Using Qualitative Site Characteristics Data in Marine Recreational Fishing Models: A New Site Aggregation Approach

Jesse Backstrom[1]

Dr. Richard Woodward

Department of Agricultural Economics

Texas A&M University

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2017 Annual Meeting, Chicago, IL, July 30 - August 1, 2017

[1] Corresponding author: Dept. of Agricultural Economics, Texas A&M University, 600 John Kimbrough Blvd, College Station, TX 77843, USA. Email: jessebackstrom@tamu.edu

Abstract

In the marine recreational fisheries demand literature, especially studies using Marine Recreational Information Program (MRIP) data, information on qualitative site characteristics is largely limited as it is often overlooked or too expensive to obtain. Most recreation demand models have incorporated a small number of site characteristics in their empirical estimation, usually limited to distance, a large component of the cost associated with traveling to a fishing site, and some form of expected catch as the only key factors explaining angler site choice. To our knowledge, we are the first to use a relatively new National Marine Fisheries Service (NMFS) data set, published supplementary to the MRIP data, in an approach to deal with site aggregation bias issues and determine the importance of numerous qualitative characteristics[2], such as the proximity to bait and tackle shops, the number of boat ramps, ample parking space, site pressures (congestion), and others in site choice models of marine recreational fishing demand. In order to use these data, a methodological contribution is made by aggregating *less* relevant sites in a new way, while an empirical contribution is made by incorporating a variety of site characteristics not previously analyzed. Preliminary estimation results are provided for anglers taking one-day trips and targeting Spotted Seatrout across several Gulf States.

Introduction

Marine recreational angling contributes significantly to the economy in the Gulf of Mexico as it attracts anglers from a variety of areas across the U.S., and bringing in money to coastal economies via the purchase of fishing gear, bait, boat, food, and other revenue-generating trip expenditures. Although these flows are direct services to society, there are also indirect flows from recreational angling having economic value for which the direct service flows are dependent on, such as the presence of a particular fish species and its catch rate, among other things, contributing to an angler's decision of taking a fishing trip and his or her site choice. To model the indirect flows, revealed preference methods such as the travel cost random utility maximization (RUM) model are often used. The travel cost RUM model is valuable because of

[2] The data used in this analysis are time invariant, as the data set with time-varying characteristics was unavailable at the time of this analysis.

its ability to capture and reveal angler demands and behavior in real markets related to the environmental good, which in this case is a particular fishery and the sites used by anglers to access it. Inferences can then be made about the value an angler places on a fishery indirectly in the form of his or her site selection, as a function of the travel distance (or cost) of taking the trip and the site-specific characteristics the angler also bases his or her decision on.

The Marine Recreational Information Program (MRIP) of the National Oceanic Atmospheric Administration (NOAA) is the way that catch and effort data are counted and reported for federal fishing sites along coastal U.S. waters. In most recreation studies using MRIP data, fishing sites have been aggregated to the county or region level, and have incorporated only a small number of site-specific characteristics in empirical estimation. These characteristics have typically been limited to the travel distance, a large component of the cost associated with traveling to a fishing site, and some form of expected catch as the only key factors explaining angler site choice.

Information on other site characteristics has largely been limited for several reasons. First, extensive qualitative information is often prohibitively expensive to obtain via traditional survey methods and was unavailable through MRIP until 2013. Second, the estimation of choice models with large choice sets can introduce a large computational burden (Parsons and Needlemen et al. 1992; Feather 1994; Kanaroglou and Ferguson 1996; Haab and Hicks 1999; Haener et al. 2004; Lemp and Kockelman 2012; von Haefen and Domanksi 2013; and Alvarez et al. 2014). It also appears that many site characteristics may have been overlooked due to a limited amount of catch data available at the site level, where aggregating to the county level alleviates this concern and eases estimation. Site aggregation and the omission of relevant site characteristics can cause significant bias to parameter estimates on travel distance and catch as each observed and unobserved characteristic is correlated with site choice through its contribution to angler utility. Measurement error in characteristics or independent variables, particularly catch rates, is an additional concern as it also will bias parameter estimates.

We make use of a relatively new National Marine Fisheries Service (NMFS) data set[3] published supplementary to the MRIP data that allows us to incorporate a variety of site characteristics not previously analyzed when modeling marine recreational fishing behavior. Although the time-varying component of the characteristics data was unavailable at the time of this analysis, the set of time-invariant characteristics is still useful and is arguably a more complete representation of relevant observable site characteristics than those considered in a study by Murdock (2006) on recreational fishing on lakes in Wisconsin. The site characteristics used in her study appear to be time-invariant, and include catch rates for various fish species, and dummy variables for physical lake characteristics, regulations, fish management, and urban, rural, and industrial development around fishing sites. We estimate choice models at the site level using a unique method of predicting expected catch rates, which allows us to investigate how the inclusion of site characteristics, such as the proximity to bait and tackle shops, the number of boat ramps, ample parking space, site pressures[4] (congestion), and others, mitigates endogeneity and resulting biased parameter estimates on the catch and travel distance variables.

In addition to estimating models with large complete choice sets we use a site aggregation method that is similar to the approaches used in Parsons and Hauber (1998) and Parsons, Plantinga, and Boyle (2000), but new to the marine recreation demand literature and

---

[3] The data can be obtained from the MRIP public access fishing site register at: https://www.st.nmfs.noaa.gov/msd/html/siteRegister.jsp.

