Valuing Water Quality as a Function of Water Quality Measures*

Kevin J. Egan

Department of Economics, University of Toledo

Joseph A. Herriges[†]and Catherine L. Kling

Department of Economics, Iowa State University

John A. Downing

Department of Ecology, Evolution and Organismal Biology, Iowa State University

January 23, 2008

Abstract

Employing a unique and rich data set of water quality attributes in conjunction with detailed household characteristics and trip information, we develop a mixed logit model of recreational lake usage. Our empirical analysis shows that individuals are responsive to the full set of water quality measures used by biologists to identify the impaired status of lakes. WTP estimates are reported based on improvements in these physical measures. This implies that cost benefit analysis based on water quality measures can be used as a direct policy tool.

Valuing Water Quality as a Function of Water Quality Measures

Abstract

   Employing a unique and rich data set of water quality attributes in conjunction with detailed household characteristics and trip information, we develop a mixed logit model of recreational lake usage. Our empirical analysis shows that individuals are responsive to the full set of water quality measures used by biologists to identify the impaired status of lakes. WTP estimates are reported based on improvements in these physical measures. This implies that cost benefit analysis based on water quality measures can be used as a direct policy tool.

More than three decades have elapsed since the passage of the 1972 Clean Water Act (CWA), yet progress toward meeting the standards set forth in the CWA has been slow in the area of nonpoint source pollution. The most recent National Water Quality Inventory (USEPA,[?]) categorizes forty-five percent of assessed lake acres in the U.S. as impaired, with the leading causes of these impairments being nutrients and siltation. Moreover, few states have developed the priority ranking of their impaired waters or determined the Total Maximum Daily Loads (TMDLs) as required under Section 303(d) of the CWA.[1] Legal actions by citizen groups have prompted renewed efforts toward developing both the priority listing and associated TMDL standards.[2] However, the task facing both the EPA and state regulatory agencies remains a daunting one. The prioritization process alone, which is all the more important given current tight budgets, requires information on the cost of remediation and the potential benefits that will flow from water quality improvements. Both types of information are in short supply. The purpose of this paper is to help fill this gap by providing information on the value of water quality improvements as a function of detailed physical attributes of the water bodies involved. The water quality values are obtained from a recreation demand model of lake usage in the state of Iowa, combining trip and socio-demographic data from the Iowa Lakes Valuation Project and an extensive list of physical water quality measures collected by Iowa State University's Limnology Laboratory[3].

Recreation demand models have long been used to value water quality improvements, but studies typically rely on limited measures of water quality. The most commonly used indicators are fish catch rates (e.g., [?], [?]). However, catch rates are themselves endogenous, depending on both fishing pressure and the abilities of the anglers, and provide only indirect measures of the underlying water quality. Physical water quality measures, such as Secchi depth and bacteria counts, are used only sparingly, in large part because of limitations in available data. Phaneuf, Kling, and Herriges [?] use fish toxin levels in their model of Great Lakes fishing, but the toxin levels were available only for a limited number of aggregate sites in the region. Parsons and Kealy [?] use dummy variables based on dissolved oxygen levels and average Secchi depth readings to capture the impact of water quality on Wisconsin lake recreation. Similarly, Parsons, Helm, and Bondelid [?] construct dummy variables indicating *high* and *medium* water quality levels for use in their analysis of recreational demand in six northeastern states. These dummy variables are based on pollution loading data and water quality models, rather than on direct measurements of the local water quality. In all of these studies, the physical water quality indicators are found to significantly impact recreation

demand, but, because of the limited nature of the measures themselves, provide only a partial picture of value associated with possible water quality improvements.Other papers that have used one or more measures of water quality include von Haefen (1993); Atasoy et al and Phaneuf (2003).

An alternative to physical measures of water quality has been the use of perceived or reported water quality measures (Adamowicz et al.; Jeon et al.). While perceived measures are likely to be the more direct drivers of behavior (McConnell), a major drawback to their use is that information can be gleaned only for an individual only on visited sites. While important questions concerning the relationships between perceived and observed measures remains, it is likely that perceptions are related to physical measures, the focus of this work.

Bockstael, Hanemann, and Strand's [?] analysis of beach usage in the Boston-Cape Cod area has perhaps one of the most extensive lists of objective physical water quality attributes included in a model of recreation: oil, fecal coliform, temperature, chemical oxygen demand (COD), and turbidity. However, the study also points out one of the frequently encountered problems in isolating the impact of individual water quality attributes - multicollinearity. Seven additional water quality measures were available to the analysts: color, pH, alkalinity, phosphorus, nitrogen, ammonia, and total coliform. These latter variables were excluded from the analysis because of correlations among the various groups of water quality measures. The five water quality variables used were chosen because they were either directly observable by recreationists or highly publicized. While these choices are certainly reasonable given limitations in the available data, the lack of direct information on how nutrient levels (phosphorus and nitrogen) impact recreational usage is unfortunate in the context of setting standards in many states, where nutrient loadings are of particular concern.

The contribution of the current paper lies in our ability to incorporate a rich set of physical water quality attributes, as well as site and household characteristics, into a model of recreational lake usage. Importantly, the full set of water quality variables used by biologists to classify lakes as impaired by the EPA, and therefore potentially in need of policy action, are included. Trip data for the study are drawn from the 2002 Iowa Lakes Survey. The survey was sent to a random sample of 8,000 Iowa households, eliciting information on their recreational visits to Iowa's 129 principal lakes, along with socio-demographic data and attitudes toward water quality issues. The unique feature of the project, however, is that a parallel inventory of the physical attributes of these lakes is being conducted by Iowa State University's Limnology Laboratory.[4] Three times a year, over the course of a five-year

project, thirteen distinct water quality measurements are being taken at each of the lakes, providing a clear physical characterization of the conditions in each lake. Moreover, because of the wide range of lake conditions in the state, Iowa is particularly well suited to identifying the impact of these physical characteristics on recreation demand. Iowa's lakes vary from a few clean lakes with up to fifteen feet of visibility to other lakes having some of the highest concentrations of nutrients in the world, and roughly half of the 129 lakes included in the study are on the EPA's list of impaired lakes. An additional unique aspect of Iowa lakes is that the diversity of land uses in the watershed contributing to them leads to a low degree of collinearity among the water quality measures. Thus, Iowa lakes provide an almost ideal "laboratory" for studying the effects of biological water quality measures on usage and value.

A second unique contribution of this study is the application of careful model specification and fitting procedures to identify the best set of explanatory variables, and their functional form, for the estimated model. Since economic theory does not provide guidance to the analyst on these issues, ex ante selection of model variables and structure will often fail to achieve the best model fit. On the other hand, specification searching on a given data set leads to the well known problem of incorrect standard error estimates (Leamer, [?]). Thus, we exploit our large sample size by splitting the sample into three equal parts. With the first sample, we extensively explore various specifications of the model, including a variety of variables and their functional form. The second sample was reserved for clean model estimation, allowing us to generate unbiased estimates of precision for all of the parameter estimates. The third split of the sample was used to perform out-of-sample prediction to provide an overall assessment of the model fit.[5]

The remainder of the paper is divided into five sections. Section 2 provides an overview of the two data sources. A repeated mixed logit model of recreational lake usage is then specified in Section 3. The mixed logit model allows for a wide variety of substitution patterns among the recreational sites and for heterogeneity among households in terms of their reaction to individual site characteristics. (See, e.g., [?],[?], and [?].) The specification search procedure and subsequent parameter estimates are reported in Section 4, along with the assessment of the out-of-sample predictions provided by the final model. In Section 5, we illustrate not only the implications of the model in terms of recreational value of meeting the objectives of the CWA (i.e., removing all of the lakes in the state from the impaired water quality list) but also how the model can be used to prioritize the remediation task. Conclusions of the paper are provided in Section 6.

Data

Two principal data sources are used in developing our model of recreational lake usage in Iowa: the 2002 Iowa Lakes Survey and the physical water quality measures collected by Iowa State University's Limnology Laboratory. The focus of the survey was on gathering baseline information on the visitation patterns to Iowa's 129 principal lakes, as well as socio-demographic data and attitudes towards water quality issues. After initial focus groups and pre-testing of the survey instrument, the final survey was administered by mail in November 2002 to 8,000 randomly selected households in the state. Standard Dillman procedures ([?]) were used to ensure a high response rate.[6] Of the 8,000 surveys mailed, 4,423 were returned. Allowing for undeliverable surveys, this corresponds to an overall response rate of sixty-two percent.

