows mean willingness to accept estimates in dollars in the sample. The next column shows a conservative population estimate: it is calculated by multiplying the mean figure in the first column by the product of sample size
(4,846) and the weight for each motorboat in the sample (242.5205).\textsuperscript{48} This provides a conservative estimate because it is based only on Wisconsin motorboaters whose primary boating destination could be identified, and who visited a site in the choice set. Also, recall that each individual in the model is observed to visit a single site once. Since the mean willingness to accept figures are per choice occasion, the population counterpart displayed in this column will be an underestimate inasmuch as boaters take multiple trips. The last column, on the other hand, is based on the total number of boating days (24,123) reported by sampled Wisconsin motorboaters, and the sample weight for each boater. Figures reported in this column may be overestimates since each day of a multiple-day trip is counted separately, or underestimates since only motorboats licensed in Wisconsin (in 1989-1990) are included in the calculations.

The two policies involving free inspections are omitted from the table since they are predicted to not impose any costs on society. This, of course, only arises because I am abstracting from the opportunity cost of time: it is not reflected in any of the willingness to accept estimates. In reality, though, boaters would demand some compensation for the inconvenience even free inspections entail. Mean willingness to accept for the other policies is less than a dollar in each case. Given the relatively low choice share of each site (few boaters are affected by either policy), and the availability of substitute sites, the low mean willingness to accept figures are to be expected.\textsuperscript{49} The reduction in total consumer surplus

\textsuperscript{48} The sample weight was determined by the Wisconsin Department of Natural Resources, and is based on the ratio of population size (motorboat licenses in the file) and the number of sampled motorboaters (Penaloza, 1991).

\textsuperscript{49} The maximum willingness to accept values in the sample are: $3.30 for the $10 inspection, $5.03 for the $20 inspection, and $6.46 for the quarantine policy at Lake Winnebago, and $2.06 for the quarantine policy at Lake Monona.
associated with any of the policies is, however, at least hundreds of thousand dollars. Closing Lake Winnebago to transient recreational boaters, for example, would not only be counter-productive in terms of increasing the invasion probabilities of other sites, but may also lead to additional societal costs of over six million dollars annually (when converted to current, 2008, value using the relevant Consumer Price Index conversion factor).

5.3 Concluding Remarks

While weather phenomena and other natural events do occasionally transport non-indigenous species, most biological invasions are directly or indirectly caused by human activity. The anthropogenic causality is particularly evident in the case of many freshwater invasive plants and animals: aquatic species generally require a transport mechanism to spread in a patchy environment (that is, to move between isolated bodies of water), and the most significant vector is often recreational boating. A notably successful aquatic invader, the zebra mussel, causes millions of dollars worth of direct economic damage and great ecological harm each year. Zebra mussels are unique among species spread by recreational boaters in several aspects. They are highly prolific, thrive under a variety of physical, chemical and geographical conditions, and can also survive several days of aerial exposure. For all practical purposes, they are impossible to eradicate from an infested water body, limiting available policy options. The North American mussel invasion started in the relatively recent past, and only a fraction of potentially habitable water bodies have been invaded thus far, lending practical significance to an analysis of zebra mussel spread. Finally, relatively

50 The plant pathogens soybean rust and citrus canker, for example, are believed to have been introduced into the United States by hurricanes, and can spread via wind and windblown rain, respectively.
reliable spatial and temporal data on the distribution of zebra mussels is available for such an analysis since this particular invasive has been the subject of considerable interest and monitoring effort.

Our capacity to move exotic species and thereby bring about biological invasions has long been recognized. The converse, that the spread of a pest may lead to changes in the behavior responsible for its invasion, is less commonly acknowledged. My objective in this dissertation was to account for this endogeneity in the case of the zebra mussel invasion by explicitly modeling the behavior of transient recreational boaters. I used data on the spatial distribution of zebra mussels and choice probabilities calculated from a multinomial logit travel cost model to construct a time- and site-specific variable that characterizes the risk of mussel transport to uninfested water bodies. This threat variable has an intuitive interpretation, and is able to capture complex spatial and temporal relationships: it reflects distance and quality aspects of recreation sites, as well as changes in the choice set and the risk set. Invasion probabilities for uninfested sites are then found by employing the threat variable in a discrete duration model. The integrated model I presented explains the currently observed spatial distribution of zebra mussels in the study area fairly well. Of more interest, however, is its ability to predict the effect various policy instruments may have on zebra mussel invasion risk: the random utility model enables us to trace the behavioral consequences of policy alternatives (or of environmental changes), and corresponding changes in the threat variable can be translated into invasion probabilities using the discrete duration model. Any perturbation that elicits a behavioral response will also affect invasion risk in this framework.
Finally, I used policy counterfactuals to demonstrate that a traditional model that does not account for behavioral adjustments would overestimate the benefits of control policies that impose a cost on motorboaters. In a worst case scenario, the unintended consequences of a policy may not only render it completely ineffective, but even counterproductive. The magnitude of efficiency loss depends on (besides the extent of utility decrease associated with visiting the policy site) the location and attributes of the policy and focus sites, and of substitute sites as well. An explicitly integrated behavioral approach is necessary to evaluate the full implications of alternative policies since potential unintended consequences would be impossible to identify (and quantify) in a more traditional framework.

