

Modeling Recreation Demand when the Access Point is Unknown¹

Yongjie Ji Joseph A. Herriges² Catherine L. Kling
Iowa State University Iowa State University Iowa State University

Preliminary Draft - Please do not Quote without Permission

October 5, 2013

Abstract

The task of modeling the recreation demand for geographically large sites, such as rivers and beaches or large parks with multiple entrances, is often challenged by incomplete information regarding the access point used by the individual. Traditionally, analysts have relied upon convenient approximations, defining travel time and travel distances on the basis of the midpoint of a river or beach segment or on the basis of the nearest access point to the site for each individual. In this paper, we instead treat the problem as one of aggregation, drawing upon and generalizing results from the aggregation literature. The resulting model yields a consistent framework for incorporating information on site characteristics and travel costs gathered at a finer level than that used to obtain trip counts. We use a series of Monte Carlo experiments to illustrate the performance of the traditional mid-point and nearest access point approximations. Our results suggest that, while the nearest access point approach provides a relatively good approximation to underlying preferences for a wide range of parameter specifications, use of the midpoint approach to calculating travel cost can lead to significant bias in the travel cost parameter and corresponding welfare calculations. Finally, we use our approach in modeling recreation demand for the major river systems in Iowa using data from the 2009 Iowa Rivers and River Corridors Survey.

¹This paper was supported by Hatch Act and State of Iowa funds. Funding for this project was also provided by the Iowa Department of Natural Resources and the U.S. Environmental Protection Agency's Science to Achieve Results (STAR) program. Although the research described in the article has been funded in part by the U.S. Environmental Protection Agency through grant R83426001, it has not been subjected to any EPA review and therefore does not necessarily reflect the views of the Agency, and no official endorsement should be inferred.

²Contact author information: 260 Heady Hall, Department of Economics, Iowa State University, Ames, IA 50011. email: jaherrig@iastate.edu. Phone: 515-294-4964

1 Introduction

Recreation demand (or travel cost) models provide one of the primary tools for valuing environmental amenities, inferring value by observing the full costs incurred by the individual or household in reaching sites in a choice set. While there are a myriad of conceptual issues in defining travel costs themselves (see, e.g., [6] and [23]), practitioners are typically content with computing these costs as the sum of out-pocket costs (usually a fixed mileage rate times round-trip travel distance) and an opportunity cost for the individual's travel time (often valued at a fixed fraction of the individual's wage rate times round-trip travel time). However, in applications where the specific access point used to visit the "site" is unknown, the appropriate way to compute the cost of access can be unclear. A prime example is river based recreation. Surveys can elicit information on the number of trips to one or more river segments during the course of a season, but typically do not acquire information on the precise access point used by the individual on each choice occasion. This makes the computation of travel costs problematic in that the analyst cannot precisely compute either the travel distance or travel time. At best, in these cases travel costs can be bounded by considering the nearest and furthest access points along the river segment. Similar problems emerge in the context of beach (e.g., [13]) and wetland (e.g., [20]) recreation, or, more generally, any large geographic regions.

A number of solutions to this problem are employed in the literature, including computing travel costs based upon the nearest access point for each individual or using the midpoint along the river segment as the assumed point of entry for everyone. The issue with these *ad hoc* approaches is that they implicitly make assumptions regarding the role of travel costs (the marginal utility of income) in the individual's decision making that are inherently inconsistent with the broader models used to represent the choice *among* river segments. For example, using the nearest access point along a river segment to compute travel cost implicitly assumes that travel cost is *the* determining factor in choosing where along a river segment to recreate (essentially assuming that the marginal utility of income is infinite), whereas in the broader models of the choice among river segments (say a RUM model) travel cost is but one of the factors in site selection (implying a finite marginal utility of income) (e.g., [13] and [20]).

In this paper, we consider an alternative approach, treating segment level trip data as the aggregation of underlying access-point level trip information. A logit structure is used to construct the choice probabilities for this aggregated data and to recover preference parameters. A series of Monte Carlo exercises are used to compare and contrast the performance

of the mid-point and nearest access point (or shortest distance) approximations used in the literature. The simulation results suggest that our model successfully recovers the underlying preference parameters, while the two traditional approaches vary in their estimation of the key travel cost parameter and subsequent welfare estimates. The shortest distance model generally provides a good approximation over a wide range of model parameterizations. In contrast, the commonly used midpoint model generates bias in both the travel cost parameter and subsequent welfare estimates that increases substantially as the number of river segments increases and travel costs become a more important determinant of behavior. In addition, we apply our approach, along with the midpoint and nearest access point approaches, using data from the 2009 Iowa River Survey. The survey was conducted in late 2009, eliciting information on the visitation patterns of 10000 randomly chosen Iowans to 73 identified river segments in the state.

The remainder of this paper is organized as follows. Section 2 provides a brief review of the literature. Our modeling approach to handling the missing access point data is then described in section 3. Section 4 describes a Monte Carlo exercise used to illustrate the scope of the bias from using either the midpoint or nearest access point approximations to travel cost. Finally, section 5 describes the 2009 Iowa Rivers project application, including both a description of the data and the resulting parameter estimates. Section 6 concludes the paper.

2 Related Literature

The problem of missing access point data is directly related to the issue of site aggregation encountered in recreation demand analysis (See, e.g., [3], [17] and [19]). Whereas missing access point data essentially forces the aggregation of possible “sites,” practitioners have often intentionally aggregated elementary sites for computational convenience. In this literature, a wide variety of aggregation schemes have been considered, including county level aggregation, activity based aggregation, aggregation of familiar or unfamiliar sites and distance-based aggregation (See, e.g., [4], [17], and [18]). The potential bias associated with aggregation is the major concern in this context. The magnitude of bias depends on the degree of the aggregation and the heterogeneity across the aggregated elementary sites. Unfortunately, the nonlinearity of the RUM model makes the direction of bias generally ambiguous with respect to both preference parameter estimates and in terms of the welfare change induced by site loss or a change in site characteristics.

Kaoru and Smith [10] were the first to analyze the effects of aggregation on preference parameter estimation and welfare measurement in the context of recreation demand. Their work suggested that models with only a mild degree of site aggregation (i.e., 35 sites aggregated to 23 or 11 composite sites) performed relatively well in characterizing recreation behavior. The results, however, were not as promising in terms of subsequent welfare calculations. For example, the welfare impact from the closure of an aggregate site was understated by more than a factor of two using either site aggregations. The estimated welfare gain from site quality improvements faired even worse, being understated by a factor of five when 11 composite sites were used (See Kaoru *et al.* [11]).

Parson and Needelman's [17] subsequent paper identified two distinct sources of bias stemming from site aggregation, one linked to the number of sites being aggregated (the so-called *size* effect) and the other tied to the degree of heterogeneity among the sites being combined. Specifically, drawing on earlier work in the transportation literature by Ben-Akiva and Lerman [1], Parsons and Needleman note that, if the utility received by individual i from choosing an elementary site j is given by

$$U_{ij} = V_{ij} + \epsilon_{ij} \quad j = 1, \dots, J, \quad (1)$$

where the ϵ_{ij} 's are distributed *i.i.d.* Gumbel with mode 0 and scale parameter μ , then the utility associated with choosing the aggregate site s ($s = 1, \dots, S$) is given by:³

$$U_{is} = \max_{j \in A_s} U_{ij} \quad (2)$$

$$= \bar{V}_{is} + \mu \ln J_s + \mu \ln B_{is} + \epsilon_{is}, \quad (3)$$

where A_s denotes the set of elementary sites associated with the aggregate site s ,

$$\bar{V}_{is} = \frac{1}{J_s} \sum_{j \in A_s} V_{ij}, \quad (4)$$

J_s denotes the number of sites associated with aggregate site s ,

$$B_{is} = \frac{1}{J_s} \sum_{j \in A_s} \exp[\mu^{-1}(V_{ij} - \bar{V}_{is})], \quad (5)$$

and the ϵ'_{is} s are again distributed *i.i.d.* Gumbel with mode 0 and scale parameter μ . Estimating a model of aggregate site choice using only average site characteristics (including travel cost) corresponds to specifying that the utility from visiting aggregate site s is given by:

$$U_{is} = \bar{V}_{is} + \epsilon_{is} \quad (6)$$

³The standard deviation of the ϵ_{ij} 's is given by $\mu\pi/\sqrt{6}$. Note that the scale parameter referred to in Parsons and Needleman corresponds to our μ^{-1} .

Comparing equations (3) and (6), it is clear that the latter specification suffers potential bias due to two omitted variables: (a) a *size* variable reflecting the number of sites in the aggregate alternative s (i.e., $\ln J_s$ in (3)) and (b) a measure of the heterogeneity of the sites being combined (reflected by $\ln B_{is}$ in (3)).⁴ The general nature of the problem does not change if, in lieu of \bar{V}_{is} , an alternative proxy (V_{is}^p) is used to characterize the aggregate site utility (e.g., by using the nearest access point or site midpoint to determine travel cost). In this case, equation (3) simply becomes

$$U_{is} = V_{is}^p + \mu \ln J_s + \mu \ln B_{is}^p + \epsilon_{is}, \quad (7)$$

where

$$B_{is}^p = \frac{1}{J_s} \sum_{j \in A_s} \exp[\mu^{-1}(V_{ij} - V_{is}^p)]. \quad (8)$$

Modeling aggregate site utility using $U_{is} = V_{is}^p + \epsilon_{is}$ would again be subject to omitted variables bias.

Parsons and Needleman provide empirical evidence as to the scope of aggregation bias. Specifically, using data on fishing trips to 1133 lakes in Wisconsin, they estimate three models: one using the full choice set as in (1), a second using site aggregates with only average site characteristics as in (6), and a third using site aggregates with both average site characteristics and a size correction (but no heterogeneity correction). Two levels of aggregation are considered (9 regions and 61 counties). The results suggest that ignoring both heterogeneity and size factors leads to significant bias in parameter estimates (except for the price coefficient) and that, while the size correction alone works well with limited aggregation, the size corrected model performs poorly when large numbers of sites are aggregated. The authors suggest minimizing heterogeneity of sites within aggregates and controlling for the number of sites in the aggregate groups. A series of subsequent papers have largely confirmed the findings in Parsons and Needleman [17] (including Kaoru, Smith, and Liu [11], Feather [3], Feather and Lupi [4] and Parson, Plantinga and Boyle [18]).