The survey sample was initially paired down to 3,859 households as follows. Those households who returned the survey from out of state were excluded (thirty-eight observations), as it is not feasible to ascertain whether these respondents have permanently left the state or simply reside elsewhere for part of the year. Respondents who did not complete the trip questions or did not specify their numbers of trips (i.e., they simply checked that they had visited a given lake) were excluded (224 observations). Lastly, any household reporting more than fifty-two total single-day trips to the 129 lakes were excluded (133 observations). In the analysis that follows, only single-day trips are included to avoid the complexity of modeling multiple-day visits. Defining the number of choice occasions as fifty-two allows for one trip per week to one of the 129 Iowa lakes. While the choice of fifty-two is arbitrary, we chose this cut-off as a conservative approach to insure our data sample is visitors to the lakes and not residents who may claim many recreational trips simply due to living on the shore of a lake.[7] This last step eliminated approximately three percent of the returned surveys. Finally, because of the large number of respondents, the overall sample was randomly divided into three segments; specification, estimation, and prediction portions, each component using just under 1,290 observations. Once the estimation stage is reached, the results will be less prone to pretest bias and the conventional standard errors for the resulting parameter estimates will be biased less by the extensive specification search. For the bias to be eliminated the standard errors in the model (unobservables in the context of the random utility model employed here) must be uncorrelated across the individuals in the three samples, and the "correct" model must be identified after the specification search (to eliminate any misspecification error). To the extent that correlations do exist (say due

to shared observable site characteristics), or the "correct" model is not identified, then the benefits from the split sample procedure used here will be diminished and the conventional standard errors for the parameter estimates obtained at the estimation stage will be biased by the specification search process. However, the bias should be reduced relative to having used a single sample for both specification and estimation.[8]

Table 1 provides summary statistics for the full sample regarding trip and the socio-demographic data obtained from the survey.[9],[10] The average number of total single-day trips for all 129 lakes is 6.44, varying from some respondents taking zero trips and others taking fifty-two trips. In total, 41.2% of the sample (1,588 respondents of the 3,859) reports zero recreation trips. Of the remaining 2,271 respondents, 61% report fewer than 10 trips, with the number of respondents reporting a particular trip level gradually declining as the number of trips increases. Only 3.7% of the respondents who report positive trips take more than 40. In general, the survey respondents are more likely to be older, male, and to be more educated than the general population. Schooling is entered as a dummy variable equaling one if the individual has attended or completed some level of post-high school education.

The physical water quality measures used in modeling recreational lake usage in Iowa were gathered by Iowa State University's Limnology Laboratory as part of the ongoing state lake monitoring program. Table 2 provides a listing of the water quality attributes and 2002 summary statistics for the 129 lakes used in our analysis. All of the physical water quality measures are the average values for the 2002 season. Samples were taken from each lake three times throughout the year, in spring/early summer, mid-summer, and late summer/fall to cover the range of seasonal variation.

Each of the water quality measures help to characterize a distinct aspect of the lake ecosystem. Secchi transparency is one of the most widely applied limnological parameters and approximately reflects the lake depth at which the bottom of the lake can still be seen. Chlorophyll is an indicator of phytoplankton plant biomass which leads to greenness in the water. Three nitrogen levels were gathered. Total nitrogen is the sum of all dissolved and particulate forms. NH3+NH4 measures ammonium nitrogen that derives from fertilizer or anaerobic conditions and can be toxic. NO3+NO2 measures the nitrate level in the water that derives from aerobic nutrient contributions. Total phosphorus is an important indicator of nutrient conditions in freshwater systems and is usually the principal limiting nutrient which determines phytoplankton growth. Silicon is important to diatoms, a key food source for aquatic organisms. The hydrogen ion concentration of the water is measured

by pH with levels below 6 indicating acid conditions and levels above 8 indicating extreme basicity. As Table 2 notes, all of the pH levels in this sample are clustered between 7.7 and 10. Alkalinity is a reflection of the buffering capacity of the water and is expressed as the concentration of calcium carbonate in the water. Plants need carbon to grow and most carbon comes from alkalinity in lakes; therefore, alkalinity is an indication of the availability of carbon to plant life. Inorganic suspended solids (ISS) consist of soil and silt suspended in the water through erosion, whereas volatile suspended solids (VSS) consists of suspended organic matter. Increases in either ISS or VSS levels decrease water clarity.

Table 2 demonstrates that there is considerable variation in water quality conditions throughout the state. For example, average Secchi depth varies from a low of 0.09 meters (or 3.5 inches) to a high of 5.67 meters (over 18 feet). Total phosphorus varies from 17 to 453 μg/L, spanning the range of concentrations seen in the world (Arbuckle and Downing, [?]). In addition to trip and water quality data, two other data sources were used. First, the travel costs, from each survey respondent's residence to each of the 129 lakes, were needed. The out-of-pocket component of travel cost was computed as the round-trip travel distance multiplied by $0.25 per mile.[11] The opportunity cost of time was calculated as one-third the estimated round-trip travel time multiplied by the respondent's average wage rate.[12] Table 3 provides summary statistics for the resulting travel cost variable. The average price of a recreational trip to a lake is $135, although perhaps a more meaningful statistic is the average price of a lake visit, $37 (i.e., the average price including only those respondents who visited each particular lake).

Second, lake site characteristics were obtained from the Iowa Department of Natural Resources [?]. Table 3 provides a summary of these site characteristics. As Table 3 indicates, the size of the lakes varies considerably, from 10 acres to 19,000 acres. Four dummy variables are included to capture different amenities at each lake. The first is a "ramp" dummy variable which equals one if the lake has a cement boat ramp, as opposed to a gravel ramp or no boat ramp at all. The second is a "wake" dummy variable that equals one if motorized vessels are allowed to travel at speeds great enough to create wakes and zero otherwise. About sixty-six percent of the lakes allow wakes, whereas thirty-four percent of lakes are "no wake" lakes. The "state park" dummy variable equals one if the lake is located adjacent to a state park, which is the case for 38.8 percent of the lakes in our study. The last dummy variable is the "handicap facilities" dummy variable, which equals one if handicap amenities are provided, such as handicap restrooms or paved ramps.[13] A concern may be that handicap facilities

would be strongly correlated with the state park dummy variable. However, while fifty of the lakes in the study are located in state parks and fifty have accessible facilities, only twenty six of these overlap. Finally, a "fish index" variable is included that varies from zero to four representing the number of fish species for which the lake is considered one of the "top 10" sites for the species in the state.[14]

The Model

The repeated mixed logit model was chosen because it exhibits many desirable properties, including that "...it allows for corner solutions, integrates the site selection and participation decisions in a utility consistent framework, and controls for the count nature of recreation demand" (Herriges and Phaneuf, [**?**]).

The model specification begins by assuming that the utility individual i receives from choosing site j on choice occasion t is of the form

$$(1) \qquad U_{ijt} = V\left(X_{ij}; \beta_i\right) + \varepsilon_{ijt}, \ \ i = 1, ..., N; \ \ j = 0, .., J; \ \ t = 1, ..., T$$

where V represents the observable portion of utility and, from the perspective of the researcher, ε_{ijt} represents the unobservable portion of utility. A mixed logit model is defined as the integration of the logit formula over the distribution of unobserved random parameters (Revelt and Train, [**?**]). If the random parameters, β_i, were known then the probability of observing individual i choosing alternative j on choice occasion t would follow the standard logit form

$$(2) \qquad L_{ijt}\left(\beta_i\right) = \frac{\exp\left(V_{ijt}\left(\beta_i\right)\right)}{\sum\limits_{k=0}^{J} \exp\left[V_{ikt}\left(\beta_i\right)\right]}.$$

Since the β_i's are unknown, the corresponding unconditional probability, $P_{ijt}\left(\theta\right)$, is obtained by integrating over an assumed probability density function for the β_i's. The unconditional probability is now a function of θ, where θ denotes the vector of hyperparameters (e.g. the mean and standard deviation) characterizing the distribution of the random parameters (i.e., the β_i's) in the population. This repeated mixed logit model assumes the random parameters are $i.i.d.$ distributed over the individuals so that

$$(3) \qquad P_{ijt}\left(\theta\right) = \int L_{ijt}\left(\beta\right) f\left(\beta|\theta\right) d\beta.$$

