Three studies on environmental valuation

by

DongGyu Yi

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Economics

Program of Study Committee:
Joseph A. Herriges, Major Professor
Catherine L. Kling
John Miranowski
Peter F. Orazem
John R. Schroeter

Iowa State University

Ames, Iowa

2014

DEDICATION

This dissertation is dedicated to my Lord, Jesus Christ, who led me at every step of the way. I would also like to thank my family for their unconditional love and support throughout the course of this work.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Herriges, and my committee members, Dr. Kling, Dr. Miranowski, Dr. Orazem, and Dr. Schroeter, for their guidance and support throughout the course of this research.

In addition, I would also like to thank my colleagues, the department faculty and staff for making my time at Iowa State University a wonderful experience.

Finally, thanks to my family, friends including TJ Kim for their encouragement and love.

GENERAL ABSTRACT

This dissertation is devoted to the study of environmental valuation with three independent topics. The first topic investigates the consistency of consumer preferences over time and revealed versus stated preference data. This study draws on data from the Iowa Lakes Project, which provides information on recreational usage patterns over several years and for approximately 130 lakes, along with detailed information on the water quality for each lake. This allows examination of the extent of convergence in how individuals respond to changing site characteristics. In addition, because of the nature of the data, this study was able to investigate the consistency of consumer preferences over time and between actual versus anticipated visitation patterns. The second topic examines how housing prices were impacted by this unexpected event, while controlling for pre-existing flood risk (as captured by 100- and 500-year floodplains). Both difference-in-differences (DID) and triple differences (DDD) techniques are used to isolate the impact of the 2008 flood. The results show prices for houses within the 100-year floodplain were discounted prior to the flood and no significant changes occurred in prices for those houses inundated by the 2008 flood. However, the results find a significant rebounding in post-flood prices in areas not actually inundated by the flood. On the other hand, the prices of properties in both the 500-year floodplains and outside the floodplains were not significantly discounted before the flood. There was a significant decrease in price after the flood, if the area was inundated. These findings imply as new information on flooding occurs, the housing market updates the risk perception for properties, as indicated by a change in housing prices. Finally, the third topic analyzes state dependence in recreational demand using panel data on lake visitation patterns from the Iowa Lakes Project. When calculating the welfare effect of an environmental

policy, the estimation can be misleading—either by ignoring state dependence or by dealing with state dependence incorrectly. To avoid this problem, this topic adopts the approach proposed by Wooldridge. This approach starts with a single site case and then extends the analysis to a multiple site setting. For the single site case, a dynamic random effect (RE) logit model is utilized. In the multiple site setting, a RE two-step nesting structure model is used, capturing state dependence in terms of overall trip taking, although not in terms of the specific sites selected. For both the single and multiple site cases, a RE Poisson model is also estimated as an alternative approach to compare the results and as a robustness check. Also, a Monte Carlo simulation exercise is used to show the biases that can arise either from neglecting state dependence entirely or from treating it incorrectly.

CHAPTER 1. CONVERGENT VALIDITY AND TIME CONSISTENCY OF PREFERENCES: EVIDENCE FROM THE IOWA LAKES RECREATION DEMAND

1.1. Introduction

The task of valuing environmental amenities is hampered by the lack of direct markets for such goods. To fill this void, researchers have turned to a variety of revealed preference (RP) and stated preference (SP) methods to infer the values of interest.[1] For example, recreation demand models are often used to estimate the value of both sites as a whole and their individual attributes (e.g., water quality) by modeling individual visitation patterns as a function of travel cost and site characteristics. The logic is that individuals reveal information about the value they place on an environmental amenity by incurring travel cost to reach sites where the amenity is found. One practical problem with this approach is that there may be little variation among the available sites in the amenity of interest, making it difficult to tease out its marginal value to consumers. Stated preference techniques can help alleviate this problem. In particular, contingent behavior (CB) surveys ask individuals how they might shift their visitation patterns in response to changes in site access, travel cost, or individual site attributes. This provides not only the variation needed to identify the marginal impact of a particular amenity, but, as noted by von Haefen and Phaneuf (2008), also avoids potential omitted variables' bias through the random assignment of survey scenarios.

The potential problem with combining stated and revealed preference data sources is that they may not be driven by the same data generating process. For example, one might be concerned that respondents to a contingent behavior survey have an incentive to overstate their additional trips to a site improved under a proposed policy scenario. By doing so, the respondent

[1] See, Freeman, Herriges, and Kling (2014) for an overview of this literature.

encourages policymakers to adopt the change, creating the *option* for future use of the improved site, and, depending upon how the policy is paid, incurring little or no direct costs.[2] Problems can also emerge when the data sources differ in a temporal dimension. For example, anticipated trips for the coming year, even without changing conditions, can differ significantly from actual trips because respondents are overly optimistic about their future recreational activities (much like one might overstate future trips to the gym). Even comparisons in actual behavior over time may be problematic, as preferences evolve or are impacted by changes over time in unobservable individual or site attributes.

There have been a number of papers in the literature examining the convergent validity of valuations based on actual versus contingent behavior responses (e.g., Azevedo *et al*. (2003), Grijalva *et al*. (2002), Loomis and Richardson (2006)). However, most of these studies are based upon a single site and focus on consumer response to changes in either travel cost or site access. The purpose of this chapter is to examine the convergent validity of nonmarket valuations (and the underlying preference parameters) along several dimensions, drawing on a unique database from the Iowa Lakes Project. The Iowa Lakes Project is a multi-year panel study of the usage patterns of Iowa households with regards to the 132 primary recreational lakes in the state. Of particular interest for this chapter are the data sets collected in 2004 and 2005 surveys. In 2004, households were asked to report:

- Actual single-day trips to each lake in 2004 (AT04);
- Expected single-day trips to each lake in 2005 under current water quality conditions (ET05); and

[2] This problem is analogous to the difficulty associated with valuing private, good using contingent valuation techniques (See, e.g., Carson and Hanemann (2005)).

• Expected single-day trips to each lake in 2005 *contingent* on water quality improvements to a subset of the lakes (CB05).

In 2005, these same households were ask to report on their actual single day trips to each of the 132 lakes (AT05). The four data sets allow for a total of six pair-wise convergent validity tests, including convergence between actual and contingent behavior responses (AT05 vs. CB05), convergence between actual and expected trips under fixed water quality conditions (AT05 vs. ET05), convergence between expected trip responses with and without water quality improvements (CB05 vs. ET05), and convergence between actual trips in differing years (AT04 vs. AT05).[3] For each pairwise comparison, convergence was examined along three dimensions:

• Convergence in individual parameters (such as the marginal utility of income);

• Joint convergence in the parameter estimates; and

• Convergence in the implied welfare measures for a series of policy scenarios.

The remainder of this chapter is organized into five additional sections. Section 1.2 briefly reviews the existing literature examining convergent validity in nonmarket valuation. Section 1.3 describes the repeated nested logit model use in the convergent validity tests and the estimation strategy used. Section 1.4 describes the Iowa Lake survey in greater detail and provides summary statistics for the various data sources. Section 1.5 provides the estimated models and the pairwise convergent validity comparisons. The chapter finishes with a summary and conclusions in Section 1.6.

[3] Comparisons are also possible between AT04 vs. CB05 and between AT04 vs. ET05, although these comparisons compound changes in years and some other factors.

1.2. Literature Review

Extensive literature has emerged over the past two decades, both in terms of combining state and revealed preferences (See, e.g., Cameron (1992) and Whitehead *et al*. (2008)) and in testing for the convergent validity of the two data sources (e.g., Carson *et al*. (1996) and Carson and Hanemann (2005)). While much of this literature has focused on contingent valuation as the source of the stated preference data, there are a number of studies drawing on contingent behavior data. Adamowicz *et al*. (1994), for example, combine revealed preference on visitation patterns to 24 recreational sites and stated preferences elicited through a choice experiment. They find that the fundamental preferences in both cases are similar, and that combining the revealed and stated preferences yields benefits in estimation. They conclude that the multicollinearity among quality attributes, which often plagues revealed preference data sources, can be ameliorated through a well-designed stated preference survey.

Englin and Cameron (1996) combine observed behavior and contingent behavior responses to an increased travel cost to estimate the demand for recreational angling. Like Adamowicz *et al*. (1994), they conclude that contingent behavior data can be a valuable supplement to observed data. Grijalva *et al*. (2002) test the validity of contingent behavior trip for three different levels of recreational site access: data before and after a policy to restrict climbing access and hypothetical changes in site access. They suggest that CB data can be a useful supplement to RP when policy proposals are historically unobservable. Loomis and Richardson (2006) investigate Rocky Mountain National Park visitation behavior from climate change to compare observed behavior and intended behavior. They also do not find any statistical difference between RP and CB.

Though the above literature supports the consistency between revealed and stated preference data, other studies reject convergent validity. Adamowicz *et al*. (1997), for example, test consistency between observed and CB data in their moose hunting demand study and, at least for some modeling configurations, reject convergent validity. von Haefen and Phaneuf (2008), using the same data source and a mixed logit framework, also reject consistency between the implied RP and SP preferences. Azevedo *et al*. (2003) combine revealed preferences and state preferences under hypothetical higher trip costs using data from the Iowa wetlands survey. They test the hypotheses of consistency between the RP and SP with four different levels of travel cost and find that all hypotheses are rejected. Whitehead *et al*. (2010) consider the consistency in the context of trips to beaches in southern North Carolina. They construct three models: a Kuhn-Tucker (KT) model, a single equation count data model of RP-SP trips and a count data demand system. Their results are mixed. Their KT and RP-SP models are convergent valid in trip prediction, while they are convergent invalid in terms of implied welfare effects. In addition, their count system model is convergent invalid with both KT and RP-SP models.

One of the limitations with most studies combining RP and CB data is that the actual trips data lack significant variation in the underlying site characteristics, either because it is not there or it is not observed, making it difficult to isolate the impact that these characteristics have on the revealed preference trips. Indeed, this is one of the primary arguments for combining RP and SP data. The problem, of course, is that convergent validity tests in this setting are essentially limited to testing convergence in the implied marginal utility of income and unable to test for convergence in the marginal willingness to pay for site attributes. One exception is Jeon and Herriges (2010), who draw on a portion of the Iowa Lakes data employed in this study. Specifically, they compare lake recreation trip prediction, based on existing variation in water quality with SP trips following a hypothetical improvement in water quality. Jeon and Herriges reject convergent validity of the two data sets. The marginal utility associated with a hypothetical higher water quality index is predicted two-thirds as great as that from actual water quality index level. This study extends the work by Jeon and Herriges in a number of directions. First, Jeon and Herriges (2010) rely on strong distributional assumptions to allow for correlations among the various data sources; whereas, in the current study, a more flexible bootstrap approach is employed. Second, Jeon and Herriges do not control for potentially omitted site characteristics variables in their analysis; whereas, this study does through the use of alternative specific constants. Third, Jeon and Herriges, unlike the current study, do not examine the consistency of preferences over time through the comparison of actual trips patterns between 2004 and 2005.

1.3. Model

The modeling framework employed in this study is the repeated nested logit (RNL) model developed by Morey *et al*. (1993). The RNL model is based on the assumption that, over the course of a season, an individual faces a series of *choice occasions* $\left(t=1,\cdots,T\right)$. On each choice occasion, the individual decides among $J+1$ alternatives, where $j=0$ denotes the alternative of staying at home and $j=1,\cdots,J$ denotes the choice of visiting one of the J recreational sites in the choice set. The current application uses $T=52$, which can be interpreted as allowing for one choice occasion weekly for the year. For data source v (v = AT04, AT05, CB05, and ET05), the utility that individual i receives from choosing alternative j on choice occasion t is assumed to take the form:

$$
\begin{aligned}
U_{ijtv} &= V_{ijv} + \varepsilon_{ijtv} \quad j = 0, \cdots, J \\
&= \begin{cases} Z_{iv}\gamma_v + \varepsilon_{i0tv} & j = 0 \\ \alpha_v + X_{jv}\beta_v + \tau_v \cdot TC_{ijv} + \xi_{jv} + \varepsilon_{ijtv} & j = 1, \cdots, J \end{cases} , \\
&= \begin{cases} Z_{iv}\gamma_v + \varepsilon_{i0tv} & j = 0 \\ \delta_{jv} + \tau_v \cdot TC_{ijv} + \varepsilon_{ijtv} & j = 1, \cdots, J \end{cases}
\end{aligned} \tag{1}
$$

where Z_{iv} denotes characteristics of individual *i*, TC_{ijv} denotes the cost of visiting site *j* (which does not vary by choice occasion), X_{jv} denotes observed attributes of site *j* (e.g., water quality), ξ_{jv} denotes attributes of site *j* not observed by the analyst (but assumed to be observed by the individual), ε_{ijtv} denotes idiosyncratic factors (again assumed to observe by the individual but not by the analyst), and

$$\delta_{jv} = \alpha_v + X_{jv}\beta_v + \xi_{jv} \quad . \tag{2}$$

The δ_{jv} represent alternative specific constants for site *j* in data source *v*, capturing both observed and unobserved site characteristics. The error vector, $\varepsilon_{i\bullet tv} = \left(\varepsilon_{i0tv}, \cdots, \varepsilon_{iJtv}\right)'$, is assumed drawn from a Generalized Extreme Value (GEV) distribution nesting together alternatives $j = 1, \cdots, J$, and is independently and identically distributed across individuals and choice occasions. As is standard in random utility models, it is assumed that the individual chooses the alternative that maximizes its utility on each choice occasion, with

$$I_{ijtv} = 1\left(U_{ijtv} > U_{iktv} \quad \forall k \neq j\right) \quad . \tag{3}$$

Given the above specification, the probability that individual *i* chooses alternative *j* during choice occasion *t* has a convenient closed form given by

$$P_{ijtv} = \begin{cases} \dfrac{\exp\left(V_{i0v}\right)}{\exp\left(V_{i0v}\right) + \left[\sum_{k=1}^{J} \exp\left(V_{ikv}/\theta_v\right)\right]^{\theta_v}} & j = 0 \\ \dfrac{\exp\left(V_{ijv}/\theta_v\right) \cdot \left[\sum_{k=1}^{J} \exp\left(V_{ikv}/\theta_v\right)\right]^{\theta_v - 1}}{\exp\left(V_{i0v}\right) + \left[\sum_{k=1}^{J} \exp\left(V_{ikv}/\theta_v\right)\right]^{\theta_v}} & j = 1, \cdots, J \end{cases}, \tag{4}$$

where $\theta_v \in \left(0,1\right]$ is often referred to as the *dissimilarity coefficient*. The smaller θ_v becomes more correlated, as the ε_{ijtv} across the trip alternatives and the better substitutes (i.e., more similar) the trip alternatives become. Conversely, the larger θ_v becomes, the more dissimilar the trip alternatives. If $\theta_v = 1$, the model reduces to a simple logit model for each choice occasion.

The overall contribution of an individual to the likelihood function becomes

$$P_{i\bullet v} = \prod_{j=0}^{J} \prod_{t=1}^{T} \left[P_{ijtv}\right]^{I_{ijtv}} \quad . \tag{5}$$

The corresponding log-likelihood function for the data source, *v*, becomes

$$LL_v = \sum_{i=1}^{N} \ln P_{i\bullet v} \quad . \tag{6}$$

As suggested by Murdock (2006), the parameters of the model are estimated in two stages. During the first stage, the parameters $\left(\gamma_v,\ \delta_{jv},\ \theta_v \text{ and } \tau_v\right)$ are estimated. The second stage involves an ordinary least squares (OLS) regression model that uses the estimated ASCs $\left(\hat{\delta}_j\right)$ as the dependent variables to estimate the parameter of site attributes and the constant in Eq. (2) $\left(\text{i.e., } \hat{\alpha}_v \text{ and } \hat{\beta}_v\right)$. If the omitted site-specific variables, $\left(\xi_{jv}\right)$, are potentially correlated with the observed site characteristics, $\left(\text{i.e., } X_{jv}\right)$, then instrumental variables techniques may be needed at this stage.

1.4. Estimation

1.4.1. Data

The main data source for the current study is the lake trip survey data from the Iowa Lakes Project. The Iowa Lakes Project is a panel study from 2002 to 2005 and 2009, supported by the Iowa Department of Natural Resources (IDNR) and the US EPA. The primary objectives of the project are to better understand the visitation pattern of Iowa residents to the primary lakes in the state and the implications this visitation pattern has in understanding the value Iowans place in lake water quality. Among five-year surveys, this study focuses on surveys for 2004 and 2005. As noted above, these surveys yield four distinct data sets regarding actual and hypothetical site visitation patterns, including

- Actual single-day trips to each lake in 2004 (AT04);
- Expected single-day trips to each lake in 2005 under current water quality conditions (ET05);
- Expected single-day trips to each lake in 2005 *contingent* on water quality improvements to a subset of the lakes (CB05);[4] and
- Actual single-day trips to each lake in 2005 (AT05).

The first three data sets were elicited in the 2004 survey, while AT05 was elicited in the 2005 survey.

[4] Specifically, the contingent behavior question asked survey respondents how their visitation pattern would change if all lakes in the state were improved to “swimmable”, an improvement for 52 of the primary lakes in the state relative to 2004 conditions.

The four data sets allow for a total of six pair-wise convergent validity tests, including convergence between actual and contingent behavior responses (AT05 vs. CB05), convergence between actual and expected trips under fixed water quality conditions (AT05 vs. ET05), convergence between expected trip responses with and without water quality improvements (CB05 vs. ET05), and convergence between actual trips in differing years (AT04 vs. AT05). Comparisons are also possible between AT04 vs. CB05 and between AT04 vs. ET05, although these comparisons compound changes in years and some other factors.

While the Iowa Lakes Project covers the 131 primary lakes in the state, this study focuses on the top 100 most visited lakes. This is achieved for two reasons. First and foremost, as described below, bootstrapped samples are used to obtain standard errors and allow for correlations across the four data sources. For rarely visited sites, these samples may have no one visiting them, making estimation of the site's corresponding alternative specific constant impossible. Second, given the large number of bootstrapped samples employed, limiting the number of sites helps to reduce the overall time required to complete the estimation task. In each survey, respondents are asked which lakes in Iowa they visit and how often. The visit consists of two different styles of trips—single-day trip and overnight trip. The analysis below focuses on single-day trips. In 2004, the average number of single-day trips to the 100 lakes is 5.89, while the average number to 131 lakes is 6.02. So, the top 100 most visited lakes account for most of the lake trip demand in Iowa, covering about 98% of Iowa lake trips.

In addition to trip data, the survey also included questions about respondents' demographic characteristics, such as age, gender, education level, household size, and income level. As for socio-demographic characteristics, among each survey's demographic questions, four characteristics—age, gender, education and household size—are selected for use. For education, the categories are simplified to use a dummy variable in the model. If the education level is equal to or higher than college graduate, then the dummy is one otherwise zero.

The data from the Iowa Lake Valuation Project also includes six available site attributes (i.e., the X_{jv}'s): (1) lake size (acres), (2) boat ramp dummy, (3) wake restrictions dummy, (4) handicap facilities dummy, (5) state park dummy, and (6) water quality index (on a scale from 1 to 10, with 7 being "swimmable"). In addition, three physical and chemical characteristics of the lakes are site characteristics: (1) secchi transparency, (2) total phosphorus, and (3) total nitrogen. Hence, the analysis in this study considers the effect of water quality on trip choice with two different methods. The first method includes X_{jv}, the site characteristics (including the water quality index), while the second method replaces the water quality index with the three chemical and physical attributes stated previously. Table 1.1 shows the summary statistics for the lake attributes as well as socio-demographic characteristics and trips.

The total number of respondents, who returned and completed the 2004 and 2005 surveys, are 4,242 and 3,993, respectively. Among the total responses, this study excludes respondents who included irrelevant answers, including no answer, but also individuals who visited any lake more than 52 times. The concern with including respondents who answered greater than 52 visits is that these respondents consist predominantly of households who live in close proximity to a certain lake. In this case, they could be residents who pass a lake on their commute to work or walking/biking along it. These kinds of demands for lakes are not under the guidelines for this study. After this first reduction, the numbers for the remaining respondents are 4,119 (AT04), 3,828 (AT05), 4,054 (ET05), and 4,084 (CB05).

There is a second reduction of the sample. Among the respondents, only those, who are in common in all four cases, are included. This is because the main objectives of this study are to test whether the recreation demand from the stated preference (including contingent survey) is different from the demand from the revealed preference, whether the demands from different levels of stated preferences in the same survey are the same, and whether the recreation demands from the revealed preferences in consecutive years are the same . The remaining sample size is 2,425 after the second reduction. In addition, there are some mismatching data and outliers related with gender and size of household. For these miscellaneous reasons, 81 persons are additionally dropped from the sample, leaving this study with a final sample size of 2,344.

For travel cost, this study basically follows the formula below:

$$Travel\ cost = round\ trip\ distance \cdot fuel\ cost + \frac{1}{3}\left(round\ trip\ time \cdot respondent's\ hourly\ wage\right).$$

Including Abidoye *et al*. (2012), many articles use this formula. In this study, PC Miler was used to compute the trip distance and time. CPI adjusted gasoline prices (dollars/gallon) divided by average fuel efficiency of U.S. light duty vehicles (miles/gallon) are used as a proxy for fuel cost (dollars/mile).[5] For hourly wage, the survey responses to total household's annual income are used. In every income category, median annual income is selected, then divided by 2,000, and finally adjusted with CPI to obtain an hourly wage.[6]

[5] Each source comes from U.S. Energy Information Administration (Midwest all grades all formulations retail gasoline prices), The Research and Innovative Technology Administration in the U.S. Department of Transportation (average fuel efficiency of U.S. light duty vehicles), Bureau of Labor Statistics in U.S. Department of Labor (annual CPI and average hourly earnings), respectively.
[6] This number comes from a 40-hour work week with two weeks of vacation annually.

1.4.2. Consistency Test

For the test of consistency, this study approaches all comparisons—both from a parametric point of view and from the view of social welfare. For the welfare estimation level, this study examines whether there is any difference in willingness to pay (WTP) or compensating variation (CV) for a specific policy between any two cases. At the parameter level, this study checks whether parameters estimated from any two of four cases are the same with each other by a parameter ratio comparison as well as by direct comparison. Most of the welfare estimation comes from the WTP, a ratio of a certain parameter to the coefficient of travel cost, such as Eq. (9) below. In this sense, parameter ratio comparison means the comparison of two data sets in terms of possible WTPs. Therefore, on one hand, parameter ratio comparison is a parameter-level test, but on the other hand, it is a welfare-estimation-level test. As a part of parameter-level comparison, this study also tests whether various sets of parameters estimated from each case are consistent with each other.

All of these consistency tests are executed with the bootstrapping technique. Bootstrapping individually makes it possible to control for correlations in trip data over time and across data types without additional assumptions regarding the specific structure of the correlation. Two-stage estimation is also adopted in the bootstrapping. Based on Eqs. (1) and (2), we can construct a bootstrapping model in the following manner.

Instead of the original sample, generate a new sample by N times random drawing with replacement from the entire sample. Then, by using Eq. (1), a set of estimates ($\hat{\gamma}^b$, $\hat{\delta}_j^b$, and $\hat{\tau}^b$) is estimated in every bootstrapping repetition.[7] This is the first stage of bootstrapping here.

---

[7] Superscript *b* denotes *b*-th repetition.

Hence, we can obtain j number of ASC estimates $\left(i.e., \hat{\delta}_1^b, \cdots, \hat{\delta}_J^b\right)$ in every first stage of bootstrapping. For the second stage estimation, this study proceeds another random sampling with the ASCs estimated in the first stage, $\hat{\delta}_j^b$, and generates a new ASCs bundle by J times random drawing with replacement from the ASC estimates. Then, by using this new ASCs bundle as a dependent variable, the classical OLS approach provides a set of estimated parameters for site characteristics ($\hat{\alpha}^b$ and $\hat{\beta}^b$) in Eq. (2). This completes the second stage of bootstrapping.

The double bootstrapping algorithm used here can be summarized as follows.

Step 1: Draw a number randomly N times from the range [1, N (=2,344)] with replacement.

Step 2: Set a new data set, which consists of the information of the respondents corresponding to the number drawn from Step 1, for all four cases—AT04, AT05, ET05, and CB05. (So, the new data set has all the same respondents across all cases.)

Step 3: Calculate the RNL. (This is the first stage of bootstrapping estimation.)

Step 4: Draw a number randomly J times from the range [1, J (=100)] with replacement.

Step 5: Set a new data set consisting of attribute information of the lakes corresponding to the number drawn from Step 4, for all the cases. (So, the new data set has all the same lakes across the four cases.)

Step 6: Use only ASCs corresponding to the number drawn from Step 4 as dependent variables to calculate the OLS for the newly chosen lakes, among the estimated ASCs in Step 3. (This is the second stage of bootstrapping estimation.)

Step 7: Repeat Step 1 through Step 6 as many as B (=500) times.

Step 8: Calculate means and standard errors of estimated parameters for each case, separately. (There will be *B* number of estimates bundles. Therefore, a researcher can calculate mean and standard error for each estimate whose sample size is *B*.)

Step 9: Calculate the difference of the estimated parameters between any two of the new four data sets to obtain means, standard errors, and the corresponding variance-covariance matrix.

From this bootstrapping process, one can test the null hypothesis—each parameter in any two cases is the same with each other. If the parameters of travel cost in the cases for AT04 and CB05 (τ_{AT04}, τ_{CB05}) are taken as an example, then the null hypothesis is $H_0 : \tau_{AT04} - \tau_{CB05} = 0$. We know the mean for ($\tau_{AT04} - \tau_{CB05}$) and its standard error. Therefore, we can verify whether the difference is statistically, significantly different from zero or not.

Furthermore, one can test joint consistency between two cases. We have the mean for all the differences and the variance-covariance matrix. Hence, we can calculate the Wald statistic for any specific group of parameters. For instance, if we denote all the parameters as estimated in AT04 and CB05 cases as vectors, β_{AT04}, β_{CB05}, respectively, then the null hypothesis of joint consistency will be $H_0 : \beta_{AT04} - \beta_{CB05} = 0$ and the Wald statistic will be like Eq. (7).

$$W_{AT04-CB05} = \left[\beta_{AT04} - \beta_{CB05}\right]^T \left[Est.\ Asy.\ Var\left(\beta_{AT04} - \beta_{CB05}\right)\right]^{-1} \left[\beta_{AT04} - \beta_{CB05}\right] . \tag{7}$$

For the final step of the consistency test, the consistency in welfare estimation between any two cases is also investigated. Among various policies, two types of policies will be discussed here. First, this study compares the WTP for improving water quality by one unit in the index. Because the contingent trip data (CB05) are assuming hypothetically higher water quality, it seems natural to make a scenario for water quality improvement. The average individual's additional WTP for improving water quality of a specific lake is a change in travel cost of the lake that will offset one unit of change in the water quality index. Hence, the utility will be the same before and after the change mathematically. The concept of the WTP for improving water quality can be expressed as

$$U_{ij}\left(\left(WQI_j+1\right),\left(TC_{ij}+WTP\right);\overline{X_j-WQI_j}\right)=U_{ij}\left(WQI_j,TC_{ij};\overline{X_j-WQI_j}\right) \quad , \tag{8}$$

where WQI_j is the water quality index level of site *j*, TC_{ij} is travel cost that individual *i* needs to go to trip site *j*, and X_j is observable attributes of site *j*. $\overline{X_j-WQI_j}$ signifies all site attributes, except for water quality index, are assumed fixed. Then, based on the utility function of Eq. (1) described in Section 1.3, the WTP can be defined as the parameter for water quality index (β_{WQI}) over the negative travel cost parameter (τ).

$$WTP=-\frac{\beta_{WQI}}{\tau} \quad . \tag{9}$$

Thus, one can calculate all WTPs for the four cases and the difference between any two of them by using Eq. (9) in every bootstrapping repetition. The corresponding means and standard errors are also obtained after all repetitions.