As I noted, several interesting implications of the analysis arise from changes in site choice probabilities that reflect boaters’ adaptive behavior. That is, they are driven by substitution patterns predicted in response to an exogenous shock in the relative desirability of various choice alternatives. The most significant contribution of my research is a straightforward and intuitive procedure for integrating the behavioral and ecological models through the human threat variable. The multinomial logit specification for the random utility model provides transparency for tracing the behavioral effects of the various shocks, and translating them into aggregate changes in invasion probabilities. The multinomial logit model, however, has well-known limitations to characterize the substitution patterns that model outcomes critically depend on. Namely, the distributional assumptions placed on the error term of the utility function that lead to the multinomial logit also result in proportional substitution: a change in the choice probability of any alternative causes the choice probabilities of all other alternatives to change proportionately in the opposite direction. This
may not be realistic in the freshwater recreation site choice context. For example, one might expect to find more substitutability between different inland lakes than between an inland lake and one of the Great Lakes, and the multinomial logit is unable to account for this. A straightforward extension, therefore, would be to specify a less restrictive behavioral model, such as a nested logit or a random parameters logit, that allows for more realistic substitution patterns.

The anthropogenic spread of aquatic invasive species typically involves successive boat launches at different water bodies, and is therefore inherently a temporal phenomenon. Data constraints did not enable me to exploit the time dimension to its full extent: I had to assume that trip probabilities emerging from a model based on the concept of a primary boating site accurately characterize recreation decisions over time. A more ambitious research project targeting freshwater invasives may involve the collection of panel data to permit modeling subtler temporal interactions. Panel data, representing repeated choices by decision makers, is the key to characterize trip frequency, habit formation or variety seeking, which may be important factors in determining the course of an invasion. In addition, information on sociodemographic characteristics, and on each boater’s attitude toward invasion-risk reduction might be desirable ingredients of such a project.
## Table 5.1. Policy and focus lakes

<table>
<thead>
<tr>
<th>Lake</th>
<th>Invasion status</th>
<th>Number of boaters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chain Lakes</td>
<td>Not infested</td>
<td>31</td>
</tr>
<tr>
<td>Fox Lake</td>
<td>Not infested</td>
<td>39</td>
</tr>
<tr>
<td>High Falls Reservoir</td>
<td>Not infested</td>
<td>33</td>
</tr>
<tr>
<td>Lake Koshkonong</td>
<td>Not infested</td>
<td>32</td>
</tr>
<tr>
<td>Lake Mendota</td>
<td>Not infested</td>
<td>115</td>
</tr>
<tr>
<td>Lake Monona</td>
<td>Infested</td>
<td>66</td>
</tr>
<tr>
<td>Lake Waubesa</td>
<td>Not infested</td>
<td>39</td>
</tr>
<tr>
<td>Lake Wisconsin</td>
<td>Not infested</td>
<td>54</td>
</tr>
<tr>
<td>Lake Wissota</td>
<td>Not infested</td>
<td>55</td>
</tr>
<tr>
<td>Lake Winnebago</td>
<td>Infested</td>
<td>240</td>
</tr>
<tr>
<td>Three Lakes Chain</td>
<td>Not infested</td>
<td>46</td>
</tr>
</tbody>
</table>
### Table 5.2. Lake Winnebago policies

The table displays policy-induced changes in baseline invasion probabilities (in percentages).