[4] Pressure data from the MRIP site register is not included in this paper as it was unavailable at the time of the analysis.

allows us to study site aggregation bias in a new dimension. To estimate this model we make use of a new and, we believe, previously unused data set from the MRIP site register. Although not included in these preliminary results, in the model specifications using the complete angler choice set we will include site level fixed effects to control for other unobservable determinants of angler choice specific to each site, and for the model specifications with partial site aggregation we will include both site and time fixed effects to control for other unobservable characteristics specific to each site and other time-specific unobservables. We will also include site pressures as a proxy for congestion once the time-varying form of the data is available from the MRIP site register.

In this paper we estimate models of the behavior of recreational fishermen in the Gulf of Mexico for 2013-2015. We focus specifically on Mississippi, Alabama, and western Florida because Texas does not participate in MRIP and Louisiana stopped participating after 2013. We use the results to explore the importance of including site characteristics in recreation demand models. If we are to include site characteristics, aggregating sites (e.g. at the county level) is problematic. Hence, we also introduce new methods to overcome the computational and data-related challenges that have traditionally led researchers to aggregate sites. For each specification we compare parameter estimates on the catch and travel distance variables in models with the new site characteristics information to those omitting it in order to determine the robustness of our estimates. If these new data are rendered valid by our identification strategy and endogeneity concerns are mitigated, then it is likely that future work using MRIP data should include the supplementary site characteristics information.

Literature Review – Choice Set Aggregation

The tradeoff between site aggregation and estimation efficiency has been brought to light in seminal studies of recreation demand in the 1990s, where site aggregation was predominantly studied in settings other than marine fishing. Kaoru and Smith (1990) were the first to investigate the implications of site aggregation in a RUM. In an application with a much larger degree of aggregation, Parsons and Needleman (1992) show that although aggregation is often used to reduce the choice set to a manageable size, it comes with losses in estimation efficiency and caution against it with evidence from their study of recreational fishing on lakes in Wisconsin. Schuhmann (1998) and Train (1998) recognize this shortcoming but to account for potential bias when aggregating fishing sites at the county level, they follow the suggestion of Ben-Akiva and Lerman (1985) by including the log of the number of fishing sites as a quality variable to obtain parameters that are equivalent across scales. Heaner et al. (2004) also find that when developing models of recreational hunting, the size of the choice alternative in spatial context matters.

There are several reasons, both stated and implicit, for why sites have frequently been aggregated in site choice models. Although McFadden (1978) proved that a sampling-of-alternatives design will provide consistent model parameters, it is becoming more common for researchers to use more behaviorally realistic discrete choice models where McFadden's proof does not hold (Nerella and Bhat 2004). Still, the sampling of alternatives can reduce computational run time substantially. In their study of residential location, Nerella and Bhat advise not to consider sample sizes that are too small. For conditional or multinomial logit models, they recommending that using an eighth of the size of the full choice set can be used at a minimum and suggest that a fourth of the size is a desirable target. They also invite future research to investigate the sampling of alternatives in different settings, such as correlation patterns between exogenous variables, to draw more definitive conclusions. In a recreational fishing study investigating the tradeoff between model run-time and the efficiency of parameter estimates, von Haefen and Domanski (2013) use samples with different sized choice sets and find there is modest efficiency loss and significant time savings for conditional logit models with a small number of alternatives. They recommend researchers use best judgement based on their specific need to determine an acceptable amount efficiency loss when reducing choice set sizes.

In another early attempt to a find a better mechanism than site aggregation, Parsons and Kealy (1992) model site choices of recreationists using Wisconsin lakes, and analyzed the effect of using randomly drawn subsets of alternatives to estimate the model parameters. Their results suggest that using draws from the complete set of alternatives might be more effective for shrinking the size of choice sets and lowering computational burden. Feather (1994) finds results that echo these findings in his model of recreational angling in Minnesota. Parsons and Hauber (1998) investigate the sensitivity of RUM parameter and welfare estimates to changes in the spatial boundary of the choice set. Their results imply that the definition of the spatial boundary of the choice set is less important than one might expect, where the 85th percentile of the trips taken have a one-way travel time of less than one hour and a median one-way travel time of approximately 20 minutes, and parameter and welfare estimates show little change as the choice of boundary is increased from 1.6 to 4.0 hours. In their study, the average number of sites in an angler’s choice set increased from approximately 350 to more than 1,500 sites over this range, and it was shown that the distant sites are, for the most part, irrelevant to the angler’s choice set.

Other early work (away from marine recreational fishing) developed new methods to reduce choice set sizes by accounting for familiar vs. unfamiliar vs. favorite sites (Parsons et al.

1999), treating nearby sites as close substitutes and more distant sites as aggregate alternatives in a nested framework (Parsons, Plantinga, and Boyle 2000), or used ways to confine the number of choices modeled such as the efficiency approach (Scrogin et al. 2010), aggregating unimportant sites (Lupi and Feather 1998), predicting choice sets (Ben-Akiva and Boccora 1995), or eliminating sites completely (Peters, Adamowicz, and Boxall 1995; and Hicks and Strand 2000). Phaneuf and Herriges (1999) implemented a similar choice set aggregation approach to Parsons and Hauber (1998), but in a Kuhn-Tucker framework, finding similar results that make modeled sites appear more unique and increasing the magnitudes of estimated welfare effects.