Second, this study also compares the average welfare estimation for the scenario that one of the 100 lakes—Big Creek Lake—is no longer accessible. In the nested logit model, the average compensating variation (CV) associated with this scenario has the well-known explicit formula as:[8]

[8] Big Creek Lake is the 16th largest lake in Iowa. It was selected arbitrarily.

$$
\begin{aligned}
CV &= \sum_{i=1}^{N} CV_i \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{-1}{\tau}\left[Max_{j\in J}\left\{V_{ij}\right\} - Max_{j\in J'}\left\{V_{ij}\right\}\right] \qquad , \quad (10) \\
&= -\frac{1}{\tau\cdot N}\sum_{i=1}^{N}\left[\ln\left\{\exp\left(V_{i0}\right)+\left[\sum_{j\in J}\exp\left(\frac{V_{ij}}{\theta}\right)\right]^{\theta}\right\}-\ln\left\{\exp\left(V_{i0}\right)+\left[\sum_{j\in J'}\exp\left(\frac{V_{ij}}{\theta}\right)\right]^{\theta}\right\}\right]
\end{aligned}
$$

where $J = \{1, 2, \cdots, 100\}$, $J' = \{2, 3, \cdots, 100\}$.

In Eq. (10), N signifies the total number of individuals (2,344), J is the set of whole alternatives, except for staying at home option (when j = 0), J' is the set of all the other alternatives, excluding alternative 1, which signifies Big Creek Lake. The last process to compare the average CVs is the bootstrapping technique explained above. In every bootstrapping repetition, one can obtain the simulated individual CVs and the average CV for the four cases. Hence, in b-th repetition, one can obtain the difference in the average CV between any two of the four cases. In the end, with B (=500) times repetition, one can calculate the means and standard errors of the CVs from the four cases and the difference in the average CV between any two of these.

1.5. Results

Table 1.2 shows the results of the first stage estimation of four cases in standard logit and RNL models. The signs of parameters for all individual characteristics mean that the older people tend to stay at home, while males and more educated persons are more willing to go on trips instead of staying at home, and that people are more inclined to go on trip as their family becomes larger. Although all the socio-demographic variables are statistically significant, except for education level in CB05, education level of all the cases and household size of CB05 are insignificant in the bootstrapping technique (Table 1.5).

The estimated coefficients of site characteristics, the estimates from the second stage, are shown in Tables 1.3 and 1.4. They are derived by the standard OLS regression formulas. The estimates in Table 1.3 are obtained when water quality index, as a representative for lakes' water quality, is inserted into the model, while those in Table 1.4 are the results when, instead of water quality index, the three factors—secchi depth, total phosphorus, and total nitrogen—are considered in the model. Four of the parameters in the model with water quality index—constant term, log of lake size, wake restriction dummy, and state park dummy—are statistically significant in all the cases. They are also significant in the bootstrapping process. The water quality index is only significant for the ET05 case. Similarly, these four water quality index parameters are also significant in the bootstrapping as well as in the basic model with the three water quality factors. On the other hand, in the model with the three water quality factors, secchi depth is also significantly positive both in the basic model and in bootstrapping. The handicap facilities dummy is significant in some cases, only when the simple logit model is applied.

Compared to the standard errors that come from the Hessian of maximum likelihood function and calculated from the variance-covariance matrix in the second stage OLS procedure, standard errors obtained from bootstrapping (500 repetitions) results in different significance levels. Tables 1.5 and 1.6 show the means and standard errors of the primary parameters derived by the first stage bootstrapping and by the second stage bootstrapping, respectively. All standard errors of estimated parameters are larger when bootstrapping is applied. It is natural that the bootstrapping procedure has larger variances because it does not make use of any special assumptions and restrictions.

As for the consistency tests, first, the results of the test on single parameter equality are shown in Tables 1.7, 1.8, and 1.9. During the first stage, the number of parameters that are significantly different between these two cases is the smallest when the combinations are AT04-ET05, AT04-CB05, and ET05-CB05 (Table 1.7). In the three combinations, the null hypothesis for parameter equality is rejected only for age. Meanwhile, other combinations have significant differences in travel cost, an important parameter to estimate welfare effect, and the dissimilarity coefficient, which determines the nesting structure with a 99% significance level. The three cases—AT04, ET05, CB05—come from the same 2004 survey. Hence, from this point of view, the preferences are more closely similar, no matter what they are (whether they are RP or SP), if they are from the survey conducted during the same time.

In the second stage with water quality index, the pair of AT04-AT05 is the only combination with no difference in every parameter. However, when excluding the constant term, all the combinations have no significant difference in all parameters (Table 1.8). Similarly, in the case of the model with three attributes on water quality, when excluding the constant term, all the combinations have no significant difference in all parameters with a 95% significance level (Table 1.9). On the whole, if both stages are considered together, then two kinds of consistency can be supported. One is that ET05 and CB05 are convergent valid with each other; the other is that RP (AT04-AT05) is relatively consistent over time. Besides age, most of the parameter ratios are statistically indifferent between any two data sets with a 95% significance level in the parameter ratio comparison in Tables 1.10, 1.11, and 1.12. Especially, there is no difference in all the parameter ratios during the second stage, even with a 90% significance level. With respect to parameter ratio comparison, the combination of AT05 and ET05 shows the biggest difference, while the pair of AT04-AT05 do the least.

Next, the results of the test on joint parameter equality tell that the pair of AT04-AT05 have the most highly, equal joint parameters (Table 1.13). This shows that the respondents have chosen to trip during two consecutive years under mostly similar preferences in their minds from the jointly parametric point of view. Meanwhile, this also signifies that the stated preferences, including simple anticipation and contingent behavior expectations, are relatively inconsistent with revealed preferences. Especially, the two kinds of SP (simple expectation and intended behavior under a certain hypothetical condition) have jointly different parameters from corresponding actual data (posterior data) and even from actual data recorded for the same survey. Furthermore, there also exits differences in joint parameters between the two SP data sets (ET05, CB05). This implies that the respondents answered to different levels of contingent questions with different preference structures, even if the questions about stated preferences are asked in the same survey. Hence, the results of joint parameter equality test support time consistency in RP, especially in terms of the second stage estimation, while those of single parameter equality support convergent validity in the same survey.

Third, the estimated WTPs for improving water quality are not statistically different in all four cases. Tables 1.14 and 1.15 show the estimated WTP all four cases and the difference in the WTP between any two of the four cases. Table 1.14 tells that respondents are willing to pay from -$1.94 to $0.83 to improve water quality by one unit of the index in one of the lakes. However, all the WTPs are statistically insignificant. Interestingly, the null hypothesis that the WTPs are the same cannot be rejected in all the pairs, even though the cases have different preference structures and have different parameters for travel cost.[9] However, this WTP consists of one

[9] The scenario that secchi depth is improved by one meter is also applied. In this case, the WTP ranges from $3.22 to $4.38 for the improvement (Table 1.18). However, this chapter cannot find any significant difference in the WTP between any two of four cases (Table 1.19). It is the same result with the case of water quality index improvement.

parameter from the first stage (travel cost) and one parameter from the second stage (water quality index). And, the coefficient of water quality index is insignificant in every case. Hence, the results of welfare estimation for the second scenario are more meaningful in the sense that it is calculated only with an estimation during the first stage.

As for the second scenario, the welfare estimation for one lake shut down case shows quite different results. The CVs range from $0.09 to $0.12 in Table 1.16. Especially, Table 1.17 shows that all differences in the CV between AT04 and ET05 data sets, and between AT04 and CB05 data sets are significantly different from zero within the 95% significant level. In the sense of the second welfare analysis, RP data set (AT04) is inconsistent with both the two SP data (ET05, CB05) conducted at the same time. These results conflict with the convergent validity from the single parameter equality test, but support time consistency of RP from the same test.[10]

1.6. Conclusions

The convergent validity test of CB data is the issue many researchers study because the data usually contain the range beyond historically observed levels and the results are still not quite clear. Some studies support that the CB data model is convergent valid; others show convergent invalidity cases. The time consistency of individual preference structures is another issue in the sense that it is the basis of using multiple times' survey data as a panel. This study tests these two kinds of consistencies with various pairs of RP and SP data, both from the parametric point of view and from the view of social welfare. Among the results, inconsistency

[10] For the same scenario, it was also applied to Saylorville Lake, the most visited lake in Iowa. The CVs range from $0.14 to $0.22 (Table 1.20). Likewise, all the differences in the CV between AT04 and ET05 data sets, and between AT04 and CB05 data sets are significantly different from zero within a 99% significant level (Table 1.21). The results also conflict with the convergent validity from single parameter equality test and support time consistency between the two RP data sets.

between CB data and RP data that comes from the same survey, that is, structural discordance between AT04 and CB05 reconfirms the results by Jeon and Herriges (2010).

Results are summarized as follows:

(1) People show a consistent preference structure from trips actually taken in consecutive years Iowa lake survey data. The pair of the two RP data sets (AT04-AT05) shows the results support time consistency in every test. This shows that people actually choose to go on single-day trips in a coherent manner over years in contrast to convergent invalidity of CB data.

(2) In some tests, respondents show different recreation demands when they face different levels of contingent questions on intended trips, even in the same questionnaire. This is another convergent invalidity between two different SP data sets. But, in other test, they do not show. In addition, on the average, people are less likely to go on trips with a hypothetically higher water quality assumption than intended trips under the status quo.

(3) Convergent validity for either contingent behavior or simple stated preference fails many tests. We could obtain consistent WTP even though convergent invalidity in preference structures exists in terms of single parametric or joint parametric equality.

This study has some limitations in the data. First, AT05 data are not post-policy data, but simply data posterior to the 2004 survey. That is, it is not data collected after the water quality improvement policy. So, in this sense, AT04 - ET05 - AT05 data sets are a good corresponding data sequence. Therefore, the data sets are more suitable for comparing the differences between RP and SP (anticipation without any hypothetical condition) data. For a rich test of convergent validity, post-policy data corresponding to AT04 - CB05 data sequence are needed. Also, more refined data can provide a better estimation. For example, in this study, the average fuel cost for Midwestern states' average fuel efficiency, U.S. light duty vehicles are used for calculating travel cost. If the cost is refined to the individual level or county level, it should draw a more precise estimation.

References

Abidoye, B.O., J.A. Herriges, and J.L. Tobias. 2012. "Controlling for Observed and Unobserved Site Characteristics in RUM Models of Recreation Demand." *American Journal of Agricultural Economics* 94(5):1070-1093.

Adamowicz, W., J. Louviere, and M. Williams. 1994. "Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities." *Journal of Environmental Economics and Management* 26:271-292.

Adamowicz, W., J. Swait, P. Boxall, J. Louviere, and M. Williams. 1997. "Perceptions versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation." *Journal of Environmental Economics and Management* 32:65-84.

Azevedo, C.D., J.A. Herriges, and C.L. Kling. 2003. "Combining Revealed and Stated Preference: Consistency Tests and Their Interpretations." *American Journal of Agricultural Economics* 85:525-537.

Berry, S. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics* 25(2):242-262.

Berry, S., J. Levinsohn, and A. Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63(4):841-890.

Berry, S., O. Linton, and A. Pakes. 2004. "Limit Theorems for Estimating the Parameters of Differentiated Product Demand Systems." *Review of Economic Studies* 71(3):613-654.

Bhat, C., and J. Guo. 2004. "A Mixed Spatially Correlated Logit Model: Formulation and Application to Residential Choice Modeling." *Transportation Research Part B* 38:147-168.

Cameron, T., 1992. “Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods.” *Land Economics* 68(3):302-317.

Carson, R.T., N.E. Flores, K.M. Martin, and J.L. Wright. 1996. “Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods.” *Land Economics* 72(1):80-99.

Carson, R.T., and W.M. Hanemann. 2005. Contingent Valuation. in K.-G. MŠler and J.R. Vincent. eds. *Handbook of Environmental Economics*. Volume 2. Elsevier.

Egan, K.J., J.A. Herriges, C.L. Kling, and J.A. Downing. 2009. “Valuing Water Quality as a Function of Water Quality Measures.” *American Journal of Agricultural Economics*. 91(1):106-123.

Englin, J., T.A. Cameron. 1996. “Augmenting Travel Cost Models with Contingent Behavior Data: Poisson Regression Analyses with Individual Panel Data.” *Environmental and Resource Economics* 7:133-147.

Freeman, A.M., J.A. Herriges, and C.L. Kling. 2014. *The Measurement of Environmental and Resource Values: Theory and Methods*. 3rd Edition. RFF Press. Forthcoming.

Grijalva, C., R.P. Berrens, A.K. Bohara, and W.D. Shaw. 2002. “Testing the Validity of Contingent Behavior Trip Responses.” *American Journal of Agricultural Economics* 84:401-414.

Jeon, Y., and J.A. Herriges. 2010. “Convergent Validity of Contingent Behavior Responses in Models of Recreation Demand.” *Environmental and Resource Economics* 45:223-250.

Loomis, J.B., and R.B. Richardson. 2006. “An External Validity Test of Intended Behavior: Comparing Revealed Preference and Intended Visitation in Response to Climate Change.” *Journal of Environmental Planning and Management* 49(4):621-630.

McFadden, D., and K.E. Train. 2000. “Mixed MNL Models for Discrete Response.” *Journal of Applied Econometrics* 15(5):447-470.

Morey, E.R., R. Rowe, and M. Watson. 1993. “A Repeated Nested Logit Model of Atlantic Salmon Fishing.” *American Journal of Agricultural Economics* 75:578-592.

Murdock, J. 2006. “Handling Unobserved Site Characteristics in Random Utility Models of Recreation Demand.” *Journal of Environmental Economics and Management* 51:1-25.

Train, K. 2003. Discrete choice methods with simulation. Cambridge University Press.

von Haefen, R.H., and D.J. Phaneuf. 2008. "Identifying Demand Parameters in the Presence of Unobservables: A Combined Revealed and Stated Preference Approach." *Journal of Environmental Economics and Management* 56:19-32.

Whitehead, J.C., S.K. Pattanayak, G.L. Van Houtven, and B.R. Gelso. 2008. "Combining Revealed and Stated Preference Data to Estimate the Nonmarket Value of Ecological Services: An Assessment of the State of the Science." *Journal of Economic Surveys* 22(5):872-908.

Whitehead, J.C., D.J. Phaneuf, C.F. Dumas, J. Herstine, J. Hill, and B. Buerger. 2010. "Convergent Validity of Revealed and Stated Recreation Behavior with Quality Change: A Comparison of Multiple and Single Site Demands." *Environmental and Resource Economics* 45:91-112.

Table 1.1. Summary statistics of survey data (100 most visited lakes)

	Mean	SD	Maximum	Minimum
2004 actual single-day trips[a]	5.889	8.488	52	0
2005 actual single-day trips[a]	6.123	9.061	52	0
2005 expected single-day trips[a]	7.807	10.104	52	0
2005 CB single-day trips[a]	6.685	10.107	52	0
Age	53.448(54.430)[b]	14.901(15.068)[b]	82.5(82.5)[b]	8.6(8.6)[b]
Gender[c]	0.675(0.675)[b]	0.469(0.469)[b]	1(1)[b]	0(0)[b]
Education[d]	0.393(0.401)[b]	0.489(0.490)[b]	1(1)[b]	0(0)[b]
Household size	2.507(2.500)[b]	1.329(1.290)[b]	10(10)[b]	0(0)[b]
Log of lake size(acre)	2.248	0.703	4.279	1.114
Boat ramp dummy	0.92	0.273	1	0
Wake restrictions dummy	0.65	0.479	1	0
Handicap facilities dummy	0.45	0.5	1	0
State park dummy	0.45	0.5	1	0
Water quality index	6.710(7.370)[e]	1.305(0.562)[e]	9(9)[e]	3(7)[e]
Secchi depth	1.257(1.368)[e]	0.862(0.768)[e]	6.086(6.086)[e]	0.303(0.304)[e]
Total phosphorus	115.39(90.07)[e]	68.23(29.85)[e]	409.40(207.15)[e]	38.14(38.14)[e]
Total nitrogen	2.812(2.495)[e]	2.790(1.792)[e]	13.664(11.570)[e]	0.686(0.686)[e]
Travel cost	85.833 (85.162)[b]	58.091 (55.952)[b]	705.864 (585.555)[b]	0.316 (0.061)[b]

a: Trips for 100 lakes
b: The number in parentheses means value for 2005 year while the number outside of parenthesis means that for 2004.
c: Male equals to one and female, zero.
d: College graduation or higher level of education equals to one, otherwise zero.
e: For water quality index, the number in parentheses denotes hypothetical level in CB survey.

Table 1.2. First stage's estimated parameters with standard logit model and with repeated nested logit model (for 100 most visited lakes)

	Standard Logit				Repeated Nested Logit			
	AT04	AT05	ET05	CB05	AT04	AT05	ET05	CB05
Age	0.014***	0.012***	0.016***	0.021***	0.010***	0.008***	0.013***	0.018***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Gender	-0.707***	-0.739***	-0.743***	-0.706***	-0.453***	-0.539***	-0.480***	-0.449***
	(0.021)	(0.021)	(0.019)	(0.020)	(0.021)	(0.021)	(0.019)	(0.020)
Education	-0.201***	-0.141***	-0.215***	-0.286***	0.075***	0.076***	0.059***	-0.008
	(0.020)	(0.020)	(0.018)	(0.019)	(0.020)	(0.020)	(0.018)	(0.018)
Household Size	-0.118***	-0.102***	-0.126***	-0.107***	-0.058***	-0.056***	-0.063***	-0.042***
	(0.007)	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)	(0.007)	(0.007)
Travel Cost	-0.059***	-0.066***	-0.057***	-0.058***	-0.020***	-0.028***	-0.018***	-0.017***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
θ	N.A.				0.286***	0.375***	0.269***	0.264***
					(0.009)	(0.010)	(0.008)	(0.009)

Note: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

Table 1.3. Second stage's OLS estimation results with water quality index (for 100 most visited lakes)

	Standard Logit				Repeated Nested Logit			
	AT04	AT05	ET05	CB05	AT04	AT05	ET05	CB05
Constant(α)	-6.743***	-6.331***	-6.446***	-4.969***	-3.165***	-3.525***	-2.729***	-2.197***
	(0.370)	(0.354)	(0.339)	(0.739)	(0.114)	(0.141)	(0.100)	(0.207)
Log of Lake size	0.793***	0.743***	0.757***	0.752***	0.262***	0.322***	0.238***	0.231***
	(0.082)	(0.078)	(0.075)	(0.079)	(0.025)	(0.031)	(0.022)	(0.022)
Boat ramp	-0.079	-0.047	-0.108	-0.180	0.011	0.014	0.004	-0.021
	(0.203)	(0.195)	(0.186)	(0.195)	(0.063)	(0.078)	(0.055)	(0.055)
Wake Restrictions	0.447***	0.467***	0.372***	0.326***	0.155***	0.205***	0.124***	0.115***
	(0.116)	(0.111)	(0.106)	(0.110)	(0.036)	(0.044)	(0.031)	(0.031)
Handicap Facilities	0.169	0.233**	0.198*	0.173	0.021	0.054	0.023	0.015
	(0.115)	(0.110)	(0.105)	(0.110)	(0.035)	(0.044)	(0.031)	(0.031)
State park	0.311***	0.285***	0.344***	0.396***	0.089**	0.099**	0.090***	0.105***
	(0.115)	(0.110)	(0.106)	(0.111)	(0.035)	(0.044)	(0.031)	(0.031)
Water Quality index	0.051	0.035	0.076**	-0.105	0.007	0.005	0.014	-0.033
	(0.041)	(0.039)	(0.038)	(0.091)	(0.013)	(0.016)	(0.011)	(0.026)

Note: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

Table 1.4. Second stage's OLS estimation results with three attributes on water quality (for 100 most visited lakes)

	Standard Logit				Repeated Nested Logit			
	AT04	AT05	ET05	CB05	AT04	AT05	ET05	CB05
Constant(α)	-6.843***	-6.497***	-6.327***	-6.157***	-3.277***	-3.674***	-2.773***	-2.587***
	(0.283)	(0.275)	(0.259)	(0.356)	(0.086)	(0.108)	(0.075)	(0.098)
Log of Lake size	0.815***	0.764***	0.780***	0.760***	0.269***	0.331***	0.245***	0.233***
	(0.078)	(0.076)	(0.072)	(0.077)	(0.024)	(0.030)	(0.021)	(0.021)
Boat ramp	-0.039	-0.032	-0.058	-0.096	0.019	0.017	0.016	0.006
	(0.195)	(0.190)	(0.178)	(0.193)	(0.059)	(0.075)	(0.051)	(0.053)
Wake Restrictions	0.400***	0.416***	0.328***	0.317***	0.141***	0.184***	0.111***	0.113***
	(0.112)	(0.109)	(0.102)	(0.114)	(0.034)	(0.043)	(0.030)	(0.031)
Handicap Facilities	0.166	0.228**	0.198**	0.180*	0.02	0.051	0.023	0.017
	(0.109)	(0.107)	(0.100)	(0.109)	(0.033)	(0.042)	(0.029)	(0.030)
State park	0.285**	0.277**	0.310***	0.351***	0.083**	0.098**	0.082***	0.091***
	(0.111)	(0.108)	(0.102)	(0.111)	(0.034)	(0.043)	(0.029)	(0.030)
Secchi depth	0.244***	0.205***	0.233***	0.189**	0.081***	0.088***	0.076***	0.062***
	(0.071)	(0.069)	(0.064)	(0.082)	(0.021)	(0.027)	(0.019)	(0.023)
Total Phosphorus	0.001	0.001	0.000	0.001	0.000	0.001*	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.002)	(0.000)	(0.000)	(0.000)	(0.001)
Total Nitrogen	0.001	-0.006	0.004	0.006	-0.001	-0.004	0.000	0.006
	(0.018)	(0.018)	(0.017)	(0.028)	(0.006)	(0.007)	(0.005)	(0.008)

Note: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

Table 1.5. Repeated nested logit estimates in the first stage using bootstrapping (for 100 most visited lakes, B=500)

	AT04	AT05	ET05	CB05
Age	0.011***	0.009***	0.013***	0.018***
	(0.002)	(0.003)	(0.002)	(0.003)
Gender	-0.455***	-0.542***	-0.484***	-0.454***
	(0.077)	(0.080)	(0.074)	(0.087)
Education	0.074	0.075	0.060	-0.009
	(0.070)	(0.074)	(0.066)	(0.073)
Household size	-0.057**	-0.056**	-0.062***	-0.040
	(0.029)	(0.026)	(0.023)	(0.027)
Travel cost	-0.020***	-0.028***	-0.018***	-0.018***
	(0.003)	(0.003)	(0.002)	(0.002)
θ	0.288***	0.378***	0.271***	0.265***
	(0.033)	(0.034)	(0.030)	(0.034)

Note: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

Table 1.6. OLS estimates in the second stage using bootstrapping (for 100 most visited lakes, B=500)

	With water quality index				With three attributes on water quality			
	AT04	AT05	ET05	CB05	AT04	AT05	ET05	CB05
Constant(α)	-3.176*** (0.291)	-3.560*** (0.299)	-2.747*** (0.263)	-2.192*** (0.367)	-3.301*** (0.293)	-3.705*** (0.289)	-2.793*** (0.247)	-2.577*** (0.259)
Log of Lake size	0.268*** (0.044)	0.331*** (0.055)	0.243*** (0.038)	0.236*** (0.042)	0.277*** (0.044)	0.339*** (0.054)	0.251*** (0.037)	0.236*** (0.040)
Boat ramp	0.007 (0.093)	0.011 (0.126)	0.005 (0.083)	-0.025 (0.074)	0.016 (0.100)	0.013 (0.137)	0.020 (0.086)	0.001 (0.076)
Wake Restrictions	0.151*** (0.048)	0.207*** (0.062)	0.125*** (0.040)	0.109*** (0.042)	0.136*** (0.046)	0.185*** (0.062)	0.111*** (0.038)	0.112*** (0.043)
Handicap Facilities	0.024 (0.045)	0.060 (0.059)	0.026 (0.038)	0.015 (0.043)	0.023 (0.043)	0.061 (0.055)	0.027 (0.035)	0.020 (0.040)
State park	0.091** (0.043)	0.099* (0.055)	0.089** (0.038)	0.106*** (0.039)	0.085** (0.042)	0.097* (0.055)	0.080** (0.036)	0.089** (0.039)
Water Quality index	0.006 (0.015)	0.005 (0.020)	0.015 (0.014)	-0.034 (0.036)	N.A.			
Secchi depth	N.A.				0.083*** (0.030)	0.089** (0.040)	0.078*** (0.027)	0.061* (0.033)
Total Phosphorus	N.A.				0.000 (0.000)	0.001 (0.000)	0.000 (0.000)	0.000 (0.001)
Total Nitrogen	N.A.				-0.001 (0.008)	-0.005 (0.010)	0.000 (0.007)	0.006 (0.011)

Note: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

Table 1.7. Results of parameter equality test for the first stage estimation

	AT04-AT05	AT04-ET05	AT04-CB05	AT05-ET05	AT05-CB05	ET05-CB05
Age	0.002 (0.002)	-0.002** (0.001)	-0.008*** (0.002)	-0.004* (0.002)	-0.010*** (0.003)	-0.005*** (0.002)
Gender	0.086 (0.064)	0.029 (0.040)	-0.001 (0.061)	-0.058 (0.064)	-0.088 (0.080)	-0.030 (0.052)
Education	-0.001 (0.068)	0.015 (0.036)	0.083 (0.057)	0.016 (0.068)	0.084 (0.074)	0.068 (0.046)
Household Size	-0.002 (0.027)	0.005 (0.017)	-0.017 (0.022)	0.006 (0.024)	-0.016 (0.026)	-0.022 (0.017)
Travel Cost	0.008*** (0.003)	-0.002 (0.002)	-0.002 (0.002)	-0.010*** (0.003)	-0.010*** (0.003)	0.000 (0.002)
θ	-0.090*** (0.032)	0.017 (0.018)	0.023 (0.027)	0.107*** (0.032)	0.113*** (0.037)	0.006 (0.020)