<table>
<thead>
<tr>
<th>Model I</th>
<th>Chain Lakes</th>
<th>Fox Lake</th>
<th>High Falls Reservoir</th>
<th>Lake Koshkonong</th>
<th>Lake Mendota</th>
<th>Lake Waubesa</th>
<th>Lake Wisconsin</th>
<th>Lake Wissota</th>
<th>Three Lakes Chain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Invasion rank&lt;sup&gt;51&lt;/sup&gt;</td>
<td>8</td>
<td>2</td>
<td>7</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>35</td>
<td>22</td>
</tr>
<tr>
<td>Invasion probability</td>
<td>0.054</td>
<td>0.124</td>
<td>0.057</td>
<td>0.075</td>
<td>0.583</td>
<td>0.066</td>
<td>0.103</td>
<td>0.023</td>
<td>0.035</td>
</tr>
<tr>
<td>Inspection $10</td>
<td>-11.57</td>
<td>-9.71</td>
<td>-7.02</td>
<td>-2.61</td>
<td>-4.32</td>
<td>-3.16</td>
<td>-5.08</td>
<td>-0.36</td>
<td>-5.96</td>
</tr>
<tr>
<td>Inspection $20</td>
<td>-3.44</td>
<td>0.15</td>
<td>-1.54</td>
<td>0.63</td>
<td>-1.18</td>
<td>-0.62</td>
<td>-1.47</td>
<td>-0.12</td>
<td>-1.79</td>
</tr>
<tr>
<td>Quarantine</td>
<td>3.99</td>
<td>8.82</td>
<td>3.26</td>
<td>3.15</td>
<td>1.19</td>
<td>1.35</td>
<td>1.41</td>
<td>0.06</td>
<td>1.84</td>
</tr>
<tr>
<td>Model III</td>
<td>Chain Lakes</td>
<td>Fox Lake</td>
<td>High Falls Reservoir</td>
<td>Lake Koshkonong</td>
<td>Lake Mendota</td>
<td>Lake Waubesa</td>
<td>Lake Wisconsin</td>
<td>Lake Wissota</td>
<td>Three Lakes Chain</td>
</tr>
<tr>
<td>Invasion rank</td>
<td>16</td>
<td>2</td>
<td>12</td>
<td>5</td>
<td>1</td>
<td>10</td>
<td>3</td>
<td>57</td>
<td>34</td>
</tr>
<tr>
<td>Invasion probability</td>
<td>0.040</td>
<td>0.104</td>
<td>0.047</td>
<td>0.075</td>
<td>0.707</td>
<td>0.058</td>
<td>0.083</td>
<td>0.018</td>
<td>0.025</td>
</tr>
<tr>
<td>Free inspection</td>
<td>-21.64</td>
<td>-23.41</td>
<td>-14.27</td>
<td>-7.97</td>
<td>-6.82</td>
<td>-7.27</td>
<td>-10.67</td>
<td>-0.69</td>
<td>-11.45</td>
</tr>
<tr>
<td>Inspection $10</td>
<td>-10.01</td>
<td>-8.45</td>
<td>-6.04</td>
<td>-2.21</td>
<td>-2.59</td>
<td>-2.70</td>
<td>-4.40</td>
<td>-0.30</td>
<td>-5.12</td>
</tr>
<tr>
<td>Inspection $20</td>
<td>-2.97</td>
<td>0.13</td>
<td>-1.32</td>
<td>0.53</td>
<td>-0.70</td>
<td>-0.53</td>
<td>-1.27</td>
<td>-0.10</td>
<td>-1.53</td>
</tr>
<tr>
<td>Quarantine</td>
<td>3.42</td>
<td>7.60</td>
<td>2.79</td>
<td>2.66</td>
<td>0.70</td>
<td>1.15</td>
<td>1.22</td>
<td>0.05</td>
<td>1.57</td>
</tr>
</tbody>
</table>

<sup>51</sup> 'Invasion rank' ranks sites in the risk set by invasion probability from highest (1) to lowest
Table 5.3. Lake Monona policies
The table displays policy-induced changes in baseline invasion probabilities (in percentages).

<table>
<thead>
<tr>
<th></th>
<th>Chain Lakes</th>
<th>Fox Lake</th>
<th>High Falls Reservoir</th>
<th>Lake Koshkonong</th>
<th>Lake Mendota</th>
<th>Lake Waubesa</th>
<th>Lake Wisconsin</th>
<th>Lake Wissota</th>
<th>Three Lakes Chain</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model I</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free inspection</td>
<td>-2.32</td>
<td>-8.11</td>
<td>-0.32</td>
<td>-12.65</td>
<td>-41.80</td>
<td>-26.03</td>
<td>-28.09</td>
<td>-0.49</td>
<td>-0.43</td>
</tr>
<tr>
<td>Quarantine</td>
<td>-0.68</td>
<td>-2.13</td>
<td>0.01</td>
<td>-4.12</td>
<td>-23.49</td>
<td>-14.60</td>
<td>-16.06</td>
<td>-0.24</td>
<td>-0.08</td>
</tr>
<tr>
<td><strong>Model III</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free inspection</td>
<td>-2.00</td>
<td>-7.05</td>
<td>-0.28</td>
<td>-10.82</td>
<td>-28.11</td>
<td>-22.66</td>
<td>-24.72</td>
<td>-0.42</td>
<td>-0.37</td>
</tr>
<tr>
<td>Quarantine</td>
<td>-0.59</td>
<td>-1.84</td>
<td>0.01</td>
<td>-3.50</td>
<td>-14.85</td>
<td>-12.60</td>
<td>-14.02</td>
<td>-0.21</td>
<td>-0.07</td>
</tr>
</tbody>
</table>
Table 5.4.  Welfare analysis
The first data column of the table shows sample mean willingness to accept estimates in 1989 dollars. The next two columns show a conservative and a non-conservative population estimate, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Mean willingness to accept</th>
<th>Total willingness to accept (low)</th>
<th>Total willingness To accept (high)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lake Winnebago</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$10 inspection</td>
<td>0.34</td>
<td>404,170</td>
<td>2,011,926</td>
</tr>
<tr>
<td>$20 inspection</td>
<td>0.50</td>
<td>586,922</td>
<td>2,921,651</td>
</tr>
<tr>
<td>Quarantine</td>
<td>0.62</td>
<td>724,427</td>
<td>3,606,138</td>
</tr>
<tr>
<td><strong>Lake Monona</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quarantine</td>
<td>0.16</td>
<td>189,921</td>
<td>945,412</td>
</tr>
</tbody>
</table>
Figure 5.1. Invasion predictions from Model I with a clarity feedback effect
Each panel shows a schematic map of sites in the choice set. Initial invasions include Lake Michigan counties, Mississippi River counties, and a single Lake Superior county.
Figure 5.2. The geographic location of policy lakes and focus lakes
Figure 5.3. Absolute changes in predicted choice shares from the Lake Winnebago quarantine policy
Figure 5.4. Percentage changes in invasion probabilities from the Lake Winnebago quarantine policy
Figure 5.5. Absolute changes in predicted choice shares from the Lake Monona quarantine policy
Figure 5.6. Percentage changes in invasion probabilities from the Lake Monona quarantine policy
References