Site aggregation is a particularly important consideration in studies using MRIP data because the complete choice set can contain several hundred to thousands of possible sites, and it is plausible that some or many of these sites (especially the more distant ones) have a near-zero probability of being chosen. Although most studies that use MRIP data have typically aggregated sites to the county level or higher, some have taken one or more additional steps to reduce the size of choice sets, mitigate site aggregation bias, or reduce computational burden. With the exception of Whitehead and Haab (1999), most other MRIP studies that aggregate sites have included the log of the number of choices within each county (Gentner 2007; Haab et al. 2012; Lovell and Carter 2014; and Alvarez et al. 2014). Hindsley, Landry, and Gentner (2011) use Marine Recreational Fisheries Statistics Survey (MRFSS) data, the MRIP parent, and develop a RUM model to investigate the problems inherent with on-site sampling and propose a correction method based on propensity scores that mitigates the sample selection bias. Whitehead and Haab (1999) eliminate sites by using distance and historical catch to determine whether a site is too far or unproductive to be considered a viable choice. Their results were found not to be significantly affected by the elimination of the non-viable choices.

Haab and Hicks (1999) mention that many issues of choice set definition remain to be solved. We believe that one of the unfortunate consequences of site aggregation is that the use of site characteristics in the model is more difficult. It may make sense to assume that the expected catch is roughly the same for all sites within a county. However, it makes substantially less sense to assume that a trip to a county is driven by the average number of parking spaces in sites within the county – that is a characteristic that affects only one site. We believe that studies using MRIP have aggregated sites partly due to choice set issues, but also partly due to data limitations with respect to catch, which is a common issue in recreational fishing demand modeling (Morey and Waldman 1998). In order to predict expected catch for a site in a given wave, catch data might be unavailable due to either no surveys conducted or no anglers present during survey periods, implying a need to obtain more data, use historical catch rates, and or use catch data from nearby sites. Although not in a marine fishing context, Murdock (2006) argues that researchers cannot expect to observe all site characteristics affecting site choice, and implements a two-stage instrumental variable approach to control for unobserved site characteristics in an application to recreational fishing in Wisconsin. Her approach involves omitting various characteristics under various specifications, but controlling for them with an alternative specific constant (or fixed effect) for every site. A benefit of this approach is that it provides a way to relax the assumption that all relevant site characteristics are observed in the available data. Using a large sample with 3581 choice occasions, and reducing the likelihood of bias, Murdock recovers policy-relevant travel cost parameters that differ from traditional models by a factor of four, even though unobserved characteristics may be correlated with travel cost.

Given this variety of findings, it is apparent that there is further research needed to understand how site aggregation reduces computational burden and how it may affect parameter estimates, how the omission of site characteristics affects parameter estimates in RUM models of marine recreational fishing, and how catch can be estimated at the site level when some sites are constrained by data limitations. With no clear consensus on how to handle these issues, we propose and implement methods to overcome each of these issues by aggregating less relevant sites in a new way, predicting expected catch rates in a new way at the site level, and incorporating new characteristics information on the relevant sites in an angler's choice set.

Literature Review – Estimating a Site Choice Model

Drawing on the econometric models of Morey et al. (1993 and 2002), Parsons and Hauber (1998), Parsons, Plantinga, and Boyle (2000), Whitehead and Haab (1999), Haab et al. (2009, 2012), Alvarez et al. (2012 and 2014), Lovell & Carter (2014), and others, we implement several travel cost RUM models to estimate site choice models of marine recreational fishing demand. Standard conditional logit models including all sites within 150 miles of an angler's home zip are estimated initially and serve as the baselines for comparisons to models with partial site aggregation. After distinguishing the statistically desirable site characteristics in angler site selection, our economic model can help to predict how angler trips change in response to changes in the fish stock (i.e., through catch rates). Our models further differ from previous marine recreational fishing studies by the mechanism of site aggregation, the calculation of expected catch per at each site, and the inclusion of new site characteristics information. The standard conditional logit model, as developed by McFadden (1973) and Manksi (1997) is as follows:

$$P_{ij} = P\left(y_i = j\right) = \frac{e^{V_{ij}}}{\sum_{k=1}^{J} e^{V_{ik}}}, \quad \text{where} \quad V_{ij} = \boldsymbol{\beta_q q}_j + \beta_d d_{ij} + \varepsilon_{ij}, \tag{1}$$

where P_{ij} is the probability that angler i will chose site j, $\boldsymbol{q_j}$ is a vector of site characteristics, d_{ij} is the travel distance (a portion of the trip cost) for angler i to get to site j, and ε_{ij} is a random error term with a Type-1 extreme value distribution. The parameters to be estimated are $\boldsymbol{\beta_q}$ and β_d, The log-likelihood equation is

$$LL(\beta) = \sum_{i=1}^{I} \sum_{j=1}^{J} d_{ij} logP_{ij} \,, \tag{2}$$

where I is the number of anglers in the sample, and d_{ij} = 1 if angler i chooses site j and 0 otherwise. Although the conditional logit remains the workhorse of discrete-choice models, it does have the well-known limitation that it fails to satisfy the independence of irrelevant alternatives (IIA) condition, meaning that the analyst believes the relative probability of an angler choosing site A over site B is independent of the attributes of all the other sites.

Although the implementation of the conditional logit model has been successful in numerous recreation demand studies, it has not been used to model angler choice at the site level when using MRIP data. The computational burden is believed to be too great when including each individual site, and estimating choice models at the county level prohibits the use of the qualitative site characteristics data. Our analysis also only concerns anglers fishing on what constitutes as a single-day trip. Hence, following Lovell & Carter (2014), we exclude sites more than 150 miles from an angler’s home zip code. In the next two sections describe our data and propose our approach.

Data

We use MRIP data published by the NMFS of the (NOAA) for this analysis. The models concern recreational angling across Mississippi, Alabama, and the western side of Florida over 2013-2015. The MRIP data come from point-intercept surveys of recreational anglers and include separate files for angler trip, catch, and fish size, but we only use the trip and catch data. Site distances data are published supplementary to the MRIP data set, and include the distance in miles from each intercept site to each surrounding zip code within 150 miles of the site.