Note: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

Table 1.8. Results of parameter equality test for the second stage estimation with water quality index

	AT04-AT05	AT04-ET05	AT04-CB05	AT05-ET05	AT05-CB05	ET05-CB05
Constant(α)	0.384 (0.336)	-0.429* (0.233)	-0.984*** (0.376)	-0.813** (0.319)	-1.368*** (0.407)	-0.555* (0.332)
Log of Lake size	-0.063 (0.062)	0.025 (0.043)	0.032 (0.046)	0.088 (0.058)	0.096 (0.061)	0.007 (0.043)
Boat ramp	-0.004 (0.141)	0.003 (0.109)	0.033 (0.107)	0.006 (0.142)	0.037 (0.145)	0.030 (0.098)
Wake Restrictions	-0.056 (0.073)	0.026 (0.053)	0.041 (0.057)	0.083 (0.067)	0.098 (0.070)	0.015 (0.051)
Handicap Facilities	-0.036 (0.073)	-0.002 (0.055)	0.009 (0.059)	0.034 (0.066)	0.046 (0.070)	0.011 (0.052)
State park	-0.008 (0.067)	0.002 (0.053)	-0.015 (0.053)	0.010 (0.060)	-0.006 (0.063)	-0.016 (0.048)
Water Quality index	0.001 (0.025)	-0.009 (0.019)	0.040 (0.038)	-0.009 (0.022)	0.039 (0.040)	0.048 (0.037)

Note: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

Table 1.9. Results of parameter equality test for the second stage estimation with three attributes on water quality

	AT04-AT05	AT04-ET05	AT04-CB05	AT05-ET05	AT05-CB05	ET05-CB05
Constant(α)	0.404 (0.317)	-0.508** (0.223)	-0.724*** (0.258)	-0.912*** (0.311)	-1.128*** (0.317)	-0.216 (0.215)
Log of Lake size	-0.062 (0.063)	0.026 (0.042)	0.041 (0.046)	0.088 (0.058)	0.103* (0.059)	0.015 (0.040)
Boat ramp	0.003 (0.156)	-0.004 (0.116)	0.015 (0.114)	-0.007 (0.155)	0.012 (0.157)	0.019 (0.104)
Wake Restrictions	-0.049 (0.072)	0.025 (0.050)	0.024 (0.057)	0.074 (0.067)	0.073 (0.071)	-0.001 (0.051)
Handicap Facilities	-0.038 (0.069)	-0.004 (0.051)	0.003 (0.056)	0.034 (0.060)	0.041 (0.065)	0.007 (0.047)
State park	-0.012 (0.066)	0.004 (0.051)	-0.005 (0.051)	0.016 (0.059)	0.007 (0.062)	-0.009 (0.047)
Secchi depth	-0.007 (0.046)	0.005 (0.037)	0.022 (0.042)	0.012 (0.046)	0.029 (0.050)	0.017 (0.040)
Total Phosphorus	0.000 (0.001)	0.000 (0.000)	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)	0.000 (0.001)
Total Nitrogen	0.004 (0.013)	-0.001 (0.010)	-0.007 (0.013)	-0.004 (0.012)	-0.011 (0.014)	-0.007 (0.012)

Note: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

Table 1.10. Results of parameter ratio comparison in the first stage estimation (ratio of given parameter to travel cost coefficient)

	AT04-AT05	AT04-ET05	AT04-CB05	AT05-ET05	AT05-CB05	ET05-CB05
Age	-0.237* (0.132)	0.180** (0.081)	0.507*** (0.157)	0.417*** (0.136)	0.744*** (0.196)	0.327** (0.142)
Gender	3.880 (3.990)	-4.069 (2.810)	-2.805 (4.312)	-7.949* (4.063)	-6.685 (5.132)	1.264 (3.547)
Education	-1.126 (3.271)	-0.410 (2.138)	-4.294 (3.299)	0.716 (3.383)	-3.168 (3.787)	-3.883 (2.778)
Household Size	0.888 (1.231)	-0.610 (0.809)	0.598 (1.170)	-1.498 (1.141)	-0.289 (1.363)	1.209 (0.981)

Note: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

Table 1.11. Results of parameter ratio comparison in the second stage estimation with water quality index (ratio of given parameter to travel cost coefficient)

	AT04-AT05	AT04-ET05	AT04-CB05	AT05-ET05	AT05-CB05	ET05-CB05
Log of Lake size	-1.755 (2.225)	0.050 (2.079)	-0.189 (2.112)	1.805 (2.062)	1.566 (2.224)	-0.239 (2.109)
Boat ramp	0.042 (5.811)	-0.110 (5.801)	-1.784 (5.632)	-0.152 (6.044)	-1.825 (6.042)	-1.674 (5.541)
Wake Restrictions	-0.272 (2.903)	-0.648 (2.691)	-1.437 (2.958)	-0.376 (2.677)	-1.166 (2.818)	-0.789 (2.823)
Handicap Facilities	0.947 (3.091)	0.228 (2.949)	-0.408 (3.207)	-0.720 (2.811)	-1.356 (3.136)	-0.636 (2.943)
State park	-1.072 (2.866)	0.371 (2.783)	1.384 (2.734)	1.443 (2.554)	2.456 (2.728)	1.013 (2.654)
Water Quality index	-0.118 (1.019)	0.512 (1.011)	-2.252 (2.123)	0.631 (0.958)	-2.134 (2.111)	-2.765 (2.106)

Note: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

Table 1.12. Results of parameter ratio comparison in the second stage estimation with three attributes on water quality (ratio of given parameter to travel cost coefficient)

	AT04-AT05	AT04-ET05	AT04-CB05	AT05-ET05	AT05-CB05	ET05-CB05
Log of Lake size	-1.924 (2.218)	0.055 (2.018)	-0.615 (1.995)	1.979 (2.030)	1.310 (2.056)	-0.669 (1.959)
Boat ramp	-0.345 (6.398)	0.280 (6.153)	-0.699 (6.019)	0.625 (6.522)	-0.354 (6.456)	-0.978 (5.829)
Wake Restrictions	-0.309 (2.840)	-0.656 (2.583)	-0.538 (2.937)	-0.347 (2.620)	-0.229 (2.924)	0.118 (2.837)
Handicap Facilities	1.015 (2.896)	0.322 (2.701)	-0.039 (3.049)	-0.694 (2.570)	-1.054 (2.938)	-0.360 (2.688)
State park	-0.827 (2.775)	0.207 (2.648)	0.788 (2.666)	1.034 (2.469)	1.615 (2.656)	0.581 (2.580)
Secchi depth	-0.991 (1.852)	0.173 (1.918)	-0.727 (2.234)	1.164 (1.923)	0.264 (2.247)	-0.900 (2.268)
Total Phosphorus	0.003 (0.023)	-0.008 (0.023)	-0.011 (0.048)	-0.011 (0.022)	-0.013 (0.047)	-0.003 (0.047)
Total Nitrogen	-0.125 (0.522)	0.029 (0.517)	0.406 (0.706)	0.155 (0.517)	0.532 (0.686)	0.377 (0.687)

Note: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

Table 1.13. Results of Wald test for the joint equality of parameters

	AT04-AT05	AT04-ET05	AT04-CB05	AT05-ET05	AT05-CB05	ET05-CB05
1st stage parameters	13.887 **	12.630 **	25.034 ***	28.451 ***	42.397 ***	17.889 ***
1st stage only ASCs	98.047 -	100.536 -	103.220 -	107.281 -	115.409 -	80.339 -
1st stage par. with ASCs	126.085 *	758.954 ***	194.598 ***	267.722 ***	166.713 ***	166.267 ***
(Second stage estimation with water quality index)						
2nd stage parameters	2.599 -	16.060 **	16.268 **	13.371 *	18.859 ***	3.998 -
2nd stage par. without	1.911 -	0.937 -	1.806 -	4.107 -	4.637 -	2.080 -
full stages parameters	16.090 -	158.370 ***	49.687 ***	78.538 ***	52.378 ***	46.021 ***
(Second stage estimation with three attributes on water quality)						
2nd stage parameters	2.765 -	16.266 *	15.024 *	13.609 -	18.168 **	2.416 -
2nd stage par. without	2.061 -	0.893 -	1.334 -	4.515 -	4.852 -	0.739 -
full stages parameters	16.730 -	168.314 ***	49.386 ***	83.548 ***	51.515 ***	46.178 ***

Note: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

Table 1.14. Results of Wald test for the joint equality of parameter ratios

	AT04-AT05	AT04-ET05	AT04-CB05	AT05-ET05	AT05-CB05	ET05-CB05
1st stage parameters	6.475 -	10.577 **	16.096 ***	17.030 ***	20.901 ***	10.645 **
1st stage only ASCs	155.519 ***	218.548 ***	196.152 ***	193.221 ***	187.582 ***	210.401 ***
1st stage par. with ASCs	163.405 ***	234.481 ***	212.113 ***	198.858 ***	197.056 ***	214.801 ***
(Second stage estimation with water quality index)						
2nd stage parameters	4.061 -	2.094 -	3.898 -	2.776 -	3.201 -	2.738 -
2nd stage par. without	0.945 -	0.510 -	1.668 -	1.674 -	3.093 -	2.225 -
full stages parameters	6.992 -	10.903 -	17.777 *	17.759 *	23.141 **	12.861 -
(Second stage estimation with three attributes on water quality)						
2nd stage parameters	4.794 -	2.210 -	2.904 -	4.341 -	2.010 -	1.204 -
2nd stage par. without	1.854 -	0.548 -	0.648 -	3.246 -	1.810 -	0.706 -
full stages parameters	7.950 -	10.857 -	16.991 -	19.390 *	22.490 **	11.932 -

Note: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

Table 1.15. Average additional WTP for improving water quality by 1 unit of index (unit: dollar)

	AT04	AT05	ET05	CB05
WTP	0.314 (0.778)	0.195 (0.720)	0.826 (0.762)	-1.939 (2.058)

Note: The numbers in parentheses are standard errors.

Table 1.16. WTP (for water quality) consistency test between any two of four cases

Difference b/t	Mean	Standard Error	Significant Level
AT04 vs AT05	0.118	(1.019)	-
AT04 vs ET05	-0.512	(1.011)	-
AT04 vs CB05	2.252	(2.123)	-
AT05 vs ET05	-0.631	(0.958)	-
AT05 vs CB05	2.134	(2.111)	-
ET05 vs CB05	2.765	(2.106)	-

Note: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

Table 1.17. Average CV for the case Big Creek Lake is no longer accessible (unit: dollar)

	AT04	AT05	ET05	CB05
WTP	0.090*** (0.014)	0.096*** (0.012)	0.122*** (0.018)	0.115*** (0.019)

Note 1: The numbers in parentheses are standard errors.
Note 2: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

Table 1.18. CV (for Big Creek Lake shut down case) consistency test between any two of four cases

Difference b/t	Mean	Standard Error	Significant Level
AT04 vs AT05	-0.006	(0.013)	-
AT04 vs ET05	-0.032	(0.009)	***
AT04 vs CB05	-0.025	(0.015)	**
AT05 vs ET05	-0.026	(0.016)	-
AT05 vs CB05	-0.019	(0.019)	-
ET05 vs CB05	0.007	(0.013)	-

Note: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

Table 1.19. Average additional WTP for improving secchi depth by 1 meter (unit: dollar)

	AT04	AT05	ET05	CB05
WTP	4.210***	3.219**	4.383***	3.483*
	(1.469)	(1.434)	(1.444)	(1.845)

Note 1: The numbers in parentheses are standard errors.
Note 2: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

Table 1.20. WTP (for secchi depth) consistency test between any two of four cases

Difference b/t	Mean	Standard Error	Significant Level
AT04 vs AT05	0.991	(1.852)	-
AT04 vs ET05	-0.173	(1.918)	-
AT04 vs CB05	0.727	(2.234)	-
AT05 vs ET05	-1.164	(1.923)	-
AT05 vs CB05	-0.264	(2.247)	-
ET05 vs CB05	0.900	(2.268)	-

Note: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

Table 1.21. Average CV for the case Saylorville Lake is no longer accessible (unit: dollar)

	AT04	AT05	ET05	CB05
WTP	0.141***	0.166***	0.182***	0.218***
	(0.021)	(0.019)	(0.026)	(0.034)

Note 1: The numbers in parentheses are standard errors.
Note 2: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

Table 1.22. CV (for Saylorville Lake shut down case) consistency test between any two of four cases

Difference b/t	Mean	Standard Error	Significant Level
AT04 vs AT05	-0.025	(0.019)	-
AT04 vs ET05	-0.041	(0.013)	***
AT04 vs CB05	-0.077	(0.024)	***
AT05 vs ET05	-0.016	(0.022)	-
AT05 vs CB05	-0.052	(0.030)	*
ET05 vs CB05	-0.036	(0.020)	*

Note: *, ** and *** are used to denote significance at the 10%, 5% and 1% levels, respectively.

CHAPTER 2. EVIDENCE OF UPDATING RISK PERCEPTION: 2008 GREAT FLOOD EFFECTS ON PROPERTY VALUES IN IOWA'S MAIN CITIES

2.1. Introduction

In 2008, one of the most severe flooding events in Iowa history, sometimes referred to as "Iowa's Katrina," occurred between June 8 and July 1. This event continued on the Upper Mississippi River in the southeastern area of Iowa for several days. The Iowa floods of 2008 caused the greatest damage to the Iowa River and Cedar River watersheds among any natural disasters in the state's recorded history. The most destructive damage from this catastrophe was reported in Cedar Rapids and Iowa City, Iowa. About ten square miles of the downtown area in Cedar Rapids were flooded, 14% of the city. The Cedar River crested at 31.12 feet, the highest level in Cedar Rapids' history, with the previous record 20 feet. In Iowa City, the arts campus of the University of Iowa was mostly destroyed. There were more than 6 million sandbags filled in Johnson County, more than were filled during Hurricane Katrina and breaking a national record.[11] The Des Moines basin was also affected by the floods. The Des Moines River crested at 31.57 feet on June 13th. Des Moines city officials issued voluntary and mandatory evacuations to the residents of around 270 homes on June 14th.[12] Hicks and Burton (2008) estimated agricultural losses from these floods at about $2.7 billion, revenue losses at more than $20 million, and anticipated total infrastructure damages at about $159 million for Iowa. Although the impact of the 1993 floods were greater for the nation as a whole, the floods of 2008 had a greater impact for Iowa. Based on estimated financial public assistance, the floods and tornadoes

---

[11] Higgins, Tim, and Mason Kerns. 2008. "Iowa City sending a gift: 250,000 extra sandbags." *Des Moines Register*, June 18, 2008.
[12] Des Moines is the capital and most populous city in the state of Iowa.

of 2008 in Iowa are regarded as the sixth largest Federal Emergency Management Agency (FEMA) disaster declaration.

This study basically begins with two questions. First, “Do housing prices reflect ex ante and/or ex post information on flood risk?” Second, “How can we extract the pure, perceived flood risk discount from the entire flood effects, including property stigmatization?” With this motivation, this study seeks to estimate the pure impact of the floods on property prices. In particular, we examine the variation in property value of certain areas before and after the floods to isolate the impact of updated perceived risk and/or increment of stigmatization that results from the floods.

There are a number of studies that deal with the effects of flooding on property values. However, most of these analyses rely on cross-sectional data and focus their attention on coastal flooding that results from hurricanes. Inland floods are rarely studied. Moreover, those studies that do employ inter-temporal data focus on the difference in prices between before and after a specific disaster. They pay little attention to spatial characteristics of the location. This research studies the effects of flooding by comparing property values not only before and after the flood, but by also taking into account various differences in risk exposure for the different floodplains. The analysis then goes one step further by investigating how realized flooding (i.e., whether a property was actually inundated by the 2008 flood) alters the changes patterns of housing prices over time. To capture these effects, this study uses a classical hedonic property price model with both difference-in-differences and triple differences techniques.

The remainder of this chapter consists of seven sections. The next section summarizes previous studies of how disasters impact residential property values. Section 2.3 develops a theoretical framework used to characterize how floods potentially impact housing prices. Section

2.4 describes the primary data sources, while Section 2.5 outlines the basic empirical model. The results are shown in Section 2.6. In Section 2.7, a parallel application for the Cedar Rapids area is introduced. Finally, summary and concluding remarks are provided in Section 2.8.

2.2. Literature Review

A number of studies investigate the effect of both natural and man-made disasters on residential property values. These disasters include the nuclear accident of 1979 at Three-Mile Island (Nelson 1981), the 1988 explosion of a chemical plant in Henderson, Nevada (Carroll *et al.* 1996), the 1999 pipeline explosion in Bellingham (Hansen *et al.* 2006), Loma Prieta earthquake of 1989 in the San Francisco Bay area (Beron *et al.* 1997), coastal flood hazards due to hurricanes (MacDonald *et al.* 1990; Bin and Polasky 2004; Hallstrom and Smith 2005), the inland floods of 1979 in Texas (Skantz and Strickland 1987) and 1993 on the Missouri and Mississippi rivers (Kousky 2010).

While many papers investigate the effects of flood risk on property values, by and large they rely on cross-sectional data. Only a few papers compare property values inside the floodplains with those outside the floodplains. Skantz and Strickland (1987) look into inland flood risks by using the 1979 floods in Texas. They do not find any significant drop in property prices after the floods. However, they do find a decline in prices stemming from an increase in insurance costs a year later. They use 133 properties, including 33 flooded homes. Bin and Polansky (2004) examine the effect of Hurricane Floyd by using difference in a difference (DID) approach. They find that the differential in the price between properties inside and outside the floodplain in Pitt County, North Carolina, increased after the hurricane. Hallstrom and Smith

(2005) and Carbone *et al.* (2006) employ a repeat-sales model, together with DID's, to estimate the new risk information conveyed by Hurricane Andrew to homeowners in a county directly hit (Dade County, Florida) versus a county that experienced a near-miss (Lee County, Florida). They find that the flood effect is noticeably consistent for both counties. After Hurricane Andrew, property prices in both counties declined by around 20%.

Kousky (2010) investigates whether a severe flood causes homeowners to update their assessment of flood risk for the property. For her analysis, Kousky targets the floods of 1993, one of the worst floods on record for the United States. She uses a hedonic property model to estimate the price differences between residential properties within and outside of the floodplain in St. Louis County, Missouri over the years 1979-2006. Then, Kousky captures the flood effect in the price differences by utilizing DID method. She finds before the flood, there was no significant discount for properties in the 500-year floodplains relative to properties outside the floodplain. However, properties in the 100-year floodplains were discounted 3.2 to 3.9%. For the period after the flood, Kousky finds property prices in the 100-year floodplains did not change significantly, while those in the 500-year floodplains decreased by between 2 and 5%. Under the National Flood Insurance Program (NFIP), home buyers are supposed to be informed of the flood risks before purchasing property within 100-year floodplains, but not the 500-year floodplains. In this sense, Kousky concludes that homeowners are likely to pay for a reduction in the probability of a disaster and that the risk information from NFIP actually influences homeowners.

This study begins with an approach similar to Kousky's (2010), but takes the analysis one step further by considering the role of actual flood inundation on changing home prices. This is the main difference from any existing studies, including Kousky's (2010). Even with the same flood, shifts in property values will be potentially sensitive to the realized flooding patterns, which can differ significantly from historical flood plain maps. Limited in access to actual damage data, this study uses geographic data on inundation as a proxy on direct damage. To estimate the effect, this research considers the actual, inundated area as the treatment group and non-inundated area as the corresponding control group. Then, by using differences techniques, we can extract the pure effects of inundation for each level of the flood risk area (that is, 100-year floodplains, 500-year floodplains, and non-flood plain areas).[13] This enables researchers to investigate how differently people update their risk assessment with direct, experienced events compared to the assessment with indirectly experienced events.

2.3. Theoretical Framework

This section draws and expands on the models developed in Palmquist (2005) and Kousky (2010). In this section, this study shows how information on flood risk can affect the price of properties using a hedonic property model. From the hedonic property models, the prices for property can be expressed as a function of the property's attributes, such as neighborhood characteristics, locational characteristics, structural characteristics, and an interesting environmental variable. In this study, as an environmental variable of interest, a consumer's subjectively assessed probability of flood is utilized.

Consumers maximize their expected utility subject to their budget constraints. Given r is the subjectively perceived probability for a flood, which represents individual assessment of disaster risk; z is a Hicksian composite good; X is property attributes vector; m is income; and $P(\cdot)$ is the

[13] Section 2.5 will provide more detail for this technique.

hedonic property price function, the basic consumer's property choice decision model can be summarized as [14]

$$\begin{aligned} &Max\, EU = rU_F\left(z_F, X\right) + (1-r)U_{NF}\left(z_{NF}, X\right) \\ &s.t.\ m = z + P(X, r) \end{aligned} \quad , \tag{11}$$

where subscripts *F* and *NF* denote the cases of flood and no flood, respectively. The numeraire *z* can be expressed as $z_{NF} = m - P$ when no flood occurs, $z_F = m - P - L$ in the case of flood, where *L* is the losses from the flood. From the implicit function theorem, the implicit price of risk can be expressed as such:

$$\frac{\partial P}{\partial r} = \frac{U_F - U_{NF}}{r\dfrac{\partial U_F}{\partial z_F} + (1-r)\dfrac{\partial U_{NF}}{\partial z_{NF}}} < 0 \quad . \tag{12}$$

From Eq. (12), it is shown that the property price is determined at the lower level as flood risk increases.

Now, we can utilize the same modeling structure to incorporate the role of information for flood risk. This information includes both statistical flood frequency of a residence (as revealed in historical floodplain maps) and an actual flood event. It can influence both the perceived probability of a flood event and the anticipated losses. For example, the perceived probability of flood risk and losses from a flood may be altered when a severe flood occurs. The direction of effect can be either negative or positive. If a residence is unexpectedly inundated from the point of view of ex ante information, then the new information will increase the perceived flood risk of the residence. Alternatively, if a residence is known as a high-risk area,

[14] This chapter assumes that consumers buy only one house to live in. This is one of basic assumptions in hedonic property models (See Palmquist 2005).

but is not inundated during an actual flood, this new information will change the consumer's perceived risk to the lower level.

Updating the subjective probability of disaster risk and losses from a disaster, according to the disaster risk information, can be expressed in the following manner. Let the subjective probability, *r*, and losses, *L*, be functions of the information (*I*). Then, the expected utility function is

$$EU = r(I)U_F(z_F, X) + [1 - r(I)]U_{NF}(z_{NF}, X) \quad , \tag{13}$$

where $z_F = m - P[X, r(I)] - L(I)$ and $z_{NF} = m - P[X, r(I)]$. In the same manner as the previous explanation, the partial derivative of the hedonic property price with respect to the information can be derived as

$$\frac{\partial P}{\partial I} = \frac{\frac{\partial r}{\partial I}(U_F - U_{NF}) - r\frac{\partial U_F}{\partial z_F} \cdot \frac{\partial L}{\partial I}}{r\frac{\partial U_F}{\partial z_F} + (1-r)\frac{\partial U_{NF}}{\partial z_{NF}}} \quad . \tag{14}$$

The sign for Eq. (14) relies on the signs of $\partial r/\partial I$ and $\partial L/\partial I$. Suppose that catastrophic flooding occurs and that a residence is unexpectedly inundated with flood losses greater than expected. Then, an individual's subjective probability of flood risk to the residence will increase after experiencing the flood, and predicted losses expected from a future flood will also increase. That is, $\partial r/\partial I > 0$ and $\partial L/\partial I > 0$. In this case, Eq. (14) will be negative. Alternatively, if the actual flood does not affect a residence with a high, historical risk of flooding and the actual damage is less than expected, then the perceived risk will likely lessen, as will anticipated losses with future floods. Hence, in this case, $\partial r/\partial I < 0$ and $\partial L/\partial I < 0$ and Eq. (14) will be positive. This shows that the hedonic property price can either increase or decrease after a flood event, according to the characteristics of the new flood information.

2.4. Data

This study uses data for single-family, detached residential properties sold between 2000 through 2012 in Des Moines, Iowa. The historical transaction data were obtained by a sales query from the Polk County assessor's website (http://web.assess.co.polk.ia.us/cgi-bin/web/tt/infoqry.cgi?tt =info/help/help-v2/sales_residential_dm). The website includes actual sales price, sales date, address, zip code, and property's structural attributes, such as land square feet, living area square feet, total number of bedrooms, number of baths, age of the house, number of fireplaces, garage dummy, and residence type dummy. Especially, for residence type, there are eleven types—'1 story', '1.5 story', '2 story', '2 story plus', '1 story unfinished attic', '1 story finished attic', 'split foyer', 'split level', 'partial construction', 'manufactured home', and 'other types'. In addition to structural characteristics, the data set also includes neighborhood characteristics, such as information on the associated school district. There are five school districts in the data set—Des Moines, Carlisle, Southeast Polk, Saydel, and Johnston. For sales price, the prices the Polk County assessor provides are in current dollars. Using the house price index, the sales prices were adjusted during the fourth quarter 2012 dollars.[15] Summary statistics for all explanatory variables and dependent variable are provided in Table 2.1.

[15] A house price index specific to Des Moines-West Des Moines Core Based Statistical Area (CBSA) was used here. This index was obtained from the Federal Housing Finance Agency website (http://www.fhfa.gov/Default.aspx? Page= 87).

The number of total transactions was 52,889. However, after cleaning the data, 1,091 transactions are dropped leaving a final sample size of 51,798.[16] The annual average number of transactions is 3,984. Even in 2008, the year when the floods occurred, the numbers of sales is 3,182.[17] So, the number of housing transactions seems sufficient to derive a statistically meaningful estimation of the hedonic price function. For a robustness check, this study uses three ways to split the entire period into pre-flood and post-flood periods: (1) Based on the beginning date of the flood, (2) dropping all observations during the flooding period, and (3) dropping all transactions that occurred within three months before and after the flooding period, i.e., transactions from March to September in 2008. The resulting sample sizes are 51,798, 51,185 and 49,711, respectively.