No closed-form solution exists for this unconditional probability and therefore simulation is required for the maximum likelihood estimates of θ.[15]

Following Herriges and Phaneuf [**?**], a dummy variable, D_j, is included which equals one for all of the recreation alternatives $(j = 1, \ldots, J)$ and equals zero for the stay-at-home option $(j = 0)$. Including the stay-at-home option allows a complete set of choices, including in the population those individuals who always "stay at home" on every choice occasion and do not visit any of the sites. It is convenient to partition the individual's utility into the stay-at-home option or choosing one of the J sites, with

$$(4) \qquad U_{ijt} = \begin{cases} \tilde{\beta}' z_i + \varepsilon_{i0t} \\ \beta_i' x_{ij} + \alpha_i + \varepsilon_{ijt}, \ j = 1, ..., J, \end{cases}$$

where α_i is the random parameter on the dummy variable, D_j, which does not appear since it equals one for $j = 1, ..., J$ and zero for $j = 0$. The observable information included in the state-at-home option, represented by z_i, is the set of socio-demographic characteristics, such as gender, age, and education. Notice that with this specification the socio-demographic factors do not influence the choice among sites, but simply the participation decision, as this information does not vary across the sites.[16] The x_{ij} matrix represents the observable information that varies across the lakes, including the travel costs for each individual to each lake, and lake characteristics such as amenities and water quality measures.

The random coefficient vectors for each individual, β_i and α_i, can be expressed as the sum of population means, b and a, and individual deviation from the means, δ_i and γ_i, which represents the individual's tastes relative to the average tastes in the population (Train, [**?**]). Therefore, we can redefine

$$(5) \qquad \beta_i' x_{ij} = b' x_{ij} + \delta_i' x_{ij}$$

$$(6) \qquad \alpha_i = a + \gamma_i$$

and then the partitioned utility is

$$(7) \qquad U_{ijt} = \begin{cases} \tilde{\beta}' z_i + \eta_{i0t} \\ b' x_{ij} + a + \eta_{ijt}, \ j = 1, ..., J, \end{cases}$$

where

$$(8) \qquad \eta_{ijt} = \begin{cases} \varepsilon_{i0t} & i = 1, ..., N; \ t = 1, ..., T \\ \delta_i' x_{ij} + \gamma_i + \varepsilon_{ijt}, & j = 1, ..., J; \ i = 1, ..., N; \ t = 1, ..., T \end{cases}$$

is the unobserved portion of utility. This unobserved portion is correlated over sites, $j = 1, ..., J$, and choice occasions, $t = 1, ..., T$, because of the common influence of the terms δ_i

and γ_i. The same preferences are used by the individual to evaluate each site at each time period. Since the unobserved portion of utility is correlated over sites and trips, the familiar IIA assumption does not apply for mixed logit models. For example, consider α_i, the random parameter on the dummy variable D_j, which equals 0 for the home option and 1 otherwise. An individual who chooses the stay-at-home option for all choice occasions would have a negative deviation from a (i.e. $\gamma_i < 0$), the mean of α_i, while someone who takes many trips would have a positive deviation from a (i.e. $\gamma_i > 0$), allowing the marginal effect of D_j to vary across individuals. However, the individual carries their α_i across choice occasions and sites.

In particular, we model the utility individual i receives from choosing lake j on choice occasion t as

$$(9) \qquad U_{ijt} = \begin{cases} \tilde{\beta}' z_i + \varepsilon_{i0t} \\ -\beta^P P_{ij} + \beta^{q'} Q_j + \beta_i^{a'} A_j + \alpha_i + \varepsilon_{ijt}, \ j = 1, ..., J \end{cases},$$

where the vector z_i consists of socio-demographic characteristics, P_{ij} is the travel cost from each Iowan's residency to each of the 129 lakes, as calculated with PCMiler. The vector Q_j denotes the physical water quality measures and A_j represents the attributes of the lake. As shown in equation (9), notice that the parameters on the lake attributes and the dummy variable, D_j, are random. All of the other variables are assumed to have fixed parameters.

Specification and Estimation

While the repeated mixed logit model provides the general framework for our analysis, it does not determine the specific variables to use in the model (e.g., which water quality measures) or the functional form they should take in equation (9) (e.g., linear versus logarithmic). Moreover, economic theory provides little guidance in terms of these choices, although valuable guidance can be obtained from a number of sources, including ecologists who understand the source and degree of impairment in a watershed, previous studies of similar water quality concerns, and focus groups in the affected population. Nonetheless definitive functional forms and choice of variables based on physical science or other sources will be rare.. In order to investigate the model specification issue, we divided the full survey sample into three portions, with one portion each dedicated to model specification, estimation and out-of-sample prediction. We begin this section with a discussion of the model specification process.

Specification

There are, of course, a large number of potential model specifications given the range of water quality, site characteristics, and household characteristics identified in Tables 1 through

3. We focus on modeling the role of water quality characteristics in determining recreation demand patterns, holding constant the manner in which both socio-demographics and other site characteristics impact preferences. Specifically, socio-demographic characteristics are assumed to enter through the stay-at-home option. They include age and household size, as well as dummy variables indicating gender and college education (See Table 1). A quadratic age term is included in the model to allow for nonlinearities in the impact of age. Site characteristics, identified in Table 3, are included with random coefficients. This is to allow for heterogeneity in individual preferences regarding site characteristics, such as wake restrictions and site facilities, observed in previous studies (e.g., Train, [?]). For example, some households may prefer to visit less developed lakes with wake restrictions in place, and thus would have a positive coefficient associated with wake restrictions. Others might be attracted to sites allowing the use of motorboats, jet skis, etc. and would have a negative coefficient associated with wake restrictions. In order to allow for a range of possible reactions to the various site characteristics, it is initially assumed that the random parameters (β_i^a) are each normally distributed with the mean and dispersion of each parameter estimated.[17]

Even restricting our attention to the water quality characteristics in Table 2, there are a large number of potential model specifications. We focus on five groups of water quality characteristics for the Q_j in equation (9):

- Secchi depth;

- Chlorophyll;

- Nutrients (Total Nitrogen and Total phosphorus);

- Suspended Solids (Inorganic and Organic); and

- Bacteria (Cyanobacteria and Total Phytoplankton).

The first four characteristic groups directly impact the visible features of the water quality, making it more likely that households respond to them. Bacteria is included because surveyed households report it to be the single most important water quality concern (See [?]).

Our initial intent was to consider three possible specifications for each of these groups of variables: inclusion linearly, inclusion logarithmically, or exclusion. However, preliminary analysis indicated that these variables individually and as groups were consistently significant

at a five percent level or better. Thus, we chose to focus on determining whether each group of factors should enter the model in a linear or logarithmic fashion. This required estimating a total of $2^5 = 32$ versions of the model. Table 4 provides a summary of the results, with the various specifications listed in terms of decreasing log-likelihood.[18] Since all of the models in table 4 (and also generally the following tables 5, 6, and 7) are with the same number of parameters it is not possible to use the likelihood ratio test.Generally, all that can be done is compare the models based on the log-likelihood values, with a lower negative value representing the model is better fitting the data. One interesting feature of the results is that the modelling ranking is lexicographic in terms of the bacteria specification, with all models that have bacteria entering logarithmically being preferred to (i.e., have a higher log-likelihood relative to) those in which bacteria enters linearly, regardless of the functional form for the remaining variables. Given the specification for bacteria, the models having chlorophyll entering logarithmically are generally preferred to those having it enter linearly. The preferred model has Secchi depth and suspended solids entering the model linearly, with the remaining variables entering in a logarithmic fashion. This model is referred to as Model A below.[19]

In addition to investigating the appropriate functional form for the five water quality groups above, we also considered the question as to which factors had the greatest impact on the overall fit of the model. Such an analysis is useful in suggesting factors that might be important to monitor in other settings. Secchi depth was assumed to be included in the model, since this measure is frequently available for freshwater lakes. We then considered which single factor had the greatest marginal impact on the log-likelihood, as well as which pair of factors together had the greatest impact on log-likelihood. The results are reported in Tables 5 and 6 respectively. Total phosphorous is the most important additional single factor, where total phytoplankton and cyanobacteria emerged as the most important pair of additional characteristics in modelling recreational site choice. It is important to emphasize that this does not mean that these characteristics are themselves the most important factors in determining site choice, but rather they serve as the best proxies for all of the water quality measures that impact recreational choices.