The site characteristics data come from the MRIP site register, which is a tool maintained by NOAA Fisheries in cooperation with regional and state partners, and serves two purposes. It acts as a public resource for anglers and others, providing detailed information about all publicly accessible fishing sites. The site register is also used by NOAA to gather other information about each site, including: how often anglers fish throughout the year, what types of fishing occur, and which site amenities attract anglers to visit. Keeping record of this information is important to help ensure that an accurate representation of fishing activity is gathered so it can be determined where to send survey enumerators and in appropriate proportions (NOAA undated).

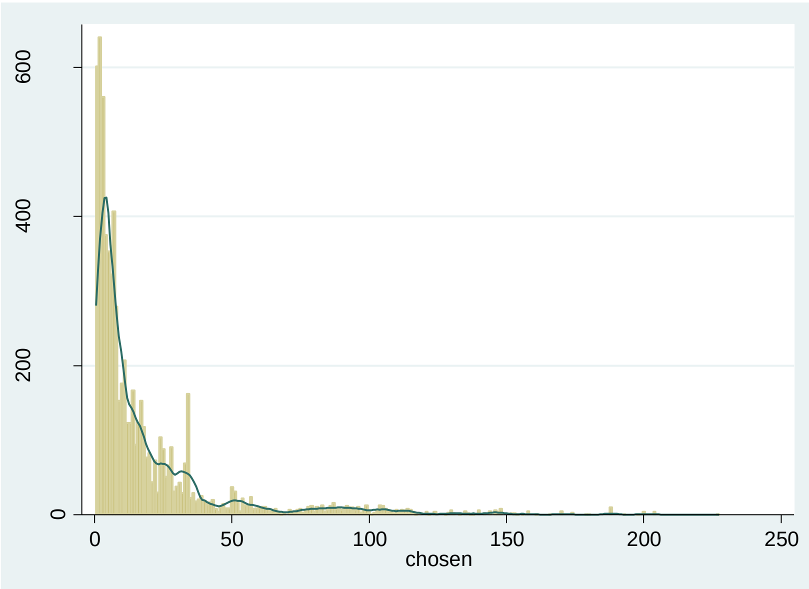
The characteristics list includes variables for the numbers of boat slips, car parking spaces, charter boats, head boats, boat ramps, and trailer parking spaces, it includes dummy variables on whether there is boat maintenance and repair, boat storage, fuel docks, bait and tackle shops, and restaurants and lodging nearby; dummy variables indicating if there are fish cleaning stations onsite, if fishing is affected by tide, if there is a fee required for access or if the site has private access, if there are major tournaments held at the site, the type of shore area and shore mode; and surveyor-type dummies including information on whether interviews can be conducted, if the site is safe for two samplers at night, and if there is night lighting. The characteristics data set used currently is time-variant, implicitly assuming that the characteristics do not change over time. Time-varying characteristics will be used in later analysis when they become available from NOAA Fisheries.

Partial Choice Set Aggregation

The complete MRIP angler choice set includes over 500 potential recreational fishing sites with either trips taken that targeted or reported catch of Spotted Seatrout, spanning from Mississippi to Collier County in southwestern Florida. In order to reduce the size of the choice set, we cannot involve simply lump sites together, because that would require unreasonable assumptions about site characteristics being the same for all sites within a given aggregate site. For example, given a county containing a total of 10 sites, it is unreasonable to aggregate these sites and attempt to estimate a model as if the site characteristics are the same for all ten sites. We propose and implement a partial site aggregation mechanism for an "outside option" in each angler's choice set. Parsons and Hauber (1998) provide evidence of such an effect, where they find that there exists some threshold distance beyond which adding more sites to the choice set has negligible effects on the estimation results. Lupi and Feather (1998) and Parson, Plantinga, and Boyle (2000) offer similar strategies for doing so, involving including popular individual sites in the choice set, and then including aggregations of other remaining groups of similar sites.

Our approach builds on the fact that anglers tend to visit one of the closest sites. Although an angler could take a day trip to any site within 150 miles of their home, for most anglers the effective choice set is limited to the sites that are nearest to the angler's home. As seen in Figure 1, 75% of angler trips in our data are to one of the nearest 25 sits and 90% of the trips are to one of the nearest 50 sites. Building on this, our proposed approach models all of the angler's site choices as between the nearest *n* sites, or one of the many sites that are more distant.

Figure 1. Histogram of Ranked Chosen Sites by Private Boat Anglers.



The y-axis shows frequencies of choices, the x-axis shows the ranks of the chosen sites.

To implement this new approach, we alter the choice set in the standard conditional logit model (1). Instead of J alternatives, all sites within 150 miles of an agler's home use a modified choice set consisting of J^*, the n nearest sites, and one alternative consisting of all of the more remote options, which is treated as a constant term. In this specification, equation (1) is rewritten:

$$P\left(y_i = j \in J^*\right) = \frac{e^{V_{ij}}}{1+\sum_{k=1}^{n} e^{V_{ik}}} \quad \text{or} \quad P\left(y_i = j \notin J^*\right) = \frac{1}{1+\sum_{k=1}^{n} e^{V_{ik}}}. \qquad (5)$$

The numerator in (5) remains the same as in the full model if $j \in J^*$, i.e. for one of the n nearest sites. If the one of the more distant sites is chosen, the numerator will collapse to one and is estimated as a constant. The denominator in this probability is what changes the most under the new specification, as it is smaller due to fewer choices.