For the flood data, this study uses geographic information system (GIS) shape data on 100-year and 500-year floodplains, and actual inundation during the 2008 floods. The year shown in each floodplain denotes statistical frequency of flooding events. So, a 100-year floodplain indicates flooding occurs, on average, in the specific area one time during a 100-year period. This does not mean there will be no flooding for 100 years, since flooding does occur on occasion. Rather, this signifies that the area has one percentage chance experiencing a flood in a given year. In this sense, each floodplain implies flood risk. Thus, we can obtain each property's locational characteristics on flood risk by checking its type of floodplain location. That is, housing units are geo-located to determine whether or not they are in the 100- and 500-year floodplains, as well as the 2008 flood inundated property area. Using ArcGIS shapefiles makes it

[16] Among 1,091 removed observations, houses with no information on certain attributes are 997. Nine of these have inappropriate information on certain house attributes. For example, sales year is earlier than year built in some cases. The number of houses with zero value in such attributes as land square feet, number of baths, and total number of bedrooms is 81. Finally, four are mobile homes, whose type is quite different from that of general residential property occupied by a single family.

[17] Table 2.2 shows the number of transactions by year.

possible to geocode each property's address and zip code to identify each location. These locational attributes on flood risk are used as dummy variables in the model.

The Iowa Flood Center (IFC) (See Figure 2.1) provided all GIS data sets on floodplains and actual inundation. The 100- and 500-year floodplains are calculated by IFC. Their flood plain maps are quite close to, but not exactly the same as, the 100- and 500-year floodplains that FEMA provides. Most of previous studies on flood effects used FEMA's 100- and/or 500-year floodplains. Although the floodplains data that FEMA provides are generally considered as official data on flood risk, IFC also calculates floodplains in the same manner as FEMA. Moreover, IFC has updated the floodplains data more recently for Des Moines than FEMA. These two floodplains data are all proxy of true floodplains.

Finally, when identifying transactions inside 100- or 500-year floodplains, this study follows FEMA's suggestion to apply a 75-meter buffer to floodplain boundary because of the flood shapefile's scale (Kousky 2010). Additionally, a 30-meter buffer is applied to the actual inundated areas. As a robustness check in terms of actually flooded areas' boundaries, this study also estimates the hedonic function with various buffers – 75m, 30m, and 3m. Detailed numbers for the observations are summarized in Table 2.3.

2.5. Estimation

This study uses both difference-in-differences (DID) and triple differences (DDD) techniques to capture the impact of the floods. The basic treatment group consists of housing units in the 100- and 500-year floodplains, while the control group consists of houses outside both floodplains. This is the same approach used by Kousky (2010). However, taking the analysis one step further, this research also looks into the effects of actual inundation. For a given flood, the impact on any two home values can differ substantially, depending on the actual damage. We do not have available information on the actual monetary damage to a housing unit. Instead, we use geographic data on inundation as a proxy dummy variable for direct damage. Moreover, inundation can be interpreted as fulfillment of flood expectations. In other words, an inundation is an actual flooding event, while floodplains are the areas whose flood risk is calculated by researchers. Floodplains are a statistically, geographically and topographically, calculated concept to show the level of flood risk. On the other hand, an inundation is the realization of corresponding risk. Therefore, an inundated floodplain in an extraordinary flood can be considered as the fulfillment of risk expectation, i.e., the case where ex ante information turns out to be correct. On the other hand, properties in a floodplain, but not-inundated during a flood event, may suggest the need to update or adjust ex ante information.

In estimating this effect, the actual inundated area is considered the treatment group and the non-inundated area is the corresponding control group here. Specifically, properties are partitioned into three distinct subgroups: (1) those within the 100-year floodplain, (2) those within the 500-year floodplain, and (3) those properties outside the 500-year floodplain. We also distinguish whether the property was actually inundated during the 2008 floods. This enables us to investigate how property values (and their associated risk assessment) change with a directly experienced flood event. Because 100- and 500-year floodplains are statistical classifications with respect to flood frequency, they are likely to overlap with the area actually damaged by a flood, but complete overlap is not a requirement. Figure 2.2 shows an example of this case. So, even in the case where two people live in same floodplain area, if one person's house is inundated by a flood, while the other's house is not, then the effects of the flood could be different, as could be the ex post perceived risk. In addition, information on the 100-year floodplains must be provided to potential buyers, while information on 500-year floodplains are not provided. From this point of view, this study can provide an additional clue on how differently people update their subjective assessment of flood risks, according to not only whether they have ex ante information, but also whether their ex ante information is well-estimated.

To extract this type of effect, this study adopts a modified DID technique. The dummy for actual inundation is subordinate to the dummy for the flood because inundation happens only when a flood occurs. Also, to compare ex ante information on flood risk with ex post flooding, this research can also focus on the post-flood inundation effect for each level of floodplains and in non-flood plain areas rather than on both pre- and post-flood inundation effects.[18] The semi-log function used as a basic functional form here is

$$\begin{aligned} y_{it} = \ln\left(P_{it}\right) = \beta_0 + \sum_j \beta_{0j} Z_{ij} + \beta_{11} 100F_i + \beta_{12} 500F_i \\ + \alpha_1 D_{Pit} + \alpha_{11} D_{Pit}\left(1 - D_{Ai}\right) 100F_i + \alpha_{12} D_{Pit}\left(1 - D_{Ai}\right) 500F_i \\ + \delta_1 D_{Pit} D_{Ai}\left(1 - 100F_i\right)\left(1 - 500F_i\right) + \delta_{11} D_{Pit} D_{Ai} 100F_i + \delta_{12} D_{Pit} D_{Ai} 500F_i + \varepsilon_{it} \end{aligned}, \quad (15)$$

where i denotes each property, t indexes transaction date, P is price of the property, Z_j signifies the jth attribute of the property, and *100F* and *500F* are dummy variables that indicate whether the property is within 100- or 500-year floodplains, respectively. D_P denotes the post-flood dummy variable 1, if t is after the 2008 flood. D_A indexes the actual inundation dummy variable 1, if the location was inundated during the 2008 flood. Therefore, the overall constant, β_0,

[18] To investigate post-flood inundation effects on each type of floodplains, all areas are partitioned into six parts exclusively – (1) inundated 500-year floodplains, (2) non-inundated 500-year floodplains, (3) inundated 100-year floodplains, (4) non-inundated 100-year floodplains, (5) inundated non-flood plain areas, and (6) non-inundated non-flood plain areas. This partition reduces one difference between inundation and floodplain. Therefore, this approach is a difference-in-differences instead of triple differences.

represents the base group, that is, non-floodplain areas before the flood. The coefficients of D_P's (α_1) mean treatment effects of flood, while the coefficients of D_A's (δ_1, δ_{11}, δ_{12}) signify direct effects of flood by being inundated in each area—100-year floodplains, 500-year floodplains, and non-flood plain areas. On the other hand, α_{11} and α_{12} imply the flood effects in non-inundated 100- and 500-year floodplains. β_0, $\beta_0 + \beta_{11}$, and $\beta_0 + \beta_{12}$, respectively, reflect each locational effect for the three kinds of areas (non-flood plain areas, 100-, and 500-year floodplains) before the flood. After the flood, there are six kinds of areas—non-inundated/inundated non-flood plain areas, 100-, and 500-year floodplains. Their locational effect will be $\beta_0 + \alpha_1$, $\beta_0 + \alpha_1 + \delta_1$, $\beta_0 + \alpha_1 + \beta_{11} + \alpha_{11}$, $\beta_0 + \alpha_1 + \beta_{11} + \delta_{11}$, $\beta_0 + \alpha_1 + \beta_{12} + \alpha_{12}$, and $\beta_0 + \alpha_1 + \beta_{12} + \delta_{12}$ in this order. Hence, we can estimate the parameter of the pure flood effect on inundated 100-year floodplains, δ_{11}, by double differencing as below:

$$\hat{\delta}_{11} = \left(\bar{y}_{\mathrm{P,A\&100F}} - \bar{y}_{\mathrm{P,0}}\right) - \left(\bar{y}_{\mathrm{0,100F}} - \bar{y}_{\mathrm{0,0}}\right) \quad , \tag{16}$$

where the first subscript for $\bar{y}$ signifies the event dummy 'P', if transactions occurred after the flood otherwise zero, and the second subscript for $\bar{y}$ means the locational dummy – zero signifies the non-flood plain areas, '100F' means 100-year floodplains, and 'A&100F' signifies actual inundated 100-year floodplains.

Meanwhile, the DDD technique is utilized as a generalization of DID. The DID model assumes no locational effect of the inundated region before the flood. So, relaxing this restriction, the inundated areas' locational trait also needs consideration to extract the flood effect. A triple differences technique can examine inundated areas' geographical effects, which means that there could exist a geographical effect of the inundated region even before the flood. From the first difference, we can eliminate the group effect between floodplains and non-floodplains. Then, second difference will rule out the time effects (i.e., pre- and post-flood) for all groups. Finally, the third difference is the difference between inundation area and non-inundated area, which separates direct damage effects and indirect damage effects. General triple differences hedonic models have the following functional form:

$$\begin{aligned} y_{it} = \ln\left(P_{it}\right) = \beta_0 + \sum_j \beta_j Z_{ij} + \beta_{11} 100F_i + \beta_{12} 500F_i + \delta_1 D_{Ai} + \delta_{11} D_{Ai} 100F_i + \delta_{12} D_{Ai} 500F_i \\ + \alpha_1 D_{Pit} + \alpha_{11} D_{Pit} 100F_i + \alpha_{12} D_{Pit} 500F_i + \alpha_2 D_{Pit} D_{Ai} \\ + \alpha_{21} D_{Pit} D_{Ai} 100F_i + \alpha_{22} D_{Pit} D_{Ai} 500F_i + \varepsilon_{it} \end{aligned} , \quad (17)$$

where the meanings for all the variables are the same with the previous DID model. Then, the OLS estimate of α_{21}, the parameters on the triple interaction term that means the pure flood effect on inundated 100-year floodplains after the flood can be expressed as

$$\widehat{\alpha}_{21} = \left[\left(\overline{y}_{\mathrm{P,A,100F}} - \overline{y}_{\mathrm{1,A,0}}\right) - \left(\overline{y}_{\mathrm{P,0,100F}} - \overline{y}_{\mathrm{P,0,0}}\right)\right] - \left[\left(\overline{y}_{\mathrm{0,A,100F}} - \overline{y}_{\mathrm{0,A,0}}\right) - \left(\overline{y}_{\mathrm{0,0,100F}} - \overline{y}_{\mathrm{0,0,0}}\right)\right], \quad (18)$$

where the first subscript for $\overline{y}$ means the event dummy is 'P', if transactions occurred after the flood, otherwise zero. The second subscript for $\overline{y}$ means the locational dummy for inundation is 'A' is the actual inundated area, otherwise zero. Finally, the third subscript signifies the locational dummy for floodplains is '100F', if the area is located in 100-year floodplains, otherwise zero. Unlike the DID model, using the DDD model, one can also estimate the pure geographical effect of inundated areas, which exists regardless of the 2008 flood, i.e., δ_1, δ_{11}, and δ_{12}. This is the benefit of using a DDD model instead of using the previous DID model.

One critical issue on this method is how to distinguish the impact of financial crisis from that of the floods in the whole impact because the periods when both events occurred overlap each other. The bankruptcy of Lehman Brothers, considered the trigger of the 2008 financial crisis, occurred on September 15, while the floods of 2008 began around June 8 and ended about July 1, 2008. To solve this problem, this research assumes the effects of the financial crisis are locationally homogeneous in each city. Although the degree of the effects might be different, according to the characteristics of the associated locations, both the treatment group and control group experienced the financial crisis, that is, the event is a common effect. In this sense, if this effect is assumed homogeneous across locations, then it can be eliminated by the DID technique.

2.6. Results

This section will show the estimation of hedonic price function and test its robustness. The basic estimation is calculated on the basis of 51,798 observations without dropping observations during the flooding. The floodplains and the actual flooded areas are calculated by applying the 75-meter and 30-meter buffers, respectively. The results of estimation are summarized in Tables 2.4 and 2.5. Table 2.4 shows the results of the DID model and Table 2.5 provides the results for the DDD model. In the DID model, 100-year floodplains have a significantly negative effect before the flood at the 95% significance level. The price decreased by 15.7%. However, after the flood, if the areas were inundated, there is no significant effect, while there is a significantly positive effect, if they were inundated during the flood. Overall, the flood effect is not significant. Flood effects on the 500-year floodplains are negative, especially for the floodplains actually inundated. The price plummeted by 76.9%, which is extremely high rate.[19] Non-floodplains areas also have a negative effect on inundation, although the scale is not as much as the 500-year floodplains. The DDD model shows similar results. All parameters in common have the same signs with the DID model, although a couple of the significances are different. In addition, there are also locational effects of inundated areas. All locational effects

[19] This extraordinary decrease rate comes mostly from the lowest priced houses. See robust tests, and Tables 2.10 and 2.11.

are negative, except for the 100-year floodplains before the flood. However, they are insignificant, which supports the assumption of the DID model.

As shown in the tables, all parameters related to housing facilities have reasonable signs and are statistically significant at the 99% significance level. The sizes of lot, living area, and basement have positive effects on house prices, while house age has a negative effect. Garage and the numbers of bathrooms, bedrooms, and fireplaces are also positive factors in property values. For fixed effects of each year, there is an obviously declining pattern in housing prices starting in 2008, likely the result of the financial crisis that began in 2008.[20] For structure type dummies, the more stories, the higher the price.[21] Both partially constructed houses and manufactured homes have lower prices than general type houses.[22]

In summary, in terms of flood risk, the post-flood dummy is not significant and even positive, which means the overall effect of floods is not harmful to property values. However, focusing on floodplains and inundated areas, we find some interesting results. First, for 500-year floodplains, there was no significant difference before the flood. However, if the area was inundated during the flood, the price was discounted significantly after the flood. If not, the price was not discounted. One interpretation of this result is that people did not reflect flood risk in house price before the flood because the information on 500-year floodplains is not necessarily provided to potential purchasers, which means most people might not have the information. However, when people experienced the inundation directly, they realized the flood risk and this is reflected in the house price thereafter. Next, for 100-year floodplains, there was already a significant discount in the price before the flood. If the area was inundated in 2008, there is no

[20] The base year related with year dummies is 2000.
[21] Structure type dummy 1, 2, and 5 mean 1.5-story, 2-story, 2-story plus, respectively. The base structure type is 1-story.
[22] These types of houses are shown in Structure type dummies 7 and 8.

change in price after the flood. However, if the home was not inundated, there is critical rebounding in price after the flood. This can also be interpreted from the risk updating point of view. The information on 100-year floodplains is mandatorily open to house buyers. Therefore, the house market already reflects the risk in price regardless of actual flood. Hence, when floods in 2008 occurred, the inundated areas do not need further discounts. However, properties in the 100-year floodplains not inundated could be reassessed because they turned out less risky than the market expected. Finally, for non-flood plain regions, similar with 500-year floodplains, if they were inundated in the floods, then there is a significant price discount, but not as much as for the 500-year floodplains. However, if the properties were not inundated, then there is no change in price. The difference in price discount between 500-year floodplains and non-flood plain areas can be explained as the difference in flood risk between 500-year floodplains and the non-flood plain areas.

Robustness Tests

To check the robustness of the results shown in Tables 2.4 and 2.5, this study considered these alternative ways of

1. Dealing with the flood period itself;
2. Defining the inundation region; and
3. Handling outliers in terms of transaction prices.

For the flood period, two more cases were estimated—dropping transactions during the flooding (June 8 to July 1, 2008), dropping three months' transactions before and after the flooding (from March to September in 2008). Tables 2.6 and 2.7 provide the resulting parameter estimates. There are no substantive changes to the results from our handling of the flood period itself. All the cases show almost the same parameter values and significance.

For the buffer of inundated areas, two more cases were estimated– 75 and 3m buffers. The results are provided in Tables 2.8 and 2.9. The pre-flood effects are almost exactly same with one another, while the post-flood effects are not as well-matched as the pre-flood effects. Especially, the dummy for 500-year floodplains inundated in 2008 flood has a somewhat inconsistent sign. However, most of the values not well-matched are statistically insignificant, and all the significant parameters show consistent signs and similar values across the various buffers. Thus, the results also show robustness with respect to the definition of the inundation region.

For the cutoff test, this study uses the ratio price to the size of the living area (hereafter 'price ratio'). Observations with the lowest 10% price ratio are dropped from the sample. Next, this research estimates the hedonic model again using this new sample (sample size is 46,619) and compares the new results with original ones. Original results show a price decrease of 76.9% (in DID estimation) or 59.1% (in DDD estimation) in post-flood inundated 500-year floodplains, which may seem unusually large. Trimming the sample's outliers, the home prices in the same areas decrease by 11.6% (in DID estimation) or even increase slightly by 4.5% (in DDD estimation) (See Tables 2.10 and 2.11). Judging from this result, the extreme price discount mostly comes from the lowest priced transactions. However, even after removing the fairly low priced transactions, a significant price discount in pre-flood 100-year floodplains and price rebound in post-flood non-inundated 100-year floodplains are still supported by the test, although the size of the effects become smaller. As a result, both price discount in the floodplains before the flood and price rebound in non-inundated 100-year floodplains after the flood are robust with respect to trimming the data. Likewise, the test also confirms that unexpected inundation (i.e., inundation in the non-floodplain region) has a significantly negative effect.

2.7. Application to Cedar Rapids

This study also applies the same investigation techniques to Cedar Rapids, the second largest city in Iowa. Cedar Rapids is one of the most severely damaged cities during the floods of 2008. Housing transactions from 2004 to 2012 were utilized and the total number of observations is 17,790. The housing transaction data set for this application is purchased from DataQuick, while the geographical shape data on floodplains and inundation are obtained from the Iowa Flood Center. For comparison of the flood effects between Cedar Rapids and Des Moines, the flood effects in Des Moines were re-estimated with the data set exactly corresponding to that of Cedar Rapids, i.e., housing transactions from 2004 to 2012 with only house attributes in common – number of bathrooms, bedrooms, living area size, lot size, house age, basement size, garage dummy, and school district dummies.

The results are shown in Tables 2.12 and 2.13. For the structural characteristics, the two cities have the same signs. Only the coefficient of house age is negative and all the other coefficients of structural attributes are positive, which accords with common sense. Year dummies in Des Moines show an obvious pattern of significantly fallen in price from 2008 to 2011 by 31.5%. On the other hand, year dummies in Cedar Rapids show a significant decrease only in 2008 and 2011 by 2.7 and 6.4%, respectively. Especially for locational characteristics, Cedar Rapids shows distinctively different flood effects from Des Moines. The main features of flood effects in Cedar Rapids compared to those in Des Moines are as follows. First, while Des

Moines has a price discount in both floodplains, if the areas were inundated during the floods, Cedar Rapids shows a positive effect of inundation after the floods. Second, Des Moines shows a significant price rebound in the 100-year floodplains after the floods, if the areas were not inundated. In contrast, Cedar Rapids shows a significant price discount in the same floodplains under the same conditions. Third, price discount in Des Moines' floodplains before the flood is also supported by the results for Cedar Rapids. However, if restricted to non-inundated 100-year floodplains before the flood, the price discount is no longer applicable to Cedar Rapids.

To check for the robustness of the flood effect on house prices in Cedar Rapids, this research tests the results sensitivity in terms of three aspects—(1) flood period, (2) definition of the inundation region, and (3) price outliers, which are the same as the main estimation for Des Moines. The results are summarized in Tables 2.14 through 2.19. With different ways of dealing with flood periods, there are neither large nor significant changes in estimation. Likewise, even if the inundated areas are defined with different distances, most of the coefficients have no change in their signs, although some change their statistical significances. Additionally, a few coefficients change their signs notably. In DID estimation, if the buffer of the inundation region is extended to 75 from 30m (base buffer), then the coefficient of post-flood inundated 100-year floodplains changes from significantly negative to insignificantly positive. Similarly, in DDD estimation, both pre-flood inundated 100-year floodplains and post-flood 100-year floodplains effects change from significantly negative to insignificantly positive. Post-flood inundated 100-year floodplains effect changes from insignificantly positive to significantly negative. When the buffer for the inundation region is set as 75m, the results change to a similar direction with Des Moines. Finally, to manage the outliers in terms of prices, 1,755 observations with the lowest 10% price/living area ratio were dropped. In this case, the coefficient for post-flood non-inundated 100-year floodplains in DID changes from significantly negative to insignificantly positive, although most of the coefficients do not change remarkably. Even for the DDD estimation, the coefficient values for the post-flood non-inundated floodplains drop dramatically, although the signs are unchanged. In this sense, it can be inferred most of the negative effects in the post-flood non-inundated floodplains stem from the bottom-level, low priced houses.

2.8. Discussion and Conclusions

This research investigates how people update their perception of flood risk by tracking property prices. The estimation of the hedonic property price model can be mainly divided into two parts—pre-flood and post-flood effects. The pre-flood effect is associated with ex ante information. That is to say, this is the effect about whether people reflect their ex ante conceptual information into the housing market without any direct experience. House buyers are informed about only flood risks for the 100-year floodplains. The results show that only the coefficient of 100-year floodplain dummy has a significantly negative sign. This implies that people reflect their flood risks in market price, according to ex ante flood information.

The post-flood effect is utilized to examine how people update their flood risks, when their learning from experience accords with ex ante information or when it does not. Inundation dummies in the 500-year floodplains and in non-flood plain areas mean new flood risk information because people did not have sufficient information about flood risks in 500-year floodplains and non-flood plain areas before the floods. The results show that the coefficients for these dummies have a significantly negative sign. This can be interpreted that people update flood risk with their direct experiences to reflect the new information into prices. The interpretation of the inundation dummy in the 100-year floodplains is different from those in the 500-year floodplains and in non-flood plain. People already have flood risk information for 100-year floodplains before the floods. Therefore, this dummy means the fulfillment of their expectations in the real world. It is not surprising information in people's risk perceptions. The results show that the coefficient of the dummy is insignificant, which accords with the interpretations.

Finally, the non-inundation dummy in 100-year floodplains has an interesting significance. From ex ante flood information, people can conjecture the 100-year floodplains will be inundated first rather than any other areas, if a great flood occurs. However, the non-inundation dummy in 100-year floodplains means there are some places in 100-year floodplains not flooded, even when some regions other than 100-year floodplains are inundated. The results show that the coefficient of the dummy is significantly positive. This implies that people revise their existing flood information on the basis of their direct experiences to make the market price rebound. Moreover, there is one critical advantage in the coefficient of non-inundation dummies against that of inundation dummies. The inundation effect includes physical damage or direct stigmatization to properties, as well as an increment of perceptual flood risk. However, due to the limitation of available data, we cannot identify each subordinate effect. On the other hand, non-inundated areas do not have direct physical damages to houses from the floods. Therefore, we can extract change of perceptual flood risk separately. The coefficient of non-inundation dummies can be interpreted as pure change of perceptual flood risk without additional physical damage effect.

In contrast with the results of estimation for Des Moines, Cedar Rapids shows opposite results for some flood effects. Cedar Rapids shows a positive inundation effect and negative non-inundation effect from the 100-year floodplains after the flood. Also, it does not show any price discount for pre-flood non-inundated 100-year floodplains. As mentioned previously, some of the lowest priced houses can explain the negative post-flood non-inundation effect in 100-year floodplains. There are also some possible scenarios that interpret the results, although the available data set cannot demonstrate the justification for the interpretations with certainty. The scenery effect is one candidate for no discount effect for pre-flood non-inundated 100-year floodplains. If a house located in a non-inundated 100-year floodplains has a beautiful landscape, then the scenery has a positive effect on price, which can offset a flood risk discount. However, there is no way to distinguish the scenery effect and flood risk effect from the whole price change with the existing available data set.

Remodeling or reconstruction is a scenario for positive inundation effect in floodplains after the floods. Suppose that a house in the floodplains was inundated during the flood and totally damaged. After this incident, it was remodeled or reconstructed and then sold at a higher price. The remodeling effect can surpass the pure negative inundation effect.

An approach to the supply side is another candidate for the same effect. The 2008 floods in Cedar Rapids was a uniquely unexpected catastrophe and many houses were destroyed. In turn, the supply of houses significantly decreased in this area. Because of a decrease in the supply, the price was pressured to increase. However, this scenario is difficult to apply here. This study assumes that the entire city is one market. Therefore, if housing supply decreases, the house price should increase, not only in floodplains, but all over the entire city. Hence, this scenario cannot explain the increase in price only in the floodplains. Also, this scenario is difficult to verify using the hedonic price model because this scenario is about the supply side.

Apart from the application to Cedar Rapids, this study has some limitations in data. First and above all, as shown in Table 2.3, the number of transactions within the two kinds of floodplains is not sufficient. In some cases, the number of observations is less than ten. If there are sufficient numbers of observations in the floodplains, then the accuracy of estimation can be improved. Second, as mentioned previously, there were no available data on physical damage to properties from the flood, which is measured in terms of money. If available, then we can extract the physical damage effects from the coefficient of inundation dummies so the pure floor risk updating is estimable, even for inundated areas. Finally, as described in Section 2.4, the floodplain GIS shapefiles used here are produced by the Iowa Flood Center. They are not exactly the same with the floodplain map files from FEMA, although they are quite close to each other. Even though this research uses various levels of buffers to check the robustness, if the FEMA data that informs house buyers are used in this study, the estimation can be more precise. To develop this study further, we can apply other alternative methods, such as matching techniques. Also, in 1993, there was another great flood in Iowa. So, if the property transaction data before and after 1993 are available, then we can compare these two flood effects and understand them more deeply.

References

Beron, K.J., J.C. Murdoch, M.A. Thayer, and W.P.M. Vijverberg. 1997. “An Analysis of the Housing Market before and after the 1989 Loma Prieta Earthquake.” *Land Economics* 73(1):101-113.

Bin, O., and J.B. Kruse. 2006. “Real Estate Market Response to Coastal Flood Hazards.” *Natural Hazards Review* 7(4):137-144.

Bin, O., and C.E. Landry. 2013. “Changes in Implicit Flood Risk Premiums: Empirical Evidence from the Housing Market.” *Journal of Environmental Economics and Management* 65:361-376.

Bin, O., and S. Polasky. 2004. “Effects of Flood Hazards on Property Values: Evidence before and after Hurricane Floyd.” *Land Economics* 80(4):490-500.

Carbone, J.C., D.G. Hallstrom, and V.K. Smith. 2006. “Can Natural Experiments Measure Behavioral Responses to Environmental Risks?” *Environmental and Resource Economics* 33(3):273-297.

Carroll, T.M., T.M. Clauretie, J. Jensen, and M. Waddoups. 1996. “The Economic Impact of a Transient Hazard on Property Values: The 1988 PEPCON Explosion in Henderson, Nevada.” *Journal of Real Estate Finance and Economics* 13(2):143-167.

Dale, L., J.C. Murdoch, M.A. Thayer, and P.A. Waddell. 1999. “Do Property Values Rebounded from Environmental Stigmas? Evidence from Dallas.” *Land Economics* 75(2):311-326.

Hallstrom, D.G., and K. Smith. 2005. “Market Responses to Hurricanes.” *Journal of Environmental Economics and Management* 50:541-561.

Hansen, J.L., E.D. Benson, and D.A. Hagen. 2006. “Environmental Hazards and Residential Property Values: Evidence from a Major Pipeline Event.” *Land Economics* 82(4):529-541.

