*Marine Resource Economics*, Volume 14, pp. 343–355
Printed in the U.S.A. All rights reserved
0738-1360/99 $3.00 + .00


# Choice Set Definition Issues in a Kuhn-Tucker Model of Recreation Demand

DANIEL J. PHANEUF
North Carolina State University

JOSEPH A. HERRIGES
Iowa State University

**Abstract** *Much of the literature on choice sets has focused on how alternative specifications of market scope and site definition impact site selection models and the resulting welfare estimates* per choice occasion. *In this paper, choice set definition issues are investigated using the Kuhn-Tucker model, which integrates the site selection and participation decisions in a unified and utility theoretic framework. This allows us to consider the impact that alternative site set definitions may have on both where individuals recreate and the numbers of trips they take. Using data from the 1997 Iowa Wetlands Survey we examine the effects on estimates and welfare measures of choice sets representing various levels of site aggregation and market scope. We find that significant differences in welfare measures arise from changing choice set definitions.*

**Key words** Choice set definition, Kuhn-Tucker model.

## Introduction

Choice set definition in recreation demand modeling is a complex issue for which economic theory provides relatively little guidance. Broadly speaking, the issue can be divided into two areas: the determination of the proper scope of the market (*i.e.*, what goods enter into an individual's choice set during the timeframe of interest) and the decision as to how sites are to be defined and/or aggregated. These decisions must be made both on a conceptual level (*e.g.*, considering whether an individual actually knows about all of the available options) and on a practical level, recognizing the limitations in empirical settings of both the available data and the ability of the specified model to handle a large number of alternatives. In determining the scope of the market, for example, attention must be given to both the geographical and horizontal extents of the market. In the case of recreational day-trips, where the price is determined largely by a site's distance from an individual's home, the geo-

---

Daniel J. Phaneuf (corresponding author) is assistant professor in the Department of Agricultural and Resource Economics at North Carolina State University, Box 8109, Raleigh, NC 27695-8109, e-mail: phaneuf-dj@are1.cals.ncsu.edu. Joseph A. Herriges is full professor in the Department of Economics at Iowa State University, Ames, IA 50011-1070, e-mail: jaherrig@iastate.edu.

The authors would like to thank Cathy Kling and Kerry Smith, as well as participants in the Choice Set Definition Workshop sponsored by the National Marine Fisheries Service and two anonymous referees, for their helpful comments and discussion regarding earlier drafts of this paper. This research was supported in part by the U.S. Environmental Protection Agency and by the Western Regional Research Project W-133. Although the research described in this article has been funded in part by the United States Environmental Protection Agency through R826615-01-1 to the authors, it has not been subject to the Agency's required peer review policy and therefore does not necessarily reflect the views of the Agency and no official endorsement should be inferred. All remaining errors are our own.

graphical scope can often be effectively limited via a feasible cut-off distance.[1] However, the horizontal extent of the market (*i.e.*, what substitute goods to include in the model) is not so cleanly defined. When modeling the demand for salt water fishing, should one include fresh water alternatives as potential substitutes? When expanding the geographical scope of the market, more potential substitute activities will inevitably fall into the expanded geographical region. Should these be included? Is it sufficient to include expenditures on horizontal substitutes simply as part of a numeraire good, or is important information lost by not including the characteristics/quality levels of these sites in the model?

There is also an information extent of the market to consider. Sites that are physically possible for the individual to visit may not, in fact, enter that individual's decision making process if they are unaware of the sites' characteristics or even existence. Horowitz (1991), for example, considers this problem in the context of job search, arguing that "…the cost of information often precludes an individual from learning about and applying for all available jobs" (p. 1239). Similar information costs and constraints potentially limit the scope of the market for the recreator as well. Finally, the definition of a site itself is a nontrivial task. In some applications, natural boundaries exist (*e.g.*, in the case of small inland lakes), whereas in others (*e.g.*, in the case of a major river system) a continuum of sites exists. Unfortunately, site definitions are driven as often as not by practical limitations in terms of the data or the model being estimated, despite the fact that different decisions on choice set can lead to significant differences in the welfare measures obtained.[2]

Conceptually and practically related to the choice set definition issues is the challenge of modeling corners in recreation demand. Corner solutions are common in this setting because individuals typically visit only a subset of the available sites, setting the demand for the remaining sites to zero. There is growing literature on methods for dealing with corner solutions in recreation demand [see Herriges, Kling and Phaneuf (1999) for a recent review]. The prevalence of corner solutions is linked to choice set definition decisions. An increase in the market scope to be analyzed will invariably lead to more corner solutions, as options are added in which not all individuals will partake. Conversely, a high degree of aggregation will decrease the number of corner solutions, as previously individual sites are lumped together, increasing the likelihood of a visit to one of the aggregated sites. At the extreme, choice set definition determines the types of model that can be feasibly estimated, which in turn influences the resulting welfare estimates. Single site or pooled models, which by definition require either a restricted scope or a high level of aggregation, will typically produce results different from models that can be estimated for a larger number of sites, such as Morey, Rowe and Watson's (1993) repeated nested logit models (RUMs) or linked models that combine site selection and participation decisions (see, *e.g.* Herriges, Kling, and Phaneuf 1999).