Issues with Expected Catch

Site choice models of marine recreational angling can suffer from several forms of endogeneity from the expected catch variable. First, it is likely that catch rates are not the only observable component determining angler site choice. Previous studies using MRIP data have omitted other important site characteristics that help characterize the quality of a site and are correlated with catch rates as they contribute to angler utility. When these characteristics are omitted, the parameter estimate on the catch variable can be biased. Controlling for the omitted site characteristics is not a valid identification strategy to recover an unbiased parameter estimate on expected catch. Ideally, an instrument for expected catch would be used, but at the time of this analysis data on an instrument such as fish populations were unavailable. Our future work will mitigate the concern for endogeneity from omitted variables by including a full set of site level fixed effects to control for another dimension that is specific to each site.

Second, selection bias is another concern in site choice models because different types of anglers may have heterogeneous concerns (or preferences) for catch rates. For example, more skilled anglers (for e.g., catch-and-release anglers), might be less concerned with catch rates since they may be more confident in their abilities to catch fish (preferring a challenge). They also may prefer sites with fewer total fish, but fish of a larger size are present, and in essence are in hunt of a trophy fish. These preferences are unobservable to us as analysts, but imply that the parameter estimate on the catch variable may not be representative of the full population of anglers. To mitigate this concern, a proxy variable for angler avidity such as the number of days fished in the last year might be used, or a random coefficient on the catch variable might also be considered. We do not investigate either of these options in our preliminary estimations.

Finally, simultaneity is plausible in site choice models if realized angler catch for the chosen wave (or period) is included in the calculation of the catch rate for that site in the current period. An angler's site choice affects the catch rate at that site after visiting and catching fish and it is also likely that anglers choose a particular site because of a good catch rate at that site. We correct for this simultaneity concern by proxying for current wave catch rates with a weighted spatial-temporal form of lagged (or past) catch rates that predicts deviations of expected catch rates from the county average one period ahead for all sites over 2013-2015.

Predicting Expected Catch

Most studies using MRIP data have calculated expected catch by aggregating sites to the county level and then used some form of lagged average catch rates at all sites within each county. We opt for a similar approach to Murdock (2006) as we believe there is reasonable correlation between nearby fishing sites and over time since fish migrate and also because anglers can travel several miles in their boats. Our approach predicts the expected catch at site *j* in period *t*, $C_{j,t}$, based on observed catch data at surrounding sites and at the same site in previous periods. Specifically, the prediction of $C_{j,t}$ starts with the average catch over the full period observed (2012-15) in the county in the the j^{th} site is located, $\bar{C}_j$. We then estimate the deviation of $C_{j,t}$ from $\bar{C}_j$ using catch observations in surrounding sites in the current period and in all sites in the previous periods, with the weight given to a particular observation declining as the distance from *j* and *t* increases in both time and space. Specifically,

$$E\left(C_{j,t}\right)=\frac{\Sigma_k \omega^{dist}\left(k,j\right)\omega^{wv}\left(k,t\right)\omega^{yr}\left(k,t\right)\left(C_k-\bar{C}_k\right)}{\Sigma_k \omega^{dist}\left(k,j\right)\omega^{wv}\left(k,t\right)\omega^{yr}\left(k,t\right)}+\bar{C}_j \quad (5)$$

where $\bar{C}_k$ is the average catch in the county of the k^{th} observation, and $\omega^{dist}(k, j)$, $\omega^{wv}(k,t)$, and $\omega^{yr}(k,t)$ are weights given for the k^{th} observation in the calculation of the $E(C_{j,t})$. We assume that the weights decline geometrically as follows: $\omega^{dist}(k, j) = exp(-\alpha_d \cdot dist(k, j))$ where $dist(k, j)$ is the linear distance between site j and the site at which of the k^{th} observation, $\omega^{wv}(k,t) = exp(-\alpha_w \cdot (t - w(k)))$ where $w(k)$ is the wave of the k^{th} observation for $t - w(k) \leq 5$ and $\omega^{y}(k,t) = \alpha_y$ for $t - w(k) = 6$.

We then choose α_d, α_w, and α_y, to minimize the sum of squared differences between predicted expected and actual catch. When these parameters were estimated in an unconstrained model, the estimated coefficients essentially equal weight given to all observations in the data. To ensure sufficient reasonable diffusion in the weights over time and space the coefficients were constrained as follows: $0.01 \leq \alpha_d \leq 0.3$; $0.05 \leq \alpha_w \leq 7$; and $0 \leq \alpha_y \leq 1$. For the catch rates in each wave 2013, we use data from the previous six waves in 2012, starting with wave one.

Preliminary Estimation Results

The baseline models for the specifications with the full choice set and the aggregated outside option included travel distance and expected catch as the only determinants of angler site choice. All specifications were estimated with robust standard errors. The output for these results can be seen in Table 1, below. In columns CL_1 and CLO_1, all coefficient estimates were significant at the 1% level. The coefficient on travel distance declined slightly in the specification with the outside option, plausibly due to the omission of distant sites in choice sets, where the average distance traveled declined slightly.