In a more recent paper, Haener *et al.* [7] suggest that, while analysts may choose not to model detailed site visitation data, they often have access to detailed site characteristics data, including travel costs. As such, they should be able to form both the *size* and *heterogeneity* correction terms in equation (3) and obtain consistent parameter estimates. Their empirical analysis, however, suggests that the size correction alone mitigates much of the aggregation bias. For their application, site heterogeneity appears to not play a significant role.

⁴Note that while the impact of these omitted variables can be mitigated by making the aggregates similar in size and minimizing the degree of heterogeneity across sites in terms of site attributes, site heterogeneity will necessarily persist in the form of heterogeneous travel costs to the elemental sites.

3 Methodology

The approach followed in the aggregation literature is based on the underlying structure of the logit model, yielding the specific *size* and *heterogeneity* correction terms identified in equation (3). The problem, however, is more general and the solution need not rely on the logistic structure. Rather than observing the elemental site visitation data, we observe only whether one of a series of sites is visited. Specifically, in the case of a single choice, let y_{isj} equal 1 when individual i visits elemental site j in aggregate site s (and equals 0 otherwise). Observing visitation data for the aggregate site s corresponds to observing $y_{is\bullet}$, where

$$y_{is\bullet} = \sum_{j=1}^{J_s} y_{isj}, \quad (9)$$

where J_s denotes the number of elemental sites in the aggregate site s . The corresponding choice probability for the aggregate site is then simply the sum of the individual choice probabilities; i.e.,

$$P_{is\bullet} = Pr[y_{is\bullet} = 1] \quad (10)$$

$$= \sum_{j=1}^{J_s} Pr[y_{isj} = 1] = \sum_{j=1}^{J_s} P_{isj}. \quad (11)$$

In this section, we begin by laying out the implications of this aggregation in the context of a simple repeated logit model of riverine recreation, linking it to the existing literature on site aggregation. We then extend to model to the nested and mixed logit settings and to the case in which site attributes are available at the elemental site level.

The Repeated Logit Model

Following Morey, Rowe and Watson [15], we model an individual's riverine recreation using a repeated random utility maximization (RUM) framework. In particular, we assume that there are T choice occasions in each year. On each choice occasion, the individual decides either to visit one of the river segments or to stay at home. There are S river segments and J_s access points along the segment. The conditional utility that household i receives from visiting river segment s ($s = 1, \dots, S$) via access point j ($j = 1, \dots, J_s$) on choice occasion t ($t = 1, \dots, T$) is assumed to take the form

$$U_{isjt} = \alpha_s + \beta C_{isj} + \epsilon_{isjt}, \quad (12)$$

where α_s is a segment specific constant reflecting all segment attributes and C_{isj} is the

travel cost of reaching access point j along river segment s for household i .⁵ The error term ϵ_{isjt} captures unobserved factors influencing the choice made by the household. Letting $s = j = 0$ denote the option of choosing to stay at home on a given choice occasion, the relevant conditional utilities can be summarized as

$$U_{isjt} = \begin{cases} \epsilon_{isjt} & \text{if } s = j = 0 \\ V_{isj} + \epsilon_{isjt} & \text{otherwise} \end{cases} \quad (13)$$

where $V_{isj} = \alpha_s + \beta C_{isj}$, and V_{i00} has been normalized to zero for the stay-at-home option. Assuming that the ϵ_{isjt} 's are *i.i.d.* type I extreme value random variables, individual i will choose to visit the segment-access point combination sj , denoted by $y_{isjt} = 1$, with the probability of

$$P_{isjt} = Pr(y_{isjt} = 1) = \frac{\exp(V_{isj})}{1 + \sum_{r=1}^S \sum_{k=1}^{J_r} \exp(V_{irk})} = P_{isj} \quad \forall t. \quad (14)$$

If the elementary choices made by households (i.e., the y_{isjt} 's) were observed, we could form the appropriate likelihood function on the basis of equation (14) and estimate the parameters of the model. Instead, information is only provided at the segment level; i.e.,

$$y_{is\bullet t} = \sum_{j=1}^{J_s} y_{isjt}. \quad (15)$$

However, we can still use equation (14) to construct the relevant choice probabilities. In particular, we have

$$P_{is\bullet t} = Pr(y_{is\bullet t} = 1) = \frac{\sum_{j=1}^{J_s} \exp(V_{isj})}{1 + \sum_{r=1}^S \sum_{k=1}^{J_k} \exp(V_{irk})} = P_{is\bullet} \quad \forall t \quad (16)$$

where $y_{is\bullet t}$ equals 1 if the individual chooses to visit the segment s at some unknown access point along this segment.⁶ These aggregate probabilities provide the basis for estimating a repeated logit model using the aggregated data and maximum likelihood estimation. In

⁵The assumption that there is single segment specific constant, rather than an alternative specific constant for each access point (i.e., an α_{sj}), implicitly assumes that there is no heterogeneity in site attributes along the river segment, an assumption that will be relaxed below. Given this assumption, the only source of heterogeneity across access points is in terms of the travel cost C_{isj} .

⁶A similar approach was suggested and applied by Kurkalova and Rabotyagov [14] in a binary model when county level, rather than farm level, data was available in an agricultural technology adoption setting.

particular, the contribution of individual i to the log-likelihood function is given by:

$$\mathcal{L}_i(\mathbf{n}_i) = \sum_{s=0}^S n_{is\bullet} \ln \left(\sum_{j=1}^{J_s} P_{isj} \right) \quad (17)$$

$$\begin{aligned} &= \sum_{s=0}^S n_{is\bullet} \ln(P_{is\bullet}) \\ &= \left\{ \sum_{s=1}^S n_{is\bullet} \ln \left[\sum_{j=1}^{J_s} \exp(V_{isj}) \right] \right\} - T \cdot \ln \left[1 + \sum_{r=1}^S \sum_{k=1}^{J_r} \exp(V_{irk}) \right], \end{aligned} \quad (18)$$

where $\mathbf{n}_i = (n_{i0\bullet}, \dots, n_{iS\bullet})$ and $n_{is\bullet} = \sum_{t=1}^T y_{is\bullet t}$ denotes the total number of times aggregate alternative s is chosen across the T choice occasions. Note that the specification of the log-likelihood function in (17) holds in general when only aggregate data are available, but that (18) holds specifically for the logit formulation of the choice probabilities. From a programming point of view, (18) provides all that is needed in terms of estimation.⁷ It is not necessary to reduce the expression further. However, it is instructive to do so. Specifically, note that we can rewrite equation (16) as

$$P_{is\bullet} = \frac{\exp(V_{is\bullet})}{1 + \sum_{n=1}^S \exp(V_{in\bullet})} \quad (19)$$

where

$$V_{is\bullet} = \ln \left[\sum_{j=1}^{J_s} \exp(V_{isj}) \right] \quad (20)$$

$$\begin{aligned} &= \ln \left[\sum_{j=1}^{J_s} \exp(\alpha_s) \exp(\beta C_{isj}) \right] \\ &= \alpha_s + \ln \left[\sum_{j=1}^{J_s} \exp(\beta C_{isj}) \right] \end{aligned} \quad (21)$$

$$= \alpha_s + \beta C_{is\bullet} \quad (22)$$

with

$$C_{is\bullet} \equiv \frac{1}{\beta} \ln \left[\sum_{j=1}^{J_s} \exp(\beta C_{isj}) \right]. \quad (23)$$

The term $C_{is\bullet}$ can be thought of as the aggregate price for segment s . Indeed, viewing (23) as a function of the access point travel costs (i.e., the C_{isj} 's), a first order Taylor-series approximation of $C_{is\bullet}$ around the mean segment travel cost yields

$$C_{is\bullet} \approx \sum_{j=1}^{J_s} P_{isj|s} C_{isj} + \frac{1}{\beta} \ln J_s \quad (24)$$

⁷This assumes, of course, that the model remains identified, an issue that is returned to below.

where $P_{isj|s} \equiv P_{isj}/P_{is\bullet}$ denotes the probability that access point j is chosen, given segment s has been selected. The segment level travel cost is just a probability weighted average of the access point travel costs, with a correction for the size (J_s) of the aggregate site.⁸ Because $C_{is\bullet}$ involves the unknown preference parameter β , there is no promising way to construct it *ex ante* for use in estimation. Consequently, the conditional indirect utility function for the aggregate site s (i.e., $V_{is\bullet}$) is no longer linear in the parameter β , as can be seen in (21), making estimation potentially more difficult.⁹

The alternative approach typically used in the literature is to replace $C_{is\bullet}$ in equation (22) with a proxy (C_{is}^p), computing travel cost of the basis of either the nearest access point ($p = \text{min}$) or the midpoint ($p = \text{mid}$) of the segment. The advantage of doing so is that the corresponding conditional indirect utility function is once again linear in its parameters, with $V_{is\bullet}^p = \alpha_s + \beta C_{is}^p$. The problem, as noted above, is that the subsequent parameter estimates will be subject to omitted variable bias, since:

$$\begin{aligned}
V_{is\bullet} &= \alpha_s + \ln \left[\sum_{j=1}^{J_s} \exp(\beta C_{isj}) \right] \\
&= \alpha_s + \beta C_{is}^p + \ln \left[\sum_{j=1}^{J_s} \exp(\beta [C_{isj} - C_{is}^p]) \right] \\
&= \alpha_s + \beta C_{is}^p + \ln(J_s) + \ln \left[\frac{1}{J_s} \sum_{j=1}^{J_s} \exp(\beta [C_{isj} - C_{is}^p]) \right] \\
&= V_{is\bullet}^p + \ln(J_s) + \ln \left[\frac{1}{J_s} \sum_{j=1}^{J_s} \exp(\beta [C_{isj} - C_{is}^p]) \right], \tag{25}
\end{aligned}$$

where the last two terms are the size and heterogeneity corrections identified in the aggregation literature.¹⁰ The minimum distance proxy, $C_{is}^{\text{min}} = \min_{j=1, \dots, J_j} \{C_{isj}\}$, has the intuitively appealing property that the omitted variable bias disappears as the marginal utility of income increases (i.e., $V_{is\bullet}^{\text{min}} \rightarrow V_{is\bullet}$ as $\beta \rightarrow -\infty$).