Our objective in this paper is to bring recent developments in the literature on corner solutions, and in particular the Kuhn-Tucker (KT) model (*e.g.*, Wales and Woodland 1983; and Phaneuf, Kling and Herriges 2000), to bear on the choice set definition debate. The KT framework is attractive for two reasons. First, much of the literature on choice set definition has focused on how alternative specifications impact site selection models and the resulting welfare estimates *per choice occasion*. For example, Parsons and Needelman (1992), Feather (1994), and Kaoru, Smith and Oiu (1995) consider how site aggregation alters welfare calculations, whereas Peters, Adamowicz and Boxall (1995) and Parsons and Hauber (1998) emphasize the

[1] See Parsons and Hauber (1998) for an excellent discussion of the use of spatial boundaries in choice set definitions, including the risk of setting these boundaries too tight.
[2] See, *e.g.*, Parsons and Needelman (1992), and Kaoru, Smith, and Oiu (1995).

importance of scope specification.[3] The Kuhn-Tucker model, however, integrates the site selection and participation decisions in a unified and utility theoretic framework. This allows us to consider the impact that alternative site set definitions may have on both where individuals recreate and the numbers of trips they take. Second, because the KT model starts with the familiar direct utility function, well-known results on aggregation in the general literature can be brought to bear in terms of both the specification of and testing for alternative aggregation schemes.

The remainder of the paper is organized as follows. The next section provides a brief overview of the KT model. This framework is then used to model the demand for recreational activities in Iowa wetlands. The underlying data set is then detailed. The subsequent section provides the empirical specification and the resulting parameter estimates using different levels of site aggregation and geographical scope. We then examine the effects that various choice set specifications have on the welfare estimates associated with changes in site characteristics and access. The final section provides a discussion and suggestions for future research.

## Kuhn-Tucker Model

Wales and Woodland (1983) and Hanemann (1978) independently suggested the Kuhn-Tucker model for estimation of consumer preferences when binding nonnegativity constraints are present in the observed data.[4] The model begins with utility maximization subject to income and nonnegativity constraints. The first order conditions, given the potential for nonconsumption of a subset of the goods, take the form of the Kuhn-Tucker conditions. Formally the consumer is assumed to solve the standard utility maximization problem

$$\max_{\mathbf{x},z} u(\mathbf{x}, z, \mathbf{q}, \gamma, \varepsilon) \quad s.t. \quad y = z + \mathbf{p}'\mathbf{x}, \quad x_i \geq 0 \tag{1}$$

where $\mathbf{x} = (x_1, \ldots, x_M)'$ is a vector of visits to available recreation sites, $\mathbf{p} = (p_1, \ldots, p_M)'$ is a vector of prices, $y$ denotes income, $z$ is a numeraire good representing spending on all other goods, $\mathbf{q} = (q_1, \ldots, q_M)'$ is a vector of attributes of the recreation sites, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_M)'$ is a vector of unobserved random components, and $\gamma$ is a vector of parameters to be estimated. Assuming the numeraire good is necessary, the Kuhn-Tucker conditions for this problem take the form

$$u_j \leq p_j u_z; \; x_j \geq 0; \; x_j\left[u_j - p_j u_z\right] = 0, \quad j = 1, \ldots, M \tag{2}$$

where $u_j$ indicates the partial derivative of utility with respect to $x_j$. Given specific assumptions on the structure of the utility function, the first order conditions in equation (2) can be rewritten as[5]

$$\varepsilon_j \leq g_j(\mathbf{x}, y, \mathbf{q}, \gamma); x_j \geq 0; x_j\left[\varepsilon_j - g_j(\mathbf{x}, y, \mathbf{q}, \gamma)\right] = 0, \quad j = 1, \ldots, M \tag{3}$$

where $g_j(\cdot)$ is a function of observed variables and parameters to be estimated, determined by the choice of functional form for utility.

---

[3] See Haab and Hicks (2000) in this issue for a review of the literature on choice set definition issues.

[4] Bockstael, Hanemann, and Strand (1986) suggest using the model for addressing corner solutions in recreation demand, while Phaneuf, Kling, and Herriges (2000) provide an application. See Phaneuf, Kling, and Herriges (2000) for a more complete description of the Kuhn-Tucker model.

[5] In particular, it is assumed that $u_{z\varepsilon} = 0$, $\partial u_j/\partial\varepsilon_k = 0 \;\; \forall k \neq j$ and $\partial u_j/\partial\varepsilon_j > 0 \; \forall j$.