Hicks, M.J., and M.L. Burton. 2008. “Preliminary Flood Damage Estimates for Iowa: Great Flood of 2008.” http://cms.bsu.edu/-/media/WWW/DepartmentalContent/Miller CollegeofBusiness/BBR/ Publications/disasterStudies/iowaFloodEstimates08.pdf.

McCluskey, J.J., and G.C. Rausser. 2003. “Stigmatized Asset Value: Is It Temporary or Long-term?” *The Review of Economics and Statistics* 85(2):276-285.

Kousky, C. 2010. “Learning from Extreme Events: Risk Perceptions after the Flood. *Land Economics* 86 (3): 395-422.

McKenzie, Russell, and John Levendis. 2008. Flood Hazards and Urban Housing Markets: The Effects of Katrina on New Orleans.” *Journal of Real Estate Finance and Economics* DOI 10.1007/s11146-009-9141-3.

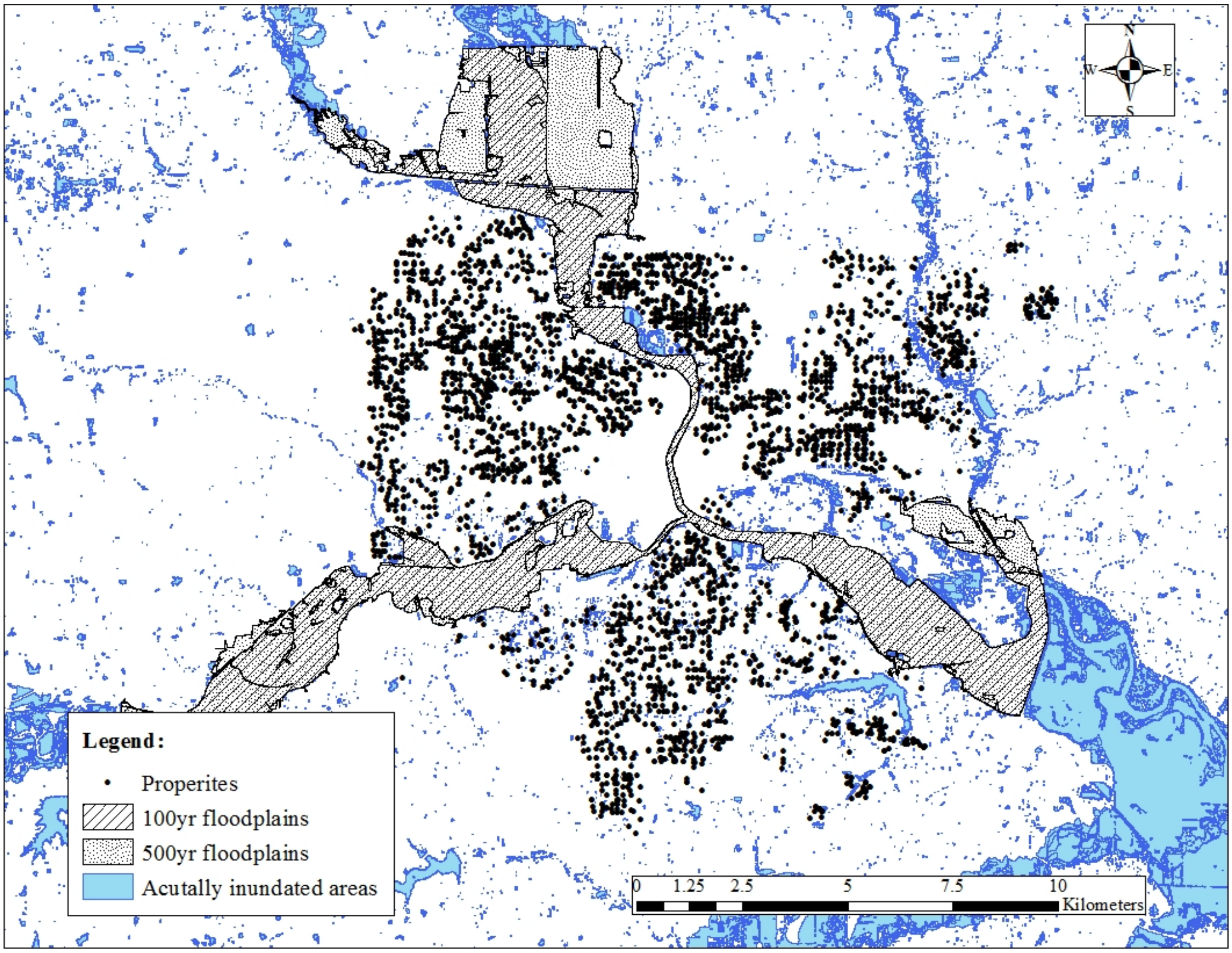
Nelson, J.P. 1981. “Three Mile Island and Residential Property Values: Empirical Analysis and Policy Implications.” *Land Economics* 57(3):363-372.

Palmquist, R.B. 2005. Property Value Models. *Handbook of Environmental Economics*. Volume 2 Chapter 16. Amsterdam:Elsevier.

Skantz, T.R., and T.H. Strickland. 1987. "House Prices and a Flood Event: An Empirical Investigation of Market Efficiency." *Journal of Real Estate Research* 2(2):75-83.

Wooldridge, J.M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd Edition. The MIT Press.

Figure 2.1. Geographical location of floodplains, flooding areas and observations in Des Moines, Iowa



Note: Floodplains and inundated areas are overlapped with each other in the order of the legend. So, if they are overlapped, only higher ordered floodplains will show up in the map.

Figure 2.2. Example of a house which is located in 100-year floodplains but was not actually inundated in 2008 flood (southwestern part of Des Moines near the Raccoon river)

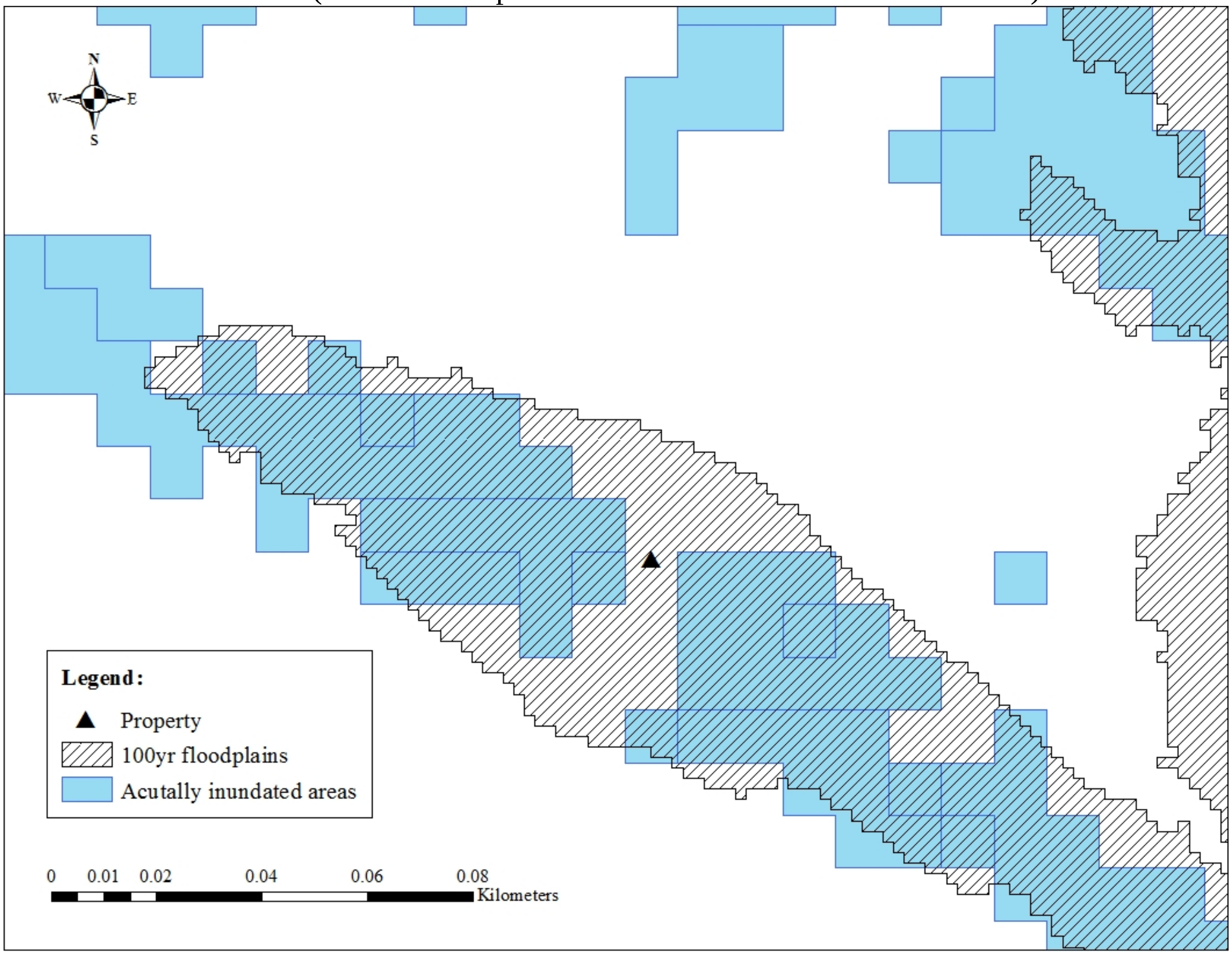


Table 2.1. Summary statistics (Total sample size: 51,802)

	Variables	Mean	Max	Min	S.D.	# of 1s
	Log of Price	11.38	15.21	6.87	0.67	n.a.
Dummies	Post-Flood	0.27	1	0	0.44	13,947
	Inundated Areas (Inun.)	0.03	1	0	0.18	1,751
	500yr Floodplains (500yr)	0.002	1	0	0.05	106
	100yr Floodplains (100yr)	0.002	1	0	0.05	120
	500yr w/o Inun.	0.002	1	0	0.04	100
	100yr w/o Inun.	0.001	1	0	0.03	59
	Non-flood plains w. Inun.	0.03	1	0	0.18	1,684
	500yr w. Inun.	0.0001	1	0	0.01	6
	100yr w. Inun.	0.001	1	0	0.03	61
	2001yr	0.08	1	0	0.28	4,388
	2002yr	0.09	1	0	0.28	4,524
	2003yr	0.09	1	0	0.29	4,638
	2004yr	0.09	1	0	0.28	4,600
	2005yr	0.10	1	0	0.29	4,923
	2006yr	0.09	1	0	0.29	4,901
	2007yr	0.08	1	0	0.27	4,096
	2008yr	0.06	1	0	0.24	3,182
	2009yr	0.07	1	0	0.25	3,369
	2010yr	0.05	1	0	0.23	2,837
	2011yr	0.05	1	0	0.22	2,688
	2012yr	0.06	1	0	0.24	3,274
	School District 1	0.01	1	0	0.11	622
	School District 2	0.01	1	0	0.12	743
	School District 3	0.0002	1	0	0.02	12
	School District 4	0.01	1	0	0.09	430
	Garage	0.77	1	0	0.42	39,970
	Building Type 1	0.08	1	0	0.27	3,951
	Building Type 2	0.13	1	0	0.33	6,518
	Building Type 3	0.03	1	0	0.16	1,371
	Building Type 4	0.04	1	0	0.19	1,862
	Building Type 5	0.03	1	0	0.17	1,541
	Building Type 6	0.04	1	0	0.20	2,158
	Building Type 7	0.001	1	0	0.03	50
	Building Type 8	0.0006	1	0	0.02	29
	Building Type 9	0.12	1	0	0.32	5,962
	Building Type 10	0.00006	1	0	0.01	3
	Land Sf.	10,855.92	3,092,847	1,584	29,928.83	n.a.
	House Age	63.41	159	1	30.16	n.a.
	Living Area Sf.	1,169.49	7,290	288	469.96	n.a.
	Total bathrooms	1.37	7.5	0.5	0.59	n.a.
	Total bedrooms	2.69	9	1	0.78	n.a.
	Total Fireplaces	0.29	4	0	0.51	n.a.
	Basement Sf.	788.97	4,421	0	352.03	n.a.

Table 2.2. Number of transaction by year

Year	# of Transactions	Year	# of Transactions
2000	4,378	2007	4,096
2001	4,388	2008	3,182
2002	4,524	2009	3,369
2003	4,638	2010	2,837
2004	4,600	2011	2,688
2005	4,923	2012	3,274
2006	4,901	Average	3,984.5

Table 2.3. Number of observations for actually inundated area with various buffers

	75m	30m	3m
500-year floodplain without inundation	38	100	105
100-year floodplain without inundation	5	59	109
Non-flood plain area with inundation	10244	1684	161
500-year floodplain with inundation	68	6	1
100-year floodplain with inundation	115	61	11

Table 2.4. Results of estimation with DID technique

Variable	Estimate	Std. Err.	|Estimate|/ S.E.
Constant	**10.787**	**0.014**	**748.50**
500-year floodplains	-0.071	0.053	1.36
100-year floodplains	**-0.157**	**0.050**	**3.16**
Post-Flood (PF)	0.021	0.017	1.22
PF*500-year floodplain without inundation	-0.036	0.112	0.32
PF*100-year floodplain without inundation	**0.236**	**0.119**	**1.98**
PF*non-flood plain area with inundation	**-0.065**	**0.021**	**3.05**
PF*500-year floodplain with inundation	**-0.769**	**0.338**	**2.27**
PF*100-year floodplain with inundation	0.0004	0.157	0.00
Year 2001 dummy	0.006	0.010	0.59
Year 2002 dummy	0.005	0.010	0.47
Year 2003 dummy	**0.038**	**0.010**	**3.80**
Year 2004 dummy	**0.020**	**0.010**	**1.99**
Year 2005 dummy	0.009	0.010	0.91
Year 2006 dummy	**0.042**	**0.010**	**4.31**
Year 2007 dummy	**0.041**	**0.010**	**3.95**
Year 2008 dummy	**-0.031**	**0.015**	**2.11**
Year 2009 dummy	**-0.105**	**0.020**	**5.22**
Year 2010 dummy	**-0.157**	**0.020**	**7.73**
Year 2011 dummy	**-0.273**	**0.020**	**13.35**
Year 2012 dummy	**-0.189**	**0.020**	**9.39**
School dummy 1	**-0.124**	**0.020**	**6.15**
School dummy 2	**-0.111**	**0.019**	**5.99**
School dummy 3	-0.009	0.136	0.07
School dummy 4	-0.027	0.023	1.14
Garage dummy	**0.290**	**0.005**	**56.30**
Structure type dummy 1	**0.033**	**0.009**	**3.60**
Structure type dummy 2	**0.036**	**0.009**	**4.10**
Structure type dummy 3	**0.036**	**0.013**	**2.65**
Structure type dummy 4	**0.040**	**0.012**	**3.36**
Structure type dummy 5	**0.121**	**0.015**	**7.87**
Structure type dummy 6	**0.041**	**0.011**	**3.84**
Structure type dummy 7	**-0.523**	**0.067**	**7.78**
Structure type dummy 8	**-0.312**	**0.088**	**3.53**
Structure type dummy 9	**0.058**	**0.007**	**7.85**
Structure type dummy 10	**2.388**	**0.273**	**8.75**
Lot size Sf./10^6	**1.895**	**0.070**	**27.08**
House age	**-0.007**	**0.0001**	**77.09**
Living area Sf./10^4	**3.680**	**0.090**	**40.81**
Number of bathrooms	**0.040**	**0.005**	**7.29**
Number of bedrooms	**0.013**	**0.004**	**3.44**
Number of Fireplaces	**0.139**	**0.005**	**27.23**
Basement Sf./10^4	**3.221**	**0.077**	**41.82**

Note: Bold numbers are significantly different from zero with 95% confidence level.

Table 2.5. Results of estimation with DDD technique

Variable	Estimate	Std. Err.	|Estimate|/ S.E.
Constant	**10.787**	**0.014**	**748.35**
100-year floodplains (100yr)	**-0.204**	**0.075**	**2.73**
500-year floodplains (500yr)	-0.067	0.054	1.25
Inundated areas (Inun.)	-0.018	0.014	1.26
100yr*Inun.	0.102	0.101	1.01
500yr*Inun.	-0.076	0.243	0.31
Post-Flood (PF)	0.020	0.017	1.18
PF*100yr	**0.284**	**0.132**	**2.16**
PF*500yr	-0.040	0.112	0.36
PF*Inun.	-0.048	0.026	1.87
PF*100yr*Inun.	-0.272	0.211	1.29
PF*500yr*Inun.	-0.591	0.425	1.39
Year 2001 dummy	0.006	0.010	0.58
Year 2002 dummy	0.005	0.010	0.46
Year 2003 dummy	**0.038**	**0.010**	**3.79**
Year 2004 dummy	**0.020**	**0.010**	**1.98**
Year 2005 dummy	0.009	0.010	0.91
Year 2006 dummy	**0.042**	**0.010**	**4.32**
Year 2007 dummy	**0.041**	**0.010**	**3.95**
Year 2008 dummy	**-0.031**	**0.015**	**2.10**
Year 2009 dummy	**-0.105**	**0.020**	**5.21**
Year 2010 dummy	**-0.157**	**0.020**	**7.72**
Year 2011 dummy	**-0.273**	**0.020**	**13.34**
Year 2012 dummy	**-0.189**	**0.020**	**9.38**
School dummy 1	**-0.125**	**0.020**	**6.18**
School dummy 2	**-0.112**	**0.019**	**6.02**
School dummy 3	-0.010	0.136	0.07
School dummy 4	-0.027	0.023	1.16
Garage dummy	**0.290**	**0.005**	**56.27**
Structure type dummy 1	**0.033**	**0.009**	**3.56**
Structure type dummy 2	**0.036**	**0.009**	**4.06**
Structure type dummy 3	**0.036**	**0.013**	**2.66**
Structure type dummy 4	**0.040**	**0.012**	**3.35**
Structure type dummy 5	**0.121**	**0.015**	**7.83**
Structure type dummy 6	**0.041**	**0.011**	**3.82**
Structure type dummy 7	**-0.524**	**0.067**	**7.79**
Structure type dummy 8	**-0.313**	**0.088**	**3.54**
Structure type dummy 9	**0.058**	**0.007**	**7.81**
Structure type dummy 10	**2.405**	**0.273**	**8.80**
Lot size Sf./10^6	**1.897**	**0.070**	**27.10**
House age	**-0.007**	**0.000**	**77.09**
Living area Sf./10^4	**3.685**	**0.090**	**40.81**
Number of bathrooms	**0.040**	**0.005**	**7.29**
Number of bedrooms	**0.013**	**0.004**	**3.43**
Number of Fireplaces	**0.139**	**0.005**	**27.25**
Basement Sf./10^4	**3.220**	**0.077**	**41.79**

Note: Bold numbers are significantly different from zero with 95% confidence level.

Table 2.6. Robustness test for the treatment of flood period (DID)

Variable	No dropping		Drop flood period		Drop 3 months	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
500-year floodplains	-0.071	0.053	-0.072	0.053	-0.081	0.054
100-year floodplains	**-0.157**	0.050	**-0.157**	0.050	**-0.158**	0.050
Post-Flood (PF)	0.021	0.017	0.009	0.019	0.015	0.029
PF*500-year floodplain without inundation	-0.036	0.112	-0.062	0.114	-0.051	0.114
PF*100-year floodplain without inundation	**0.236**	0.119	**0.238**	0.119	0.239	0.122
PF*non-flood plain area with inundation	**-0.065**	0.021	**-0.069**	0.022	**-0.066**	0.023
PF*500-year floodplain with inundation	**-0.769**	0.338	**-0.769**	0.338	**-0.758**	0.338
PF*100-year floodplain with inundation	0.0004	0.157	0.005	0.158	-0.052	0.174

Note: Bold numbers are significantly different from zero with 95% confidence level.

Table 2.7. Robustness test for the treatment of flood period (DDD)

Variable	No dropping		Drop flood period		Drop 3 months	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
100-year floodplains (100yr)	**-0.204**	0.075	**-0.205**	0.075	**-0.209**	0.076
500-year floodplains (500yr)	-0.067	0.054	-0.068	0.054	-0.077	0.055
Inundated areas (Inun.)	-0.018	0.014	-0.018	0.014	-0.019	0.014
100yr*Inun.	0.102	0.101	0.102	0.101	0.108	0.102
500yr*Inun.	-0.076	0.243	-0.076	0.243	-0.063	0.243
Post-Flood (PF)	0.020	0.017	0.009	0.019	0.015	0.029
PF*100yr	**0.284**	0.132	**0.285**	0.132	**0.290**	0.135
PF*500yr	-0.040	0.112	-0.066	0.114	-0.054	0.115
PF*Inun.	-0.048	0.026	**-0.051**	0.026	-0.047	0.027
PF*100yr*Inun.	-0.272	0.211	-0.265	0.212	-0.333	0.226
PF*500yr*Inun.	-0.591	0.425	-0.562	0.426	-0.577	0.426

Note: Bold numbers are significantly different from zero with 95% confidence level.

Table 2.8. Robustness test for area buffer (DID)

Variable	75m buffer		30m buffer (Base)		3m buffer	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
500-year floodplains	-0.072	0.053	-0.071	0.053	-0.071	0.053
100-year floodplains	**-0.157**	0.050	**-0.157**	0.050	**-0.156**	0.050
Post-Flood (PF)	0.028	0.017	0.021	0.017	0.019	0.017
PF*500-year floodplain without inundation	0.075	0.166	-0.036	0.112	-0.093	0.108
PF*100-year floodplain without inundation	0.292	0.338	**0.236**	**0.119**	0.148	0.105
PF*non-flood plain area with inundation	**-0.047**	**0.010**	**-0.065**	**0.021**	**-0.137**	**0.065**
PF*500-year floodplain with inundation	-0.201	0.129	**-0.769**	**0.338**	0.235	0.277
PF*100-year floodplain with inundation	0.137	0.104	0.000	0.157		

Note 1: Bold numbers are significantly different from zero with 95% confidence level.
Note 2: For 3m-buffer, the number of observations for “PF*500-year floodplain with inundation” is zero. So, for these two buffers, I merge both 500-year, and 100-year floodplains into floodplains.

Table 2.9. Robustness test for area buffer (DDD)

Variable	75m buffer		30m buffer (Base)		3m buffer	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
100-year floodplains (100yr)	-0.426	0.273	**-0.204**	**0.075**	**-0.159**	**0.052**
500-year floodplains (500yr)	-0.099	0.088	-0.067	0.054	-0.067	0.053
Inundated areas (Inun.)	**-0.019**	**0.006**	-0.018	0.014	-0.013	0.046
100yr*Inun.	0.293	0.278	0.102	0.101	0.036	0.181
500yr*Inun.	0.056	0.110	-0.076	0.243	-0.359	0.478
Post-Flood (PF)	0.024	0.017	0.020	0.017	0.019	0.017
PF*100yr	0.559	0.431	**0.284**	**0.132**	0.150	0.106
PF*500yr	0.102	0.180	-0.040	0.112	-0.097	0.108
PF*Inun.	**-0.029**	**0.012**	-0.048	0.026	-0.124	0.079
PF*100yr*Inun.	-0.398	0.444	-0.272	0.211	0.188	0.346
PF*500yr*Inun.	-0.284	0.225	-0.591	0.425		

Note 1: Bold numbers are significantly different from zero with 95% confidence level.
Note 2: For 3m-buffer, the number of observations for "PF*500-year floodplain with inundation" is zero. So, for these two buffers, I merge both 500-year, and 100-year floodplains into floodplains.

Table 2.10. Robustness test for price ratio cutoff (DID)

Variable	Estimate	Std. Err.	|Estimate| / Std. Err.	Estimate (Base)
500-year floodplains	-0.014	0.033	0.430	-0.071
100-year floodplains	**-0.089**	**0.031**	**2.829**	**-0.157**
Post-Flood (PF)	0.013	0.011	1.178	0.021
PF*500-year floodplain without inundation	0.025	0.071	0.346	-0.036
PF*100-year floodplain without inundation	0.067	0.075	0.898	**0.236**
PF*non-flood plain area with inundation	**-0.037**	**0.014**	**2.598**	**-0.065**
PF*500-year floodplain with inundation	-0.116	0.289	0.403	**-0.769**
PF*100-year floodplain with inundation	0.127	0.106	1.191	0.000

Note: Bold numbers are significantly different from zero with 95% confidence level.

Table 2.11. Robustness test for price ratio cutoff (DDD)

Variable	Estimate	Std. Err.	|Estimate| / Std. Err.	Estimate (Base)
100-year floodplains (100yr)	-0.056	0.047	1.193	**-0.204**
500-year floodplains (500yr)	-0.008	0.034	0.245	-0.067
Inundated areas (Inun.)	-0.001	0.009	0.132	-0.018
100yr*Inun.	-0.059	0.064	0.934	0.102
500yr*Inun.	-0.149	0.169	0.877	-0.076
Post-Flood (PF)	0.013	0.011	1.178	0.020
PF*100yr	0.034	0.082	0.413	**0.284**
PF*500yr	0.019	0.071	0.261	-0.040
PF*Inun.	**-0.036**	**0.017**	**2.145**	-0.048
PF*100yr*Inun.	0.156	0.138	1.130	-0.272
PF*500yr*Inun.	0.045	0.339	0.133	-0.591

Note: Bold numbers are significantly different from zero with 95% confidence level.

Table 2.12. Estimation of Cedar Rapids' hedonic price model (DID)

Variable	Estimate	Std. Err.	|Estimate| / S.E.	Estimate (Des Moines)
Constant	**11.139**	**0.021**	**538.934**	**10.699**
500-year floodplains	**-0.062**	**0.018**	**-3.522**	-0.047
100-year floodplains	**-0.080**	**0.021**	**-3.786**	**-0.204**
Post-Flood (PF)	-0.001	0.017	-0.068	0.024
PF*500-year floodplain without inundation	**-0.153**	**0.025**	**-6.194**	-0.030
PF*100-year floodplain without inundation	**-0.130**	**0.045**	**-2.880**	**0.279**
PF*non-flood plain area with inundation	0.007	0.017	0.429	**-0.066**
PF*500-year floodplain with inundation	-0.060	0.036	-1.654	**-0.801**
PF*100-year floodplain with inundation	**-0.161**	**0.025**	**-6.462**	0.088
Year 2005 dummy	-0.013	0.014	-0.906	-0.013
Year 2006 dummy	-0.026	0.014	-1.835	0.019
Year 2007 dummy	-0.027	0.014	-1.918	0.017
Year 2008 dummy	**-0.054**	**0.018**	**-3.027**	**-0.057**
Year 2009 dummy	**-0.046**	**0.022**	**-2.065**	**-0.134**
Year 2010 dummy	0.019	0.022	0.866	**-0.185**
Year 2011 dummy	-0.044	0.022	-1.955	**-0.298**
Year 2012 dummy	0.026	0.022	1.137	**-0.217**
Garage dummy	**0.147**	**0.008**	**19.496**	**0.322**
Lot size Sf./10^6	0.285	0.349	0.815	**1.956**
House age	**-0.006**	**0.000**	**-45.731**	**-0.007**
Living area Sf./10^4	**3.025**	**0.074**	**40.612**	**4.681**
Number of bathrooms	**0.038**	**0.006**	**6.619**	**0.061**
Number of bedrooms	**0.037**	**0.004**	**8.942**	**0.010**
Basement Sf./10^4	**2.290**	**0.084**	**27.288**	**3.314**
School dummy 1	0.0003	0.012	0.024	**-0.132**
School dummy 2	-0.020	0.012	-1.770	**-0.125**
School dummy 3		n.a.		-0.062
School dummy 4				0.012

Note: Bold numbers are significantly different from zero with 95% confidence level.