Table 1. Standard Conditional Logit Models with All Sites (CL) and Outside Option (CLO)

VARIABLES	CL_1	CLO_1	CL_2	CLO_2
Distance (Miles)	-0.0811***	-0.0680***	-0.0828***	-0.0695***
	(0.00124)	(0.00127)	(0.00130)	(0.00135)
Expected Catch	0.287***	0.348***	0.412***	0.549***
	(0.0267)	(0.0180)	(0.0301)	(0.0177)
#Boat Slips			0.00153***	0.00134***
			(0.000226)	(0.000235)
#Car Parking Spaces			-0.00133***	-0.00118***
			(0.000265)	(0.000258)
#Boat Ramps			0.150***	0.186***
			(0.00829)	(0.00773)
#Trailer Parking Spaces			0.0197***	0.0178***
			(0.000423)	(0.000380)
Boat Maintenance/Repair			-0.240***	-0.278***
			(0.0903)	(0.0883)
Boat Storage			-0.667***	-0.523***
			(0.0740)	(0.0699)
Fish Cleaning Stations			-0.102***	-0.0422
			(0.0338)	(0.0316)
Fuel Dock			0.237***	0.222***
			(0.0643)	(0.0639)
Fee			-0.409***	-0.339***
			(0.0369)	(0.0347)
Night Lighting			-0.307***	-0.254***
			(0.0347)	(0.0331)
Lodging			-2.282***	-1.903***
			(0.118)	(0.106)
Major Tournaments			1.006***	0.889***
			(0.0360)	(0.0347)
Restaurants			-2.238***	-2.359***
			(0.134)	(0.116)
Bait Shop			-1.065*	-0.841
			(0.599)	(0.594)
Tackle Shops			-0.525***	-0.344***
			(0.0718)	(0.0664)
Bait_Tackle			1.865***	1.488**
			(0.601)	(0.596)
Restaurant_Lodging			1.873***	1.943***
			(0.139)	(0.117)
#Anglers	6996	6996	6996	6996
Total Sites	432	433	432	433
Avg #Choices	144.1	46.15	144.1	46.15
Max #Choices	224	50	224	50
Min #Choices	5	5	5	5
Run Time (Seconds)	297.9	89.58	300.8	148.6
Robust SEs	YES	YES	YES	YES

*** p<0.01, ** p<0.05, * p<0.10

The coefficient on expected catch increased in magnitude in the specification with the outside option. In columns CL_2 and CLO_2, we added the full set of time-invariant site characteristics to each choice set specification, along with interaction terms concerning the correlations between bait and tackle shops and restaurants and lodging. The coefficient estimates on travel distance and expected catch remained significant at the 1% level, but the coefficient on travel distance increased slightly and the coefficient on expected catch increased significantly in the model with the full choice set. In column CLO_2 for the specification with the outside option, we see that both the coefficients on travel distance and expected catch remained significant. The coefficient on travel distance remained at roughly the same magnitude as the baseline model, but the coefficient on expected catch increased significantly more relative to the increase seen in the specification with the full choice set.

It might be expected that site characteristics are correlated with more popular sites. Meaning that sites that are frequented more might be more likely to have better characteristics or vice versa. After estimating the baseline models, we viewed correlation tables for all characteristics to get an idea for which characteristics are, or might be, correlated and or collinear. It appeared that several groupings of characteristics were highly correlated with one another, particularly the service-type characteristics such as boat storage, boat maintenance and repair, and fuel docks, and nearby amenities such as bait and tackle shops and restaurants and lodging. Since high correlations between these variables were present (i.e. greater than 0.6), in the next model specification in columns CL_3 and CLO_3 we decided to include only one variable from each of these groupings, keeping boat storage, bait shops, and restaurants as the variables capturing this variation. Each of these specifications can be seen in Table 2, below.

Table 2. More Standard Conditional Logit Models with All Sites (CL) and Outside Option (CLO)

VARIABLES	CL_3	CLO_3	CL_4	CLO_4
Distance (Miles)	-0.0834***	-0.0695***	-0.0828***	-0.0698***
	(0.00133)	(0.00131)	(0.00130)	(0.00132)
Expected Catch	0.464***	0.550***	0.494***	0.544***
	(0.0291)	(0.0176)	(0.0287)	(0.0178)
#Car Parking Spaces	-0.000167	0.243***		0.239***
	(0.000207)	(0.00609)		(0.00639)
#Boat Ramps	0.192***	0.0149***	0.198***	0.0144***
	(0.00690)	(0.000315)	(0.00611)	(0.000308)
#Trailer Parking Spaces	0.0164***		0.0158***	-0.307***
	(0.000341)		(0.000348)	(0.0498)
Boat Storage	-0.245***	-0.448***		-0.407***
	(0.0520)	(0.0348)		(0.0353)
Fish Cleaning Stations	-0.301***	0.732***		0.706***
	(0.0352)	(0.0310)		(0.0309)
Fee	-0.480***		-0.530***	0.208***
	(0.0373)		(0.0384)	(0.0337)
Night Lighting	-0.0729**			-0.536***
	(0.0311)			(0.0339)
Major Tournaments	0.966***	-0.699***	0.886***	
	(0.0354)	(0.0366)	(0.0333)	
Restaurants	-0.661***	0.133***	-0.682***	
	(0.0390)	(0.0290)	(0.0389)	
Bait Shop	0.222***		-0.0376	
	(0.0415)		(0.0299)	
#Anglers	6996	6996	6996	6996
Total Sites	432	433	432	433
Avg #Choices	144.1	46.15	144.1	46.15
Max #Choices	224	50	224	50
Min #Choices	5	5	5	5
Run Time (Seconds)	293.1	93.04	292.5	97.94
Robust SEs	YES	YES	YES	YES

*** $p<0.01$, ** $p<0.05$, * $p<0.10$

Relative to the output in Table 1, the coefficient on travel distance in CL_3 maintained the same significance and increased just slightly relative to the previous two specifications with the full choice set, and the coefficient on expected catch remained significant and increased as well. The coefficients on all of the other characteristics were significant at the 1% level. For CLO_3, the coefficients on travel distance and expected catch remained significant at the 1%

level and stayed nearly the same as in CLO_2. Fewer site characteristics included in this specification however, were significant, and as a result, we dropped the number of trailer parking spaces, access fee, night lighting, and bait shop from this specification. Similar specifications were run in columns CL_4 and CLO_4, but including a different set of site characteristics. Importantly, the coefficients on travel distance remained significant and nearly the same. The coefficients on expected catch also remained significant, but increased in magnitude relative to the estimates in CL_3 and CLO_3.