Assuming the random variables are distributed via the density function $f_\varepsilon(\varepsilon)$, the probability of observing an individual's outcome in the data can be constructed from equation (3). For example, if the first $k$ goods are positively consumed, the probability of this outcome is given by

$$\int_{-\infty}^{g_{k+1}} \cdots \int_{-\infty}^{g_M} f_\varepsilon(g_1, ..., g_k, \varepsilon_{k+1}, ..., \varepsilon_M) abs|J_k| d\varepsilon_{k+1}, ..., d\varepsilon_M \tag{4}$$

where $J_k$ is a Jacobean transformation term. A probability as in equation (4) can be computed for each individual in the sample, and maximum likelihood used to recover estimates of the parameter vector. Because of the nonnegativity constraints, the demand system, and hence the indirect utility function of interest for welfare analysis, is nondifferentiable. If there are $M$ recreation sites available, the individual will have $2^M$ different combinations of sites that can be visited, including the possibility of not visiting any recreation sites during the season. Let

$$\Omega \equiv \{\emptyset, \{1\}, \{2\}, \ldots, \{M\}, \{1, 2\}, \ldots, \{1, M\}, \ldots, \{1, 2, \ldots M\}\} \tag{5}$$

denote the collection of all possible demand patterns (*i.e.*, subsets of $I = \{1, 2, \ldots, M\}$) and $v_\omega(\mathbf{p}_\omega, \mathbf{q}, y, \gamma, \varepsilon)$ denote the indirect utility function when the individual is restricted to the commodities indexed by $\omega \in \Omega$ (*i.e.*, $x_j = 0 \forall j \notin \omega$). The individual's unconditional indirect utility function is then given by

$$v(\mathbf{p}, \mathbf{q}, y, \gamma, \varepsilon) = \max_{\omega \in \Omega}\{v_\omega(\mathbf{p}_\omega, \mathbf{q}, y, \gamma, \varepsilon)\}. \tag{6}$$

As a side note, the structure of preferences in equation (6) highlights the conceptual similarities between the RUM and KT models. In each case preferences are characterized up to an unobserved error term. It is assumed consumers make a choice among discrete alternatives. In RUMs, consumers chose which site to visit on a given choice occasion, while in the KT model they chose the visitation pattern ($\omega \in \Omega$) for the season. The models differ in that the RUM restricts the analysis to a single choice occasion; thus, choices involving multiple sites are not possible and scale (the number of trips to each site) information is not incorporated. The KT model uses additional information, adding the scale dimension and allowing multiple trips to various sites. In this sense, the KT model can be seen as a generalization of the RUM.

## The Iowa Wetlands Data

The data used in this application come from the 1997 Iowa Wetlands Survey conducted at Iowa State University. The purpose of the survey was to gather information on how Iowans use wetlands in the state, as well as their attitudes towards wetland preservation/restoration programs.[6] The survey included a variety of questions soliciting actual and hypothetical use of wetlands, as well as contingent valuation

[6] While Iowa wetlands obviously do not fall into category of marine resources, the choice set definition challenges associated with this data set are similar to those facing analysts using marine recreation data. The survey itself is part of a larger project to examine the value of wetlands in Iowa. For details on the survey process and discussion of the project's wider goals, including wetland definitions and discussion of the importance of wetland conservation, see Azevedo (1999) or Herriges, Kling, and Azevdo (1999).

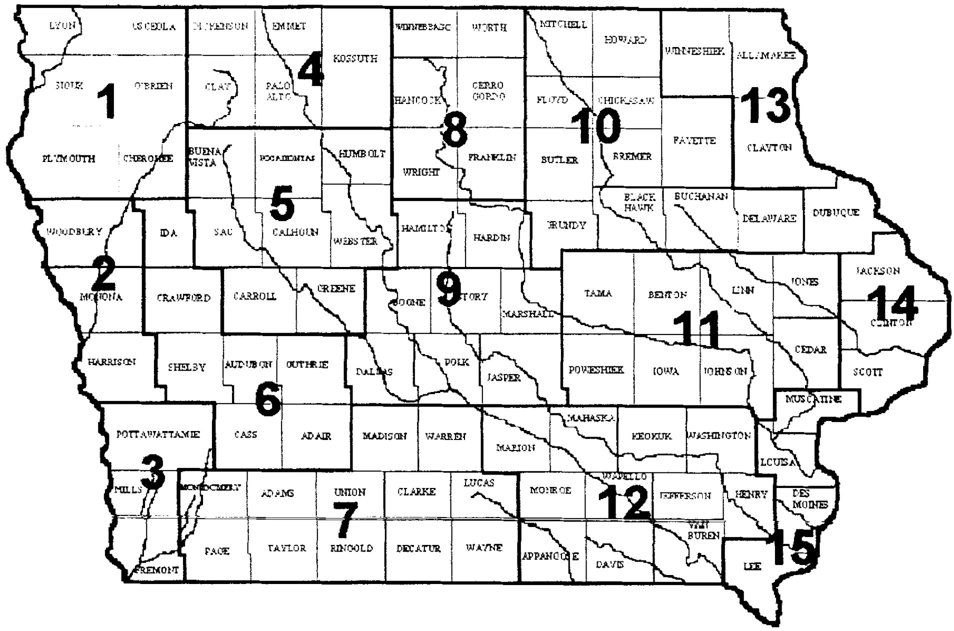


**Figure 1**. Iowa Wetland Zones

and behavior questions. Finally, detailed demographic characteristics and information for constructing travel prices were gathered. This study focuses on the visitation data. The behavioral data are augmented by pheasant count data, provided by the Iowa Department of Natural Resources.