Finally, the analysis above assumes that the conditional utilities (V_{isjt}) derived from the

⁸Substituting (24) into (22), $V_{is\bullet}$ can also be written as $V_{is\bullet} = \alpha_s^* + \beta C_{is\bullet}^*$ where $\alpha_s^* \equiv \alpha_s + \ln J_s$ and $C_{is\bullet}^* = \sum_{j=1}^{J_s} P_{isj|s} C_{isj}$ is simply the probability weighted average of the access point travel costs. Note that two commonly used proxies for aggregate travel cost take a similar form, weighing the access point travel cost. The nearest travel cost proxy puts all the weight on a single site (i.e., the closest one), whereas the average travel cost proxy gives equal weight to all sites. The aggregate probability approach uses information about the relative cost of sites, and the marginal utility of income, to construct these weights.

⁹It is, however, the case that $V_{is\bullet}$ is still linear in the segment specific constants α_s . As a result, the model will still be mean fitting (i.e., the actual segment shares will precisely equal the mean fitted shares) and the contraction mapping algorithm outlined in Murdock [16] can still be used in estimation.

¹⁰Here we have normalized the scale parameter $\mu = 1$.

elemental access sites differ only terms of travel cost, sharing a common segment specific constant. As argued in Appendix A to this paper, more general structures allowing for access point specific constants (i.e., with $V_{isj} = \alpha_{sj} + \beta C_{isj}$) or access-point level characteristics (e.g., $V_{isj} = \alpha_s + \beta C_{isj} + \delta Z_{sj}$) are identified even when only segment level counts are available. However, these models are only *poorly* identified, relying entirely on the logit structure to distinguish α_{sj} or δ from a segment level ASC.¹¹ This weakness is potentially exacerbated by limited variation in site attributes within a segment. In the extreme, if there is no variation in access point attributes, their impact would be captured by the segment specific constant α_s .

The Nested Logit Model

While the focus of our Monte Carlo and empirical analysis below is on the mixed logit generalization of the logit model, in this section we touch briefly on the implications of aggregated data for the more traditional nested logit models. We consider two nested logit specifications.

Specification 1: Trip Nest

In the first specification, all of the segments (and their associated access points) are grouped together in a single nest. In this case, the choice probability for access point j becomes:

$$P_{isj} = \exp(\tilde{V}_{isj}) \left[\sum_{r=1}^S \sum_{k=1}^{J_r} \exp(\tilde{V}_{irk}) \right]^{\theta-1} \left\{ 1 + \left[\sum_{r=1}^S \sum_{k=1}^{J_r} \exp(\tilde{V}_{irk}) \right]^{\theta} \right\}. \quad (26)$$

where

$$\tilde{V}_{isj} = \frac{V_{isj}}{\theta} = \frac{\alpha_s}{\theta} + \frac{\beta C_{isj}}{\theta} = \tilde{\alpha}_s + \tilde{\beta} C_{isj}. \quad (27)$$

As shown in Appendix A, the choice probability for aggregate site s retain the general nested logit structure, with

$$P_{is\bullet} = \exp(\tilde{V}_{is\bullet}) \left[\sum_{r=1}^S \sum_{k=1}^{J_r} \exp(\tilde{V}_{ir\bullet}) \right]^{\theta-1} \left\{ 1 + \left[\sum_{r=1}^S \sum_{k=1}^{J_r} \exp(\tilde{V}_{ir\bullet}) \right]^{\theta} \right\} \quad (28)$$

where

$$\tilde{V}_{is\bullet} = \tilde{\alpha}_s + \tilde{\beta} \tilde{C}_{is\bullet}. \quad (29)$$

¹¹One need only consider the alternative linear probability model to see this. In the case of the linear probability model, the segment level probability becomes a function of the sum of the access point ASC's or attributes, precluding them from being distinguished from an overall segment level alternative specific constant.

with

$$\tilde{C}_{is\bullet} = \frac{1}{\tilde{\beta}} \ln \left[\sum_{j=1}^{J_s} \exp(\tilde{\beta} C_{isj}) \right]. \quad (30)$$

Appendix A also provides technical details regarding the identification of the underlying parameters of the model (i.e., α_s , β , and θ).

Specification 2: Segment Nests

In the second specification, the access points within each segment form distinct nests. In this case, (26) is replaced with

$$P_{isj} = \exp(\check{V}_{isj}) \left[\sum_{k=1}^{J_s} \exp(\check{V}_{isk}) \right]^{\theta_s - 1} \left\{ 1 + \sum_{r=1}^S \left[\sum_{k=1}^{J_r} \exp(\check{V}_{irk}) \right]^{\theta_r} \right\}, \quad (31)$$

where

$$\check{V}_{isj} = \frac{V_{isj}}{\theta_s} = \frac{\alpha_s}{\theta_s} + \frac{\beta C_{isj}}{\theta_s} = \check{\alpha}_s + \check{\beta}_s C_{isj}. \quad (32)$$

Now (as shown in Appendix A), the choice probability for segment s becomes

$$P_{is\bullet} = \exp(\check{V}_{is\bullet}) \left\{ 1 + \sum_{r=1}^S \exp(\check{V}_{ir\bullet}) \right\}^{-1} \quad (33)$$

where

$$\check{V}_{is\bullet} = \alpha_s + \beta \check{C}_{is\bullet} \quad (34)$$

with

$$\check{C}_{is\bullet} = \frac{1}{\check{\beta}_s} \ln \left[\sum_{j=1}^{J_s} \exp(\check{\beta}_s C_{isj}) \right]. \quad (35)$$

Again, Appendix A provides technical details regarding the identification of the underlying parameters of the model (i.e., α_s , β , and θ_s). It is worth noting that, unlike in the previous case, the segment choice probabilities look like a standard logit model. Identification of θ_s (which is equivalent to identification of $\check{\beta}_s$ given β) hinges on the structure of the nonlinear relationship in (35). One observation regarding (35) is that as $\theta_s \rightarrow 0$, the travel cost index $\check{C}_{is\bullet} \rightarrow C_{is}^{min}$, which reduces the variability needed to identify θ_s .

Normal Error Component Logit Mixture Models

Normal Error Component Logit Mixture (NECLM) models have become a popular alternative to nested logit model as a means of inducing correlation patterns among alternatives in the

choice set (See, e.g., Herriges and Phaneuf [9] and Walker *et al.* [25]). Unobservable factors, shared by one or more of the alternatives, are introduced into the conditional utility functions in the form of normally distributed error components. One of the advantages the approach provides is the ability to create more complex and overlapping nests, rather than relying on the usual tree structure assumed by nested logit. At the same time, as Walker *et al.* [25] note, model identification can be more difficult to establish and spurious results can be obtained for models that are not identified if care is not taken in simulating the requisite choice probabilities. In this section, we discuss the identification of several NECLM model in the context of aggregate choice data. Though aggregate choice probabilities analogous to (16) hold, the identification of parameters in the model becomes difficult to establish. The requisite conditions for identification developed in Walker *et al.* [25] are employed.

Specification 1: Trip Nest

An error component structure similar to specification 1 of the nested logit model in the previous section would replace (13) with

$$\tilde{U}_{isjt} = \begin{cases} \epsilon_{isjt} & \text{if } s = j = 0 \\ V_{isj} + \tau_{it} + \epsilon_{isjt} & \text{otherwise} \end{cases} \quad (36)$$

where $\tau_{it} \sim N(0, \sigma^2)$ and ϵ_{isjt} is distributed *i.i.d.* Gumbel with mode zero and scale parameter μ . In this case, the corresponding utility for the aggregate segment level alternative becomes:

$$\begin{aligned} \tilde{U}_{is\bullet t} &= \max_{j=1, \dots, J_s} \tilde{U}_{isjt} \\ &= \max_{j=1, \dots, J_s} (V_{isjt} + \tau_{it} + \epsilon_{isjt}) \\ &= \left(\max_{j=1, \dots, J_s} [V_{isjt} + \epsilon_{isjt}] \right) + \tau_{it} \\ &= V_{is\bullet} + \tau_{it} + \epsilon_{is\bullet t} \end{aligned} \quad (37)$$

where $\epsilon_{is\bullet t}$ is distributed *iid* Gumbel with mode zero and scale parameter μ . The unconditional choice probabilities have the same form as those for the disaggregate data, *except* that the aggregate cost variable becomes $C_{is\bullet}$ in equation (30); i.e.,

$$P_{is\bullet} = \int \frac{\exp(V_{is\bullet} + \tau)}{1 + \sum_{n=1}^S \exp(V_{in\bullet} + \tau)} f(\tau) d\tau \quad (38)$$

where $f(\tau)$ denotes the pdf for τ_{it} . Appendix A provides additional discussion regarding identification of the parameters of this model (i.e., α_s , β and σ).¹²

Specification 2: Segment Nests

An error component structure similar to specification 2 of the nested logit model in the previous section would replace (13) with

$$\tilde{U}_{isjt} = \begin{cases} \epsilon_{isjt} & \text{if } s = j = 0 \\ V_{isj} + \tau_{ist} + \epsilon_{isjt} & \text{otherwise} \end{cases} \quad (39)$$

where $\tau_{ist} \sim N(0, \sigma_s^2)$ and ϵ_{isjt} is distributed *i.i.d.* Gumbel with mode zero and scale parameter μ . In this case, the corresponding utility for the aggregate segment level alternative becomes:

$$\begin{aligned} \tilde{U}_{is\bullet t} &= \max_{j=1, \dots, J_s} \tilde{U}_{isjt} \\ &= \max_{j=1, \dots, J_s} (V_{isjt} + \tau_{ist} + \epsilon_{isjt}) \\ &= V_{is\bullet} + \tau_{ist} + \epsilon_{is\bullet t} \\ &= \alpha_s + \beta C_{is\bullet} + \tau_{ist} + \epsilon_{is\bullet t} \end{aligned} \quad (40)$$

where $\epsilon_{is\bullet t}$ is distributed *i.i.d.* Gumbel with mode zero and scale parameter μ . The model in equation (40) is analogous to the *alternative-specific variance model* considered by Walker *et al.* [25], which the authors establish is identified. The only difference in this context is that $C_{is\bullet}$ is a nonlinear function of the model parameter β .