Appendix A  The Questionnaire of the Wisconsin Recreational Boating Study

Special Recreational Boating Study

For this study, we are interested only in your personal experience with the boat mentioned in the cover letter. Please answer all questions with that boat in mind.

1. What type of hull does this boat have? (Circle one)
   - Open
   - Cabin
   - Pontoon
   - Other (please specify)

2. What type of propulsion does it have? (Circle one)
   - Inboard
   - Outboard
   - Inboard/outboard
   - Other powered (specify)
   - Sail
   - Sail with power
   - Other non-powered (Specify)

   If this boat has a gasoline motor, what is the horsepower of the primary motor on this boat?

   ______ horsepower

3. Please describe this boat:
   - Overall length: ______ feet
   - Beam (width): ______ feet

4. Did you use this boat at any time from April 21 (Saturday) through May 4 (Friday) in Wisconsin? (Circle one)
   - Yes
   - No... (PLEASE STOP AND SEND THE QUESTIONNAIRE BACK TO US)

1
5. Here's a calendar showing this 2-week boating period. On which days, if any, from April 21 through May 4 did you use the boat mentioned in our letter? Please circle all the days that you boated.

<table>
<thead>
<tr>
<th>Sun</th>
<th>Mon</th>
<th>Tues</th>
<th>Wed</th>
<th>Thur</th>
<th>Fri</th>
<th>Sat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>27</td>
</tr>
<tr>
<td>28</td>
<td>29</td>
<td>30</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

6. Now think about all the gasoline that you used in this boat from April 21 through May 4. How many gallons of gasoline did you use? (Make an estimate but please be as accurate as possible.)

I used _______ gallons of gasoline in this boat from April 21 through May 4.

About how much money did you spend on gasoline for this boat during this time?

I spent _______ on gasoline.

7. Did you use this boat on any inland lakes, rivers, or streams in Wisconsin from April 21 through May 4?

(Circle one)

Yes ________________________________ 1

No... (PLEASE GO TO QUESTION 9) ____________________ 2

8. In the spaces below, name the Wisconsin counties (cities or towns if you don't know the counties) where you used this boat on inland lakes, rivers, and streams during the past 2 weeks. Also record the number of days you boated in each county during this 2-week period (if you only boated part of a day, count that as a full day). Refer to the map on the back of this cover letter.

Name of county (city or town) | Number of days this boat was used on Wisconsin INLAND waters
County 1: ___________________________ 3
County 2: ___________________________ 4
County 3: ___________________________ 5

What is the name of the lake, river, or stream on which you did most of your boating during this time? ___________________________ (water where I boated the most)

9. Did you use this boat along any of the Wisconsin sections of Lake Michigan or Lake Superior coastline during this 2-week period?

(Circle one)

Yes ________________________________ 1

No... (PLEASE GO TO QUESTION 11) ____________________ 2

10. In the spaces below, name the Wisconsin counties (cities or towns if you don't know the counties) where you used this boat from the Great Lakes shores you used this boat during the past 2 weeks. Also record the number of days you boated off each county during this 2-week period (if you only boated part of a day, count that as a full day). Refer to the map on the back of the cover letter.

Name of county (city or town) | Number of days this boat was used on Wisconsin Great Lakes waters
County 1: ___________________________ 6
County 2: ___________________________ 7
County 3: ___________________________ 8
11. Here is a list of activities that you may have been involved in while boating. For the past 2-week period, please circle the number next to those activities which were a part of your boating experience.