A sample of 6,000 Iowa households was drawn from the general population and from state hunting and fishing license holders and sent a mail survey, from which 3,131 useable surveys were returned.[7] As part of the survey, each individual was provided a copy of the map in figure 1, dividing the state into fifteen zones.[8] Individuals were asked to record the number of trips made to wetlands in each of the zones during 1997. For this application, 2891 respondents are used, of whom roughly two-thirds visited a wetland in the state during 1997.

Given the site visitation data, the next task is to define choice sets for the models to be applied. In part, this specification depends on the goals of the empirical study. We may wish to consider the demand for and policies affecting wetlands in the entire state. Conversely, we may be interested in a particular resource in the state, such as the Des Moines lobe of the Prairie Pothole Region. The Prairie Pothole Region is a large, fairly unique section of North American, encompassing parts of Iowa, Minnesota, the Dakotas, and the Canadian plains provinces. The area is

---

[7] A series of focus groups and a pre-test of 600 Iowa households were used to develop the survey instrument prior to its final administration to the full sample of 6,000 households. The sample was stratified to insure users were included in the final survey sample, with 4,000 households drawn from the general residential population and 2,000 households drawn from fishing and hunting license holders. An overall survey response rate of 58% was achieved among the deliverable surveys. See Azevedo (1999) for additional details.

[8] While the zones were specified along county boundaries, they were also selected so as to reflect broadly homogeneous wetland types within the state, such as the riverine wetlands along the eastern and western borders of the state (*i.e.*, zones 1, 2, 3, 13, 14, and 15) versus the prairie pothole wetlands in north-central Iowa (zones 4, 5, and 8).

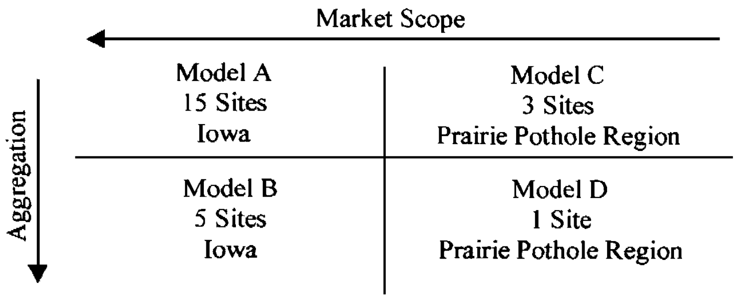


**Figure 2**. Choice Set Definition Options

dotted with indentations, in otherwise flat landscapes, that are wet for at least part of the year. This type of wetland is ideal habitat for many types of wildlife, including ducks and pheasants (most of the continents ducks breed in this area), and has importance at both the continental and local levels. The Iowa portion of the Prairie Pothole Region corresponds roughly to zones 4, 5, and 8 in figure 1. Choice set definitions may also be made based on the desire to limit the dimension of the models applied.[9] This may lead one to consider restricted levels of market scope and increased site aggregation in an empirical model.

In order to illustrate the impact of these choice set decisions, we consider four models in this study, summarized by the varying degrees of scope and aggregation depicted in figure 2. Model A represents the largest scope combined with the lowest level of aggregation, modeling the demand for recreation in the entire state and defining sites as the fifteen zones.[10] Model B is conceptually similar to A, considering demand for recreation in the entire state, but with sites aggregated such that individuals chose from among five “mega-zones”. Corresponding to figure 1, the aggregate sites are defined as {1,2,3}, {4,5,8}, {6,7,12}, {9,10,11}, and {13,14,15}. Care has been taken to aggregate sites exhibiting similar geographical features, with the Prairie Pothole Region and east and west riverine wetland regions being grouped respectively. The final two definitions consider limiting the scope of the choice set, focusing on demand for recreation in the Prairie Pothole sites. Model C considers the demand for trips to the three disaggregate Prairie Pothole sites (zones 4, 5, and 8), while Model D combines these sites into a single good, resulting in a one-site model.[11] In the restricted scope models, expenditures on visits to the other sites are

---

[9] For example, in our application of the KT model to the demand for Wisconsin Great Lakes fishing trips (see Phaneuf, Kling, and Herriges 2000), we aggregate data on twenty-two possible destinations to four sites. This was done primarily to reduce the dimension of the model being estimated.

[10] We note that, due the nature of the original survey instrument, we only have available data from sites that have already been substantially aggregated. Thus, it is not all together proper to call Model A a disaggregate model, except in relation to the others.

[11] It is worth noting again that while the aggregation restrictions used in the KT model are similar to those often used in RUM’s, the scope restrictions are somewhat different. Analysts employing the standard RUM framework have typically restricted the choice set for the individual by some criteria, allowing the choice sets to potentially be of different magnitudes across individuals, while in this paper the KT model’s choice set has been restricted to be the same for all individuals. Thus, it is the geographic scope of the resource to be examined in detail, rather than the *individual’s* geographic scope. Another perspective on the scope restrictions considered in this paper comes from noting that the Prairie Pothole wetlands are different from the riverine and small pond wetlands elsewhere in the state. Thus, the scope restriction employed here is analogous to say excluding inland fisheries when studying marine fisheries along the East Coast. Finally, we note that, conceptually, there is nothing preventing the KT model from employing individual specific scope restrictions. The computing coding would simply be more complex.

included, but only as a component of the numeraire good. In this sense the restricted scope models represent higher degrees of horizontal aggregation, leaving out characteristics of the excluded sites. In the following sections we apply the KT model to each of these choice set definitions.