Discussion

We believe that we are the first to estimate a site choice model of marine recreational fishing demand using MRIP data at the site level. The coefficient estimates of models with varying choice set specifications provide some interesting results. With a large sample size of 6996 day-trip anglers, we believe that the potential for endogeneity to bias the travel distance variable is somewhat limited, but there is still a small downward change when estimating the model with the outside option that needs additional investigation. The coefficients on expected catch however, changed significantly across all specifications, including when more characteristics are included in the specification with the full choice set, and also when estimating the models with the outside option.

It is apparent that each specification needs additional investigation to determine whether these changes are due to an endogenous component of catch, or if coefficient estimates on catch are becoming more precise with the inclusion of additional site characteristics and or specifying the model with an outside option. If we are to make claims that our coefficient estimates on travel distance and catch (especially) are unbiased, we need to control for all characteristics (observed with error and unobserved) in order to say we have a valid identification strategy. Including the site FEs is a potential falsification test that we can use to show that our results are robust. It is also another way to check if our method for predicting expected catch provides sufficient variation across sites and whether it is a reasonable proxy for actual catch.

Relative to the specifications with the full choice set, model run times improved significantly when estimating the specifications with an outside option. As shown at the bottom of Tables 1 and 2, estimation with the outside option took significantly less than half the time it took to run the same or similar models with specified with the full choice set, as the average number of choices per angler was cut from 144.1 to 46.15. A closer look at the changes in the coefficient estimates across these specifications is needed in order to weigh tradeoff between model run time and estimation efficiency and precision.

Conclusions and Further Work

There appears to be a level of subjectivity involved in travel cost models with respect to the selection of a choice set as most studies of recreation demand can realistically model site choice using full choice sets that are quite large. Modeling site choice can be difficult as choice sets can be endogenously determined, plausibly due to possible patterns of angler substitution among sites. Haab and Hicks (1997) argue that researchers typically assume that the angler's choice set is the same as the set of alternatives included in the survey instrument or sometimes what is most convenient to them, which stains credibility. In other previous studies, including too few or too many sites can be a shortcoming that results in incorrect or inefficient parameter estimates on the travel distance (or travel cost) and or catch variables.

In most cases, anglers are not likely to know or think seriously about more than a few sites, which implies it is up to the analyst to determine a reasonable method for determining appropriate choice set sizes. For this reason, researchers have estimated models using much narrower choice sets in a variety of ways, sometimes relying on surveys responses to obtain information on actual choice set perception or awareness of anglers (Parson, Massey, and Tomasi 1999). We believe that in the marine recreation context, an appropriate method for determining reasonable choice sets has not been implemented previously. Aggregating sites to the county or region level is not sufficient as it disables the use of site specific characteristics that are important to an angler's choice decision. If sufficient care is not taken when determining choice sets, parameter and welfare estimates can be significantly biased with respect to the true underlying choice set.

Using a rich data set of site characteristics, our preliminary estimates show that the travel cost parameter does not change significantly across a variety of specifications, particularly when including only the travel distance and catch variables under the complete choice set and outside option specifications, and when including the travel distance and catch variables and an array of site characteristics under the same two choice set specifications. The parameter on the catch variable however, changes significantly across specifications, which may be attributable to a downward bias from the omission of the site characteristics, or due to method of predicting expected catch being inadequate. Further investigation will involve incorporating the time-varying site characteristics data as well as the pressure data, adding fixed effects to each specification, and re-examining our method for predicting lagged catch. We will also calculate marginal effects for each specification, and in addition we would like to obtain more detailed cost information on angler trips in order to estimate welfare effects.

References

Alvarez, S., S. Larkin, J. Whitehead, and T. Haab. 2012 (Working Paper). "Substitution, Damages, and Compensation for Anglers due to Oil Spills: The Case of the Deepwater Horizon." *Working Paper – Presented at the 2012 AAEA Meeting*.

Alvarez, S., S. Larkin, J. Whitehead, and T. Haab. 2014. "A Revealed Preference Approach to Valuing Non-Market Recreational Fishing Losses from the Deepwater Horizon Oil Spill." *Journal of Environmental Management*. 145: 199-209.

Ben-Akiva, M., and B. Boccara. 1995. "Discrete Choice Models with Latent Choice Sets." *International Journal of Research in Marketing.* 12: 9-24.

Ben-Akiva, M, and S.R. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*, 1st ed. Boston, MA: MIT Press.

Chen, M., and F. Lupi. 2009 (Working Paper). "Does Endogeneity of Site Facilities in Recreation Demand Models Lead to Statistical Endogeneity?" *Working Paper – Presented at the 2009 AAEA Meeting.*

Feather, P. 1994. "Sampling and Aggregation Issues in Random Utility Model Estimation." *American Journal of Agricultural Economics.* 76(4): 772-780.

Gentner, B. 2007. "Sensitivity of Angler Benefit Estimates from a Model of Recreation Demand to the Definition of the Substitute Sites Considered by the Angler." *Fishery Bulletin* 105: 161-167.

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(2): 127-147.