(Circle all that apply)

A. Fishing from boat .............................................. 1
B. Cruising/sailing to destination .................................. 2
C. Water skiing .......................................................... 3
D. Swimming ............................................................ 4
E. Other enjoyment boating (other than above) .............. 5
F. Something else? Specify .......................................... 6

Please write the letter of the activity from question 11 which you spent the most time on during this 2-week period: ________ (activity I spent the most time on)

12. While you were boating did you have contact with any DNR warden or other local boating law enforcement official?

(Circle all that apply)

Yes ................................................................. 1
No ........................................................................ 2

13. Did others on the water interfere with your activity in any way? Please tell us what happened (use an extra sheet of paper if necessary):

NOW THINK ABOUT YOUR BOATING EXPERIENCES FOR THE LAST DAY THAT YOU BOATED DURING THIS 2-WEEK PERIOD. The following questions should be answered with this day in mind.

14. On the last day you boated from April 21 through May 4, how crowded did you feel while boating?

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all crowded</td>
<td>Slightly crowded</td>
<td>Moderately crowded</td>
<td>Extremely crowded</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

15. How satisfied were you with your boating on this day?

(Circle one)

Poor ................................................................. 1
Fair, things didn't work out very well ......... 2
Good, but a number of things could have been better . 3
Very good, but some things could have been better . 4
Excellent, only minor problems .................. 5
Perfect ............................................................ 6

16. On this day of boating, during what hours were you on the water?

Started: _________ am/pm  Finished: _________ am/pm
17. Including yourself, how many people were in your boating party?
   There were _______ people in my boating party.

18. About how much did you and ALL members of your group spend on the following items (including all money spent preparing for the occasion as well as during it)?
   _______ Food (groceries, etc.)
   _______ Lodging
   _______ Restaurants
   _______ Package liquor, wine, beer
   _______ Amusements
   _______ Auto
   _______ Sporting goods
   _______ Clothing and related goods
   _______ Gifty/souvenirs
   _______ Temporary slip/mooring rental
   _______ Other (Describe ____________________________ )

19. In what state and county do you live?
   __________________________ state __________________________ county

What is your zip code? __________

Sometimes we need to follow up on questionnaires to get more information. If we need to follow up, what number should we dial and who should we ask for?

________________________ Area Code/Phone # __________ first name

Thank you for taking the time to complete this survey. Please fold it so our address appears on the outside and return it to us right away. No postage is needed.

Have a safe and enjoyable boating season!

This study is being conducted by the Wisconsin Department of Natural Resources Bureau of Research

SPECIAL RECREATIONAL BOATING SURVEY ’95/’96
Wisconsin Department of Natural Resources
P.O. Box 7921
Madison, WI 53707

4
Appendix B  Corrected Second-Step Standard Errors

Two-step econometric procedures yield consistent estimates of second-stage parameters under fairly general conditions, but lead to biased second-step standard errors (Murphy & Topel, 1985). In general, the corrected second-step asymptotic covariance matrix, $V_2^*$, is calculated as

$$V_2^* = V_2 + V_2 \left( CV_1 C' - RV_1 C' - CV_1 R' \right) V_2,$$

(B.1)

where

$$C = E \left[ \left( \frac{\partial LL(\kappa)}{\partial \kappa} \right) \left( \frac{\partial LL(\kappa)}{\partial \lambda'} \right) \right],$$

and

$$R = E \left[ \left( \frac{\partial LL(\kappa)}{\partial \kappa} \right) \left( \frac{\partial LL(\lambda)}{\partial \lambda'} \right) \right].$$

$V_1$ is the asymptotic covariance matrix of $\hat{\lambda}$, the first-step parameters, based on $LL(\lambda)$, and $V_2$ is the asymptotic covariance matrix of $\hat{\kappa}$, the second-step parameters, based on $LL(\kappa)$ and given $\lambda$ (Greene, 2000). Note that because the expected second derivatives matrix of the likelihood function is the covariance matrix of its first derivatives vector, $V_1$ and $V_2$ may also be expressed as

$$V_1 = \left\{ E \left[ \left( \frac{\partial LL(\lambda)}{\partial \lambda} \right) \left( \frac{\partial LL(\lambda)}{\partial \lambda'} \right) \right] \right\}^{-1},$$

and

$$V_2 = \left\{ E \left[ \left( \frac{\partial LL(\kappa)}{\partial \kappa} \right) \left( \frac{\partial LL(\kappa)}{\partial \kappa'} \right) \right] \right\}^{-1}.$$

---

52 Step in this section refers to one of the two models, while stage refers to one of the two stages of the estimation procedure used in the random utility model.

53 $R$ is not to be confused with the human threat variable, $R_{jt}$, which is always written with subscripts specifying the site and time period it pertains to. Likewise, $C$ is not the same as $C_{nj}$, the travel cost to site $j$ for person $n$. 

150
In practice,

\[
\hat{C} = \frac{1}{n} \sum_{n=1}^{N} \left( \frac{\partial LL_n(\kappa)}{\partial \hat{K}} \right) \left( \frac{\partial LL_n(\kappa)}{\partial \hat{\lambda}'} \right)
\]

\[
\hat{R} = \frac{1}{n} \sum_{n=1}^{N} \left( \frac{\partial LL_n(\kappa)}{\partial \hat{K}} \right) \left( \frac{\partial LL_n(\lambda)}{\partial \hat{\lambda}'} \right),
\]

and any of the conventional asymptotic covariance matrix estimators can be used for \( V_1 \) and \( V_2 \). The terms involving \( R \), the covariance of the two gradients, disappear if the random components of the two models are independent.