## Empirical Model and Results

Estimation of the KT model requires specification of the functional form for utility and the choice of distribution for the error terms. Following Phaneuf, Kling, and Herriges (2000) we choose as our utility function a version of the linear expenditure system. The consumer's direct utility function is given by

$$u(\mathbf{x}, \mathbf{q}, \varepsilon) = \sum_{j=1}^{M} \Psi_j(q_j, \varepsilon_j) \ln(x_j + \theta) + \ln(z) \tag{7}$$

where $\Psi_j$ is a quality index give by $\Psi_j(q_j, \varepsilon_j) = \exp(\delta_0 + \delta_1 ph_j + \varepsilon_j)$, and $ph_j$ is a site quality variable equal to the pheasant count in the $j$th site if the individual indicated possession of a hunting or fishing license and equal to zero otherwise. The price of visiting zone $j$ for individual $i$ ($p_{ij}$) was constructed by first establishing the roundtrip travel distance ($d_{ij}$) and travel time ($t_{ij}$) from their residence to the center of wetland zone $j$ using the software package PCMiler. The price was then constructed as $p_{ij} = 0.22d_{ij} + (0.33w_i)t_{ij}$, where $w_i$ denotes the individual's marginal wage rate. Simple averages were used to construct price and quality variables for the aggregate mega-zones.

The linear expenditure system is a somewhat restrictive specification for utility. The structure implies regime-specific demand equations of the form

$$x_j = -\theta + \frac{\Psi_j}{1 + \sum_{k \in \omega} \Psi_k} \frac{1}{p_j} \left( y + \sum_{k \in \omega} p_k \theta \right), \quad j \in \omega \tag{8}$$

which is limiting in the types of substitution patterns captured between sites. This specification, however, combined with the assumption that the random terms are distributed independent and identical extreme value, makes it feasible to estimate relatively large dimensional models.[12] The probability of observing an individual in demand regime ω has a closed form given by

$$\pi_\omega = \exp\left( -\sum_{j \in \omega} \frac{g_j}{v} \right) \times \exp\left[ -\sum_{i=1}^{M} \exp\left( -\frac{g_j}{v} \right) \right] \times abs \; \left| J_\omega \right| \tag{9}$$

[12] In addition, the level of restriction implied by the LES system is similar to the use of linear functional forms in most applications of random utility models. The KT model provides the added benefit that welfare measures reflect seasonal, rather than the loosely defined choice occasion, measures. The extreme value distribution was chosen in this case for simplicity. A generalized extreme value distribution could have been employed as well, at the expense of more time consuming welfare calculation procedures. Expressions for the Jacobean transformation terms, as well as example programs written in GAUSS for estimation and welfare calculations of various dimensional KT models using the LES utility function, are available from the authors upon request.

where

$$g_j = \ln\left[\frac{p_j(x_j + \theta)}{y - \sum_{i=1}^{M} p_i x_i}\right] - \delta_0 - \delta_1 ph_j \tag{10}$$

$J_\omega$ denotes the Jacobean transformation from $(\varepsilon_1, \ldots, \varepsilon_M)'$ to $(x_1, \ldots, x_M)'$, and $v$ is a scale parameter in the extreme value distribution. A probability term such as in equation (9) can be calculated for each individual in the sample, and maximum likelihood used to recover estimates of the utility function parameters.

The results from estimating each of the four models using the entire survey sample are presented in table 1. All of the parameters are found to be significantly different from zero at 1% critical level. In and of themselves the parameters estimates are not interesting, except to note that the parameter associated with pheasant counts ($\delta_1$) is positive and statistically significant. This indicates that pheasant counts positively influence both overall utility and the number of trips to a given region. A comparison across the models suggest that the scale and pheasant count parameter estimates are relatively stable, while there are noticeable differences in the estimates of $\theta$ and $\delta_0$.

Of greater interest are the welfare implications of the four models, presented in table 2. We examine three scenarios, each reflecting different potential policy concerns. Scenario I examines the effect of policies that would increase pheasant counts statewide by 20%, whereas under Scenario II the pheasant counts are altered only in the Prairie Pothole Region. Scenario III attempts to assess the recreational value of the Prairie Pothole Region as a whole by examining the welfare effects of eliminating the resource. As previously noted, the various levels of scope and aggregation in the four models may affect the reported welfare measures.

The results in table 2 indicate that, for this application, site aggregation consistently reduces the estimated welfare effects, regardless of the scope specification or the scenario being considered. The reductions range from just over 80% in the case of Scenario III (for Models C versus D) to 58% in the case of Scenario I (again for

**Table 1**
Estimation Results—Statewide Sample

| | Model | | | |
|---|---|---|---|---|
| Parameters | Model A: Statewide Scope 15 Zones | Model B: Statewide Scope 5 Mega-zones | Model C: Prairie Pothole 3 Zones | Model D: Prairie Pothole 1 Mega-zone |
| $\theta$ | 5.92 | 6.17 | 6.86 | 7.39 |
| | (0.15) | (0.16) | (0.44) | (0.46) |
| $\delta_0$ | –6.03 | –5.30 | –5.91 | –5.14 |
| | (0.04) | (0.03) | (0.11) | (0.08) |
| $\delta_1$ | 0.009 | 0.007 | 0.008 | 0.006 |
| | (< 0.001) | (< 0.001) | (0.001) | (0.001) |
| $v$ | 0.58 | 0.54 | 0.51 | 0.46 |
| | (0.01) | (0.01) | (0.02) | (0.01) |

Note: Standard errors in parenthesis. All parameter estimates are significantly different from zero at 1% level.