4 Monte Carlo Simulation

The goal of this section is to evaluate the performance of the two standard travel cost proxies (i.e., the midpoint and shortest distance measures), relative to using the aggregated travel cost index $C_{is\bullet}$, in recovering preference parameters and calculating welfare change associated with the loss of a river segment. The simulation scenarios vary the river and parameter configurations along four dimensions:

- Price responsiveness: As noted above, the shortest distance proxy becomes more appropriate as travel cost becomes the dominant consideration in site selection. Thus we

¹²Similar results apply if it is assumed that τ_{it} is constant over time with $\tau_{it} = \tau_i \sim N(0, \sigma^2)$.

would expect the use of this proxy to create less bias as β increases. We consider the three levels of β shown in Table 1.

- Number of river segments: We consider three levels for the number of aggregate river segments S ($S=5,10$ and 20).
- River/Population configurations: Ferguson and Kanaroglou [5] note that the shape of the spatial object (river segments in this paper) and the spatial distribution of households will affect the heterogeneity among the aggregated sites, though they did not examine it's specific impact on estimation results. We consider four possible configurations for river segments and population centers. In two of the configurations, the rivers are assumed to be straight segments, 50 miles in length, whereas in the other two configurations the rivers are 50 miles long, but kinked at the midpoint. Population is either uniformly distributed or centers around the first two segments. Thus we have four possibilities along this dimension:
 - B: The base scenario with straight river segments and no population centers
 - K: Kinked river segments, with no population centers
 - P: Straight river segments, with population centers
 - C: The combination of kinked river segments and population centers
- Water quality: We consider two types of conditional utility functions. The first consists of segment specific constants along with travel cost (i.e., as depicted in equation 12). The second includes an additional term, representing say water quality, along the lines of $V_{isj} = \alpha_s + \beta C_{isj} + \delta Z_{sj}$, with Z_{sj} representing water quality.

A total of 72 ($2 \times 4 \times 3 \times 3$) Monte Carlo scenarios were considered, with 100 replications for each scenario. In each case, a simple logit structure is employed, though similar results are obtained when nesting among sites is allowed. Details of the data generation process are provided in Appendix B.

For each scenario, three models are used to recover the underlying preference parameters: two based on standard proxies for the aggregate site travel cost (i.e., the midpoint and shortest distance proxies) and one based on the aggregated choice probabilities. In addition, we consider the performance of the models in estimating the welfare costs associated with the closure of river segment 1 using the standard log-sum formula. In the context of the simple logit structure, this reduces to

$$CV_1 = \frac{1}{\beta} \ln(1 - P_{i1\bullet}) \approx \frac{-P_{i1\bullet}}{\beta} \quad (41)$$

Since all three models have the basic logit structure with segment level alternative specific constants, they are mean fitting (i.e., the fitted choice probabilities will equal the average observed choice shares). Any bias in CV_1 will be driven by bias in the travel cost parameter.

4.1 Simulation Results

Table 2 summarizes the results for the first 36 of the Monte Carlo experiments; i.e., those without access-point level attributes (labeled here as water quality). We focus our attention on the travel cost parameter β , as it is the main determinant of subsequent welfare measures, with the alternative specific constants (the α_j 's) changing to insure that the model is mean-fitting. The first half of Table 2 provides the mean absolute percentage error in the estimated travel cost parameter, β . Several results emerge here. First, the aggregated choice probabilities approach successfully recovers the underlying travel cost parameter, with a mean absolute percentage error of 0.3 percent or less. This should not be surprising, since it represents the true data generating process in this case. The aggregation of choice data to the segment level represents a loss of information, and hence efficiency, but does not alter the underlying model. Second, the shortest distance proxy also performs quite well, with a mean absolute percentage error that is typically less than 5 percent. However, the midpoint proxy does not do as well, particularly when both the number of segments and the travel cost parameters are large. When $S = 20$ and $\beta = -0.1$, the mean absolute percentage error ranges between 15.8 and 19.6 percent. As we would expect, bias in the travel cost parameter translates directly into bias for the corresponding welfare measures. Table 2b compares the mean absolute percentage error for CV_1 across the same 36 Monte Carlo experiments. Again, the aggregated choice probabilities approach successfully recovers the true compensating variation associated with the loss of site 1, and shortest distance proxy works reasonably well. However, the midpoint proxy results in mean absolute percentage errors exceeding twenty-five percent.

Table 3 reports similar summary statistics for the cases in which water quality is included in households' recreational utility function at the access point level. Before proceeding with describing the results in Table 3, two facts should be noted. First, there are two versions of the shortest distance model. In version 1, the water quality associated with the nearest site for each individual is used as their water quality measure, whereas in version 2 the average water quality over the entire segment is used. Second, for two of the specifications (i.e., the midpoint model and the shortest distance model, version 2), the water quality does not vary by individual and, hence, β_w is not identified (being collinear with the alternative specific

constants).

Starting with the results in Table 3a for the travel cost coefficient, the findings are similar to those in Table 2a. The aggregated choice probability model does a good job in recovering the underlying price coefficient, with the nearest distance proxy performing reasonably well. The midpoint proxy again suffers from the largest bias, with the mean absolute percentage error being higher when the price coefficient is larger and when there are more segments in the choice set. In Table 3b, we see that the aggregated choice probability model recovers the water quality parameter reasonably well, with a mean absolute percentage error that is typically less than five percent, though it reaches as high as 9.8 percent. Interestingly, the error rate appears to be highest when the price coefficient is small, perhaps because in that situation the water quality factor becomes a more dominant determinant of the individual's choice. The shortest distance proxy, in contrast, does a poor job in recovering β_w , with the mean percentage error typically exceeding fifty percent. Finally, as Table 3c indicates, the inclusion of access point water quality attributes does not change the basic conclusions in terms of the estimated compensating variation associated with losing a site (CV_1). This is not surprising, as CV_1 is largely driven by the travel cost coefficient. Again, we find that the aggregated choice probability approach has the lowest mean absolute percentage error, followed by the shortest distance proxy and then the midpoint proxy approach.

5 An Application to Iowa Rivers

This section of the paper provides an application of the aggregated choice probability approach to the study of recreational river usage in Iowa. The primary data source for our analysis is the 2009 Iowa River Survey, funded by the Iowa Department of Natural Resources and the USEPA. The purpose of the survey was to gather baseline information about riverine recreation along 73 key river and stream segments in the state, depicted in Figure 1. The survey, conducted by mail, elicited data from each respondent regarding their total number of trips in 2009 to each of the river segments, as well as information regarding the individual's socio-demographic characteristics. However, information is not available regarding the specific point used by recreationists to access a given river or stream segment. With the segments ranging in length from 26 to 121 miles, considerable uncertainty exists in the imputed travel costs to the segment. We use the aggregated choice probability model to implicitly construct a travel cost index for each segment. The results are compared to models estimated using both the shortest distance and midpoint specifications commonly employed

in the literature.¹³

5.1 Models

A total of three models are estimated using the Iowa Rivers data. All of the models are based on repeated logit version of the normal error component logit mixture (NECLM). Specifically, we employ a structure similar to (36), but with $\tau_{it} = \tau_i \sim \mathcal{N}(0, \sigma_\tau^2)$. This creates a nesting of all trip alternatives and a common error term inducing correlation across choice occasions. The models start from the same basic structure for access-point level utility, with

$$\tilde{U}_{isjt} = \begin{cases} \gamma S_i + \epsilon_{isjt} & \text{if } s = j = 0 \\ \alpha_s + \beta C_{isj} + \tau_i + \epsilon_{isjt} & \text{otherwise} \end{cases} \quad (42)$$

where S_i denotes socio-demographic characteristics of individual i , potentially influencing their propensity to take trips. The three models differ in terms of how they handle aggregation, with one employing aggregated choice probabilities, while the other two employ the shortest distance and midpoint travel cost proxies, respectively.¹⁴ Following Murdock [16], a second stage regression of the alternative specific constants on segment characteristics is used to examine the role of site characteristics on recreation demand. Specifically, we run the second stage regression:

$$\hat{\alpha}_s = \alpha_0 + \delta Z_s + \xi_s \quad (43)$$

where Z_s denotes a vector of observable site characteristics for segment s .

5.2 The Iowa Rivers Data

After focus groups and pre-testing of the survey instrument, the 2009 Iowa Rivers Survey was mailed to a total of 10,000 Iowa households, beginning in November of 2009. Multiple mailings of the survey, as well as a postcard reminder and an incentive of \$12 for completing the survey, were used to increase survey response. Among all the surveys mailed, 4758 surveys were returned, for an overall response rate of 49% among deliverable surveys. Of the returned surveys, 4137 are employed in the analysis below.¹⁵ Table 4 provides a summary of

¹³A copy of the survey instrument is available from the authors upon request.

¹⁴A fourth model, employing aggregated choice probabilities *and* allowing for access level ASC's, was also estimated and yielded qualitatively similar results to those obtained using the specification in (42).