In the present context, let \( \lambda \) represent the \( l \times 1 \) vector of first-stage utility function parameters in equation 4.2, and \( x_{nj} \) the conformable vector containing the explanatory variables entering the utility function for person \( n \) and site \( j \). Similarly, let \( \kappa \) represent the \( k \times 1 \) parameter vector of the hazard index, and \( w_{jt} \) the vector of its explanatory variables for site \( j \) and time period \( t \). Thus,

\[
V_{nj} = \delta_j + \gamma C_{nj} + \theta S_j^t + \phi S_j^t d_n^z + \beta B_n d_j^z = \lambda' x_{nj},
\]

\[
H_{jt} = \mu + \rho R_{jt} = \kappa' w_{jt},
\]

where the hazard index may be modified to include the environmental threat variable(s) when appropriate. The log-likelihood function of the first model is

\[
LL(\lambda) = \sum_{n=1}^{N} \sum_{j=1}^{J} y_{nj} \ln P_{nj},
\]

and the log-likelihood function of the second model is

\[
LL(\kappa) = \sum_{j=1}^{J} \left\{ \left( \sum_{l \in \mathcal{D}_j} \ln(1 - \pi_{jl-1}) \right) + z_{j,ij} \ln(\pi_{j,ij}) \right\}.
\]
Next, I express each log-likelihood derivative used in the calculation of $\hat{C}$ and $\hat{R}$. To proceed, note the following equalities (which will be used throughout this section to simplify results):

$$P_{nj} = \frac{e^{\lambda x_{nj}}}{\sum_{k=1}^{J} e^{\lambda x_{nk}}} ,$$

$$R_{ji} = \sum_{i=1}^{N} \sum_{n=1}^{N} P_{mi} P_{nj} ,$$

$$\pi_{ji} = \frac{e^{\lambda x_{ji}}}{1 + e^{\lambda x_{ji}}} .$$

Using the expressions for the hazard index and for invasion probability, the second-step log-likelihood function may be written as

$$LL(\kappa) = \sum_{j=1}^{J} \left\{ \sum_{\{i < j \leq t_j\}} \ln \left( 1 - \frac{e^{\mu P R_{ji-1}}}{1 + e^{\mu P R_{ji-1}}} \right) \right\} + z_{jt} \ln \left( \frac{e^{\mu P R_{ji-1}}}{1 + e^{\mu P R_{ji-1}}} \right) .$$

Before taking its derivative with respect to the parameters of the first model, note that the functional dependence on those parameters is through the threat variable. To simplify results, the exact form of this dependence is suppressed for now:

$$\frac{\partial LL(\kappa)}{\partial \lambda} = \sum_{j=1}^{J} \left\{ \sum_{\{i < j \leq t_j\}} \frac{1}{1 - \pi_{jt-1}} \left( -\left[ 1 + e^{H_{ji-1}} \right] e^{H_{ji-1}} \rho \frac{\partial R_{ji-1}}{\partial \lambda} - e^{H_{ji-1}} e^{H_{ji-1}} \rho \frac{\partial R_{ji-1}}{\partial \lambda} \right) \right\}$$

$$+ z_{jt} \frac{1}{\pi_{jt-1}} \left\{ \left[ 1 + e^{H_{ji-1}} \right] e^{H_{ji-1}} \rho \frac{\partial R_{ji-1}}{\partial \lambda} - e^{H_{ji-1}} e^{H_{ji-1}} \rho \frac{\partial R_{ji-1}}{\partial \lambda} \right\} .$$
\[
\frac{\partial LL(\kappa)}{\partial \lambda} = \sum_{j=1}^{J} \left\{ \sum_{ \{ \theta \neq c_{j1} \} } \frac{1}{1 - \pi_{j-1}} \left( -e^{H_{j-1}} \rho \frac{\partial R_{j-1}}{\partial \lambda} \left[ 1 + e^{H_{j-1}} - e^{H_{j-1}} \right] \right) \right\}
\]

\[
\quad + z_{j_{1}} \frac{1}{\pi_{j_{1}-1}} \left( e^{H_{j_{1}-1}} \rho \frac{\partial R_{j_{1}-1}}{\partial \lambda} \left[ 1 + e^{H_{j_{1}-1}} - e^{H_{j_{1}-1}} \right] \right)
\]

\[
\frac{\partial LL(\kappa)}{\partial \lambda} = \sum_{j=1}^{J} \left\{ \sum_{ \{ \theta \neq c_{j1} \} } \frac{1}{1 + e^{H_{j-1}}} \left( -e^{H_{j-1}} \rho \frac{\partial R_{j-1}}{\partial \lambda} \left[ 1 + e^{H_{j-1}} - e^{H_{j-1}} \right] \right) \right\}
\]