**Table 2**
Welfare Estimates—Statewide Sample

| | Model | | | |
|---|---|---|---|---|
| Welfare Scenarios | Model A: Statewide Scope 15 Zones | Model B: Statewide Scope 5 Mega-zones | Model C: Prairie Pothole 3 Zones | Model D: Prairie Pothole 1 Mega-zone |
| I. 20% increase in pheasant counts at all sites | $543 | $273 (27) | $141 (27) | $58 (13) |
| II. 20% increase in prairie pothole pheasant counts | $154 | $73 (8) | $141 (27) | $58 (13) |
| III. Loss of prairie pothole region | $208 | $156 (73) | $126 (28) | $93 (21) |

Notes: Welfare measures are in dollars per respondent per year. Standard errors on the welfare measures were constructed using a bootstrap procedure. Standard errors were not constructed for the 15-good model due to time limitations in completing the manuscript for this special issue. Computation of the standard errors in case, while technically feasible, remains time consuming.

Models C versus D). These findings are consistent with earlier studies by Kaoru, Smith, and Oiu (1995) and Feather (1994), though Parsons and Needelman (1992) typically found the aggregation bias to go in the opposite direction.

Limiting the geographical scope of the model also results in reduced welfare predictions. This is what one would expect for Scenario I. Models C and D, which restrict the scope of the market to the Prairie Pothole region, produce substantially lower welfare measures when compared to their statewide counterparts simply because they ignore the benefits of the improved pheasant counts outside of the Prairie Pothole Region. This is a direct result of the lack of characteristic data for the excluded sites, for which expenditures on are only included in the numeraire.

On the other hand, the explanation for the differences in the magnitudes between Scenarios II and III is less obvious. One *a priori* belief is that by limiting the scope of the market to the Prairie Pothole Region, we are excluding all other possible substitute wetland sites, making the modeled wetland sites more unique in the household's choice set. This would in turn increase the magnitudes of the welfare loss stemming from their elimination. Indeed, this is exactly what Parsons and Hauber (1998) found when they used spatial boundaries to limit the choice set. A second *a priori* belief is that the differences between the full and limited scope models should be small, since the numeraire good allows inclusion of expenditures on the nonmodeled wetland sites, and the quality change occurs in all cases for sites which are fully modeled.[13] The results seem to lean towards the second interpretation. In particular, the welfare estimates from a limited scope models (C and D) are *smaller* in magnitude than those from their full scope counterparts (A and B respectively). Interestingly as well, the effect of this *horizontal* aggregation on sites into the numeraire good is consistent with the direction of the effect for the *geographical* aggregation mentioned above.

---

[13] We appreciate the comments of an anonymous reviewer, who clarified this second point.

This would seem to be the opposite of the findings of Parsons and Hauber (1998). There is, however, a key distinction between the scope restrictions employed by these authors and the one being considered here. In their paper, geographical scope is defined uniquely for each individual in the sample, with sites included or excluded from the *individual's* choice set based on their distance from the individual's home. As a result, the sites that are first excluded from the choice set are those that are the furthest from the individual's home and, typically, represent low probability trips. This form of scope restriction is helpful when the analyst has available data on trips a large number of sites and wishes to reduces the dimensionality of the estimation problem. Parsons and Hauber (1998) show that excluding far flung sites from the choice set has little impact on the welfare estimates, with these sites are assigned a very low probability in the RUM framework.

The scope restrictions reflected in table 2, however, are quite different. In moving from Models A and B to Models C and D, respectively, we are not excluding remote sites. On the contrary, for the majority of the sample we are excluding their primary wetland visitation sites, since most of the sample lives outside of the Prairie Pothole Region. This mimics what might occur in an empirical setting in which data are available on trips to a specific region, including visits by individual that live far from the sites of interest, but data are unavailable on sites that are the primary recreation areas for these remote individuals. By restricting the geographical scope, we are relying on both relatively few wetland sites to capture preferences for wetlands and a relatively small proportion of the sample, since individuals from outside the Prairie Pothole Region will be at corners in the KT model.

The findings in Parsons and Hauber (1998) partially emerge, however, if we restrict our analysis to individuals living in the Prairie Pothole Region. For these individuals, limiting the analysis to the Prairie Pothole sites (as in Models C and D) is comparable to using the spatial boundaries of Parsons and Hauber (1998). The resulting parameter estimates and welfare predictions are provided in tables 3 and 4, respectively. For this restricted sample, we again find that aggregation reduces the estimated welfare effects in Scenarios I through III, with the reductions ranging from 10% to 61%. However, the scope effects now move in the opposite directions.