¹⁵A total of 176 returned surveys were unusable because the respondents did not provide their numbers of visits to the river segments depicted in Figure 1. An additional 445 respondents reported taking more than

the demographic characteristics of the survey respondents. The demographic statistics show that, on the average, a respondent to the survey tends to be an older, female Iowan with college level or equivalent education.¹⁶

Table 5 provides an overall summary of the data on trips to each of the river segments, both in terms of the percentage of the respondents who report visiting a given river segment and in terms of total numbers of trips to the segments. As the data indicate, the segments vary considerably in terms of popularity. River segment 71 (the Mississippi River between Clinton and Muscatine, Iowa) is the most popular, visited a total of 1591 times by just under six percent of the sample. At the other end of the spectrum, segment 37 (Big Cedar Creek, in northwest Iowa) is the least popular, visited only 20 times by 0.24 percent of the sample. In total, forty-eight percent of the sample visit at least one river segment during the course of the year, with an average number of trips per year of over six.

The travel cost variables (C_{isj} , C_{is}^{min} , and C_{is}^{mid}) must be constructed for each access point and river segment. There are several issues in doing so. First, a complete set of access points are not available for the Iowa rivers and streams.¹⁷ In the current application, we divide each of the river segments into approximately twenty-mile sub-segments, defining “access points” in terms of the mid-point of each of these sub-segment. This process results in a total of 300 access-points. The numbers of sub-segments per segment ranges from one for river segment 1 (Rock River in northwest Iowa) to eight for river segment 56 (for portions of the Wapsipinicon River in eastern Iowa). Second, given these access points, travel cost must be calculated. PCMiller is used to compute both the round-trip distance (d_{isj}) and travel time (t_{isj}) between the individual’s home and the relevant access point. Travel costs are then computed as $C_{isj} = g \cdot d_{isj} + (w_i/3) \cdot t_{isj}$, where g is per mile vehicle cost and w_i denotes the individual wage rate.¹⁸ As indicated at the top of Table 6, the average round

52 trips to individual river segments and were excluded from the sample. The focus of our analysis is on day-trips to the river segments. Setting the maximum numbers of trips to 52 allows for one trip per week. While this specific cut-off is arbitrary, the goal here is to focus on day trips and to exclude individuals who report large numbers of trips simply because they live on or near a specific river segment. Similar cutoffs have been used in other recreational studies (e.g., Egan *et al.* [2]) and were not found to significantly impact the results of the analysis.

¹⁶According to US Census, approximately 32 percent of adult Iowans are over 60, whereas in the survey sample this figure is somewhat higher at 36 percent. Likewise, among respondents the percentages of females (70 versus 50.4) and those with college degrees (69 versus 24.9) are higher than in the general Iowa population. The average family size of 2.4 is virtually the same in the sample as in the Iowa population as a whole.

¹⁷Indeed, for a number of activities, such as hunting, bird watching, etc., it is not clear what criteria to use in defining access points.

¹⁸The value of g was set to 54 cents per mile based on the 2009 AAA annual driving cost for an average sedan with 15,000 miles per year driving. The wage rate w was set at household income divide by 2000 times the number of adults in the household.

trip travel cost is approximately \$162.85, ranging from close to zero to almost \$500.

The final data category consists of river site characteristics (i.e., the Z_s 's) summarized in Table 6. The following river characteristics were constructed:

- *LENGTH* indicates the length of the river segment;
- *CANOE* indicates the percentage of the river segment that is considered canoeable, as defined by the Iowa Department of Natural Resources (IDNR);
- *OUTCROPPING* is a count of the number of outcroppings along the river segment, thought to contribute to the scenic nature of the river;
- *WATERBODY*, *WETLAND FOREST*, *GRASS*, *CROP*, and *DEVELOPED* indicate the percentage of the river corridor (defined as 75 meters on either side of the center-line for the river) that is water, wetland, forest, grassland, cropland and developed (industrial, commercial or residential) land, respectively;
- *IWQI* denotes a water quality index developed by the IDNR;¹⁹
- *MIWQI* is a dummy variable equal to one for river segments for which the *IWQI* is not available;
- *FISH* denotes the number of fish species found along the river segment;²⁰ and
- *MFISH* is a dummy variable that equals one for river segments without fish species data.

As Table 6 indicates, just over sixty percent of the Iowa river segments are canoeable. Not surprisingly, cropland is the largest form of land cover along the river segments (close to forty percent), with forested land being the second most common at under thirty percent. In terms of our two primary water quality measures, there are several important factors to note. First, both measures are available for only a fraction of the segments or sub-segments. Second, water quality along river segments is particularly difficult to capture, as rivers quality levels are only measured at selective sites. Even short distances from the monitoring site, water quality can be substantially different, depending on the river currents. Finally, the trips included a wide range of activities, from birdwatching to swimming, with the water

¹⁹This water quality index differs from USEPA's national water quality index. One drawback of the *IWQI* measure is that it is available for less than 70 percent of the river segments.

²⁰This information was provided directly by the IDNR.

quality measure being more or less salient depending upon the specific activity engaged in. With these concerns in mind, we consider below the use of turbidity as an alternative proxy for overall water quality. Turbidity is a measure of the cloudiness of water and, as such, is readily visible to recreators. As with *IWQI*, turbidity is only available for a fraction of river segments, so *MTURBIDITY*, a dummy variable for segments missing turbidity data, is also included in the model.

5.3 Results

The models estimated using the Iowa Rivers data are summarized in Tables 7 and 8. As indicated in Section 5.1, all three of these models employ a normal error component logit mixture (NECLM) structure and are estimated in two stages. The first stage involves estimating the alternative specific constants (i.e., ASC's α_s), the travel cost coefficient (β), and the parameters associated with the sociodemographic factors thought to influence the individual's propensity to stay at home (γ), as well as the variance of the trip error component (i.e., σ_τ^2).

Given the large number of ASC's (73), we refrain from reporting all of them here. However, the ASC's can be used as an indicator of the relative "appeal" of each site, controlling for travel cost. That is, all else equal, a site with a larger ASC is preferred to a site with a lower ASC. Table 7 illustrates the implied ranking of the river segments on the basis of the estimated ASC's from each model. The second column in Table 7 also provides the ranking of sites by total visitation. Not surprisingly, all of the top river segments based on visitation rates are near population centers. For example, segments 23 and 24 rank 6th and 3rd, respectively, in terms of total visitation and are located near the state's largest city, Des Moines. However, the rankings of these segments drop substantially once travel cost is controlled for. Segment 24's rank falls to 9th or 10th on the basis of its estimated ASC in models 1 through 3, while segment 23 falls out of the list of top ten sites altogether. For models 1, 2 and 3, the relative "appeal" of the river segments do differ, though they share some common features. Eight of the top ten segments and six of the bottom ten segments are the same across the three models.

Table 8a provides the other stage 1 parameter estimates. As expected, the travel cost coefficient is negative and statistically significant under all three specifications. Indeed, β varies relatively little across the three models. The three models also generally agree as to the impact of age and boat ownership, with older individuals and those without a boat being

more likely to stay at home. On the other hand, differences do emerge in terms of other factors. For example both females and college educated individuals are found to be more likely to take trips in models 2 and 3, whereas these factors are statistically insignificant in model 1. Also, larger households are significantly less likely to take trips according to model 1. Perhaps most importantly, the variance of the trip error component (i.e., σ_τ^2) is substantially larger in model 1 than in models 2 and 3, indicating a greater similarity across (and correlation among) the utilities received from trip options.²¹ This is analogous to a greater degree of “nesting” among the trip options in a nested logit setting. As will be seen below, this has implications for estimated welfare effects.

The second stage associated with estimating models 1 through 3 involves regressing the estimated segment level ASC’s on segment characteristics, as depicted in equation (43). The results are reported in Table 8b. The parameter estimates again suggest some consistency across the three specifications. In all three cases, river segments that are canoeable, relatively wide (i.e., with a larger value for *WATERBODY*), and contain a larger number of fish species are more appealing (i.e., have a larger ASC). Not surprisingly, the segments associated with the border rivers (i.e., the Mississippi and Missouri Rivers) are also more popular, as these rivers provide opportunities for activities (such as power boating) not available for most other river segments. Longer river segments are significantly less appealing according to Model 1, whereas they are significantly more appealing according to Model 3. Finally, while the parameter on Iowa’s water quality index (*IWQI*) has the expected sign, it is not statistically significant in any of the models. This is not surprising. As suggested above, *IWQI* is likely to be a poor measure of the perceived water quality along the river segments. First of all, it is measured for only a portion of the river segments and then only at specific monitoring sites. Moreover, given the wide range of activities associated with the trips being reported in the survey (from hiking and bird watching, to swimming and fishing), the *IWQI* is likely to be salient only for a fraction of the reported trips.

At the bottom of Table 8b, we provide the predicted welfare implications associated with closure of river segment 71 (the most visited site) and the closure of all 73 river segments. The three models provide significantly different welfare estimates. The model based on aggregate choice probabilities for the river segments (i.e., model 1) yields a substantially higher welfare estimate under both scenarios. The compensating variation associated with closing site 71 is over thirty percent higher in model 1 than in model 2 (which uses the shortest distance proxy for travel cost), and nearly twenty percent high than in model 3. The former result is not surprising, since model 2 is guaranteed to understate the travel cost associated with

²¹The correlation coefficient could be calculated through the formula, $\sigma_\tau^2/(\sigma_\tau^2 + \pi^2/6)$, where $\pi^2/6$ is the variance of a standard logistic distribution.

visiting a river segment and, hence, undervalue the resulting lost trips. The differences are even larger when closing all seventy-three river segments.

Given the limitations associated with the Iowa water quality index (*IWQI*), Table 9 provides results from an alternative specification for the Stage 2 model of the ASC's (i.e., equation 43), in which *IWQI* is replaced with *TURBIDITY*. The qualitative findings for most variables are similar to those from Table 8b. However, in the case of the aggregate choice probability model, turbidity is found to be a negative and statistically significant factor. On the hand, while fish stocks still have a positive coefficient, they are no longer a statistically significant factor. These results highlight the importance of viewing all three water quality measures (*IWQI*, *TURBIDITY*, and *FISH*) as proxies for water quality in the river segments, rather than direct causal influences in determining the appeal of a given site.