\[
\quad + z_{j_{1}} \frac{1 + e^{H_{j_{1}-1}}}{e^{H_{j_{1}-1}}} \left( e^{H_{j_{1}-1}} \rho \frac{\partial R_{j_{1}-1}}{\partial \lambda} \left[ 1 + e^{H_{j_{1}-1}} - e^{H_{j_{1}-1}} \right] \right)
\]

\[
\frac{\partial LL(\kappa)}{\partial \lambda} = \sum_{j=1}^{J} \left\{ - \sum_{ \{ \theta \neq c_{j1} \} } \left( e^{H_{j-1}} \rho \frac{\partial R_{j-1}}{\partial \lambda} \right) + z_{j_{1}} \left( \frac{1}{1 + e^{H_{j_{1}-1}}} \right) \rho \frac{\partial R_{j_{1}-1}}{\partial \lambda} \right\}
\]

\[
\frac{\partial LL(\kappa)}{\partial \lambda} = \sum_{j=1}^{J} \left[ z_{j_{1}} \left( 1 - \pi_{j_{1}-1} \right) \rho \frac{\partial R_{j_{1}-1}}{\partial \lambda} - \sum_{ \{ \theta \neq c_{j1} \} } \pi_{j_{1}-1} \rho \frac{\partial R_{j_{1}-1}}{\partial \lambda} \right].
\]
As noted, the above equation is a function of the derivative of the human threat variable with respect to the random utility model parameters. Recall that the threat variable may be expressed as

\[ R_{jt} = \sum_{i=I_1}^{N} \sum_{n=1}^{j} \frac{e^{x_{ni}}}{\sum_{k=1}^{J} e^{x_{nk}}} \left( \frac{e^{x_{nj}}}{\sum_{k=1}^{J} e^{x_{nk}}} \right) \]

\[ R_{jt} = \sum_{i=I_1}^{N} \sum_{n=1}^{j} \frac{e^{x(x_{ni}+x_{nj})}}{\left( \sum_{k=1}^{J} e^{x_{nk}} \right)^2} . \]

Therefore, its derivative with respect to the first-step parameters becomes

\[ \frac{\partial R_{jt}}{\partial \lambda} = \sum_{i=I_1}^{N} \sum_{n=1}^{j} \left[ \left( \sum_{k=1}^{J} e^{x_{nk}} \right)^2 e^{x(x_{ni}+x_{nj})} (x_{ni} + x_{nj}) - 2e^{x(x_{ni}+x_{nj})} \left( \sum_{k=1}^{J} e^{x_{nk}} \right) \left( \sum_{k=1}^{J} e^{x_{nk}} x_{nk} \right) \right] \left( \sum_{k=1}^{J} e^{x_{nk}} \right)^4 . \]

To simplify, let \( m = \sum_{k=1}^{J} e^{x_{nk}} \). Then,

\[ \frac{\partial R_{jt}}{\partial \lambda} = \sum_{i=I_1}^{N} \sum_{n=1}^{j} \left[ m^2 e^{x(x_{ni}+x_{nj})} (x_{ni} + x_{nj}) - 2me^{x(x_{ni}+x_{nj})} \left( \sum_{k=1}^{J} e^{x_{nk}} x_{nk} \right) \right] \left( \sum_{k=1}^{J} m^4 \right) . \]

\[ \frac{\partial R_{jt}}{\partial \lambda} = \sum_{i=I_1}^{N} \sum_{n=1}^{j} \left[ \frac{1}{m^4} (x_{ni} + x_{nj}) - \frac{2}{m^3} \left( \sum_{k=1}^{J} e^{x_{nk}} x_{nk} \right) \right] e^{x(x_{ni}+x_{nj})} . \]
Substituting this into the expression for \( \frac{\partial LL(\kappa)}{\partial \lambda} \), we arrive at

\[
\frac{\partial LL(\kappa)}{\partial \lambda} = \sum_{j=1}^{J} \left\{ \sum_{i \in I_{j-1}} \frac{1}{1 - \pi_{j-1}} \right\} \rho \sum_{i \in I_{j-1}} \sum_{n=1}^{N} \left[ \frac{1}{m^2} \left( x_{ni} + x_{nj} \right) - \frac{2}{m^3} \left( \sum_{k=1}^{J} e^{x_{nk}} x_{nk} \right) \right] e^{x_{nj}}.
\]