**Table 3**
Estimation Results—Prairie Pothole Sample

| | Model | | | |
|---|---|---|---|---|
| Parameters | Model A: Statewide Scope 15 Zones | Model B: Statewide Scope 5 Mega-zones | Model C: Prairie Pothole 3 Zones | Model D: Prairie Pothole 1 Mega-zone |
| $\theta$ | 7.09 | 7.21 | 7.32 | 6.70 |
| | (0.52) | (0.56) | (0.77) | (0.89) |
| $\delta_0$ | –5.79 | –5.12 | –5.81 | –5.22 |
| | (0.11) | (0.10) | (0.16) | (0.16) |
| $\delta_1$ | 0.009 | 0.007 | 0.011 | 0.008 |
| | (0.001) | (0.001) | (0.002) | (0.002) |
| $v$ | 0.52 | 0.54 | 0.58 | 0.56 |
| | (0.02) | (0.02) | (0.03) | (0.04) |

Note: Standard errors in parenthesis. All parameter estimates are significantly different from zero at 1% level.

**Table 4**
Welfare Estimates—Prairie Pothole Sample

| Welfare Scenarios | Model A: Statewide Scope 15 Zones | Model B: Statewide Scope 5 Mega-zones | Model C: Prairie Pothole 3 Zones | Model D: Prairie Pothole 1 Mega-zone |
|---|---|---|---|---|
| I. 20% increase in pheasant counts at all sites | $674 | $264 (71) | $322 (88) | $126 (42) |
| II. 20% increase in prairie pothole pheasant counts | $185 | $90 (26) | $322 (88) | $126 (42) |
| III. Loss of prairie pothole region | $321 | $299 (49) | $560 (128) | $446 (96) |

Notes: Welfare measures are in dollars per respondent per year. Standard errors on the welfare measures were constructed using a bootstrap procedure.

For Scenario I, we continue to find that the limited scope models understate the gains from statewide improvements in the pheasant counts. However, for Scenarios II and III, ignoring the substitute sites outside of the Prairie Pothole Region (as we do in Models C and D) results in welfare estimates that are up to 72% higher than if we include these substitute sites in the model.

## Conclusions and Directions

The results presented above in many ways confirm the difficult nature of determining choice set definitions. In general, there is no obvious answer, leaving the analyst to make decisions based on his best judgement. *A priori*, we had expected site aggregation would not significantly affect the reported welfare measures, since the aggregation was done over fairly homogeneous resources. Yet, for this application aggregation consistently led to reduced welfare estimates. Conversely, our expectations were that scope limitation would affect welfare measures. Specifically, in limiting the scope of the market we are reducing the number of explicitly modeled substitutes, which would in turn make the modeled sites appear more unique, increasing the magnitudes of the estimated welfare effects. This result emerged when the scope restrictions were analogous to those employed by Parsons and Hauber (1998), providing spatial boundaries that eliminated remote sites. However, in the full sample, when the geographical scope restrictions eliminated sites frequently visited by much of the population, the welfare estimates were biased downwards. This perhaps supports the hypothesis that including horizontal substitutes in a numeraire good does not lead to large upward bias in welfare estimates, provided we are only considering changes in the attributes of explicitly modeled sites.

In the end, the choice set issue is also a data and data collection issue requiring pragmatic decisions by the analyst. The specification of choice sets remains as much art as science. We can never hope to gather information about all possible substitutes for all individuals. However, there are perhaps a few simple guidelines that can be followed. First, when gathering data on resource use, every effort should be made to survey not only resource users, but also nonusers. This will enable us to model the "nonparticipation" decision, allowing an aggregation of all horizontal substitutes for the resource of interest. Next, it may be possible to geographically segment the sample population and identify the most likely substitutes for the resource of interest for each segment. This could then be included in the choice sets for the specific sub-sample. This would, or course, require econometric methods capable of handling this heterogeneous specification, proving a direction for further research in the KT model.

Finally, we note that, while the above analysis provides the first empirical investigation into site set definition using the KT framework rather than a single choice occasion RUM model, there are a number of other avenues for future research using the KT model. First, as suggested above, some of the results may be driven in part by the linear expenditure system's functional form, rather than underlying preferences. It would be useful to revisit this problem using a more flexible function form. Second, because analysis has generally focused on single choice occasion RUM models rather than fully utility consistent systems models, little attention has been paid to the micro foundations of aggregation in recreation demand. As we have done in this study, aggregation is typically accomplished via ad hoc averaging of component-site prices and quality measures. It is likely that different aggregation decisions and/or calculation of aggregate prices and quality would affect welfare results. Future research may call on the mature literature addressing aggregation in other areas of consumer choice (see, *e.g.*, Varian 1992, section 9.3, Deaton and Muellbauer 1980, part 2, or Blackorby, Primont, and Russell 1978) that can be readily applied within the KT framework. For example, Lupi and Feather (1998) suggest that there may be advantages to the aggregation of "collateral" sites to keep estimation tractable while allowing a larger market scope. This could be accomplished in the KT model by specifying the collateral sites as homothetically separable from the sites of primary interest, allowing for a theoretically consistent two-stage budgeting model. Under this specification, income would be first allocated between, say, remote trips, local trips and other goods, with a second stage modeling the allocation of local trip expenditures among the local sites. Consistent price and quantity indices could be constructed for the various commodity bundles, rather than relying upon average prices and total trips.[14] Furthermore, the assumption of homothetic separability could be explicitly tested.