6 Concluding Remarks

The task of modeling recreation demand is often complicated by incomplete information regarding specifically where an individual travels to in visiting a geographically large site. While midpoint and shortest distance travel cost measures are often used as proxies for the unobserved travel cost, the resulting parameter estimates are likely to suffer from omitted variables bias. In this paper, we suggest that the problem be viewed instead as one of implicit site aggregation. In general, the probability of any aggregate site being visited is simply the sum of the probabilities that its component sites would be chosen. Using an underlying logit structure, the existing aggregation literature provides the specific functional form for the aggregated site choice probabilities, as well as an explicit characterization of the omitted variables that arise when using proxy variables for site attributes (including travel cost). We show that, in the context of the RUM model with a full set of alternative specific constants, the appropriate travel cost for an aggregate site is a probability weighted average of the travel cost to the component sites. We also generalize the existing aggregation models to include Normal Error Component Logit Mixture (NECLM) models that have become increasingly popular in the literature, paying particular attention to concerns regarding model identification in these settings.

A Monte Carlo exercise illustrates that the use of travel cost proxies can potentially lead to significant bias in characterizing recreation demand. In particular, while the nearest access point approach provides a relatively good approximation to underlying preferences for a wide range of parameter specifications, use of the midpoint approach to calculating travel cost can

lead to significant bias in the travel cost parameter and corresponding welfare calculations. Finally, an application is provided drawing on data from the 2009 Iowa Rivers and Rivers Corridors Survey. We find that the use of either the midpoint or shortest distance travel cost proxies yields a substantially smaller estimate of the welfare impacts of site loss than what is obtained using a model based on aggregated site probabilities.

The aggregation approach outlined above is not without its limitations. First, it assumes that the recreationist views the segment as composed of distinct sub-sites (i.e., the access points) to be chosen from in the same manner in which the segments themselves are selected. In the extreme, of course, this would yield an infinite number of alternatives in the choice set. However, at the scale considered in the Iowa Rivers application, the subdivision of the states primary rivers into 20 miles segments does not seem unreasonable. Second, and perhaps more importantly, the proposed solution to the lack of access-point data is a structural one, predicated on the logit model in constructing the segment level price index C_{is} . While extending the logit structure from the segment level of choice to the choice of access points within a segment seems natural, it is nonetheless an assumption. It is unclear how robust the parameter estimates would be to departures from the assumed logit model. At the same time, the logit-based probability weights given to the access point costs are likely closer to the underlying weights given by the individual than weights implied by the standard proxies used in the literature, as they at least taken into account the relative costs of the access points and the individual's marginal utility of income.

References

- [1] Ben-Akiva, M., and S. Lerman (1985) "Discrete Choice Analysis: Theory and Application to Travel Demand", Cambridge: The MIT Press.
- [2] Egan, K.J., Herriges, J.A., Kling, C.L. and Downing, J.A. (2009) "Valuing water quality as a Function of Water Quality Measures" *American Journal of Agricultural Economics*, Vol. 91, No. 1, pp. 106-123.
- [3] Feather, P. (1994), "Sampling and Aggregation Issues in Random Utility Model Estimation," *American Journal of Agricultural Economics*, **76**, pp. 926-33.
- [4] Feather, P. and Lupi, F. (1998) "Using partial aggregation to reduce bias in random utility travel cost models" *Water Resources Research* Vol. 34, No. 12, pp. 3595-3603.
- [5] Ferguson, M. R., and P. S. Kanaroglou (1998). "Representing the Shape and Orientation of Destinations in Spatial Choice Models," *Geographical Analysis* **30(2)**, pp. 119-137.

- [6] Haab, T., and K. E. McConnell (2002), *Valuing Environmental and Natural Resources: The Econometrics of Non-Market Valuation*. Northampton, MA:Edward Elgar.
- [7] Haener, M. K., Boxall, P. C., and Adamowicz, W. L., (2004) "Aggregation bias in Recreation site choice models: resolving the resolution problem," *Land Economics*, **80(4)**, pp. 561-574.
- [8] Herriges, J., and C. Kling. (1997) "The Performance of Nested Logit Models When Welfare Estimation is the Goal," *The American Journal of Agricultural Economics*, 79: 792-802.
- [9] Herriges, J. A., and D. J. Phaneuf, (2002) "Inducing Patterns Correlation and Substitution in Repeated Logit Model of Recreation Demand," *American Journal of Agricultural Economics*, **84(4)**: 1076-1090
- [10] Kaoru Yoshiaki, and V. K. Smith. (1990), "'Black Mayonnaise' and Marine Recreation: Methodological Issues in Valuing a Cleanup." Marine Policy Center, Woods Hole Oceanographic Institution, Woods Hole, MA.
- [11] Kaoru, Y., V.K. Smith, and J. Liu (1995), "Using Random Utility Models to Estimate the Value of Estuarine Resources," *American Journal of Agricultural Economics*, Vol. 77, No. 2, pp 141-151.
- [12] Klaiber, H. Allen, and Roger H. von Haefen. (2008) "Incorporating Random Coefficients and Alternative Specific Constants into Discrete Choice Models: Implications for InSample Fit and Welfare Estimates," Working Paper
- [13] Kling, C., and C. Thomson (1996), "The Implications of Model Specification for Welfare Estimation in Nested Logit Models," *American Journal of Agricultural Economics*, **78**, pp. 103-114.
- [14] Kurkalova, L.A. and S.S. Rabotyagov (2006), "Estimation of a Binary Choice Model with Grouped Choice Data," *Economics Letters* **90(2)**: 170-175.
- [15] Morey, Robert D. Rowe and Michael Watson (1993), "A Repeated Nested-Logit Model of Atlantic Salmon Fishing", *American Journal of Agricultural Economics*, **75(3)**, pp. 578-592.
- [16] Murdock, Jennifer. (2006) "Handling unobserved site characteristics in random utility models of recreation demand", *Journal of Environmental Economics and Management*, Vol. 51, No.1 pp. 1-25.

- [17] Parsons, G., and M. Needelman (1992), "Site Aggregation in a Random Utility Model of Recreation," *Land Economics*, **68**: 418-33.
- [18] Parson, G., Plantinga, A.J. and Boyle, K.J, (2000) "Narrow Choice Sets in a Random Utility Model of Recreation Demand", *Land Economics*, Vol. 76, No. 1, pp. 86 - 99.
- [19] Parsons,G.R., and Hauber, A.B. (1998), "Spatial Boundaries and Choice Set Definition in a Random Utility Model of Recreation Demand" *Land Economics*, **74**:3248.
- [20] Phaneuf, D. J., and J. A. Herriges (2000), "Choice Set Definition Issues in a Kuhn-Tucker Model of Recreation Demand," *Marine Resource Economics*, **14**, pp. 343-55.
- [21] Phaneuf, D. J., and J. A. Herriges (2002), "Inducing Patterns Correlation and Substitution in Repeated Logit Model of Recreation Demand," *American Journal of Agricultural Economics*, **14**, No.4, pp. 1076-1090.
- [22] Phaneuf, D. J.,. and Smith, V. K. (2005) " Recreation Demand Models" *Handbook of Environmental Economics* Edited by K. G. Mäler and J. R. Vincent, Vol. 2, pp. 672 - 751.
- [23] Smith, V. K., W. H. Desvousges and M. P. McGivney (1983), "The Opportunity Cost of Travel Time in Recreation Demand Models," *Land Economics*, Vol. 59, No. 3 , 259-278.
- [24] Train, K.,(2003), "Discrete Choice Methods with Simulation," Massachusetts: Cambridge University Press
- [25] 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* Vol. 22 pp 1095-1125.