The other derivative needed in the calculation of \( \hat{C} \) is that of the second-step log-likelihood function with respect to the second-step parameters:

\[
\frac{\partial LL(\kappa)}{\partial \kappa} = \sum_{j=1}^{J} \left\{ \sum_{i \in I_{j-1}} \frac{1}{1 - \pi_{j-1}} \right\} \rho \sum_{i \in I_{j-1}} \sum_{n=1}^{N} \left[ \frac{1}{m^2} \left( x_{ni} + x_{nj} \right) - \frac{2}{m^3} \left( \sum_{k=1}^{J} e^{x_{nk}} x_{nk} \right) \right] e^{x_{nj}}.
\]

which, using the steps employed in deriving \( \frac{\partial LL(\kappa)}{\partial \lambda} \), simplifies to

\[
\frac{\partial LL(\kappa)}{\partial \kappa} = \sum_{j=1}^{J} \left\{ \sum_{i \in I_{j-1}} \frac{1}{1 - \pi_{j-1}} \right\} \rho \sum_{i \in I_{j-1}} \sum_{n=1}^{N} \left[ \frac{1}{m^2} \left( x_{ni} + x_{nj} \right) - \frac{2}{m^3} \left( \sum_{k=1}^{J} e^{x_{nk}} x_{nk} \right) \right] e^{x_{nj}}.
\]

Finally, the first-step log-likelihood function, and its derivative with respect to the first-step parameters are

\[
LL(\lambda) = \sum_{n=1}^{N} \sum_{j=1}^{J} y_{nj} \ln P_{nj},
\]
\[ LL(\lambda) = \sum_{n=1}^{N} \sum_{j=1}^{J} y_{nj} \ln \left( \frac{e^{\lambda x_{nj}}}{\sum_{k=1}^{J} e^{\lambda x_{nk}}} \right), \]

\[ \frac{\partial LL(\lambda)}{\partial \lambda} = \sum_{n=1}^{N} \sum_{j=1}^{J} y_{nj} \left( \frac{\sum_{k=1}^{J} e^{\lambda x_{nk}} x_{nj}}{\sum_{k=1}^{J} e^{\lambda x_{nk}}} \right) \left[ \frac{\sum_{k=1}^{J} e^{\lambda x_{nk}} x_{nj} - \left( \sum_{k=1}^{J} e^{\lambda x_{nk}} x_{nk} \right) e^{\lambda x_{nj}}}{\left( \sum_{k=1}^{J} e^{\lambda x_{nk}} \right)^2} \right] \]

\[ \frac{\partial LL(\lambda)}{\partial \lambda} = \sum_{n=1}^{N} \sum_{j=1}^{J} y_{nj} \left( x_{nj} - \frac{\sum_{k=1}^{J} e^{\lambda x_{nk}} x_{nk}}{\sum_{k=1}^{J} e^{\lambda x_{nk}}} \right) \]

\[ \frac{\partial LL(\lambda)}{\partial \lambda} = \sum_{n=1}^{N} \sum_{j=1}^{J} y_{nj} x_{nj} - \sum_{n=1}^{N} \sum_{j=1}^{J} y_{nj} \sum_{k=1}^{J} P_{nk} x_{nk} \cdot \]

\[ \frac{\partial LL(\lambda)}{\partial \lambda} = \sum_{n=1}^{N} \sum_{j=1}^{J} \left( y_{nj} - P_{nj} \right) x_{nj} \cdot \]

Note that in the present application, \( \sum_{j=1}^{J} y_{nj} = 1 \) (each boater visits a single site), hence

\[ \frac{\partial LL(\lambda)}{\partial \lambda} = \sum_{n=1}^{N} \sum_{j=1}^{J} y_{nj} x_{nj} - \sum_{n=1}^{N} \sum_{k=1}^{J} P_{nk} x_{nk} \]

Corrected standard errors calculated via equation B.1 for Model I parameters are presented in the third data column of Table B.1. The corrected standard errors imply that the parameter estimate for the human threat variable is not statistically significant. This result may appear somewhat surprising. As I discuss in Chapter 4, in-sample predictions fairly successfully replicate the observed spread of zebra mussels. The driving force behind those
predictions is the threat variable – the constant term in the hazard index does not create a hazard difference between various sites. The uncorrected standard errors are based on the sandwich (robust covariance matrix) estimator, and, though incorrect, their bias could be expected to be relatively small considering the large sample size and the generally significant parameter estimates shown in Table 4.1. Note, however, that several other parameters not displayed in the table are also estimated in the first stage of the random utility model – namely, the $J-1$ alternative specific constants. These affect $V_2^*$ via both $V_1$ and $C$.\textsuperscript{54} The findings presented in this appendix are likely driven by the fact that the precision of some of the alternative specific constants is low. Given that estimating a full set of alternative specific constants is needed for the unbiasedness of other parameter estimates, there is no easy way around this problem.

\textsuperscript{54} The covariance of the two gradients converges to zero, and the terms containing \( R \) have no effect on the value of \( V_2^* \).
Table B.1.  Two-step statistical inference

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Parameter estimate</th>
<th>Uncorrected standard error</th>
<th>Corrected standard error</th>
<th>Corrected t-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-4.3424</td>
<td>0.1557</td>
<td>0.5029</td>
<td>-8.6340</td>
</tr>
<tr>
<td>Human Threat</td>
<td>0.1037</td>
<td>0.0204</td>
<td>0.2104</td>
<td>0.4926</td>
</tr>
</tbody>
</table>