## References

Azevedo, C.A. 1999. Linking Contingent Valuation and Travel Cost Data. Ph.D. dissertation, Department of Economics, Iowa State University.

Blackorby, C., D. Primont, and R.R. Russell. 1978. *Duality, Separability, and Functional Structure: Theory and Economic Applications*. New York: Elsevier North Holland, Inc.

Bockstael, N.E., W.M. Hanemann, and I.E. Strand. 1986. *Measuring the Benefits of Water*

[14] Shaw and Shonkwiler (1999) have suggested an alternative approach to constructing consistent site aggregates, relying instead on Hicksian separability. In particular, as travel costs are typically assumed to be proportional to round-trip travel time and/or travel distance, the price of visiting sites moves essentially in a fixed proportion with changes in costs per mile, satisfying the conditions for Hicksian aggregation. Their suggestion is to use total miles driven as the means of aggregating multiple trips, rather than total number of trips.

*Quality Improvements Using Recreation Demand Models,* Vol. 2, U.S. Environmental Protection Agency, Office of Policy Analysis, Washington, D.C.

Deaton, A., and J. Muellbauer. 1980. *Economics and Consumer Behavior.* Cambridge: Cambridge University Press.

Feather, P.M. 1994. Sampling and Aggregation Issues in Random Utility Model Estimation. *American Journal of Agricultural Economics* 76:772–80.

Haab, T.C., and R.L. Hicks. 1997. Accounting for Choice Set Endogeneity in Random Utility Models of Recreation Demand. *Journal of Environmental Economics and Management* 34:127–47.

___. 2000. A Survey of Literature Concerning Choice Set Definition in Recreation Demand. *Marine Resource Economics* 14(4):271–81

Hanemann, W.M. 1978. A Methodological and Empirical Study of the Recreation Benefits from Water Quality Improvement. Ph.D. dissertation, Department of Economics, Harvard University.

Herriges, J.A., C.L. Kling, and C. Azevedo. 1999. Ask a Hypothetical Question, Get a Valuable Answer? Iowa State University Working paper, November, 1999.

Herriges, J.A., C.L. Kling, and D.J. Phaneuf. 1999. Corner Solution Models in Recreation Demand: A Comparison of Competing Frameworks. *Valuing Recreation and the Environment: Revealed Preference Methods in Theory and Practice*, J.A. Herriges and C.L. Kling, eds., Aldershot: Edward Elgar Publishing, Ltd.

Horowitz, J.L. 1991. Modeling the Choice of Choice Set in Discrete-Choice Random-Utility Models. *Environment and Planning A* 23:1237–46.

Kaoru, Y., V.K. Smith, and J.L. Oiu. 1995. Using Random Utility Models to Estimate the Recreational Value of Estuarine Resources. *American Journal of Agricultural Economics* 77:141–51.

Lupi, F., and P.M. Feather. 1998. Using Partial Site Aggregation to Reduce Bias in Random Utility Travel Cost Models. *Water Resources Research* 34:3595–603.

Morey, E.R., R.D. Rowe, and M. Watson. 1993. A Repeated Nested Logit Model of Atlantic Salmon Fishing. *American Journal of Agricultural Economics* 75:578–92.

Parsons, G.R., and A.B. Hauber. 1998. Spatial Boundaries and Choice Set Definition in Random Utility Models of Recreation Demand. *Land Economics* 74:32–48.

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

Parsons, G.R., and M.S. Needelman. 1992. Site Aggregation in a Random Utility Model of Recreation. *Land Economics* 68:418–33.

Parsons, G.R., A.J. Platinga, and K.J. Boyle. 1999. Narrow Choice Sets in Random Utility Models of Recreation Demand. *Land Economics* In press.

Peters, T., W.L. Adamowicz, and P.C. Boxall. 1995. Influence of Choice Set Considerations in Modeling the Benefits of Improved Water Quality. *Water Resources Research* 31:1781–87.

Phaneuf, D.J., C.L. Kling, and J.A. Herriges. 1998. Valuing Water Quality Improvements Using Revealed Preference Methods When Corner Solutions are Present. *American Journal of Agricultural Economics* 80:1025–31.

___. 2000. Estimation and Welfare Calculations in a Generalized Corner Solution Model with an Application to Recreation Demand. *Review of Economics and Statistics* 82:83–92.

Shaw, W.D., and J.S. Shonkwiler. 1999. Brand Choice and Purchase Frequency Revisited: An Application to Recreation Behavior. *American Journal of Agricultural Economics* In press.

Varian, H.R. 1992. *Microeconomic Analysis*, 3rd edition. New York: W. W. Norton & Company, Inc.

Wales, T.J., and A.D. Woodland. 1983. Estimation of Consumer Demand Systems with Binding Nonnegativity Constraints. *Journal of Econometrics* 21:263–85.