Table 2.13. Estimation of Cedar Rapids' hedonic price model (DDD)

Variable	Estimate	Std. Err.	|Estimate| / S.E.	Estimate (Des Moines)
Constant	**11.139**	**0.021**	**538.964**	**10.699**
500-year floodplains (500yr)	**-0.054**	**0.020**	**-2.680**	-0.046
100-year floodplains (100yr)	0.042	0.052	0.798	**-0.309**
Inundated areas (Inun.)	0.003	0.016	0.204	-0.017
500yr*Inun.	-0.036	0.041	-0.869	-0.028
100yr*Inun.	**-0.148**	**0.059**	**-2.513**	0.245
Post-Flood (PF)	-0.001	0.017	-0.079	0.023
PF*500yr	**-0.162**	**0.027**	**-6.088**	-0.032
PF*100yr	**-0.252**	**0.066**	**-3.841**	**0.385**
PF*Inun.	0.004	0.023	0.175	-0.049
PF*500yr*Inun.	**0.121**	**0.057**	**2.111**	-0.677
PF*100yr*Inun.	0.108	0.074	1.467	-0.370
Year 2005 dummy	-0.013	0.014	-0.918	-0.013
Year 2006 dummy	-0.026	0.014	-1.854	0.019
Year 2007 dummy	-0.027	0.014	-1.914	0.017
Year 2008 dummy	**-0.054**	**0.018**	**-3.008**	**-0.056**
Year 2009 dummy	**-0.046**	**0.022**	**-2.049**	**-0.133**
Year 2010 dummy	0.020	0.022	0.878	**-0.184**
Year 2011 dummy	-0.044	0.022	-1.942	**-0.298**
Year 2012 dummy	0.026	0.022	1.151	**-0.217**
Garage dummy	**0.147**	**0.008**	**19.488**	**0.322**
Lot size Sf./10^6	0.284	0.349	0.813	**1.956**
House age	**-0.006**	**0.000**	**-45.718**	**-0.007**
Living area Sf./10^4	**3.023**	**0.075**	**40.492**	**4.683**
Number of bathrooms	**0.037**	**0.006**	**6.547**	**0.061**
Number of bedrooms	**0.037**	**0.004**	**8.970**	**0.010**
Basement Sf./10^4	**2.288**	**0.084**	**27.250**	**3.315**
School dummy 1	0.0006	0.012	0.055	**-0.134**
School dummy 2	-0.020	0.012	-1.763	**-0.125**
School dummy 3		n.a.		-0.062
School dummy 4				0.012

Note: Bold numbers are significantly different from zero with 95% confidence level.

Table 2.14. Robustness test for the treatment of flood period in Cedar Rapids (DID)

Variable	No dropping		Drop flood period		Drop 3 months	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
500-year floodplains	**-0.062**	0.018	**-0.062**	0.018	**-0.064**	0.018
100-year floodplains	**-0.080**	0.021	**-0.081**	0.021	**-0.074**	0.022
Post-Flood (PF)	-0.001	0.017	-0.007	0.018	-0.001	0.031
PF*500-year floodplain without inundation	**-0.153**	0.025	**-0.155**	0.025	**-0.138**	0.025
PF*100-year floodplain without inundation	**-0.130**	0.045	**-0.129**	0.045	**-0.125**	0.046
PF*non-flood plain area with inundation	0.007	0.017	0.007	0.017	0.009	0.018
PF*500-year floodplain with inundation	-0.060	0.036	-0.060	0.037	-0.042	0.037
PF*100-year floodplain with inundation	**-0.161**	0.025	**-0.161**	0.025	**-0.137**	0.025

Note: Bold numbers are significantly different from zero with 95% confidence level.

Table 2.15. Robustness test for the treatment of flood period in Cedar Rapids (DDD)

Variable	No dropping		Drop flood period		Drop 3 months	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
500-year floodplains (500yr)	**-0.054**	0.020	**-0.054**	0.020	**-0.055**	0.021
100-year floodplains (100yr)	0.042	0.052	0.042	0.052	0.036	0.052
Inundated areas (Inun.)	0.003	0.016	0.004	0.017	0.009	0.017
500yr*Inun.	-0.036	0.041	-0.036	0.041	-0.043	0.042
100yr*Inun.	**-0.148**	0.059	**-0.148**	0.059	**-0.140**	0.059
Post-Flood (PF)	-0.001	0.017	-0.007	0.018	-0.001	0.031
PF*500yr	**-0.162**	0.027	**-0.163**	0.027	**-0.148**	0.027
PF*100yr	**-0.252**	0.066	**-0.252**	0.066	**-0.236**	0.066
PF*Inun.	0.004	0.023	0.004	0.024	0.001	0.024
PF*500yr*Inun.	**0.121**	0.057	**0.123**	0.057	**0.129**	0.058
PF*100yr*Inun.	0.108	0.074	0.109	0.074	0.118	0.075

Note: Bold numbers are significantly different from zero with 95% confidence level.

Table 2.16. Robustness test for area buffer in Cedar Rapids (DID)

Variable	75m	30m (Base)	3m
500-year floodplains	**-0.062**	**-0.062**	**-0.062**
100-year floodplains	**-0.081**	**-0.080**	**-0.080**
Post-Flood (PF)	-0.002	-0.001	0.000
PF*500-year floodplain without inundation	**-0.195**	**-0.153**	**-0.138**
PF*100-year floodplain without inundation	0.254	**-0.130**	**-0.236**
PF*non-flood plain area with inundation	0.005	0.007	0.071
PF*500-year floodplain with inundation	**-0.102**	-0.060	0.078
PF*100-year floodplain with inundation	**-0.160**	**-0.161**	**-0.099**

Note: Bold numbers are significantly different from zero with 95% confidence level.

Table 2.17. Robustness test for area buffer in Cedar Rapids (DDD)

Variable	75m	30m (Base)	3m
500-year floodplains (500yr)	-0.015	**-0.054**	**-0.063**
100-year floodplains (100yr)	-0.144	0.042	**-0.082**
Inundated areas (Inun.)	**0.021**	0.003	0.034
500yr*Inun.	**-0.080**	-0.036	0.019
100yr*Inun.	0.049	**-0.148**	-0.030
Post-Flood (PF)	0.004	-0.001	0.000
PF*500yr	**-0.242**	**-0.162**	**-0.136**
PF*100yr	0.317	**-0.252**	**-0.234**
PF*Inun.	-0.015	0.004	0.037
PF*500yr*Inun.	**0.168**	**0.121**	0.126
PF*100yr*Inun.	**-0.468**	0.108	0.095

Note: Bold numbers are significantly different from zero with 95% confidence level.

Table 2.18. Robustness test for price ratio cutoff in Cedar Rapids (DID)

Variable	Estimate	t-value	Estimate (Base)
500-year floodplains	**-0.072**	-6.630	**-0.062**
100-year floodplains	**-0.035**	-2.720	**-0.080**
Post-Flood (PF)	0.023	2.300	-0.001
PF*500-year floodplain without inundation	**-0.001**	-0.040	**-0.153**
PF*100-year floodplain without inundation	0.026	0.940	**-0.130**
PF*non-flood plain area with inundation	0.013	1.330	0.007
PF*500-year floodplain with inundation	**0.020**	0.880	-0.060
PF*100-year floodplain with inundation	**-0.003**	-0.180	**-0.161**

Note: Bold numbers are significantly different from zero with 95% confidence level.

Table 2.19. Robustness test for price ratio cutoff in Cedar Rapids (DDD)

Variable	Estimate	t-value	Estimate (Base)
500-year floodplains (500yr)	**-0.066**	-5.410	**-0.054**
100-year floodplains (100yr)	**0.077**	2.370	0.042
Inundated areas (Inun.)	**0.026**	2.850	0.003
500yr*Inun.	-0.047	-1.820	-0.036
100yr*Inun.	**-0.157**	-4.340	**-0.148**
Post-Flood (PF)	**0.024**	2.500	-0.001
PF*500yr	-0.007	-0.440	**-0.162**
PF*100yr	**-0.086**	-2.110	**-0.252**
PF*Inun.	-0.013	-1.010	0.004
PF*500yr*Inun.	0.054	1.520	**0.121**
PF*100yr*Inun.	**0.114**	2.500	0.108

Note: Bold numbers are significantly different from zero with 95% confidence level.

CHAPTER 3. IS CHOICE BEHAVIOR IN RECREATION DEMAND HABIT-FORMING OR VARIETY-SEEKING?

3.1. Introduction

There is extensive literature studying habit formation and variety seeking in the context of labor markets and food demand. However, state-dependent behavior has rarely been investigated in modeling recreation demand.[23] Yet, individuals may choose to enjoy a particular recreation site repeatedly on the basis of their experiences, instead of trying to find new sites. Once they obtain information regarding a given recreation site—such as its scenery, congestion, and facilities—through an initial visit, they may make use of this information to decide if the site is their favorite place and visit it repeatedly. Alternatively, some individuals may want to enjoy a variety of recreational sites. In this case, people may choose a different site for each choice occasion. Understanding the extent of either type of behavior is important from a policy perspective because it impacts both the estimated value of existing sites and the ability of policy-makers to induce changes in behavior. For example, if habit formation is strong, it may be difficult to induce individuals to change their visitation patterns in response to an environmental policy. This study examines whether there is state dependence in recreation demand and, if so, what form it takes. If there is positive state dependence, then recreational trips can be considered a habit-forming good. On the other hand, recreational trips can also be thought of as a variety-seeking good, if there is a negative state dependence.

Methodologically, the major concern when dealing with state dependence models is the so-called "initial conditions problem." The initial conditions problem arises because any lagged

[23] Hereafter, the word 'state dependence' is used as a concept that includes both habit-forming property and variety-seeking property.

dependent variable included in a model to capture state dependence is likely correlated with unobserved individual characteristics also impacting current decisions; i.e., the lagged dependent variables are likely to be endogenous. Treating these lagged variables as exogenous variables can lead to significant bias and a mis-characterization of state dependence.

The objective of this study is to examine the role of state dependence in recreation demand using panel data from the Iowa Lakes Project. For a solution of initial conditions problems, this research employs the two dominant approaches found in the literature. One is Heckman's (1981b) approximation and the other is Wooldridge's (2005) solution. This study begins with a single site case and extends the analysis to a multiple site setting. For the single site case, a dynamic random effect (RE) logit model is used. In the multiple site setting, a RE two-step nesting structure model is used, capturing state dependence in terms of overall trip taking, although not in terms of the specific sites selected. For both the single and multiple site cases, a RE Poisson model is also estimated as an alternative approach to compare the results and as a robustness check. Finally, a Monte Carlo simulation exercise is also used to show the biases that can arise either from neglecting state dependence entirely or from treating it incorrectly.

The remainder of this study is divided into six additional sections. Section 3.2 provides a brief review of the literature. Section 3.3 describes the basic initial conditions problem and the solutions proposed by Heckman (1981b) and Wooldridge (2005). The specific models employed in the analysis are provided in Section 3.4, followed by a description of the Iowa Lakes data used in the empirical analysis in Section 3.5. Section 3.6 presents results, while Section 3.7 concludes the investigation.

3.2. Literature Review

The concept of state dependence traces to Pollak (1970). He explains that consumption of a good is habit-forming, if present preferences depend on past consumption. Thus, habit formation means that past consumption strengthens the tendency to consume the same good over time. On the other hand, if past consumption affects current and future consumptions oppositely, then the consumption of that good is variety-seeking.

The field, which has most actively studied state dependence, is food or drug consumption. These studies have been conducted with various models and data sets. The most commonly used models are dynamic versions of the linear expenditure system, the translog model, the almost ideal demand system, and recently—discrete choice models. Most of the empirical studies in food consumption find evidence of habit formation. However, some of the study results do not support habit formation (Alessie and Kapteyn 1991; Meghir and Weber 1996; Dynan 2000) or only obtained mixed results (Pollak and Wales 1969; Edgerton *et al.* 1996; Holt and Goodwin 1997).

Representatively, Thunström (2010) estimates the strength of state dependence associated with breakfast cereal consumption and its heterogeneity across households by using a mixed multinomial logit model with a detailed micro-level dataset. She shows that breakfast cereal consumption is generally highly habitual. Thunström also finds that some households can be characterized as variety-seeking, that the strength of habit persistence is similar across income and educational groups. It seems to be weaker for households with several adults and children compared with one-adult households.

State dependence has also been actively studied in the labor market literature. Heckman (1981a) points out the importance of distinguishing true state dependence by taking some studies on the labor market as examples. He also suggests his approximation to solve the initial condition problem in the extension of labor market analysis (Heckman 1981b).[24] Recently, Oguzoglu (2010) examines whether a work limiting disability influences employment decisions by using the Household, Income, and Labor Dynamics in Australia (HILDA) Survey and finds strong state dependence in employment choices. He also shows that the full effect of a disability shock may be more severe than its initial impact on current employment outcomes, especially for low skilled individuals, because of the persistent nature of employment behavior by using simulations.

Haan (2010) estimates an inter-temporal discrete choice model of female labor supply to capture the effects of state dependence on the labor supply behavior in terms of both labor market participation (extensive margin) and working hours (intensive margin) on the basis of panel data from the German Socio Economic Panel (SOEP). His results show that state dependence has a significantly positive impact on labor market participation and a lower, but, in general, still significant impact on working hours. Haan also used his model to investigate the short- and long-run labor supply effects of a fundamental reform of the German income tax system and show that the labor supply responses are significantly higher in the long-run than in the short run.

Around the same time, Prowse (2012) also examines a similar topic to Haan's, again supporting the notion of state dependence in labor market decisions. He investigates state dependencies in women's labor supply behaviors using a dynamic multinomial mixed logit

[24] His study will be introduced in more detail in Section 3.3.

model with the British Household Panel Survey (BHPS). Prowse shows that significant state dependence is present in both full-time and part-time employment, and that state dependencies are overestimated, if persistent unobservables are ignored, and underestimated, if an overly restrictive form of persistence is imposed.

On the other hand, for recreation demand, only a small number of studies on habit formation have been conducted. Adamowicz (1994) seems the first who investigates whether there is any state dependence—either habit-forming or variety-seeking patterns—in recreation demand. He models recreational site choice decisions that contain previous experiences with the site as a characteristic or attribute. Adamowicz uses a rational dynamic model as well as one static model and naïve models. Based on the comparison of dynamic and static models, Adamowicz suggests dynamic elements influence recreational choice. He also shows that either a naïve or rational model with previous consumption as an attribute improves over the static model and that this effect can be a significant factor in welfare analysis.

Although Haab (2003) does not use the word 'habit formation' or 'variety-seeking pattern,' he also studies state dependence in recreation demand with the concept of temporal dependence. Haab shows a single-site demand model that allows for temporal correlation between choice occasions, but can be estimated on seasonal summaries of behavior data. He uses five models, such as Poisson, zero inflated Poisson, negative binomial, first-order binary Markov, and zero inflated binary Markov with the concept of first-order Markov chain to strongly support the temporally correlated model.

Swait *et al.* (2004) study the effect of temporal dependence on welfare measures. They take note of the fact that static, cross-sectional discrete choice models will be biased if the underlying preference includes temporal dependence. They apply the discrete choice model that includes consideration of prior behaviors and past attribute perceptions to the case of recreational fishing site choice and participation, comparing with the static version. They find that their time-series model provides a richer behavioral characterization of site choice and that significant differences exist between cross-section and time-series welfare measures.

Smith (2005) uses a mixed logit model combined with a state dependence parameterization to distinguish state dependence from heterogeneity in repeated decisions. He studies fishing location choices of commercial sea urchin divers in California and finds that true state dependence is an important determinant of location choice. Smith avoids initial value problems by using the data whose periods start with the very beginning year when divers in northern California started to enter the region to catch sea urchins. Using this type of data, he can assume the initial values to be exogenous.

3.3. Econometric Methodology

For the estimation, this study uses the maximum likelihood estimation (MLE) technique. This section will cover the MLE process this research follows and also explain how to deal with initial condition problems by introducing Heckman (1981b) and Wooldridge (2005). We begin with an introduction of the initial condition problem with the binary choice model. Then, both Heckman and Wooldridge's methods will be introduced.

3.3.1. Initial Condition Problem

Consider a typical, repeated logit structure in the recreational demand model for a single site. An individual's contribution to the likelihood function is given by:

$$L = f\left(N_0, N_1, \cdots, N_Y \mid Z, TC\right) \quad , \tag{19}$$

where N_y is number of trips in year, y (for $y = 0, \cdots, Y$), Z is individual characteristics, and TC is travel cost. On the other hand, the probability function is

$$f\left(N_1, N_2, \cdots, N_Y \mid N_0, Z, TC, \alpha\right) = \prod_{y=1}^{Y} f_y\left(N_y \mid N_{y-1}, Z, TC, \alpha\right) = \prod_{y=1}^{Y}\prod_{t=1}^{T}\left[e^{V_G}\Big/\left(1+e^{V_G}\right)\right]^{I_{ty}} \quad , \tag{20}$$

where t is each occasion whose total number is T in a year, I_{ty} is an index function one, if the individual chooses to go on a trip on occasion t in year y. V_G is a representative utility for going on a trip and the year subscript is equal to zero when the year is the initial year.[25] For the representative utility for going on a trip, specify it as $V_G = \delta + \beta N_{-1} + \tau TC + \alpha$. α is an unobservable individual random effect. The main issue to be solved is how to extract Eq. (19) by using Eq. (20).

When a model includes a lagged dependent variable as one of explanatory variables, the initial condition problem becomes essential because the lagged variable is fundamentally endogenous, which means it depends on the unobserved heterogeneity—individual random effect term (α) in this study. If researchers deal with N_0 as exogenous variables, then $f_0\left(N_0 \mid Z, TC, \alpha\right)$ can be set as 1. In this case, Eq. (20) can be used directly to build the likelihood function for year, y, as:

[25] The specification is explained in detail in Section 3.4.1.

$$
\begin{aligned}
L &= f\left(N_0, N_1, \cdots, N_{\mathrm{Y}} \mid Z, TC\right) \\
&= \int f\left(N_0, N_1, \cdots, N_{\mathrm{Y}} \mid Z, TC, \alpha\right) g(\alpha)\, d\alpha \\
&= \int f_0\left(N_0 \mid Z, TC, \alpha\right) \prod_{y=1}^{Y} f_y\left(N_y \mid N_{y-1}, Z, TC, \alpha\right) g(\alpha)\, d\alpha \quad . \\
&= \int \prod_{y=1}^{Y} f_y\left(N_y \mid N_{y-1}, Z, TC, \alpha\right) g(\alpha)\, d\alpha \\
&= \int \prod_{y=1}^{Y} \left[e^{V_G} \Big/ \left(1 + e^{V_G}\right) \right]^{N_y} g(\alpha)\, d\alpha
\end{aligned}
$$

This is the case for the strict exogeneity assumption. However, this assumption is unlikely to hold in most settings and may result in seriously biased and inconsistent parameter estimates, if the initial observations are determined by the evolution of observed and unobserved characteristics in the past. In most empirical studies, researchers cannot verify that their initial observations are exogenous because the initial values are also a part of individuals' historically entangled decision processes.[26] Thus, as a rule, a naïve model with the exogeneity assumption yields biased estimation.

One possible method to deal with this problem is to use a fixed effect model. A fixed effect model treats every unobservable individual effect as a parameter to be estimated. The parameter can avoid restricting the distribution of unobservable individual effects because the conditional distribution of unobserved individual effects does not play a role in the estimation. However, it faces the incidental parameters problem, which leads to a severely biased estimation. This is particularly the case for panel data in which the number of time periods is small.

[26] One exception exists when the initial value is the actual starting value, i.e., if a researcher has a trip for a certain site data set from the period when the site first opened. In this setting, there is no initial value problem because the initial value is really exogenous (See Smith (2005)).

To solve the initial condition problem when exogeneity cannot be assumed. Two approaches have emerged, proposed by Heckman (1981b) and Wooldridge (2005). They will be discussed next.

3.3.2. Heckman's Approach

At the time, Heckman (1981a) highlighted a problem in the standard approach to measure state dependence. He emphasized the identification of the effects of two different factors in a choice occasion—heterogeneity and state dependence. Heckman also distinguished the two different concepts to estimate state dependence correctly. According to Heckman, true state dependence arises when preferences, prices, or constraints relevant to future choices are altered, as a consequence of experiencing an event. On the other hand, spurious state dependence emerges when individuals may differ in certain unmeasured variables that influence their probability of experiencing the event, but are not influenced by the experience of the event itself. Improper treatment of unmeasured variables can give rise to a conditional relationship between future and past experiences like true state dependence, if these variables are correlated over time. The habit-forming property in a good is related with structural dependence. Hence, if a researcher fails to distinguish true state dependence from spurious state dependence, the resulting parameter estimates will be biased.

To solve this problem, Heckman suggests approximating the distribution of the initial observed values conditional on unobserved individual heterogeneity and available pre-sample information, that is, available strictly exogenous explanatory variables. This is the method to determine the density of $\left(N_0, N_1, \cdots, N_Y\right)$ given $\left(Z, TC\right)$. The likelihood function consists of conditional density functions of choosing the number of trips in a year and conditional density function of the initial value—the number of trips in the initial year. If researchers specify $f\left(N_0 \mid Z, TC, \alpha\right)$, then

$$f\left(N_0, N_1, \cdots, N_Y \mid Z, TC, \alpha\right) = f\left(N_1, \cdots, N_Y \mid N_0, Z, TC, \alpha\right) \cdot f\left(N_0 \mid Z, TC, \alpha\right).$$

Next, if the density function $g\left(\alpha \mid Z, TC\right)$ is specified, then the density that researchers want to obtain can be calculated by integration.

$$\begin{aligned} f\left(N_0, N_1, \cdots, N_Y \mid Z, TC\right) &= \int f\left(N_0, N_1, \cdots, N_Y \mid Z, TC, \alpha\right) g\left(\alpha \mid Z, TC\right) d\alpha \\ &= \int f\left(N_1, \cdots, N_Y \mid N_0, Z, TC, \alpha\right) \cdot f\left(N_0 \mid Z, TC, \alpha\right) g\left(\alpha \mid Z, TC\right) d\alpha \end{aligned}, \quad (21)$$

where $g\left(\alpha \mid Z, TC\right)$ is the density function of an individual random effect given observed individual characteristics and travel cost.

Researchers know the density function $f\left(N_1, \cdots, N_Y \mid N_0, Z, TC, \alpha\right)$ from the property of the logit model and also assume $g\left(\alpha \mid Z, TC\right)$, when they set up the basic model. The key, missing component is the density function $f\left(N_0 \mid Z, TC, \alpha\right)$. Once this is specified, researchers can form the appropriate likelihood function and obtain consistent parameter estimates. For this specification, Heckman suggests the utility in period 0 as an approximation to be a linear combination of pre-sample variables, individual random effect term, and disturbance term.

In this study, the concept he suggests can be summarized as follows. The utility, when an individual chooses to take a trip on occasion *t* in initial year, is[27]

$$U_{it0} = V_{i0} + \varepsilon_{it0} = \delta_0 + \gamma_0 X_i + \theta_0 \alpha_i + \varepsilon_{it0} \quad ,$$

[27] So far, this research omits subscript *i* for simple notation in this section. However, from now on, this chapter will use subscript *i* to calculate full maximum likelihood functions.

where X_i is strongly exogenous variables, α_i is normally distributed random effect term with mean zero and variance σ_α^2, and ε_{it0} is a disturbance term that follows i.i.d. Type I extreme value.

The corresponding conditional density function in initial year is

$$f_{i0}\left(N_{i,0} \mid \alpha_i\right) = \left[e^{V_{i0}} \Big/ \left(1+e^{V_{i0}}\right)\right]^{N_{i0}} \quad .$$

The corresponding likelihood function for the entire time period is

$$L_i = \int \left[\left\{ e^{V_{i0}} \Big/ \left(1+e^{V_{i0}}\right) \right\}^{N_{i0}} \prod_{y=1}^{Y} \left\{ e^{V_{iy}} \Big/ \left(1+e^{V_{iy}}\right) \right\}^{N_{iy}} \right] g\left(\alpha_i\right) d\alpha_i \quad .$$

The final log likelihood function for all individuals is

$$LL = \sum_{i=1}^{N} \ln \left[\int \left[\left\{ e^{V_{i0}} \Big/ \left(1+e^{V_{i0}}\right) \right\}^{N_{i0}} \prod_{y=1}^{Y} \left\{ e^{V_{iy}} \Big/ \left(1+e^{V_{iy}}\right) \right\}^{N_{iy}} \right] g\left(\alpha_i\right) d\alpha_i \right] \quad . \qquad (22)$$

3.3.3. Wooldridge's Approach

Rather than specifying conditional distribution of initial values, Wooldridge (2005) suggests that researchers specify an auxiliary distribution of the unobserved individual effects conditioned on the initial values and time invariant exogenous variables that may include mean values of explanatory variables. That is, Wooldridge suggests specifying $h\left(\alpha_i \mid Z_i, TC_i, N_{i,0}\right)$. With this specification, researchers can use the density, $f_i\left(N_{i,1}, \cdots, N_{i,Y} \mid N_{i,0}, Z_{iy}, TC_{iy}\right)$, because they already have the density

$$f_i\left(N_{i,1}, \cdots, N_{i,Y} \mid N_{i,0}, Z_{iy}, TC_{iy}, \alpha_i\right).$$

Now, if we apply Wooldridge's approach to the model in this study, then the individual random effect term is decomposed as

$$\alpha_i = \xi_0 + \xi_1 N_{i,0} + \xi_2 Z_i + \xi_3 TC_i + \psi_i,\ \ \psi_i \sim N\left(0, \sigma_\psi^2\right),\ \ \mathrm{Z}_i = \left(Z_{i1}, \cdots, \mathrm{Z}_{iY}\right),\ \ TC_i = \left(TC_{i1}, \cdots, TC_{iY}\right),$$

where ψ_i is an exogenous individual random effect, meaning this term is uncorrelated with the initial observation, $N_{i,0}$. Z_i is time-variant individual characteristics and TC_i is travel cost during the entire period.