Figure 1: Iowa River Segments Map

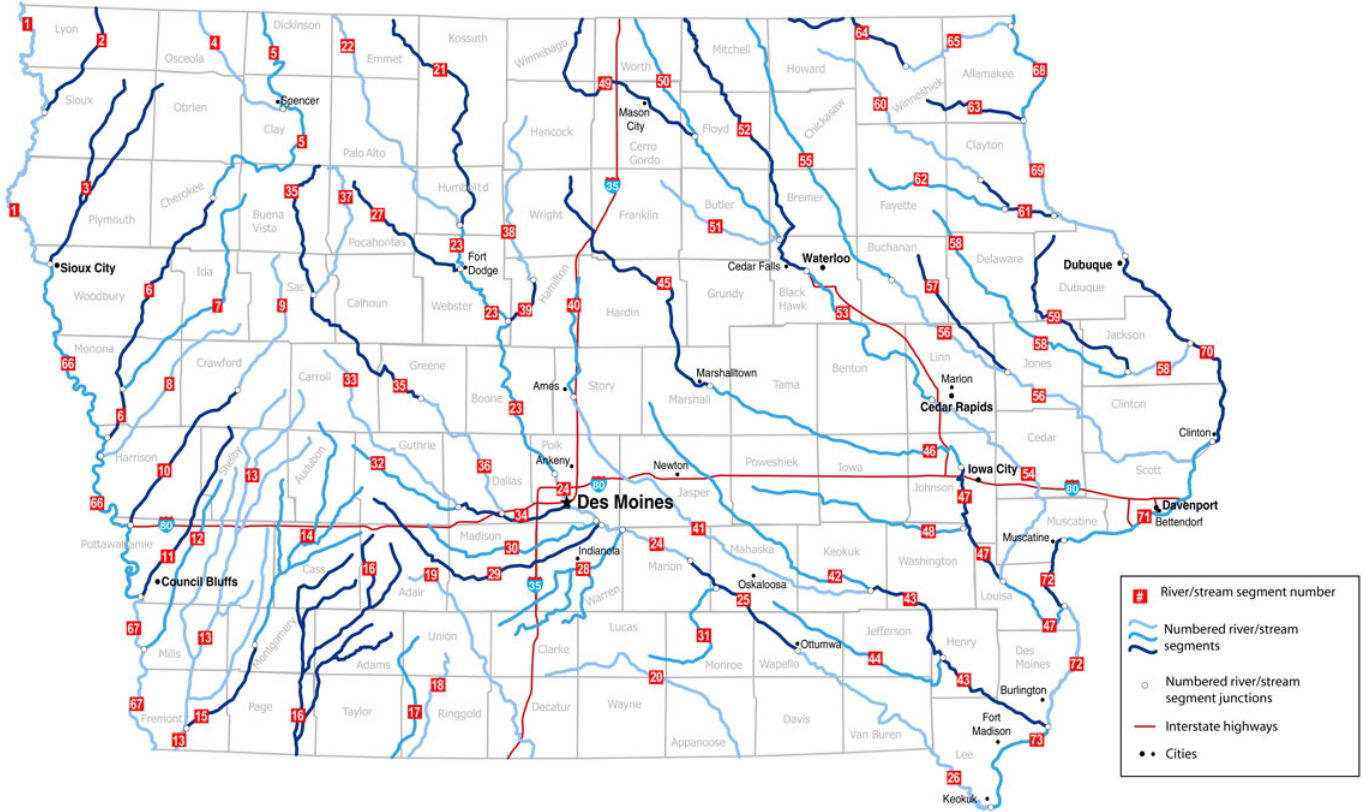


Table 1 Description of Monte Carlo Designs

Experiment Design	Description	Number of Variations
Price responsiveness (i.e., value of β)	$\beta = \{-0.01, -0.05, -0.09\}$ w/o Water $\beta = \{-0.02, -0.05, -0.1\}$ w/ Water	3
Number of river segments	s=5,10,20	3
Population and river characteristics	Base (B) Population Center (P) Nonlinear (Kinked) Rivers (K) Combined Population centers + Kinked Rivers (C)	4
Water Quality	Included (w/ Water) Not included (w/o Water)	2
Total		72

Table 2a Mean Absolute Percentage Error in Estimated β (w/o Water Quality)

Pop./River Config.	Number of Segments	Agg. Choice Prob.			Midpoint Proxy			Shortest Distance Proxy		
		$\beta = -0.01$	-0.05	-0.10	-0.01	-0.05	-0.10	-0.01	-0.05	-0.10
B	5	0.1	0.1	0.1	1.2	5.9	8.5	0.3	0.1	0.0
	10	0.1	0.1	0.1	1.2	6.5	13.7	1.2	1.0	1.0
	20	0.1	0.1	0.0	1.4	6.8	15.8	0.2	1.5	1.6
K	5	0.2	0.1	0.1	1.8	5.4	13	0.3	0.3	0.3
	10	0.3	0.1	0.1	1.6	8.2	17.9	0.6	0.5	1.8
	20	0.1	0.1	0.2	1.3	8.2	17.3	0.5	0.7	0.6
P	5	0.1	0.1	0.1	1.4	7.0	13.6	0.7	4.0	5.1
	10	0.2	0.1	0.2	1.9	7.4	14.9	0.5	2.4	4.1
	20	0.1	0.1	0.1	1.2	9.0	19.6	1.5	3.0	3.5
C	5	0.1	0.1	0.1	0.4	6.1	8.6	0.6	2.4	2.7
	10	0.1	0.1	0.2	0.9	8.1	15.4	1.0	0.5	2.6
	20	0.1	0.1	0.1	1.5	9.4	18.1	0.6	1.0	1.2

Table 2b Mean Absolute Percentage Error in Estimated CV_1 (w/o Water Quality)

Pop./River Config.	Number of Segments	Agg. Choice Prob.			Midpoint Proxy			Shortest Distance Proxy		
		$\beta = -0.01$	-0.05	-0.10	-0.01	-0.05	-0.10	-0.01	-0.05	-0.10
B	5	0.1	0.1	0.1	1.3	6.5	10.9	0.3	0.3	1.0
	10	0.1	0.1	0.1	1.5	7.2	16.1	1.5	1.1	1.0
	20	0.2	0.1	0.1	1.1	7.9	20.8	0.1	1.7	2.5
K	5	0.2	0.0	0.1	1.8	7.5	12.4	0.3	0.8	0.5
	10	0.1	0.3	0.4	1.4	9.1	16.9	0.4	0.1	1.1
	20	0.1	0.1	0.2	1.3	8.2	17.3	0.5	0.7	0.6
P	5	0.1	0.2	0.1	1.6	8.3	15.5	0.9	4.7	5.4
	10	0.2	0.2	0.3	2.0	8.5	17.7	0.5	3.0	5.0
	20	0.4	0.1	0.2	0.9	10.5	25.4	1.2	3.3	4.4
C	5	0.5	0.2	0.1	0.5	4.7	7.0	1.0	2.7	2.7
	10	0.2	0.1	0.1	1.0	10.0	19.6	0.9	0.6	3.2
	20	0.3	0.1	0.3	1.3	10.8	23.9	0.4	0.7	1.4

Table 3a Mean Absolute Percentage Error in Estimated β (w/ Water Quality)

Pop./River Config.	Number of Segments	Agg. Choice Prob.			Midpoint Proxy			Shortest Dist. Proxy ver.1			Shortest Dist. Proxy ver.2		
		$\beta = -0.01$	-0.05	-0.10	-0.01	-0.05	-0.10	-0.01	-0.05	-0.10	-0.01	-0.05	-0.10
B	5	0.1	0.1	0.1	1.6	4.2	9.3	1.0	1.7	1.8	0.7	0.4	0.5
	10	0.0	0.1	0.1	1.9	5.4	10.7	0.1	2.6	4.7	0.6	2.1	4.6
	20	0.1	0.1	0.1	2.5	6.8	15.1	1.3	2.4	3.9	0.9	1.8	3.2
K	5	0.1	0.1	0.1	1.3	4.7	13.3	0.6	0.6	1.9	0.4	0.3	1.8
	10	0.2	0.1	0.1	2.6	6.2	14.7	0.2	0.2	1.5	0.2	0.1	1.5
	20	0.1	0.1	0.1	3.1	9.0	18.9	0.2	0.5	1.2	0.2	0.3	0.9
P	5	0.2	0.2	0.1	1.7	6.0	13	0.3	0.8	3.1	1.8	0.2	3.8
	10	0.1	0.1	0.1	2.2	6.4	16.8	0.7	0.6	3.8	1.1	0.3	3.1
	20	0.1	0.1	0.1	2.3	8.0	19.4	1.9	3.3	4.8	1.4	2.5	4.0
C	5	0.0	0.3	0.2	2.5	7.0	14.1	0.1	2.5	3.5	0.3	1.8	3.6
	10	0.2	0.1	0.1	3.7	5.5	10.5	0.3	0.5	1.7	0.1	0.7	2.0
	20	0.1	0.1	0.1	3.9	9.3	22.6	0.4	0.8	2.3	0.4	0.8	2.3

Table 3b Mean Absolute Percentage Error in Estimated β_w (w/ Water Quality)

Pop./River Config.	Number of Segments	Agg. Choice Prob.			Midpoint Proxy			Shortest Dist. Proxy ver.1			Shortest Dist. Proxy ver.2		
		$\beta = -0.01$	-0.05	-0.10	-0.01	-0.05	-0.10	-0.01	-0.05	-0.10	-0.01	-0.05	-0.10
B	5	5.4	3.2	1	-	-	-	66	23.8	28.5	-	-	-
	10	0.2	0.5	0.1	-	-	-	90.4	81	82	-	-	-
	20	4	0.6	0.1	-	-	-	142	200.7	233.1	-	-	-
K	5	0.3	0.1	0.6	-	-	-	84.4	52.8	24.2	-	-	-
	10	1.2	1.1	0.6	-	-	-	98.7	91.3	80.8	-	-	-
	20	6.2	0.2	0.8	-	-	-	123.2	136.7	157.3	-	-	-
P	5	9.8	0.2	0.2	-	-	-	62.5	28.8	5.2	-	-	-
	10	0.8	0.7	0.2	-	-	-	89.5	63.6	38.2	-	-	-
	20	6.8	1.3	0.7	-	-	-	151.4	204.8	220.3	-	-	-
C	5	2.9	0.2	1	-	-	-	75.8	48.5	34.7	-	-	-
	10	3.2	0.3	0.2	-	-	-	95.2	82.4	76.1	-	-	-
	20	1.5	0.2	0.4	-	-	-	109.6	119.1	126.3	-	-	-

Table 3c Mean Absolute Percentage Error in Estimated CV_1 (w/ Water Quality)

Pop./River Config.	Number of Segments	Agg. Choice Prob.			Midpoint Proxy			Shortest Dist. Proxy ver.1			Shortest Dist. Proxy ver.2		
		$\beta = -0.01$	-0.05	-0.10	-0.01	-0.05	-0.10	-0.01	-0.05	-0.10	-0.01	-0.05	-0.10
B	5	0.2	0.1	0.1	1.9	6.4	9.2	1	0.9	0.2	0.5	1.8	0.7
	10	0.2	0.2	0.1	2.1	4.9	6.5	0.1	3.3	7.1	0.5	2.4	6.5
	20	0.4	0.7	0.4	3	7.7	16.9	1.7	2.9	3.4	1.2	2.5	2.5
K	5	0.2	0.2	0.1	2.8	2.1	5.1	0.7	0.4	0.7	0.2	0.6	1.8
	10	0.1	0.2	0.1	2.8	6.6	21.7	0.3	0.3	1.9	0.3	0.2	2.3
	20	0.1	0.3	0.3	3	9.5	24.3	0.2	0.2	1.5	0.1	0.5	0.7
P	5	0.3	0.3	0.2	2.1	7.1	14.2	0.3	1	3.1	2.2	0.5	2.5
	10	0.4	0.2	0.2	2.7	8.1	21.2	0.5	0.8	3.7	0.9	0.2	2.8
	20	0.1	0.2	0.2	2.4	9.9	26.5	1.9	3.8	6.8	1.3	2.5	4.3
C	5	0.1	0.2	0.1	2.2	7.8	12.9	0.1	2.6	2.8	0.2	2.1	3.1
	10	0.3	0.1	0.3	3.8	4.8	5.8	0.4	1.2	3.2	0.2	1.3	3.4
	20	0.1	0.3	0.1	4.1	10.7	29.3	0.3	0.8	2.5	0.4	0.8	2.5