Tables 5 and 6 indicate the marginal benefit from gathering additional water quality measures. Concerning the cost to obtain these measures, the ranking from easiest to obtain to the most difficult and costly to obtain is: 1) Secchi transparency, 2) nutrients (total nitrogen and total phosphorus), 3) chlorophyll and the suspended solids (inorganic and volatile), and

4) bacteria (cyanobateria and total phytoplankton). Focusing on the five groups of water quality measures we have considered (bolded in tables 5 and 6), Secchi transparency is clearly the best one measure to include, as it is easy to obtain and the most important single measure. The bacteria levels are the next most important measures, however, they are also the most costly. Our data indicates collecting the bacteria measures is worthwhile as the other groups of measures are considerably behind in terms of the log-likelihood values. The nutrients are relatively easy to collect and they are the most important additional measures after Secchi transparency and the bacteria measures. Therefore, if possible we recommend the inclusion of the nutrient levels. In our informal cost-benefit analysis chlorophyll and the suspended solids fair the worst. They are relatively more expensive to collect and also the least important measures.

Finally, the specification search process above assumes that the random parameters associated with site characteristics are each normally distributed. While this assumption is common practice in the literature, there is no *a priori* basis for this choice. Moreover, a number of authors have noted potential problems with unbounded parameter distributions (such as the normal) leading to implausible results for some portion of the population.[20] As a final stage in the specification search process, we consider alternative choices for the random parameters in our model.[21] Specifically, we grouped the site characteristic parameters into three categories: (i) lake size parameter, (ii) the five parameters associated with the discrete variables for paved ramps, handicap facilities, state parks, wake restrictions, and the fish index, and (iii) participation parameter α_i. For each set of parameters, we consider both normal and triangular distributions, for a total of eight possible sets of distributional assumptions. While a myriad of other distributions are of course possible (e.g., lognormal, uniform and truncated normals), we focus on the triangular distribution. The advantage of the triangular distribution is that it allows for both positive and negative parameter values and retains the basic shape of the normal distribution (i.e., with a definative mode and a density that diminishes away from that mode), while avoiding the unbounded nature of the normal distribution. As a basis of comparison, we also include a model in which all but the participation parameter are fixed (i.e., nonrandom). This structure mimics the standard nested logit structure (the participation parameter is normally distributed instead of the closed-form extreme value distribution), nesting together all the trip options as separate from the stay-at-home option, and is refered to as Model A0 below. In considering the various distributional assumptions, we also retain the "prefered" functional form specification

for the water quality variables identified in Table 4. Clearly, the specification search could be repeated for each distribution scenario (requiring an additional 8*32=256 models to be estimated), but such a task is beyond the scope of the current paper. Our focus is instead on illustrating the potential impact that such distributional assumptions can have on the resulting parameter estimates and the overall performance of the model in terms of prediction. As an aside, we note that the water quality parameters themselves were generally not substantially impacted by the distributional choices in terms of size, sign, or statistical significance.

Table 7 provides the comparison of the nine distribution scenarios in terms of maximum log-likelihood values.[22] Interestingly, the standard normal specification's performance is the second worst of the nine scenarios, dominating only model A0. Generally, triangular distributions appear to fit the data better in terms of both the lake size and discrete size characteristic parameters, with the results being more mixed for the participation parameter α_i. The preferred specification includes random triangular distributions for all of the site characteristics and a normal distribution for the participation parameter. In the remainder of this paper we refer to this specification as Model A1. Since Model A1 best fits the data, we consider it to be the best model for estimation, welfare analysis, and prediction. In the next section we turn to the estimation of Model A1 and other models for comparison.

Estimation Results

Given the results from the specification search, four models were used in the second third of the sample: Models A, A1, and A0 (varying the specification for the random parameters of the model) and a model including only Secchi depth as a measure of water quality (referred to as Model B hereafter).[23] We include Model B to illustrate the consequences of relying on a single measure of water quality, in this case one that is often available to analysts. The resulting parameter estimates are presented in three Tables, 8a, 8b, and 8c. For all of the models, the coefficients for the socio-demographic data, price, and the random coefficients on the amenities are given in Table 8a. Table 8b lists the coefficients for the physical water quality measures in all four models. Table 8c provides the estimated dispersion coefficients for the random parameters in the models. The dispersion coefficients are the estimated standard deviations for the normally distributed parameters, and the estimated spread coefficients for the triangularly distributed parameters.

In Table 8a, all of the coefficients are significant at the one percent level, except for the mean fish-index coefficients and a few socio-demographic coefficients in model A0. Note

that the socio-demographic data are included in the conditional indirect utility for the stay-at-home option. Therefore in model A, males, higher-educated individuals, and smaller households are all more likely to take a trip to a lake. Also, as age increases individuals are more likely to choose the stay-at-home option and therefore less likely to take trips. However, the size and sign of the age, school and household size variables are sensitive to the model specification used. The price coefficient is negative as expected and virtually identical in all four models.

Turning to the site amenities, again all of the parameters are of the expected sign. As the size of a lake increases, has a cement boat ramp, gains handicap facilities, or is adjacent to a state park, the average number of visits to the site increases. Notice, however, the large dispersion estimates in Table 8c. For example, in model A the dispersion on the size of the lake indicates 11 percent of the population prefers a smaller lake, possibly someone who enjoys a more private experience. The large dispersion on the "wake" dummy variable seems particularly appropriate given the potentially conflicting interests of anglers and recreational boaters. Anglers would possibly prefer "no wake" lakes, while recreational boaters would obviously prefer lakes that allow wakes. It seems the population favors a lake that allows wakes by almost a two to one margin, with 63 percent preferring a lake that allows wakes and 37 percent preferring a "no wake" lake. Lastly, the mean of α_i, the trip dummy variable, is negative, indicating that on average the respondents receive higher utility from the stay-at-home option, which is expected considering the average number of trips is 6.4 out of a possible 52 choice occasions.

The physical water quality coefficients are reported in Table 8b and are relatively stable across the various models. For all four models, the effect of Secchi depth is positive, while in both models A and A2 organic and inorganic (volatile) suspended solids have a negative impact (although ISS and VSS are not statistically significant at the 1% level for five of the six coefficients across the model specifications), indicating the respondents strongly value water clarity. However, the coefficient on logged chlorophyll is positive, suggesting that on average respondents do not mind some "greenish" water. The negative coefficient on logged total phosphorus, the most likely principal limiting nutrient, indicates higher algae growth leads to fewer recreational trips. High logged total nitrogen levels also have a statistically significant and negative impact on recreational utility associated with a site. The log of the possibly toxic cyanobacteria is negative, while the log of total Phytoplankton is positive, indicating that as cyanobacteria is a larger percentage of the total Phytoplankton in the

lake, recreators are less likely to visit.

Out-of-Sample Prediction

The final third of the survey sample was used to examine the out-of-sample performance of the models estimated in the previous subsection. For each model, the predicted number of total trips taken per year was computed for each household and used to calculate the sample's associated root mean square error (RMSE) in predicting the actual numbers of trips taken.[24] In addition, these calculations were made for the two subsamples of nonparticipants and participants, where the latter is defined as a household having visited at least one of the lake sites in 2002. The results are summarized in Table 9.

Unfortunately, none of the models perform particularly well in predicting the actual participation patterns in the third sample. The two models employing unbounded distributions for their random parameters (A and B) overstate total trips by nearly 80%, predicting total trips of roughly eleven, when the average number of trips is 6.27. Using bounded triangular distributions for the random parameters (model A1) not only provides a better fit to the model, but also reduces the overprediction out-of-sample by almost 60%. The simplest of the three models (model A0), while providing the worst fit from a log-likelihood perspective, comes closest to predicting the participation rate in the third sample, underpredicting total trips at 5.53 versus the actual numbers of trips at 6.27. For all of the model specifications the RMSE's are quite high, though this is not particularly surprising given the wide range of unobserved factors likely affecting the recreational usage patterns of individual households.