The likelihood function, when individual i chooses to go on a trip in the whole time period, is

$$L_i = \int \left[\prod_{y=1}^{Y} \left\{ e^{V_{iy}} \Big/ \left(1 + e^{V_{iy}}\right) \right\}^{N_{iy}} \right] g\left(\psi_i\right) d\psi_i ,$$

where

$$\begin{aligned} V_{iy} &= \delta_y + \beta N_{i,y-1} + Z_{iy}\gamma + \tau TC_{iy} + \alpha_i \\ &= \delta_y + \beta N_{i,y-1} + Z_{iy}\gamma + \tau TC_{iy} + \xi_0 + \xi_1 N_{i,1} + Z_i \xi_2 + TC_i \xi_3 + \psi_i, \quad \psi_i \sim N\left(0, \sigma_\psi^2\right) \end{aligned}.$$

The resulting log likelihood function can be then be expressed as

$$\begin{aligned} LL &= \sum_{i=1}^{N} \ln \left[\int_{-\infty}^{\infty} \left[\prod_{y=1}^{Y} f_i\left(N_{i,y} \mid N_{i,y-1}, Z_{iy}, TC_{iy}, \alpha_i\right) h\left(\alpha_i \mid Z_i, TC_i, N_{i,0}\right) \right] g\left(\psi_i\right) d\psi_i \right] \\ &= \sum_{i=1}^{N} \ln \left[\int \left[\prod_{y=1}^{Y} \left\{ e^{V_{iy}} \Big/ \left(1 + e^{V_{iy}}\right) \right\}^{N_{iy}} \right] g\left(\psi_i\right) d\psi_i \right] \end{aligned}. \qquad (23)$$

This approach not only enables researchers to avoid the initial condition problem, but it is also much simpler and easier to use, compared to Heckman's method. Wooldridge's method can be implemented using existing computer packages for various random effect models by adding initial values and time-variant explanatory variables over the entire period as a set of covariates for each year. For this reason, Wooldridge's approach has become more common relative to Heckman's approach.

3.4. Model

This study uses dynamic discrete choice models with a repeated mixed logit (RXL) structure. A repeated binary mixed logit model for a single site choice is adopted as the first step. Then, a nesting structure model is introduced to make use of the binary mixed logit in the case of multiple alternatives. Also, this study follows Wooldridge's method to avoid the initial condition problem. As mentioned in Section 3.3, Wooldridge's method provides tractable likelihoods the same as the standard random-effect models. Especially in this study, the two-step nested logit model is adopted, whose second stage estimation has the same structure as Wooldridge's binary choice model. As a result, we can extend Wooldridge's binary choice examples to choices from multiple options.

3.4.1. Model 1 – Binary Choice Model

As the first step, consider a binary choice model, based on the random utility maximization (RUM) hypothesis. Suppose that there is only one recreation site and that individuals need to choose between two options on a given choice occasion—take a trip to the site or not. The RUM specification assumes that an individual will choose the option that yields the greatest utility, thus revealing which option provides a higher utility. If we denote the utility from going on a trip by U_G and the utility from choosing not to go on a trip by U_N, then the probability (P_G) that the individual chooses to take a trip on a certain occasion is $P_G = \text{Prob}\left[U_G > U_N\right]$.

The utility in RUM models can be interpreted as a function that consists of a representative part (V) and a stochastic part (ε). Then, we can re-express the probability P_G in the following manner:

$$U = \begin{cases} U_N \\ U_G \end{cases} = \begin{cases} V_N + \varepsilon_N \\ V_G + \varepsilon_G \end{cases},$$

$$P_G = \text{Prob}\left[U_G > U_N\right] = \text{Prob}\left[V_G + \varepsilon_G > V_N + \varepsilon_N\right] = \text{Prob}\left[\varepsilon_N < \varepsilon_G + V_G - V_N\right] \quad .$$

In the estimation, by assuming that the stochastic part follows a Type I extreme value distribution, the logit models yield the closed form probabilities as follows:

$$P_G = \frac{e^{V_G}}{e^{V_G} + e^{V_N}} \ , \ P_N = 1 - P_G = \frac{e^{V_N}}{e^{V_G} + e^{V_N}} \quad .$$

Also, by specifying the function with observed individual and/or site attributes, researchers can estimate the representative utility (V) with observed choice data using the closed form for the probabilities. This study uses the binary choice model by adding lagged dependent variables (previous choice) to the representative utility function to examine the effect of state dependence. Thus, the representative utility functions are expressed as $V_N = \gamma Z$, $V_G = \delta + \beta N_{-1} + \tau TC + \alpha$, where Z is socio-demographic attributes not related with any travelling site and almost time-invariant, TC is travel cost, N_{-1} is the total number of trips in the previous year, δ is a constant, and α is the unobserved individual random effect. Following Wooldridge's method, this α will consist of the initial value, time-variant explanatory variables—travel cost here—for all years and exogenous random effect term. Hence, the final version of the utility function when individual i chooses to go on a trip or not to go on occasion t in year y will be for $i=1,\cdots,I,\ y=1,\cdots,Y,\ t=1,\cdots,T$,

$$U=\begin{cases} U_{ityN}=V_{iyN}+\varepsilon_{ityN}=\gamma Z_{iy}+\varepsilon_{ityN} \\ U_{ityG}=V_{iyG}+\varepsilon_{ityG}=\delta+\beta N_{i,y-1,G}+\tau TC_{iy}+\alpha_i+\varepsilon_{ityG} \end{cases}, \tag{24}$$

where $\alpha_i=\xi_1 N_{i,0,G}+TC_i\xi_2+\psi_i,\ \psi_i\sim N\left(0,\sigma_\psi^2\right),\ TC_i=\left(TC_{i1},\cdots,TC_{iY}\right)$.

3.4.2. Model 2 – Nesting Structure Model

Now, as the next step, this study also considers a nesting structure model using a lagged variable of the total number of choosing the no-trip option. The nesting structure is described as

$$\begin{matrix} U_{i0ty} = \\ U_{ijty} = \end{matrix} \underbrace{\boxed{\begin{matrix}\beta N_{i0,y-1}+Z_{iy}\gamma+\alpha_i \\ \end{matrix}}}_{\textit{Vary across nests}} \underbrace{\boxed{\begin{matrix}\\ \delta_j+\tau TC_{ijy}\end{matrix}}}_{\textit{Vary within nests}} \begin{matrix} +\varepsilon_{i0ty} = V_{i0y}+\varepsilon_{i0ty} \\ +\varepsilon_{ijty} = V_{ijy}+\varepsilon_{ijty}\end{matrix} \quad ,$$

where $N_{i0,y-1}$ is the total number for choosing no-trip option in y year, Z_{iy} is socio-demographic attributes, δ_j is the alternative specific constant for site j, and TC_{ijy} is travel cost of j-th site in y year.

In a participation decision, the full process of choice decision consists of two stages. The first stage is the site selection decision (i.e., the probability of which site to visit conditional on taking a trip). Hence, this second step is exactly the same as the classical conditional logit site selection model. The second stage decision is whether to go on a trip or not. The choice frame for the second stage is the same for typical binary choice models, which means that individuals choose between the no-trip option and the going-trip nest.

Formally, let

$$U_{ijyt} = \begin{cases} V_{i0y} + \varepsilon_{i0yt} = -\delta_{1y} + \beta N_{i0,y-1} + Z_{iy}\gamma + \alpha_i + \varepsilon_{i0yt} \\ V_{ijy} + \varepsilon_{ijyt} = \delta_{jy} - \delta_{1y} + \tau TC_{ijy} + \varepsilon_{i0yt} \end{cases}, \tag{25}$$

where $\alpha_i = \xi_1 N_{i0,0} + IV_i \xi_2 + \psi_i$, $\psi_i \sim N\left(0, \sigma^2\right)$ and *IV* is the inclusive value. The inclusive value is the expected utility of taking a trip from the analyst's perspective. It is defined as the log-sum of the exponential fitted representative utility for going on a trip to a lake. Subtract δ_{1y} from all utilities and divide all *V*s by the dissimilarity coefficient, *λ*. Then, the inclusive values (*IV*) for each year will be

$$IV_{iy} = \ln\left(\sum_{j=1}^{J} \exp\left(\hat{\tilde{V}}_{ijy}\right)\right),\ IV_i = \left(IV_{i1}, \cdots, IV_{iY}\right) \quad , \tag{26}$$

where $\hat{\tilde{V}}$ is a fitted value of $\tilde{V}$ and $\tilde{V}_{ijy} = \tilde{\Delta}_{jy} + \tilde{\tau}_y TC_{ijy}$, where $\tilde{\Delta}_{jy} = \Delta_{jy} / \lambda$, $\tilde{\tau}_y = \tau_y / \lambda$, $\Delta_{jy} = \delta_{jy} - \delta_{1y}$, and $\Delta_{1y} = 0$.

Then, the probability that an agent *i* chooses to go on a trip in the second stage is

$$Q_{ity,Trip} = \frac{\exp\left(\lambda IV_{iy}\right)}{\exp\left(-\delta_{1y} + \beta N_{i0,y-1} + \gamma Z_{iy} + \alpha_i\right) + \exp\left(\lambda IV_{iy}\right)} = \frac{\exp\left(\delta_{1y} + \lambda IV_{iy}\right)}{\exp\left(\beta N_{i0,y-1} + \gamma Z_{iy} + \alpha_i\right) + \exp\left(\delta_{1y} + \lambda IV_{iy}\right)}$$

and the probability to choose site, *j*, conditional on 'trip' nest is

$$P_{ijty|\text{Trip}} = \frac{\exp\left(\tilde{V}_{ijy}\right)}{\exp\left(IV_{iy}\right)} \quad . \tag{27}$$

The probability of taking a trip to site *j* (P_{ijty}) is

$$P_{ijty} = \Pr\left[Going\ Trip\right]\cdot\Pr\left[Choosing\ \ jth\ site\,|\,Going\ Trip\right] = P_{ijty|\mathrm{Trip}}\cdot Q_{ity,Trip}$$

$$= \frac{e^{V_{ijy}/\lambda}\cdot\left[\sum_{j=1}^{J} e^{V_{ijy}/\lambda}\right]^{\lambda-1}}{e^{V_{i0y}} + \left[\sum_{j=1}^{J} e^{V_{ijy}/\lambda}\right]^{\lambda}} \qquad . \qquad (28)$$

$$= \frac{\exp\left(\widetilde{V}_{ijy}\right)}{\exp\left(IV_{iy}\right)}\cdot\frac{\exp\left(\delta_{1y} + \lambda IV_{iy}\right)}{\exp\left(\beta N_{i0,\mathrm{y}-1} + \gamma Z_{iy} + \alpha_i\right) + \exp\left(\delta_{1y} + \lambda IV_{iy}\right)}$$

Now, the probability that an individual decides not to go on a trip in the second stage is

$$Q_{ity,No-trip} = \frac{\exp\left(-\delta_{1y} + \beta N_{i0,\mathrm{y}-1} + \gamma Z_{iy} + \alpha_i\right)}{\exp\left(-\delta_{1y} + \beta N_{i0,\mathrm{y}-1} + \gamma Z_{iy} + \alpha_i\right) + \exp\left(\lambda IV_{iy}\right)} = \frac{\exp\left(\beta N_{i0,\mathrm{y}-1} + \gamma Z_{iy} + \alpha_i\right)}{\exp\left(\beta N_{i0,\mathrm{y}-1} + \gamma Z_{iy} + \alpha_i\right) + \exp\left(\delta_{1y} + \lambda IV_{iy}\right)}$$

.

So, the probability of no trip (P_0) is

$$P_{i0ty} = \Pr\left[No\ Trip\right]\cdot\Pr\left[Choosing\ \ no-trip\,|\,No\ Trip\right] = P_{i0ty|No-trip}\cdot Q_{ity,No-trip} = Q_{ity,No-trip}$$

$$= \frac{e^{V_{i0y}}}{e^{V_{i0y}} + \left[\sum_{j=1}^{J} e^{V_{ijy}/\lambda}\right]^{\lambda}}$$

$$= \frac{\exp\left(-\delta_{1y} + \beta N_{i0,\mathrm{y}-1} + \gamma Z_{iy} + \alpha_i\right)}{\exp\left(-\delta_{1y} + \beta N_{i0,\mathrm{y}-1} + \gamma Z_{iy} + \alpha_i\right) + \exp\left(\lambda IV_{iy}\right)} \qquad . \qquad (29)$$

$$= \frac{\exp\left(\beta N_{i0,\mathrm{y}-1} + \gamma Z_{iy} + \alpha_i\right)}{\exp\left(\beta N_{i0,\mathrm{y}-1} + \gamma Z_{iy} + \alpha_i\right) + \exp\left(\delta_{1y} + \lambda IV_{iy}\right)}$$

During the second stage, the main time-variant explanatory variable will be inclusive values. One advantage to using inclusive values as a time-variant explanatory variable is that the values include not only a variation of travel cost, but also a variation of the site specific attribute for each year, i.e., annual alternative specific constant. For this reason, inclusive values are more time-variant than travel cost. Therefore, this will help estimating more accurately. Hence, multiple site models can be simplified to a binary choice model by using nests. The only difference between this second stage of Model 2 and Model 1 is the lagged dependent variables for no-trip option are used here, while lagged dependent variables for going-trip option are used in the first model. Hence, in the nesting structure model, the coefficient of state dependence, β, is about the effects of how often an individual stays at home.[28]

Based on the probabilities explained previously, this study estimates all parameters by using a two-step process. During the first stage, the site choice probabilities are estimated, including $\tilde{\tau}_y$, $\tilde{\Delta}_{jy}$. In turn, these can be utilized to construct fitted values for the inclusive values. During the second stage, simulated maximum likelihood estimation techniques are used to estimate the participation parameters, β, γ, ξ_1, ξ_2, σ, δ_{1y}, λ.

3.4.3. Exponential Model – RE Poisson

This research also uses a random effect (RE) Poisson model to compare the results from the first and the second models. The lake visitation pattern naturally fits within the Poisson modeling framework, with the number of visitations consisting of nonnegative integers. There are many zeros over all occasions. As shown in Table 3.1 for the Iowa Lakes Project, the average numbers of trips in a year is around six. The larger the number of trips, the rarer in the data. So, exponential distributions can be a good approximation of lake visitation patterns. Although Poisson distribution has a strong property of equal-dispersion, we can loosen this trait by mixing

[28] Here, I only allow for state dependence in terms of how often an individual stays at home or equivalently in terms of overall trip taking behavior, not for state dependence with respect to which specific sites are visited. This level of state dependence would be good for further research.

a random effect term. Suppose that individual i's total number of trips in year y $\left(N_{iy}\right)$ conditional on $\left(N_{i,y-1},\cdots,N_{i0},Z_i,TC_i,\alpha_i\right)$ has a Poisson distribution with mean

$$E\left(N_{iy} \mid N_{i,y-1},\cdots,N_{i0},Z_i,TC_i,\alpha_i\right)=\alpha_i \exp\left[\delta+g\left(N_{i,y-1}\right)\beta+Z_{iy}\gamma+\tau TC_{iy}\right] \quad , \tag{30}$$

where α_i is the individual RE. Applying Wooldridge's approach here, the RE term will be $\alpha_i=\psi_i \exp\left[g\left(N_{i0}\right)\xi_1+TC_i\xi_2\right]$, where ψ_i is an unobserved exogenous heterogeneity. The function g allows the lagged dependent variable to appear in a flexible way. This study defines the function either as $g\left(N_{i,y-1}\right)=\left\{I\left(0<N_{i,y-1}\leq M\right),I\left(M<N_{i,y-1}\right)\right\}$, where I is an index function and M is the mean for trip numbers or the lagged variable. For the multiple site case, travel cost TC_{iy} is weighted travel cost calculated by the following Eq. (31) :

$$TC_{iy}=\sum_{j=1}^{J}\widehat{P}_{ijy|\mathrm{Trip}}\cdot TC_{ijy} \quad , \tag{31}$$

where $\widehat{P}_{ijy|\mathrm{Trip}}$ is the fitted value of probability to go jth site conditional on 'trip' nest.

STATA provides two different distributions for ψ_i – Gamma distribution with mean one and variance, v, and Normal distribution with mean zero and variance σ_v^2. If the RE term comes from the Gamma distribution, then the entire unobserved term will be similar with a negative binomial distribution, which has overdispersion attributes. If the normally distributed RE term is imposed, the equal-dispersion property can be relaxed by adding panel-level variance.

3.5. Data

This study uses the Iowa lakes trip survey data from 2002 to 2005 as part of the Iowa Lake Valuation Project. The Iowa Lake Valuation Project is a panel study from 2002 to 2005 and

2009, supported by the IDNR and the US EPA. The primary goal of the project was to gather information regarding the visitation patterns of Iowa residents to the primary recreational lakes in the state of Iowa. The data from the Iowa Lake Valuation Project also provide site attributes, such as lake size, boat ramp dummy, wake restrictions dummy, handicap facilities dummy, state park dummy, water quality index, etc. The survey also includes questions about respondents' demographic information, such as age, gender, education level, household size, and their income level.

Among these demographic characteristics, five—age, age squared, gender, education, and household size—are selected for this study. For education, the categories are simplified to use a dummy variable in the model. If the education level is equal to or higher than college graduate, then the dummy is one, otherwise zero. Basic statistics are summarized in Table 3.1.

The total numbers of respondents who returned and completed the surveys are 4,254, 5,277, 4,242 and 3,993, respectively, for years 2002, 2003, 2004, and 2005. Among them, this research excludes (1) respondents who failed to provide trip data and (2) individuals whose number of trips to any lake is more than 52. The concern with including respondents, who answered more than 52 visits, is these consist predominantly of households who live in close proximity to a certain lake. In this case, they could be residents who pass a lake on their commute to work or take a walk or bike along it. These kinds of demands for lakes are quite different from this study’s focus. Finally, this research selects only common respondents to all surveys from 2002 to 2005 to use a balanced panel for convenience. As a result, the final remaining sample size is 1,287, who responded annually to the survey.

Travel cost (TC) are computed as follows:

$$Travel\ cost = round\ trip\ distance \cdot fuel\ cost + \frac{1}{3}\left(round\ trip\ time \cdot respondent's\ hourly\ wage\right)\ .$$

In this study, PC Miler was used to compute the trip distance and time. CPI adjusted gasoline prices (dollars/gallon) divided by average fuel efficiency of U.S. light duty vehicles (miles/gallon) for use as a proxy for fuel cost (dollars/mile).[29] For hourly wage, the survey responses to household's annual income are used. In every income category, median annual income is selected and divided by 2,000 to yield a wage rate. Finally, the wage rates are adjusted using the CPI for a common year.[30]

3.6. Estimation

3.6.1. Monte Carlo Simulation

Before the estimation on empirical recreation demand, a simulation exercise is used to illustrate the initial condition problem in a dynamic discrete choice setting. Specifically,a pseudo-data set is generated and compiles four kinds of models with the data—(1)static model, (2) naïve dynamic model, (3) Heckman's model, and (4) Wooldridge's model. A static model refers to a repeated mixed logit without lagged variables, while a naïve model refers to a dynamic repeated mixed logit *assuming* strictly exogenous initial values. Heckman's model and Wooldridge's model employ Heckman's approximation and Wooldridge's method, respectively.

---

[29] Each source comes from U.S. Energy Information Administration (Midwest all grades all formulations retail gasoline prices), The Research and Innovative Technology Administration in the U.S. Department of Transportation (average fuel efficiency of U.S. light duty vehicles), Bureau of Labor Statistics in U.S. Department of Labor (annual CPI and average hourly earnings), respectively.

[30] The number 2,000 comes from a 40-hour work week with two weeks of vacation annually.[31] Denote the initial year by zero (subscript y = 0). So, the observable periods are when y = 0, 1, 2, 3, including the initial year. In this sense, the last year observable is when Y = 3. On the other hand, denote unobservable periods by negative number (y = -1, -2, … , -19). Thus, the year when the site opened is when y = -19 and the total number of years is 23.

To generate a simple pseudo-data set, suppose that:

1. only one individual attribute is observable;
2. only one recreation site exists (J = 1);
3. the size of the panel sample is one thousand (N = 1,000);
4. the number of occasions in a year is fifty-two (T = 52);
5. the panel period over which we observe both individual attributes and travel cost is three years (Y = 3) (four year if initial year is included); and
6. the total period is twenty three years (that is, the site opened nineteen years before the initial observable year).[31]

The individual attribute and travel cost are assumed drawn from a uniform distribution while unobserved individual heterogeneity is assumed to be drawn from a normal distribution. Specifically, $Z_{iy} \sim iid\ Uniform(0,1)$ and $P_{iy} = 0.5 + 39.5 \cdot \pi_{iy}$, where $\pi_{iy} \sim iid\ Uniform(0,1)$ and $\alpha_i \sim iid\ N\left(0, 1.5^2\right)$. The number of an individual's trips in the starting year (i.e., y = -19) is assumed drawn from a binomial distribution with the probability of a trip on each of 52 choice occasions is 0.08. After the starting year, the decision rule for every occasion follows the probability of mixed logit with the following utility function.

$$U_{ijty} = \begin{cases} \varepsilon_{i0ty} \\ V_{i1y} + \varepsilon_{i1ty} \end{cases} = \begin{cases} \varepsilon_{i0ty} \\ \delta + \beta N_{i,y-1} + \gamma Z_{iy} + \tau P_{iy} + \alpha_i + \varepsilon_{i1ty} \end{cases}, \tag{32}$$

where $\varepsilon_{i0ty}, \varepsilon_{i1ty} \sim Type\ I\ extreme\ value$ for $i = 1, \cdots, N,\ t = 1, \cdots, T,\ y = 1, \cdots, Y$.

[31] Denote the initial year by zero (subscript y = 0). So, the observable periods are when y = 0, 1, 2, 3, including the initial year. In this sense, the last year observable is when Y = 3. On the other hand, denote unobservable periods by negative number (y = -1, -2, … , -19). Thus, the year when the site opened is when y = -19 and the total number of years is 23.

With the pseudo-data set described above, we obtain a set of parameter estimates. Finally, by repeating the generation of the pseudo-data set and estimating the associated parameters, we can obtain the average and the standard error of parameter estimates. This research completes this Monte Carlo simulation with 500 repetitions.

The results of the simulation show that a static model provides seriously-biased estimates, except for the coefficient of the individual specific characteristics, Z_{iy}. Although the estimated coefficient of individual attributes is within two-standard error significance interval, the estimates for all four models have large variances compared to the size of the estimates. Also, the results show that the naïve model overestimates the state dependence, when habit-forming behaviors exist (Table 3.2). On the other hand, the state dependence tends to be underestimated in the naïve model, when variety-seeking behaviors exist, although the estimate is not significantly different from the true parameter using a 95% confidence level (Table 3.3). Both Heckman and Wooldridge's methods provide parameter estimates close to the true parameters.

3.6.2. Single Site Model

This study's Model 1 introduces a method to estimate the state dependence effect is the binary dynamic discrete choice model, including only one lake in the choice set. Saylorville Lake, the most visited lake in Iowa, is used first as an example of the binary dynamic discrete choice model. Then, single site models are presented for (1) the three most visited lakes, (2) three mid-range lakes in terms of visitation, and (3) the three least visited lakes in the state. This is completed to determine whether there is any noticeable pattern in parameter estimates across the usage pattern groups.

For the Saylorville Lake case, recreational demand is examined using four models—(1) dynamic mixed logit suggested by Wooldridge, (2) naïve dynamic mixed logit, which assumes that initial values are exogenous, (3) static mixed logit that does not consider lagged variables, and (4) dynamic RE Poisson suggested by Wooldridge (Table 3.4). For a comparison of the three groups (most visited, mid-range, and bottom-range lakes), only Wooldridge's dynamic mixed logit is estimated (Tables 3.5, 3.6, and 3.7).

From Table 3.4, there is no significant state dependence in recreational demand for Saylorville Lake. Although the naïve model shows weak inertia, both Wooldridge's mixed logit and Poisson models show indistinct variety-seeking patterns. No models show any significant estimate for lagged variables. However, the coefficient of the initial value is significantly positive for both the RE logit and RE Poisson models. All four models consistently show a convex relationship in age, and a stronger tendency to go on trip for males and big families.[32]

From Tables 3.5, 3.6, and 3.7, this study shows results of estimation for nine lakes. For all nine lakes, the lagged dependent variables have a negative effect, although most are insignificant. On the other hand, initial value has a significantly positive effect at the 95% significance level in all cases. Also, as the number of visitation is smaller, the coefficient of initial value tends larger. For travel cost, its effect in each year should be the sum of the coefficient of travel cost for the entire period and travel cost for the corresponding year (i.e., one among TC 2003, TC 2004, TC 2005). So, the overall effect of travel cost is negative. We can also confirm that the estimation becomes less accurate as people visit less by comparing standard errors in each group.

[32] For individual characteristics, RE logit models place those terms in the Utility function of 'no-trip' option, while RE Poisson places them in the function of trip numbers (Refer to Eqs.(24) and (30)).

3.6.3. Nesting Structure Model

The nesting structure model developed in this study focuses on state dependence in terms of overall trip taking propensity. We start by estimating a nested logit model that focuses on trips to the three most visited lakes first and then extends the number of lakes to the 100 most visited lakes. Table 3.8 shows the results of the first stage conditional logit, which comes from Eq. (27). All estimates of the first stage are statistically significant at the 95% confidence level. Then, by using estimates of the first stage, we can form the inclusive values required for the second stage estimates provided in Tables 3.9 and 3.10. In Table 3.9, there are two cases – unconstrained and constrained. Without any constraint, the dissimilarity coefficient, λ, is estimated as a negative value. However, the dissimilarity coefficient cannot be negative. It should lie in the unit interval by definition. For this reason, this research imposes a constraint for dissimilarity,

$$\lambda = \exp(\kappa) / \{1 + \exp(\kappa)\} \ ,$$

where κ is a certain parameter in the constrained case. In either case, the coefficient of lagged variables is significantly negative, which means an individual who took many trips in a given year is likely to take fewer trips in subsequent years (i.e., chose ‘no-trip’ option more often than to go on a trip).[33] On the other hand, the coefficient of initial value has the sign opposite of the lagged variables. The results for the RE Poisson model in Table 3.10 show the same signs for the key parameters (lagged variable, initial value, travel cost, etc.) with the RE logit model in Table 3.9. However, the interpretation is somewhat different. If the coefficient of lagged variables is negative, this implies an agent tends to choose to go on a trip next year not as many as this year.

[33] In this study, we use the ‘no-trip’ option as a representative for any option other than ‘going trip’. It can be ‘staying home’, but also can be ‘visiting a friend’s home’ or ‘going to the theater’ and so on.