Haab, T.C., and R.L. Hicks. 1999. "Choice Set Considerations in Models of Recreation Demand: History and Current State of the Art." *Marine Resource Economics.* 14(4): 271-281.

Haab, T.C., and K.E. McConnell. 2002. *Valuing Environmental and Natural Resources: The Econometrics of Non-Market Valuation*, 1st. ed. Northampton: Edward Elgar Publishing, Inc.

Haab, T., R. Hicks, K. Schnier, and J.C. Whitehead. 2012. "Angler Heterogeneity and Species-Specific Demand for Recreational Fishing in the Southeast United States." *Marine Resource Economics.* 27(3): 229-251.

Haener, M.K., P.C. Boxall, W.L. Adamowicz, and D.H. Kuhnke. 2004. "Aggregation Bias in Recreation Site Choice Models: Resolving the Resolution Problem." *Land Economics.* 80(4): 561-574.

Hicks, R.L., and I.E. Strand. 2000. "The Extent of Information: Its Relevance for Random Utility Models." *Land Economics.* 76(3): 374-385.

Hindsley, P., C.E. Landry, and B. Gentner. 2011. "Addressing Onsite Sampling in Recreation Site Choice Models." *Journal of Environmental Economics and Management*. 62: 95-110.

Kanaroglou, P.S., and M.R. Ferguson. 1996. "Discrete Spatial Choice Models for Aggregate Destinations." *Journal of Regional Science*. 36(2): 271-290.

Kaoru, Y., and V.K. Smith. 1990. "'Black Mayonnaise' and Marine Recreation: Methodological Issues in Valuing a Cleanup." *RFF Discussion Paper QE91-02.* Resources for the Future, Washington, DC.

Lovell, S.J., and D.W. Carter. 2014. "The Use of Sampling Weights in Regression Models of Recreational Fishing-site Choices." *Fishery Bulletin*. 112: 243-252.

Lupi, F., and P.M. Feather. 1998. "Using Partial Aggregation to Reduce Bias in Random Utility Travel Cost Models." *Water Resources Research.* 34(12): 3595-3603.

Manksi, C.F. 1977. "The Structure of Random Utility Models." *Theory and Decisions.* 8: 229-245.

McFadden, D. 1973. "Conditional Logit Analysis of Qualitative Choice Behaviour." In P. Zarembka, ed. *Frontiers in Econometrics.* New York: Academic Press, pp. 105-142.

McFadden, D. 1978. "Modeling the Choice of Residential Location." *Transportation Research Record.* 673: 72-77.

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: 578-592.

Morey, E.R., and D.M. Waldman. 1998. "Measurement Error in Recreation Demand Models: the Joint Estimation of Participation, Site Choice, and Site Characteristics." *Journal of Environmental Economics and Management*. 35(3): 262–276.

Morey, E.R., W.S. Breffle, R.D. Rowe, and D.M. Waldman. 2002. "Estimating Recreational Trout Fishing Damages in Montana's Clark Fork River Basin: Summary of a Natural Resource Damage Assessment." *Journal of Environmental Management.* 66: 159-170.

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

Nerella, S., and C.R. Bhat. 2004. "Numerical Analysis of Effect of Sampling Alternatives in Discrete Choice Models." *Transportation Research Record.* 1894: 11-19.

NOAA. Undated. 'How to Use the Site Register'. Available at: https://www.st.nmfs.noaa.gov/recreational-fisheries/data-and-documentation/site-register [Accessed November 29, 2016].

Parsons, G.R., and A.B. Hauber. 1998. "Spatial Boundaries and Choice Set Definition in a Random Utility Model of Recreation Demand." *Land Economics.* 74(1): 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(1): 93-106.

Parsons, G.R., D.M. Massey, and T. Tomasi. 1999. "Familiar and Favorite Sites in a Random Utility Model of Beach Recreation." *Marine Resource Economics.* 14(4): 299-315.

Parsons, G.R., and M.S. Needleman. 1992. "Site Aggregation in a Random Utility Model of Recreation." *Land Economics*. 68(4): 418-433.

Parsons, G.R., A.J. Plantinga, and K.J. Boyle. 2000. "Narrow Choice Sets in a Random Utility Framework of Recreation Demand." *Land Economics.* 76(1): 86-99.

Peters, T., W.L. Adamowicz, and P.C. Boxall. 1995. "Influence of Choice Set Considerations in Modeling the Benefits from Improved Water Quality." *Water Policy and Economics.* 31(7): 1781-1787.

Phaneuf, D.J., and J.A. Herriges. 1999. "Choice Set Definition Issues in a Kuhn-Tucker Model of Recreation Demand." *Marine Resource Economics.* 14(4): 343-355.

Schmidheiny, K., and M. Brülhart. 2011. "On the Equivalence of Location Choice Models: Conditional Logit, Nested Logit and Poisson." *Journal of Urban Economics.* 69: 214-222.

Schuhmann. P.W. 1998. "Deriving Species-Specific Benefits Measures for Expected Catch Improvements in a Random Utility Framework." *Marine Resource Economics.* 13(1): 1-21.

Scrogin, D., R. Hofler, K. Boyle, and J.W. Milon. 2010. "An Efficiency Approach to Choice Set Formation: Theory and Application to Recreational Destination Choice." *Applied Economics.* 42: 333-350.

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

Von Haefen, R., and A. Domanski. 2013 (Working Paper). "Estimating Mixed Logit Models with Large Choice Sets. *Working Paper – January 2013.*

Whitehead, J.C., and T.C. Haab. 1999. "Southeast Marine Recreational Fishery Statistical Survey: Distance and Catch Based Choice Sets." *Marine Resource Economics*. 14(4): 283-298.