Focusing the analysis on the two subpopulations of participants and nonparticipants suggests that poor out-of-sample performance of the various models lies primarily in their poor prediction of nonparticipation. All three of the random parameter models (A, B, and A1) predict total trips for participants reasonably well, with model A1 performing the best in terms of RMSE. Indeed, the model mimicing the simpler nested logit structure (Model A0) performs the worst for participants, underpredicting total trips by over forty percent. In contrast, all of the models perform poorly in predicting nonparticipation. Models A and B predict average total trips of over nine per year for individuals who actually take no trips. Models A1 and A0 perform better, but still miss the mark in charactizing participation for this subsample. The results suggest that participants and nonparticipants are perhaps more distinct than the standard mixed logit structure allows for. Employing bounded distributions for the random parameters of the model appears to alleviate, but not eliminate, the problems in the overfitting of the mixed logit model in the current setting. Further research into this

problem, though beyond the scope of the current analysis, might consider less symmetry in the participation parameter α_i or a finite mixture model allowing for a discrete mass of individuals in the nonparticipant category (See [**?**]).

Welfare Calculations

Given the random parameters β_i, the conditional compensating variation associated with a change in water quality from Q to Q' for individual i on choice occasion t is

$$CV_{it}\left(\beta_i\right) = \frac{-1}{\beta^p}\left\{\ln\left[\sum_{j=0}^{J}\exp\left(V_{ijt}\left[Q';\beta_i\right]\right)\right] - \ln\left[\sum_{j=0}^{J}\exp\left(V_{ijt}\left[Q;\beta_i\right]\right)\right]\right\}$$

which is the compensating variation for the standard logit model. The unconditional compensating variation does not have a closed form, but it can be simulated by

$$CV_{it} = \frac{1}{R}\sum_{r=1}^{R}\frac{-1}{\beta^p}\left\{\ln\left[\sum_{j=0}^{J}\exp\left(V_{ijt}\left[Q';\beta_i^r\right]\right)\right] - \ln\left[\sum_{j=0}^{J}\exp\left(V_{ijt}\left[Q;\beta_i^r\right]\right)\right]\right\}$$

where R is the number of draws and r represents a particular draw from its distribution. The simulation process involves drawing values of β_i and then calculating the resulting compensating variation for each vector of draws, and finally averaging over the results for many draws.[25]

The average annual compensating variations and predicted changes in total annual trips per household are calculated for three water quality improvement scenarios using the parameter estimates from the four models reported in the previous section. The first scenario improves all 129 lakes to the physical water quality of West Okoboji Lake, the clearest, least impacted lake in the state. Table 10 compares the physical water quality of West Okoboji Lake with the average of the other 128 lakes. All of West Okoboji Lake's measures are considerably improved over the other 128. For example, West Okoboji Lake has slightly over five times the water clarity, measured by Secchi depth, of the other lakes. The second scenario is a less ambitious, though more realistic, plan of improving nine lakes to the water quality of West Okoboji Lake (see Table 10 for comparison). The state is divided into nine zones with one lake in each zone, allowing every Iowan to be within a couple of hours of a lake with superior water quality. The nine lakes were chosen based on recommendations by the Iowa Department of Natural Resources for possible candidates of a clean-up project. The third and final scenario is also a policy-oriented improvement. Currently of the 129 lakes, 65 are officially listed on the EPA's impaired waters list. TMDLs are being developed for these lakes and by 2009 the plans must be in place to improve the water quality at these lakes enough to remove them from the list. Therefore, in this scenario, the 65 impaired lakes would

be improved to the median physical water quality levels of the 64 non-impaired lakes. Table 11 compares the median values for the non-impaired lakes to the averages of the impaired lakes. The contrast between scenarios 2 and 3 is then between concentrating on major water quality improvements for a few targeted lakes versus more modest improvements to all the major impaired lakes in the state.

The resulting compensating variations and changes in total trips under each scenario are reported in Table 12. We consider the welfare results from Model A1 to be the best results for policy analysis. For comparison purposes, we present welfare results from Models A, A0, and B as well. Starting with the first scenario, we see both a substantial change in household welfare and in the predicted numbers of trips taken to lakes. For Model A, the CV is nearly $250 per Iowa household. This size of a welfare change seems reasonable given the numbers of trips taken in a typical year under baseline conditions. The travel cost savings alone from having closer (high quality) sites at which to recreate would justify much of these welfare gains.[26] Models B and A1 yield similar, though somewhat smaller, welfare gains (ranging from just under $190 to just over $200 per household). Not surprisingly, Model A0, which predicts many fewer trips per year, yields a lower welfare gain per household, though still substantial at $67 per year. The total trip changes exhibit a similar pattern, with models A, B, and A1 all predicting a 3.5 to 4.1 increase in trips per year, whereas model A0 yields a smaller increase in total trips of 1.8 per year.

Shifting to the second and third scenarios, we find similar patterns in terms of how the four models predict both welfare and total trip changes under these scenarios. The simpler nested logit-style model of A0 suggests smaller welfare and trip changes when compared with the more complex mixed logit models. Another message from Table 12, however, is consistent across the four specifications. That is, the benefits from improving a few lakes to a level of high water quality exceeds the benefits from modest improvements to all the impaired lakes. The CV benefits from scenario 2 are typically two and a half to four times those from scenario 3. The annual compensating variation for scenario 2 is $29 for each Iowa household (using model A1). As expected, this estimate is 15.3 percent of the value if all lakes were improved to the level of West Okoboji, even though the scenario involves improving only 7.0 percent of the lakes. This suggests location of the improved lakes is important and, to maximize Iowan's benefit from improving a few lakes, policymakers should consider dispersing them throughout the state.

Finally, comparing Model B and Model A across the three scenarios, indicates that

Model B, with only Secchi transparency included for water quality, reasonably proxies for the inclusion of all eight water quality measures when the water quality improvement is large. For the first scenario, with all 128 lakes improved to the level of West Okoboji, Model B's welfare value is only different from Model A by 17.6% ($203.55 vs. $246.98). For the third scenario, which is a moderate water quality improvement, Model B's welfare value is 54.0% less than Model A ($7.02 vs. $15.27). Possibly, for moderate or small water quality improvements, Secchi transparency alone does not capture as well the value from reduced bacteria or nutrient levels.

Conclusions

The first-year survey of the Iowa Lakes Project gathered information about the recreational behavior of Iowans at 129 of Iowa's principal lakes. These data were combined with extensive physical water quality measures from the same set of lakes gathered by the Iowa State University Limnology Lab. Our analysis, which employs the repeated mixed logit framework, shows that individuals are responsive to physical water quality measures, and it is possible to base willingness-to-pay calculations on improvements in these physical measures. In particular we considered three improvement scenarios, with the results suggesting that Iowans value more highly a few lakes with superior water quality rather than all recreational lakes that have only adequate levels (i.e., sufficient to not be listed as impaired by the Environmental Protection Agency).

A number of important practical findings come directly from this work. Limnologists and other water quality researchers should be interested in the results of this paper, since the general belief is that visitors care about water clarity as measured by Secchi depth (how many meters beneath the surface of the water a Secchi dish is visible) or water quality in general. By estimating the partial effects of a list of physical measures, we have determined which measures significantly affect recreationists' behavior. Limnologists and water resource managers can use this information about what physical lake attributes visitors' trip behavior responds to in designing projects for water quality improvements. Our results indicate water clarity is very important as evidenced by the Secchi dish and suspended solids parameters. Also, high concentrations of nutrients, in general, are found to decrease recreational trips.

The findings of this study also have direct relevance for environmental protection managers and citizens concerned with water quality in that they can be used to prioritize clean-up activities to generate the greatest recreational benefits for a given expenditure. Not only can the findings be used to determine which lakes to target and in what order to clean them but

also the most efficient levels of improvement can be identified.