However, the coefficient is not significant in this study, while the initial value has a positive effect in choosing to go on a trip. Meanwhile, the overdispersion parameter, v, is significantly greater than zero. This means that the RE Poisson model is definitely different from the pooled Poisson model.[34]

These results can be also confirmed with the nesting structure model. Tables 3.11 through 13 are related to 100 lakes. Even if the number of sites is extended from 3 to 100, there is no substantial difference between these two cases. Also, in the RE Poisson model, we apply both the original number and the more flexible function mentioned in Section 3.4.3 for lagged variables and initial values. Tables 3.12 and 3.13 provide the results for each case. They have consistent results in terms of sign and significance.

3.7. Discussion and Conclusions

This study investigates whether there is any state-dependent pattern in recreational demand. Especially, this research devises a method to apply Wooldridge's binary choice model to the multiple-choice case by the adopting nesting structure. The main findings from this study are summarized as:

(1) In single site cases, both RE logit and RE Poisson show a negative coefficient for the lagged variables and a significantly positive coefficient for the initial value in choosing 'trip' option.

(2) In the multiple site cases, the RE Poisson model provides a negative coefficient of lagged variables and a significantly positive coefficient of initial value, consistent with

[34] Although we do not include the results in this study, we used a normal distribution for the RE term as a robustness check. The results are fairly close to the Gamma distribution case.

the single site cases. Moreover, the negative effect of lagged variables becomes more significant as the number of sites increases.

(3) In multiple site cases, the RE two-step nested logit shows a significantly negative coefficient of lagged variables and a positive coefficient of initial value in choosing 'no-trip' option, which looks opposite the results from the other models. However, this does not necessarily mean contradiction because the negative effect in choosing 'no-trip' is not equivalent to a positive effect in choosing 'trip'. Rather, this can support a variety-seeking pattern in the whole recreational choice pool, including both 'going lakes' option and 'no-trip' option. The results of both RE nested logit and RE Poisson can be interpreted that people have a variety-seeking tendency in the options for their recreation.

(4) In all cases, the coefficient of initial value has an opposite sign from the coefficient of lagged variables and its scale is much greater than that for the coefficient of lagged variables. The initial value is lagged value in the first period. In this sense, the offset effect of the initial value in calculating the total effect of lagged variables should be considered.

(5) The lakes visited more often tend to show smaller absolute values both in lagged variables and in initial values.

As these main findings show, all results from this study support a variety-seeking choice pattern. However, there are some limitations to this study. First, as mentioned in the main findings, to confirm variety-seeking behaviors, we need to verify the total effects taking into account the offset effect of the initial value. Second, the data set has only three years of observations. To obtain more accurate estimation, a longer time period is desired. Three years appear relatively short because Wooldridge's approach shows a better performance for the panels of longer periods (Akay 2012). In addition, for the multiple sites case, this study only groups all going trip options (nesting structure model) or aggregates the sum of trips and travel cost for all sites (RE Poisson model). Generalizing the binary RE logit to the multinomial RE logit in consideration of correlation among sites is left for a future study.

References

Abidoye, B.O., J.A. Herriges, and J.L. Tobias. 2012. "Controlling for Observed and Unobserved Site Characteristics in RUM Models of Recreation Demand." *American Journal of Agricultural Economics* 94(5):1070-1093.

Adamowicz, W.L. 1994. "Habit Formation and Variety Seeking in a Discrete Choice Model of Recreation Demand." *Journal of Agricultural and Resource Economics* 19(1):19-31.

Akay, A. 2012. "Finite-sample Comparison of Alternative Methods for Estimating Dynamic Panel Data Models." *Journal of Applied Econometrics* 27:1189-1204.

Alessie, R., and A. Kapteyn. 1991. "Habit Formation, Interdependent Preferences and Demographic Effects in the Almost Ideal Demand System." *Economic Journal* 101(406):404-419.

Arulampalam, W., and M.B. Stewart. 2009. "Simplified Implementation of the Heckman Estimator of the Dynamic Probit Model and a Comparison with Alternative Estimators." *Oxford Bulletin of Economics and Statistics* 71(5):659-681.

Cameron, A.C., and P.K. Trivedi. 2010. *Microeconometrics Using Stata*. Revised Edition. Stat Press.

Casella, G., and R.L. Berger. 2002. *Statistical Inference*. 2nd Edition. Duxbury.

Daunfeldt, S., J. Nordström, and L. Thunström. 2011. Habit Formation in Food Consumption. *The Oxford Handbook of the Economics of Food Consumption and Policy* Ch.31, pp. 770-790. Oxford University Press.

Dynan, K.E. 2000. "Habit Formation in Consumer Preferences: Evidence from Panel Data." *American Economic Review* 90(3):391-406.

Ederton, D.L., B. Assarsson, A. Hummelmose, I.P. Laurila, K. Rickertsen, and P.H. Vale. 1996. *The Econometrics of Demand Systems: With Applications to Food Demand in the Nordic Countries*. Dordrecht: Kluwer Academic Publishers.

Greene, W.H. 2008. *Econometric Analysis*. 6th Edition. Pearson Prentice Hall.

Haab, T.C. 2003. “Temporal Correlation in Recreation Demand Models with Limited Data.” *Journal of Environmental Economics and Management* 45:195-212.

Haab, T.C., and K.E. McConnell. 2002. *Valuing Environmental and Natural Resources*. Edward Elgar.

Haan, P. 2010. “A Multi-state Model of State Dependence in Labor Supply: Intertemporal Labor Supply Effects of a Shift from Joint to Individual Taxation.” *Labour Economics* 17:323-335.

Heckman, J.J. 1981a. “Heterogeneity and State Dependence”. In S. Rosen (ed.). *Studies in Labor Markets*, pp. 91-139. University of Chicago Press. Chicago. IL.

———. 1981b. “The Incidental Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-Discrete Data Stochastic Process.” In: *Structural Analysis of Discrete data with Econometric Applications*, pp. 179-195. Ed. Manski, C.F. and D. McFadden. MIT Press. Cambridge. Massachusetts.

Herriges, J.A., and D.J. Phaneuf. 2002. “Inducing Patterns of Correlation and Substitution in Repeated Logit Models of Recreation Demand.” *American Journal of Agricultural Economics* 84(4):1076-1090.

Holt, M.T., and B.K. Goodwin. 1997. “Generalized Habit Formation in an Inverse Almost Ideal Demand System: An Application to Meat Expenditures in the U.S.” *Empirical Economics* 22:293-320.

Johannesson, M., and D. Lundin. 2002. “The Impact of Physician Preferences on the Prescription of New Drugs.” *SSE/EFI Working Paper Series in Economics and Finance* No. 460.

Meghir, C., and G. Weber. 1996. “Intertemporal Nonseparability or Borrowing Restrictions? A Disaggregate Analysis Using a U.S. Consumption Panel.” *Econometrica* 64(5):1151-1181.

Oguzoglu, U. 2010. “Disability and Multi-state Labour Force Choices with State Dependence.” *IZA Discussion Paper* No. 5408.

Orme, C.D. 1997. “The Initial Conditions Problem and Two-step Estimation in Discrete Panel Data Models.” Mimeo. University of Manchester.

———. 2001. "Two-step Inference in Dynamic Non-linear Panel Data Models." Mimeo. University of Manchester.

Papadimitriou, E. 2012. "Theory and Models of Pedestrian Crossing Behavior along Urban Trips." *Transportation Research Part F* 15:75-94.

Pollak, R.A. 1970. "Habit Formation and Dynamic Demand Functions." *Journal of Political Economy* 78:745-763.

Pollak, R.A., and T.J. Wales. 1969. "Estimation of the Linear Expenditure System." *Econometrica* 37:629-750.

Prowse, V. 2012. "Modeling Employment Dynamics with State Dependence and Unobserved Heterogeneity." *Munich Personal RePEc Archive Paper* No. 38038.

Seetharaman, P.B. 2004. "Modeling Multiple Source of State Dependence in Random Utility Models: A Distributed Lag Approach." *Marketing Science* 23(2):263-271.

Smith, M.D. 2005. "State Dependence and Heterogeneity in Fishing Location Choice." *Journal of Environmental Economics and Management* 50:319-340.

Swait, J., W. Adamowicz, and M. van Bueren. 2004. "Choice and Temporal Welfare Impacts: Incorporating History into Discrete Choice Models." *Journal of Environmental Economics and Management* 47:94-116.

Thunström, L. 2010. "Preference Heterogeneity and Habit Persistence: The Case of Breakfast Cereal Consumption." *Journal of Agricultural Economics* 61(1):76-96.

Train, K. 2009. *Discrete Choice Methods with Simulation*. Cambridge University Press. Cambridge.

Walker, J.L, M. Ben-Akiva, and D. Bolduc. 2007. "Identification of Parameters in Normal Error Component Logit-Mixture (NECLM) Models." *Journal of Applied Econometrics* 22:1095-1125.

Wooldridge, J. 2005. "Simple Solutions to the Initial Conditions Problem in Dynamic Nonlinear Panel Data Models with Unobserved Heterogeneity." *Journal of Applied Econometrics* 20:39-54.

Table 3.1. Summary statistics of the data set

Variables	Mean	Std. Dev.	Min	Max
Number of trips	6.32	9.18	0	52
Age	53.19	14.61	9	83
Gender	0.71	0.45	0	1
Education	0.74	0.44	0	1
Household (HH) size	2.59	1.32	0	10
Travel cost (dollar in 2002)	30.32	10.42	1.83	86.00

Table 3.2. Comparison with estimates of three different dynamic models and those of one static model by Monte Carlo simulation when habit-forming behavior exists (with 500 repetitions)

	True Parameter	Heckman's			Wooldridge's		
		Est.	Std. err.	MSE	Est.	Std. err.	MSE
ASC	-4.5	**-4.435**	0.180	0.037	**-4.409**	0.371	0.146
Lagged variable	0.2	**0.188**	0.011	0.000	**0.189**	0.023	0.001
Individual attribute	0.3	0.295	0.222	0.049	0.288	0.205	0.042
Travel cost	-0.2	**-0.198**	0.010	0.000	**-0.200**	0.010	0.000
Sigma	1.5	**1.817**	0.192	0.137	**1.274**	0.405	0.215
	True Parameter	Naïve Dynamic			Static		
		Est.	Std. err.	MSE	Est.	Std. err.	MSE
ASC	-4.5	**-4.325**	0.158	0.056	-7.689	0.571	10.495
Lagged variable	0.2	0.264	0.013	0.004		N.A.	
Individual attribute	0.3	0.311	0.218	0.048	0.193	0.414	0.183
Travel cost	-0.2	**-0.211**	0.010	0.000	-0.126	0.017	0.006
Sigma	1.5	**1.295**	0.195	0.080	6.156	0.655	22.104

Note: Numbers are bold when they satisfy both conditions that there is no significant difference between true parameter and estimated one within two-standard error bound and that the estimates are significantly different from zero within two-standard error bound.

Table 3.3. Comparison with estimates of three different dynamic models and those of one static model by Monte Carlo simulation when variety-seeking behavior exits (with 500 repetitions)

	True Parameter	Heckman's			Wooldridge's		
		Est.	Std. err.	MSE	Est.	Std. err.	MSE
ASC	-4.5	**-4.473**	0.189	0.036	**-4.505**	0.417	0.174
Lagged variable	-0.2	**-0.239**	0.075	0.007	**-0.197**	0.061	0.004
Individual attribute	0.3	0.309	0.241	0.058	0.320	0.252	0.064
Travel cost	-0.2	**-0.200**	0.012	0.000	**-0.202**	0.013	0.000
Sigma	1.5	**1.613**	0.219	0.061	**1.362**	0.597	0.376
	True Parameter	Naïve Dynamic			Static		
		Est.	Std. err.	MSE	Est.	Std. err.	MSE
ASC	-4.5	**-4.512**	0.187	0.035	**-4.477**	0.187	0.035
Lagged variable	-0.2	**-0.157**	0.065	0.006		N.A.	
Individual attribute	0.3	0.315	0.236	0.056	0.322	0.250	0.063
Travel cost	-0.2	**-0.203**	0.013	0.000	**-0.205**	0.013	0.000
Sigma	1.5	**1.499**	0.517	0.267	**1.454**	0.410	0.170

Note: Numbers are bold when they satisfy both conditions that there is no significant difference between true parameter and estimated one within two-standard error bound and that the estimates are significantly different from zero within two-standard error bound.

Table 3.4. Results of binary random effect (RE) Model for Saylorville Lake from 2003 to 2005 (initial year is 2002)

Variables	Dynamic Logit (W)		Dynamic Logit (N)		Static Logit		Dynamic Poisson (W)	
	Est.	S.E.	Est.	S.E.	Est.	S.E.	Est.	S.E.
Constant	**-8.071**	0.761	**-7.922**	0.793	**-7.935**	0.804	**-2.236**	0.679
Travel cost	0.005	0.004	**-0.032**	0.003	**-0.032**	0.003	0.005	0.004
Lagged trips	-0.011	0.010	0.013	0.010	n.a.		-0.005	0.009
Age	**-0.074**	0.028	**-0.069**	0.029	**-0.067**	0.029	**0.055**	0.025
Age^2	**0.001**	0.0003	**0.001**	0.0003	**0.001**	0.0003	**-0.001**	0.0002
Gender	**-0.371**	0.203	**-0.464**	0.228	**-0.465**	0.232	**0.202**	0.169
Education	0.003	0.133	-0.036	0.138	-0.035	0.138	-0.054	0.121
HH size	**-0.121**	0.046	**-0.137**	0.047	**-0.135**	0.047	**0.111**	0.043
TC2003	-0.001	0.006					0.001	0.006
TC2004	**-0.017**	0.007	n.a.		n.a.		**-0.018**	0.006
TC2005	**-0.019**	0.007					**-0.011**	0.006
Initial value	**0.438**	0.035					**0.415**	0.053
Variation (σ or v)	**2.049**	0.111	**2.381**	0.128	**2.451**	0.133	**4.584**	0.466

Note 1: Numbers are bold when they satisfy both conditions that there is no significant difference between true parameter and estimated one within two-standard error bound and that the estimates are significantly different from zero within two-standard error bound.
Note 2: In the first raw of the title, 'W' means Wooldridge's method while 'N' represents a naïve dynamic model.
Note 3: In the 'Variation' raw, σ means standard deviation of normally distributed unobserved exogenous heterogeneity, v means overdispersion parameter in a RE Poisson model.

Table 3.5. Results of Wooldridge's single-site dynamic RE logit Model for three most visited lakes from 2003 to 2005 (initial year is 2002)

Variables	Saylorville Lake (1)		Coralville Lake (2)		Clear Lake (3)	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
Constant	**-7.971**	**0.793**	**-6.992**	**0.911**	**-6.077**	**0.817**
Travel cost (TC)	0.005	0.004	-0.007	0.005	-0.002	0.004
Lagged trips	-0.011	0.010	**-0.027**	**0.007**	-0.001	0.008
Age	**-0.076**	**-0.028**	-0.015	-0.034	0.082	-0.027
Age^2	**0.0009**	**-0.0003**	0.0004	-0.0003	**-0.0006**	**-0.0002**
Gender	-0.373	-0.205	**-0.504**	**-0.245**	**-1.072**	**-0.293**
Education	0.019	-0.132	-0.259	-0.157	-0.235	-0.201
HH size	**-0.121**	**-0.046**	0.050	-0.064	0.102	-0.067
TC 2003	-0.001	0.006	-0.001	0.007	0.006	0.007
TC 2004	**-0.017**	**0.007**	-0.004	0.008	-0.006	0.007
TC 2005	**-0.019**	**0.007**	-0.014	0.007	-0.014	0.008
Initial value	**0.429**	**0.037**	**0.412**	**0.017**	**0.350**	**0.019**
σ	**2.004**	**0.125**	**2.333**	**0.115**	**2.981**	**0.178**

Note 1: Numbers are bold when they satisfy both conditions that there is no significant difference between true parameter and estimated one within two-standard error bound and that the estimates are significantly different from zero within two-standard error bound.
Note 2: In the first raw of the title, the number in parentheses means the rank of visitation number among 130 lakes.

Table 3.6. Results of Wooldridge's single-site dynamic RE logit Model for three mid-range lakes in visitation from 2003 to 2005 (initial year is 2002)

Variables	Poll Miller Park Lake (61)		Little Wall Lake (62)		Lake Miami (63)	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
Constant	**-27.934**	**8.888**	**-18.224**	**3.390**	**-10.741**	**2.434**
Travel cost (TC)	**-0.037**	**0.015**	0.003	0.003	**-0.039**	**0.012**
Lagged trips	-0.051	0.043	-0.194	0.132	**-0.227**	**0.054**
Age	**-0.746**	**-0.342**	**-0.266**	**-0.112**	0.102	-0.094
Age^2	**0.007**	**-0.003**	**0.002**	**-0.001**	-0.0004	-0.0009
Gender	-0.056	-1.033	-0.972	-0.564	**-1.498**	**-0.680**
Education	-0.005	-0.538	0.144	-0.416	-0.127	-0.562
HH size	0.194	-0.162	-0.074	-0.170	0.132	-0.211
TC 2003	-0.030	0.037	**-0.067**	**0.012**	0.002	0.015
TC 2004	-0.024	0.032	0.011	0.009	0.011	0.014
TC 2005	-0.020	0.032	0.002	0.009	0.021	0.013
Initial value	**0.974**	**0.160**	**2.681**	**0.672**	**2.591**	**0.313**
σ	**3.833**	**0.799**	**2.880**	**0.367**	**4.010**	**0.541**

Note 1: Numbers are bold when they satisfy both conditions that there is no significant difference between true parameter and estimated one within two-standard error bound and that the estimates are significantly different from zero within two-standard error bound.
Note 2: In the first raw of the title, the number in parentheses means the rank of visitation number among 130 lakes.

Table 3.7. Results of Wooldridge's single-site dynamic RE logit Model for three bottom-range lakes in visitation from 2003 to 2005 (initial year is 2002)

Variables	Crawford Creek Impoundment (111)		Mitchell Lake (112)		Nine Eagles Lake (112)	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
Constant	-7.144	4.016	-60.154	8106.790	**-16.135**	**5.703**
Travel cost (TC)	0.020	0.022	0.014	0.015	0.034	0.021
Lagged trips	-0.188	0.224	-0.276	0.308	-0.444	0.308
Age	0.125	-0.126	-0.254	-0.302	-0.203	-0.221
Age^2	-0.001	-0.001	0.003	-0.003	0.002	-0.002
Gender	-2.063	-1.429	-43.011	-8106.735	-0.610	-0.973
Education	-0.249	-1.061	-0.275	-1.194	0.196	-0.790
HH size	-0.258	-0.351	0.019	-0.407	-0.291	-0.219
TC 2003	**-0.088**	**0.037**	**-0.046**	**0.018**	-0.031	0.016
TC 2004	-0.029	0.026	-0.001	0.024	-0.005	0.021
TC 2005	-0.034	0.026	0.006	0.024	-0.026	0.017
Initial value	**3.197**	**1.271**	**0.834**	**0.300**	**2.909**	**1.050**
σ	**3.017**	**0.712**	**2.363**	**0.918**	**3.030**	**0.713**

Note 1: Numbers are bold when they satisfy both conditions that there is no significant difference between true parameter and estimated one within two-standard error bound and that the estimates are significantly different from zero within two-standard error bound.
Note 2: In the first raw of the title, the number in parentheses means the rank of visitation number among 130 lakes.

Table 3.8. Results of 1st stage conditional logit for three most visited lakes from 2003 to 2005 (initial year is 2002)

Variables	1st-stage		
	Estimates	Standard Error	Est./S.E.
ASC2_2003	-0.244	0.120	-2.04
ASC2_2004	-0.537	0.107	-5.04
ASC2_2005	-0.582	0.124	-4.71
ASC3_2003	-0.254	0.102	-2.49
ASC3_2004	-0.574	0.091	-6.31
ASC3_2005	-0.417	0.102	-4.08
Travel cost/λ_2003	-0.064	0.002	-28.70
Travel cost/λ _2004	-0.045	0.002	-27.21
Travel cost/λ _2005	-0.059	0.002	-27.25

Table 3.9. Results of 2nd stage repeated binomial mixed logit for three most visited lakes from 2003 to 2005 (initial year is 2002)

Variables	Unconstrained			Constrained		
	Est.	S.E.	Est./S.E.	Est.	S.E.	Est./S.E.
ASC1_2003	6.341	0.564	11.25	6.844	0.557	12.29
ASC1_2004	6.201	0.566	10.96	6.700	0.559	12.00
ASC1_2005	6.247	0.568	11.00	6.756	0.561	12.05
Lagged variable	-0.013	0.004	-2.94	-0.013	0.004	-2.93
Dissimilarity coef. (λ)	-0.029	0.051	-0.58	7.22E-07	3.78E-04	0.002
Age	0.020	0.017	1.16	0.018	0.017	1.06
Age^2	-1.42E-06	1.56E-04	-0.01	1.27E-05	1.56E-04	0.08
Gender	-0.781	0.153	-5.09	-0.796	0.148	-5.37
Education	-0.132	0.097	-1.36	-0.128	0.097	-1.32
HH size	-0.005	0.034	-0.15	-0.002	0.033	-0.06
Inclusive value_2003	-0.078	0.096	-0.82	-0.268	0.084	-3.20
Inclusive value_2004	-0.078	0.139	-0.56	0.204	0.123	1.67
Inclusive value_2005	-0.403	0.100	-4.03	-0.402	0.097	-4.16
Initial value	0.223	0.008	28.29	0.235	0.008	28.62
σ	1.875	0.076	24.64	1.846	0.070	26.46

Note: 'Constrained' means when the constraint, λ=exp(parameter)/[1+exp(parameter)], is assigned in estimation.

Table 3.10. Results of RE Poisson for pooled data of three most visited lakes from 2003 to 2005 (initial year is 2002)

Variables	RE Poisson		
	Estimates	Standard Error	Est./S.E.
ASC1_2003	0.594	0.435	1.36
Year dummy_2004	-0.142	0.040	-3.50
Year dummy_2005	-0.090	0.039	-2.35
Lagged variable	-0.006	0.004	-1.69
Weighted travel cost	0.004	0.004	0.98
Age	-0.025	0.015	-1.63
Age^2	1.37E-04	1.42E-04	0.96
Gender	0.356	0.124	2.87
Education	0.050	0.087	0.58
HH size	0.011	0.030	0.35
Weighted travel cost_2003	-0.003	0.006	-0.54
Weighted travel cost_2004	-0.005	0.006	-0.83
Weighted travel cost_2005	-0.014	0.006	-2.28
Initial value	0.264	0.024	11.01
v	3.151	0.208	15.14

Note: 'v' is the overdispersion parameter from Gamma distribution. From the likelihood-ratio (LR) test of v = 0, the probability of v = 0 is less than 0.1%.

Table 3.11. Results of 2nd stage repeated binomial mixed logit for 100 most visited lakes from 2003 to 2005 (initial year is 2002)

Variables	Unconstrained			Constrained		
	Est.	S.E.	Est./S.E.	Est.	S.E.	Est./S.E.
ASC1_2003	3.193	0.420	7.60	3.200	0.422	7.58
ASC1_2004	3.112	0.421	7.40	3.086	0.422	7.31
ASC1_2005	3.137	0.420	7.46	3.124	0.422	7.40
Lagged variable	-0.018	0.001	-14.04	-0.018	0.001	-14.00
Dissimilarity coef. (λ)	-0.031	0.020	-1.57	2.02E-05	5.19E-04	0.04
Age	0.042	0.008	5.13	0.039	0.008	4.80
Age^2	-2.46E-04	7.52E-05	-3.27	-2.09E-04	7.54E-05	-2.77
Gender	-0.650	0.112	-5.79	-0.645	0.113	-5.73
Education	-0.046	0.054	-0.86	-0.027	0.054	-0.50
HH size	-0.031	0.017	-1.83	-0.033	0.017	-1.96
Inclusive value_2003	-0.127	0.078	-1.62	-0.119	0.078	-1.52
Inclusive value_2004	-0.064	0.080	-0.81	-0.068	0.080	-0.85
Inclusive value_2005	-0.092	0.079	-1.16	-0.063	0.079	-0.80
Initial value	0.152	0.006	26.65	0.153	0.006	26.46
σ	1.683	0.039	42.99	1.689	0.039	42.86

Note: 'Constrained' means when the constraint, λ=exp(κ)/[1+exp(κ)], is assigned in estimation.

Table 3.12. Results of RE Poisson for pooled data of 100 most visited lakes from 2003 to 2005 (initial year is 2002)

Variables	RE Poisson		
	Estimates	Standard Error	Est./S.E.
ASC1_2003	2.078	0.232	8.94
Year dummy_2004	-0.075	0.016	-4.61
Year dummy_2005	-0.046	0.016	-2.88
Lagged variable	-0.009	0.001	-9.65
Weighted travel cost	0.0001	0.002	0.04
Age	-0.026	0.007	-3.80
Age^2	1.43E-04	6.40E-05	2.23
Gender	0.362	0.082	4.40
Education	0.001	0.045	0.01
HH size	0.015	0.014	1.08
Weighted travel cost_2003	-0.012	0.007	-1.85
Weighted travel cost_2004	0.008	0.007	1.14
Weighted travel cost_2005	-0.011	0.008	-1.51
Initial value	0.101	0.005	19.14
v	1.623	0.073	22.12

Note: 'v' is the overdispersion parameter from Gamma distribution. From the likelihood-ratio (LR) test of v = 0, the probability of v = 0 is less than 0.1%.

Table 3.13. Results of RE Poisson with a function (*g*) of lagged variables for pooled data of 100 most visited lakes from 2003 to 2005 (initial year is 2002)

Variables	RE Poisson		
	Estimates	Standard Error	Est./S.E.
ASC1_2003	1.727	0.232	7.44
Year dummy_2004	-0.087	0.016	-5.40
Year dummy_2005	-0.047	0.016	-2.92
g1(1)	-0.090	0.038	-2.36
g1(2)	-0.225	0.038	-5.86
Weighted travel cost	-0.001	0.002	-0.40
Age	-0.026	0.007	-3.88
Age^2	1.69E-04	6.37E-05	2.66
Gender	0.368	0.081	4.54
Education	-0.020	0.045	-0.44
HH size	0.020	0.014	1.42
Weighted travel cost_2003	-0.010	0.006	-1.53
Weighted travel cost_2004	0.007	0.007	0.96
Weighted travel cost_2005	-0.014	0.007	-1.88
g0(1)	0.930	0.090	10.29
g0(2)	2.239	0.094	23.80
v	1.565	0.073	21.57

Note: 'v' is the overdispersion parameter from Gamma distribution. From the likelihood-ratio (LR) test of $v = 0$, the probability of $v = 0$ is less than 0.1%.