Table 4: Summary Statistics Demographic Characteristics (N=4137)

Variable	Description	Mean	Std.Dev
Age1	Dummy Variable for Age Group: 18-25	0.01	0.09
Age2	Dummy Variable for Age Group: 26-34)	0.06	0.24
Age3	Dummy Variable for Age Group: 35-49)	0.21	0.41
Age4	Dummy Variable for Age Group: 50-59)	0.25	0.43
Age5	Dummy Variable for Age Group: 60-75)	0.31	0.46
Age6	Dummy Variable for Age Group: 76-)	0.15	0.36
Female	Dummy Variable for Females	0.70	0.46
College	Dummy Variable for College Degree	0.69	0.46
Size	Number of Adults	1.88	0.65
Kids	Number of Children	0.55	1.00
Employed	Dummy Variable for Employed	0.59	0.49
Student	Dummy Variable for Students	0.01	0.07
Retired	Dummy Variable for Retirees	0.36	0.48
Boat	Dummy Variable for owning a boat	0.23	0.42

Table 5: Trip Summary Statistics (N=4137)

Segment #	% Visitors	Total Trips	Segment #	% Visitors	Total Trips
1	1.93%	369	38	0.58%	116
2	1.21%	253	39	0.75%	68
3	0.94%	193	40	1.35%	293
4	0.29%	76	41	1.04%	232
5	1.57%	290	42	0.56%	113
6	1.16%	143	43	1.31%	303
7	0.31%	52	44	0.19%	31
8	0.12%	23	45	2.05%	533
9	0.29%	59	46	3.87%	800
10	0.34%	61	47	3.63%	737
11	0.56%	105	48	0.48%	76
12	0.22%	37	49	1.60%	462
13	0.75%	183	50	2.20%	381
14	0.65%	80	51	0.58%	116
15	0.58%	113	52	3.75%	1071
16	0.60%	154	53	3.92%	877
17	0.17%	25	54	3.05%	653
18	0.41%	60	55	1.76%	308
19	0.34%	59	56	3.60%	742
20	0.99%	171	57	0.56%	85
21	1.09%	314	58	3.41%	659
22	1.28%	194	59	0.73%	84
23	5.25%	979	60	1.55%	248
24	7.23%	1513	61	1.47%	142
25	3.02%	675	62	1.47%	232
26	1.69%	399	63	1.33%	111
27	0.46%	84	64	2.51%	314
28	0.60%	123	65	1.43%	208
29	0.77%	175	66	3.87%	884
30	0.58%	104	67	1.33%	245
31	0.24%	58	68	4.33%	676
32	0.58%	82	69	7.03%	1246
33	1.28%	193	70	6.24%	1578
34	2.30%	468	71	5.73%	1591
35	0.92%	234	72	2.34%	483
36	1.16%	272	73	2.34%	740
37	0.24%	20	Overall	47.76%	6.24

Table 6 Summary Statistics of River Attributes

Variable	Mean	Std.Dev	Min	Max
C_{isj}	162.85	79.15	0.08	485.14
73 river segments				
LENGTH	83.2	33.30	26.9	161.8
CANOE	62.82	27.13	0.00	99.00
OUTCROPPING	18.77	31.66	0.00	141.00
WATERBODY	0.22	0.30	0.00	1.00
WETLAND	0.01	0.01	0.00	0.05
FOREST	0.24	0.19	0.00	0.71
GRASS	0.11	0.09	0	0.40
CROP	0.39	0.29	0	0.92
DEVELOPED	0.02	0.04	0	0.24
IWQI	31.55	23.00	0	75
MIWQI	0.32	0.47	0	1
FISH	30.29	18.52	0	71
MFISH	0.12	0.33	0	1

Table 7: Ranking of River Segments

	Rank	Segment Numbers			
		Visitation	Model 1 Agg. Prob. (vers. 1)	Model 2 Shortest Distance	Model 3 Midpoint Proxy
Top 10	1	71	73	70	70
	2	70	68	73	66
	3	24	69	69	73
	4	69	70	71	68
	5	52	2	68	69
	6	23	71	66	71
	7	66	25	26	52
	8	53	66	47	72
	9	46	72	24	26
	10	56	24	72	24
Bottom 10	64	18	44	37	37
	65	9	30	44	8
	66	19	19	30	44
	67	31	8	8	17
	68	7	42	31	48
	69	12	59	39	12
	70	44	37	48	30
	71	17	17	17	39
	72	8	7	57	57
	73	37	12	12	31

Table 8a Estimation Result of Nested Logit Specifications (Stage 1)

Variable	Agg. Choice Prob. ver1		Shortest Dist. Proxy		Midpoint Proxy	
	Est.	Std.Dev	Est.	Std.Dev	Est.	Std.Dev
Travel Cost Variable						
TC	-0.035***	0.001	-0.034***	0.001	-0.033***	0.001
Demographics						
AGE 2	2.526***	0.365	2.827***	0.598	4.534***	0.449
AGE 3	2.904***	0.300	2.581***	0.576	4.460***	0.425
AGE 4	3.795***	0.288	3.077***	0.581	4.897***	0.426
AGE 5	4.679***	0.332	3.674***	0.589	5.400***	0.436
AGE 6	6.110***	0.422	4.712***	0.615	6.672***	0.458
FEMALE	0.284	0.157	-0.561***	0.116	-0.495***	0.101
COLLEGE	-0.057	0.146	-0.514***	0.110	-0.625***	0.100
SIZE	0.589***	0.028	-0.055	0.090	0.050	0.078
KIDS	0.347***	0.080	-0.014	0.070	-0.057	0.063
EMPLOYED	6.941***	0.262	0.453	0.290	0.873***	0.275
STUDENT	0.880*	0.446	2.478***	0.835	3.902***	0.754
RETIRED	7.218***	0.337	0.439	0.321	0.745***	0.303
BOAT	-1.899***	0.179	-1.913***	0.144	-1.840***	0.136
Nest Variable						
σ_τ	3.758***	0.105	2.920***	0.054	3.058***	0.051

1. Bootstrapped standard errors are reported, bootstrapping individuals.
2. *, **, *** represent significant levels at 10%, 5% and 1% respectively

Table 8b Estimation Result of Nested Logit Specifications (Stage 2)

Variable	Agg. Choice Prob. ver1		Shortest Dist. Proxy		Midpoint Proxy	
	Est.	Std.Dev	Est.	Std.Dev	Est.	Std.Dev
Site Attributes						
LENGTH	-0.008**	0.004	-0.001	0.004	0.009**	0.004
CANOE	0.010**	0.005	0.009*	0.005	0.008*	0.005
OUTCROPPING	0.003	0.004	0.004	0.004	0.001	0.004
WATERBODY	1.522**	0.782	1.688**	0.827	1.857**	0.746
WETLAND	13.338	12.324	13.887	12.643	19.386	11.902
GRASSLAND	1.257	1.572	1.619	1.549	0.974	1.387
CROPLAND	-0.210	0.729	-0.041	0.740	-0.024	0.689
DEVELOPED	-0.902	3.815	-0.289	3.877	-0.157	3.795
IWQI	0.006	0.014	0.007	0.013	0.005	0.013
MIWQI	0.078	0.677	0.180	0.641	0.231	0.677
FISH	0.014*	0.007	0.013*	0.008	0.012*	0.007
MFISH	0.578	0.360	0.562	0.362	0.356	0.354
BORDER	1.362***	0.479	1.253***	0.483	1.304***	0.452
CONSTANT	3.301***	1.130	-5.433***	1.218	-3.277***	1.119
Welfare Calculation (\$/year/household)						
Loss of River 71	17.03		12.92		14.30	
Loss of All Rivers	668.94		361.37		420.23	

1. *, **, *** represent significant levels at 10%, 5% and 1% respectively
2. The CV figures are based on 1000 numerical simulations using equation 41

Table 9 Stage 2 Estimation Results Using Turbidity as Water Quality Proxy

Variable	Agg. Choice Prob. ver1		Shortest Dist. Proxy		Midpoint Proxy	
	Est.	Std.Dev	Est.	Std.Dev	Est.	Std.Dev
Site Attributes						
LENGTH	-0.008**	0.004	-0.001	0.004	0.009***	0.003
CANOE	0.010**	0.004	0.009*	0.005	0.009**	0.004
OUTCROP	0.002	0.004	0.003	0.004	0.001	0.004
WATERBODY	1.510**	0.693	1.685**	0.810	1.817**	0.778
WETLAND	11.734	11.968	12.319	12.060	17.085	10.669
GRASSLAND	1.125	1.402	1.541	1.407	0.873	1.314
CROPLAND	-0.105	0.644	0.013	0.712	0.097	0.690
DEVELOPED	-1.543	3.795	-0.872	3.635	-0.800	3.637
TURBIDITY	-0.004*	0.002	-0.003	0.002	-0.004	0.003
MTURBIDTY	-0.553*	0.321	-0.444	0.325	-0.398	0.324
FISH	0.010	0.007	0.010	0.008	0.008	0.007
MFISH	0.622*	0.343	0.590	0.373	0.408	0.326
BORDER	1.457***	0.424	1.324***	0.446	1.408***	0.447
CONSTANT	3.876***	0.666	-4.827***	0.762	-2.736***	0.739
Observations	73		73		73	
Adjusted R ²	0.653		0.600		0.679	

1. *, **, *** represent significant levels at 10%, 5% and 1% respectively

2. The CV figures are based on 1000 numerical simulations using equation 41