Finally, our investigations into the econometric specification for the mixed logit model suggests that the results (both welfare implications and out-of-sample predictive ability) can be sensitive to the choice of the underlying distributions employed for the random parameters of the model. In general, both the model fit and predictive ability improved in the current application when the random parameters were bounded (in our case triangular), rather than unbounded. All of the models were found to do poorly in predicting the behavior of nonparticipants in our third sample. This suggests that additional research is needed to distinguish and characterize nonparticipants, perhaps by allowing for less symmetry in the random parameters of the model or some discrete distributions in the mixture models.

Table 1. 2002 Iowa Lakes Survey Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.
Total Day Trips	6.44	10.22	0	52
Male	0.68	0.46	0	1
Age	53.34	16.09	15	82
School	0.66	0.46	0	1
Household Size	2.61	1.30	1	12

Table 2. Water Quality Variables and 2002 Summary Statistics

Variable	Mean	Std. Dev.	Min.	Max.
Secchi Depth (m)	1.17	0.92	0.09	5.67
Chlorophyll (ug/l)	41	38	2	183
NH3+NH4 (ug/l)	292	159	72	955
NO3+NO2 (mg/l)	1.20	2.54	0.07	14.13
Total Nitrogen (mg/l)	2.20	2.52	0.55	13.37
Total Phosphorus (ug/l)	106	81	17	453
Silicon (mg/l)	4.56	3.24	0.95	16.31
pH	8.50	0.33	7.76	10.03
Alkalinity (mg/l)	142	41	74	286
Inorganic SS (mg/l)	9.4	17.9	0.6	177.6
Volatile SS (mg/l)	9.4	7.9	1.6	49.9
Cyanobacteria (mg/l)	295.8	833.1	0.01	7178.1
Total Bacteria (mg/l)	304.8	835.2	3.99	7178.6

Table 3. Summary Statistics for Lake Site Characteristics

Variable	Mean	Std. Dev.	Min.	Max.
Travel Cost	135.09	29.23	93.90	238.61
Acres	672	2,120	10	19,000
Ramp	0.86	0.35	0	1
Wake	0.66	0.47	0	1
State Park	0.39	0.49	0	1
Handicap Facilities	0.39	0.49	0	1
Fish Index	1.01	1.14	0	4

Body Math

Table 4: Model Specifications Listed in Decreasing Log-Likelihood Values

		Variable Group Specification			
Secchi Depth	Chlorophyll	Nitr. and Phosp.	Susp. Solids	Bacteria	Log-Likelihood
Linear	Log	Log	Linear	Log	-39010.2
Linear	Log	Log	Log	Log	-39016.2
Linear	Log	Linear	Log	Log	-39029.7
Log	Log	Log	Linear	Log	-39032.9
Linear	Log	Linear	Linear	Log	-39034.7
Log	Log	Log	Log	Log	-39037.7
Linear	Linear	Linear	Log	Log	-39045.6
Linear	Linear	Linear	Linear	Log	-39056.9
Log	Log	Linear	Log	Log	-39063.4
Log	Log	Linear	Linear	Log	-39065.6
Log	Linear	Linear	Log	Log	-39067.3
Log	Linear	Linear	Linear	Log	-39071.1
Linear	Linear	Log	Log	Log	-39087.6
Linear	Linear	Log	Linear	Log	-39090.1
Log	Linear	Log	Linear	Log	-39097.5
Log	Linear	Log	Log	Log	-39104.0
Linear	Log	Log	Linear	Linear	-39172.1
Linear	Log	Log	Log	Linear	-39183.0
Linear	Log	Linear	Log	Linear	-39196.9
Linear	Log	Linear	Linear	Linear	-39198.5
Log	Log	Log	Linear	Linear	-39201.3
Log	Log	Log	Log	Linear	-39208.4
Linear	Linear	Linear	Log	Linear	-39218.0
Linear	Linear	Log	Log	Linear	-39219.6
Linear	Linear	Log	Linear	Linear	-39227.1
Log	Linear	Log	Linear	Linear	-39231.9
Log	Log	Linear	Log	Linear	-39232.7
Log	Log	Linear	Linear	Linear	-39235.9
Linear	Linear	Linear	Linear	Linear	-39236.3
Log	Linear	Log	Log	Linear	-39240.9
Log	Linear	Linear	Log	Linear	-39242.5
Log	Linear	Linear	Linear	Linear	-39248.3

Body Math

Table 5: Most Important Single Additional Water Quality Measure

Secchi Transparency and	Log-Likelihood	Change in the Log-likelihood[a]
Log(Total phosphorus)	-39275.4	33.6
ISS	-39279.9	29.1
Log(Chlorophyll)	**−39282.6**	**26.4**
Log(Cyanobacteria)	-39288.8	20.2
Log(Total Nitrogen)	-39303.9	5.1
Log(Total Phytoplankton)	-39307.4	1.6
VSS	-39308.1	0.90
nothing	-39309.0	

[a]The change in the log-likelihood from adding Secchi transparency versus no water quality measures is 47.5 (-39356.5 vs. -39309.0).

Table 6: Most Important Pair of Additional Water Quality Measures

Secchi Transparency and	Log-Likelihood	Change in the Log-likelihood
Log(Cyanobacteria)/Log(TotalPhytoplankton)	**−39173.1**	**135.9**
Log(Chlorophyll) / Log(Total Phosphorus)	-39194.8	114.2
Log(Chlorophyll) / ISS	-39243.2	65.8
Log(Cyanobacteria) / Log(Total Phosphorus)	-39250.9	58.1
Log(Cyanobacteria) / ISS	-39262.4	46.6
Log(Total phosphorus) /ISS	-39262.5	46.5
Log(Chlorophyll) / VSS	-39270.4	38.6
Log(Chlorophyll) / Log(Total Nitrogen)	-39271.9	37.1
Log(Cyanobacteria) / Log(Chlorophyll)	-39272.5	36.5
Log(Totalphosphorus)/Log(TotalNitrogen)	**−39274.2**	**34.8**
Log(Total phosphorus) / Log(Total Phytoplankton)	-39274.5	34.5
Log(Total Phosphorus) / VSS	-39274.5	34.5
Log(Chlorophyll) / Log(Total Phytoplankton)	-39275.8	33.2
VSS/ISS	**−39277.0**	**32.0**
Log(Total Phytoplankton) / ISS	-39278.2	30.8
Log(Total Nitrogen) / ISS	-39278.5	30.5
Log(Cyanobacteria) / Log(Total Nitrogen)	-39282.0	27.0
Log(Cyanobacteria) / VSS	-39288.2	20.8
Log(Total Phytoplankton) / Log(Total Nitrogen)	-39303.0	6.0
Log(Total Nitrogen) / VSS	-39304.0	5.0
Log(Total Phytoplankton) / VSS	-39307.3	1.7
nothing	-39309.0	

Table 7: Log-Likelihood Comparison Across Distribution Scenarios

	Site Characteristic Group Specification		
Lake size	Discrete Site Vars.	Participation Parameter (α_i)	Log-Likelihood
Triangular	Triangular	Normal	-38677.5
Triangular	Normal	Normal	-38688.6
Normal	Triangular	Triangular	-38700.1
Triangular	Triangular	Triangular	-38701.3
Triangular	Normal	Triangular	-38704.1
Normal	Normal	Triangular	-38704.7
Normal	Triangular	Normal	-38716.1
Normal	Normal	Normal	-38718.6
Fixed	Fixed	Normal	-40946.8

Repeated Mixed Logit Model Parameter Estimates[a]
(Std. Errs in Parentheses)

Variable	Model A	Model B: Secchi Only	Model C: Linear WQ	Model A1	Model A0:
			Distributional Assumptions on Random Parameters		
	All Normal	All Normal	All Normal	Triang/Normal	Fixed/Normal
Male	-7.92^*	-6.65^*	-9.36^*	-6.63^*	-5.33^*
	(0.56)	(0.45)	(0.52)	(0.53)	(1.60)
Age	1.13^*	-0.23^*	-0.73^*	-0.35^*	-0.43
	(0.09)	(0.08)	(0.09)	(0.09)	(0.27)
Age^2	-0.007^*	0.006^*	0.009^*	0.006^*	0.008^*
	(0.001)	(0.001)	(0.0008)	(0.001)	(0.003)
School	-2.11^*	5.81^*	-2.59^*	5.70^*	-2.62
	(0.55)	(0.53)	(0.56)	(0.54)	(1.66)
Household	0.75^*	-0.68^*	0.36	-1.88^*	-1.41
	(0.20)	(0.18)	(0.17)	(0.22)	(0.66)
Price	-0.49^*	-0.49^*	-0.49^*	-0.49^*	-0.50^*
	(0.0011)	(0.001)	(0.001)	(0.001)	(0.001)
Log(Acres)[b]	4.69^*	4.02^*	4.07^*	4.62^*	4.13^*
	(0.09)	(0.07)	(0.08)	(0.08)	(0.04)
Ramp[b]	12.49^*	11.51^*	13.17^*	22.01^*	5.33^*
	(0.93)	(0.88)	(1.01)	(1.18)	(0.35)
Facilities[b]	2.07^*	1.37^*	1.12^*	1.42^*	2.20^*
	(0.29)	(0.26)	(0.30)	(0.29)	(0.12)
State Park[b]	3.90^*	4.04^*	3.53^*	2.51^*	2.78^*
	(0.32)	(0.31)	(0.33)	(0.34)	(0.15)
Wake[b]	3.79^*	3.88^*	1.60^*	2.08^*	3.25^*
	(0.31)	(0.30)	(0.31)	(0.33)	(0.13)
Fish Index[b]	0.04	-0.12	-0.94^*	-0.09	0.01
	(0.12)	(0.12)	(0.12)	(0.11)	(0.05)
α[b]	-11.90^*	-11.80^*	-11.93^*	-12.46^*	-11.78^*
	(0.05)	(0.04)	(0.05)	(0.05)	(0.08)

Body Math [*] Significant at 1% level.

Body Math [a] All of the parameters are scaled by 10, except α (which is unscaled).

Body Math [b]The coefficients reported for these variables are the means of their respective random parameters. The dispersion estimates for the random parameters are provided in table 8c.

Table 8b. Repeated Mixed Logit Model Parameter Estimates[a]
(Std. Errs. in Parentheses)

Variable	Model A	Model B: Secchi Only	Model C: Linear WQ	Model A1	Model A0
Secchi Depth (m)	2.40*	1.77*	1.72*	2.49*	1.38*
	(0.10)	(0.06)	(0.08)	(0.09)	(0.07)
Log(Chlorophyll (ug/l))	2.37*		2.24*	2.42*	1.52*
	(0.16)		(0.41)	(0.17)	(0.12)
Log(Total Nitrogen (mg/l))	−1.16*		−0.18*	−0.84*	−0.44*
	(0.10)		(0.02)	(0.09)	(0.07)
Log(Total Phosphorus (ug/l))	−2.83*		−0.02*	−2.87*	−2.30*
	(0.17)		(0.002)	(0.16)	(0.12)
Inorganic SS (mg/l)	−0.006		−0.01	0.001	−0.014*
	(0.005)		(0.005)	(0.005)	(0.004)
Volatile SS (mg/l)	−0.02		−0.01*	−0.01	−0.007
	(0.02)		(0.02)	(0.02)	(0.01)
Log(Cyanobacteria)	−2.44*		−.001	−2.39*	−1.86*
	(0.12)		(0.008)	(0.12)	(0.72)
Log(Total Phytoplankton)	3.44*		0.002	3.28*	2.19*
	(0.15)		(0.008)	(0.14)	(0.93)
LogLik	-37,627	-37,759	-37,743	-37,568	-39,960

Body Math *Significant at the 1% level.

Body Math [a] All of the parameters are scaled by 10.

Table 8c. Dispersion Coefficients for Repeated Mixed Logit Model[a,b]
(Std. Errs in Parentheses)

Variable	Model A	Model B: Secchi Only	Model C: Linear WQ	Model A1	Model A0:
Log(Acres)	3.81*	3.00*	3.40*	7.43*	
	(0.07)	(0.06)	(0.07)	(0.02)	
Ramp	16.86*	17.30*	18.99*	60.44*	
	(0.90)	(0.65)	(0.89)	(1.87)	
Facilities	13.46*	13.95*	14.22*	29.17*	
	(0.34)	(0.29)	(0.31)	(0.62)	
State Park	12.35*	13.99*	12.88*	36.77*	
	(0.28)	(0.27)	(0.33)	(0.92)	
Wake	11.73*	13.98*	14.10*	32.47*	
	(0.27)	(0.34)	(0.35)	(0.90)	
Fish Index	6.36*	5.41*	4.62*	10.50*	
	(0.12)	(0.11)	(0.14)	(0.27)	
α	2.76*	2.76*	3.01*	2.93*	2.29*
	(0.04)	(0.03)	(0.04)	(0.04)	(0.06)

[b]The dispersion coefficients are the estimated standard deviations for the normally distributed paramete
and the estimated spread coefficients for the triangularly distributed parameters.

Table 9. Comparing Predicted Trips across Model Specifications[a]

	Actual	Model A	Model B	Model C	Model A1	Model A0
Mean Trips	6.27	10.97	11.11	10.94	8.27	5.53
		(10.73)	(10.87)	(10.65)	(9.78)	(9.60)
Nonparticipants	0	9.29	9.53	9.57	7.12	4.45
		(10.33)	(10.65)	(10.40)	(8.00)	(5.38)
Participants	10.77	12.17	12.24	11.92	9.09	6.31
		(11.00)	(11.03)	(10.83)	(10.88)	(11.73)

[a] Root mean square error reported in parentheses.

Table 10. West Okoboji Lake vs. the other 128 Lakes

	West Okoboji Lake	Averages of the other 128 Lakes	Averages of the 9 Zone Lakes
Secchi Depth (m)	5.67	1.14	1.23
Chlorophyll (ug/l)	2.63	41.23	40.13
Total Nitrogen (mg/l)	0.86	2.21	3.64
Total phosphorus (ug/l)	21.28	106.31	91.11
Inorganic SS (mg/l)	1.00	9.49	9.52
Volatile SS (mg/l)	1.79	9.40	8.42
Cyanobacteria (mg/l)	22.00	297.98	948.00
Total Phytoplankton (mg/l)	24.06	307.04	953.37

Table 11. 64 Non-impaired Lakes vs. the 65 Impaired Lakes

	Median of the 64 Non-impaired Lakes	Averages of the 65 Impaired Lakes
Secchi Depth (m)	1.27	0.81
Chlorophyll (ug/l)	23.25	56.67
Total Nitrogen (mg/l)	1.11	2.31
Total phosphorus (ug/l)	58.79	139.91
Inorganic SS (mg/l)	3.51	14.78
Volatile SS (mg/l)	6.02	12.93
Cyanobacteria (mg/l)	42.96	516.96
Total Phytoplankton (mg/l)	47.48	528.65

Table 12. Annual Unconditional Compensating Variation Estimates

a. CV Per Iowa Household Per Year			
Model	All 128 Lakes Improved to W. Okb.	9 Zone Lakes Improved to W. Okb.	65 Impaired Lakes Improved to Median
Model A	$246.98	$40.25	$15.27
Model B	$203.55	$33.02	$7.02
Model C	$226.61	$33.44	$11.61
Model A1	$188.88	$28.92	$10.08
Model A0	$67.30	$7.71	$2.99

b. Increase in Total Trips			
Average CV	All 128 Lakes Improved to W. Okb.	9 Zone Lakes Improved to W. Okb.	65 Impaired Lakes Improved to Median
Model A	4.13	0.68	0.22
Model B	3.49	0.57	0.10
Model C	3.71	0.56	0.17
Model A1	3.51	0.55	0.15
Model A0	1.78	0.22	0.06

Table 13. Annual Conditional Compensating Variation Estimates

	a. CV Per Iowa Household Per Year		
	All 128 Lakes	9 Zone Lakes	65 Impaired Lakes
Model	Improved to W. Okb.	Improved to W. Okb.	Improved to Median
Model A	$148.70	$18.88	$10.64
Model B	$122.26	$17.14	$6.17
Model C	$105.59	$14.41	$5.41
Model A1	$147.46	$18.87	$9.82
Model A0	$87.91	$10.82	$9.19
	b. Increase in Total Trips		
	All 128 Lakes	9 Zone Lakes	65 Impaired Lakes
Average CV	Improved to W. Okb.	Improved to W. Okb.	Improved to Median
Model A	3.63	0.51	0.28
Model B	2.91	0.42	0.13
Model C	1.37	0.19	0.06
Model A1	3.26	0.46	0.24
Model A0	2.25	0.28	0.25

References

Arbuckle, K.E. and J.A. Downing (2001). "The influence of watershed land use on lake N:P in a predominantly agricultural landscape." *Limnology and Oceanography* **46**: 970-975.

Azevedo, C.D., K.J. Egan, J.A. Herriges, and C.L. Kling (2003). *Iowa Lakes Valuation Project: Summary and Findings from Year One.* Final Report to the Iowa Department of Natural Resources, August.

Bockstael, N.E., W.M. Hanemann, and I.E. Strand (1986). *Measuring the Benefits of Water Quality Improvements Using Recreation Demand Models*, report presented to the Environmental Protection Agency under cooperative agreement CR-811043-01-0, Washington, D.C.

Chen, H. Z., F. Lupi, and J. P. Hoehn (1999). "An Empirical Assessment of Multinomial Probit and Logit Models for Recreation Demand," in Herriges, J.A., and C. L. Kling, *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice*, Cheltenham, UK: Edward Elgar, pp. 141-161.

Creel, M.D., and J.B. Loomis (1990). "Theoretical and Empirical Advantages of Truncated Count Data Estimators for Analysis of Deer Hunting in California." *American Journal of Agricultural Economics* **72**(2): 434-41.

Creel, M.D., and J.B. Loomis (1992). "Modeling Hunting Demand in the Presence of a Bag Limit, with Tests of Alternative Specifications." *Journal of Environmental Economics and Management* **22**(2): 99-113.

Dillman, D. A. (1978) *Mail and Telephone Surveys – The Total Design Method*, New York: Wiley.

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

Hensher, D., and W. Greene (2003). The Mixed Logit Model: The State of Practice." *Transportation* **30**: 133-176.

Hess, S., K.E. Train, and J.W. Polak (2006). "On the use of a Modified Latin Hypercube Sampling (MLHS) in the Estimation of a Mixed Logit Model for Vehicle Choice." Transportation Research: Part B: Methodological, 40 (2) (February): 147-63.

Iowa Department of Natural Resources (2004a), *Fishing Guide for Iowa Lakes*.

Iowa Department of Natural Resources (2004b), *Iowa Conservationist,* **63**(2): 7-25.

Leamer, E.E. (1983). "Let's Take the Con out of Econometrics." *American Economic Review* **73**(1): 31-43.

McFadden, D., and K. E. Train (2000). "Mixed MNL Models for Discrete Response." *Journal of Applied Econometrics* **15**(September/October): 447-70.

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

Parsons, G.R., E. C. Helm, and T. Bondelid (2003). "Measuring the Economic Benefits of Water Quality Improvements to Recreational Users in Six Northeastern States: An Application of the Random Utility Maximization Model." Working paper, University of Delaware, July.

Parsons, G.R., and M.J. Kealy (1992). "Randomly Drawn Opportunity Sets in a Random Utility Model of Lake Recreation." *Land Economics* **68**(1): 93-106.

Phaneuf, D.J., C.L. Kling, and J.A. Herriges (2000). "Estimation and Welfare Calculations in a Generalized Corner Solution Model with an Application to Recreation Demand." *The Review of Economics and Statistics* **82**(1) (February): 83-92.

Revelt, D., and K. Train (1998). "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level." The Review of Economics and Statistics 80: 647-57.

Train, K. (1998). "Recreation Demand Models with Taste Differences Over People." *Land Economics* **74** (2) (May): 230-239.

U.S. Environmental Protection Agency (USEPA) (2000). "Nutrient Criteria Technical Guidance Manual: Lakes and Reservoirs." Office of Water, Office of Science and Technology,

Report EPA-822-B00-001, Washington, D.C.

von Haefen, R.H., D.M. Massey, and W.L. Adamowicz (2005). "Serial Nonparticipation in Repeated Discrete Choice Models." American Journal of Agricultural Economics 87(4): 1061-1076.

Notes

[1]TMDLs specify the amount of a pollutant that a water body can receive and still meet existing water quality standards.

[2]As of March 2003, there have been approximately 40 legal actions taken against the USEPA in 38 states concerning the implementation of Section 303(d) of the CWA.

[3]More specifically, physical (e.g. Secchi depth), chemical (e.g. total phosphorus), and biological data (e.g. cyanobacteria) is collected, but for brevity we use the term physical measures to collectively represent all three categories.

[4]The limnological study is funded by the Iowa Department of Natural Resources.

[5]We are aware of only two other recreation demand studies that have adopted this procedure, see Creel and Loomis ([?], [?]) and Haener and Adamowicz.

[6]Complete details of the survey design and implementation can be found in [?].

[7]Sensitivity analysis, raising the allowable number of trips per year to ninety-five, indicated that the results do not change qualitatively, including the welfare results.

[8]The authors would like to thank Spencer Banzhaf for pointing out the concern with the correlated standard errors, and one of the reviewers for pointing out the concern with the misspecification error.

[9]The unit of observation is the household. Although, we also include the survey respondent's *individual* socio-demographic data.

[10]As noted by one of the reviewers, it is useful to have in mind the recreational activities the survey respondents participate in at the lakes. One question in the survey asked the respondents, "What activities do you or members of your household typically participate in during your lake visits? (Check all that apply)." The top six activities received the following percentages: recreational boating (58.8%), fishing (47.4%), picnicking (46.6%), nature appreciation/wildlife viewing (44.9%), swimming/beach use (37.7%), and camping (28.5%), N = 3,859.

[11]The $0.25 per mile is used as a relatively conservative estimate of gasoline and depreca-

tion costs per mile of driving. This estimate is generally less than most official government reimbursement rates. Concerning the computation of the round-trip travel distance and time, we used the software package, *PCMiler (Streets Version 17)*.

[12]The "average wage rate" is calculated for all respondents as their household's income divided by 2,000 (assuming total annual hours worked is 40 hours per week for 50 weeks).

[13]As one of the reviewers points out, the handicap facilities variable probably in general proxies for easy accessibility of the lake.

[14]The candidate fish species are bluegill, crappie, largemouth and smallmouth bass, catfish, bullhead, walleye, sunfish, yellow perch, and northern pike . See [?] for details.

[15]Modified Latin Hypercube Sampling (MLHS) is used in the simulation process (Hess, Train, and Polak, [?]). Hess, Train, and Polak demonstrate that MLHS provides better coverage of the distribution space for the same number of draws versus Halton sequences or any of the Halton sequence variations (such as shuffled Halton sequences).

[16]It is possible to interact the socio-demographic data with the sites if one believed, for example, that age would affect which lake was chosen.

[17]Later in this section we discuss alternative distributional assumptions for the β_i^a's.

[18]The number of draws used in the simulation is 200 for the results in Tables 4, 5, and 6.

[19]We also estimated exploratory models utilizing the quadratic function for the water quality measures. The results indicate that a model with Secchi transparency and the suspended solids entered quadratically (and the nutrients and bacteria levels remaining logged) fits better than model A. However, we do not formally include the quadratic function in our analysis, as the versions of the model would increase from 32 to 243 ($=3^5$).

[20]See, e.g., Revelt and Train [?], endnote 14 for a discussion of this issue.

[21]Hensher and Greene [?] provide an excellent discussion of the alternative distributional assumptions commonly used in mixed logit models and their limitations.

[22]For the results in Table 7, 750 draws are used in the simulation. Changing the distributional assumptions across specifications required a large number of draws to obtain stability

in the comparisons. The estimation results in Table 8 are also with 750 draws.

[23]An additional model (Model D), including all thirteen water quality characteristics listed in Table 2, was also estimated and is available from the authors upon request. The results from this expanded model are not substantially different from the results for Model A, with Silicon, pH and Alkalinity also found to be statistically significant factors in recreation site choice.

[24]We calculate RMSE as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left(\widehat{Y_i} - Y_i\right)^2}{N}},$$

where Y_i is the total trips for each person, and $\widehat{Y_i}$ is the fitted total trips for each person. Our definition of RMSE precludes a decomposition into a bias and variance portion.

[25]We use 1,000 draws in the welfare simulations.

[26]West Okoboji Lake, while one of the most popular lakes in the state currently, is far from most population centers in Iowa.