Title	Modelling the decision making behaviour of coarse anglers in Ireland
Author(s)	Deely, John
Publication Date	2020-02-20
Publisher	NUI Galway
Item record	http://hdl.handle.net/10379/15797

Modelling the Decision-Making Behaviour of Coarse Anglers in Ireland

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

John Deely Bcomm, M.Econ.Sc.

Supervised by
Dr. Stephen Hynes (NUI Galway)
Dr. John Curtis (ERSI)

Head of Economics: Professor Paddy Gillespie

Discipline of Economics
J.E. Cairnes School of Business and Economics
December 2019

Contents

List of Figures

List of Tables

Declaration

I declare that this thesis, submitted to the National University of Ireland, Galway for the degree of Doctor of Philosophy (Ph.D.), has not been submitted for a degree at this or any other university. All research contained herewith is entirely my own and the use of all material from other sources has been properly and fully acknowledged.

John Deely

Abbreviations

AIC	Akaike Information Criteria
ASC	Alternative Specific Constants
BIC	Bayesian Information Criteria
CB	Contingent Behaviour
CC	Complete Case
CL	Conditional Logit
CS	Consumer Surplus
CV	Compensating Variation
IFI	Inland Fisheries Ireland
IIA	Independence of Irrelevant Alternatives
IID	Independent and Identically Distributed
KS	Kolmogorov–Smirnov
LOCF	Last Observation Carried Forward
MAR	Missing At Random
MCAR	Missing Completely At Random
MI	Multiple Imputation
MICE	Multiple Imputations by Chained Equations
MNAR	Not Missing At Random
NSAD	National Strategy for Angling Development
PO	Per Observation Mean Substitution
PP	Per Person Mean Imputation
RMSE	Root Mean Squared Error
RPL	Random Parameter Logit
RUM	Random Utility Model
SEs	Standard Errors
TCM	Travel Cost Models
WTP	Willingness To Pay

Acknowledgments

First and foremost, I would like to express my gratitude to my supervisors Dr. Stephen Hynes and Dr. John Curtis for all they have done for me over the last four years. Their knowledge, experience and generosity, in both time and spirt, has been of immeasurable benefit to me. Their guidance ensured the successful completion of this thesis. For that, I am unimaginably grateful. Beyond the academic, each meeting left me feeling lighter, which is sometimes all that is needed when mired by what can be a seemingly unending endeavour.

It has been my immense pleasure to be able to undertake my studies in the Discipline of Economics in the National University of Ireland, Galway. Every member of staff I have encountered has enhanced this experience through their personal demeanour and professional excellence. A special thanks must be extended to the members of my graduate research committee, Alan Ahearne, Stephen McNena and Eamon O'Shea for all their guidance and support.

I would also like to thank all the members Inland Fisheries Ireland for their expressed interest in the project and the insightful questions and comments whenever we met. Three members of the Inland Fisheries Ireland deserve a special mention. Firstly, Suzanne Campion whose support has meant so much to me, to Shane O'Reilly for all his help with the survey development and, most importantly, Paul O'Reilly as without his help I would never have been able to complete this project.

To all my fellow PhD candidates whose advice, support and merriment have been a delight during this period of my life. Although so many should be thanked a few people must be mentioned, Danny Norton, the cartographer extraordinaire, Rebecca Corless, for her eagle eyed proofing, Tom Gillespie, Suzanne Van Osch, Edward Knapp, Eoin McGurk, Jason Harold and all the team at Semru, of which I am proud to be a member.

To my parents, Rita and Dennis and my brother Patrick I am eternally grateful for all you have done for me. The breath and width of the English language has yet to find the words that each of you deserve for what you have always done for me. Above all else, it is the love and support that each member of my family has so graciously given me that I most treasure. There was not a moment ye were not there for me, for that and everything else, I wish to express how grateful I am.

Finally, my deepest thanks to my girlfriend Kate Dawson. Your support has been, without question, the most vital and valuable contributor to this process. You have been kind, loving and supporting throughout, even when your own work has been trying. Your help has extended far beyond mere words of encouragement. Difficult moments were far less daunting not only by your presence but also by the insight and assistance you provided. By the time you read these words you will have finished your PhD too, but always remember I finished mine first!

Funding from Inland Fisheries Ireland is gratefully acknowledged.

Abstract

Coarse angling is a vibrant and important aspect of recreational angling in Ireland. Both Irish anglers and tourist anglers contribute directly and indirectly to the Irish economy through their participant in this pastime and sport. In order to increase demand for this activity, as proposed by the national strategy for angling development, an understanding of the drivers of participation must be comprehensively understood. This is the primary aim of this thesis.

Efficient enhancement of a site requires an understanding of why one site is chosen over another. Some characteristics may be desirable to anglers, whereas other may lessen the sites' attractiveness. In chapter 2, site choice models are applied to data using anglers' own perception of key site attributes to determine which site characteristics play a significant role in site selection. As it is hypothesised that not all anglers will have the same preferences for each site attribute, two different forms of the logit model are used. The first, the conditional logit, assumes homogeneity across preferences. The second, the random parameter logit, allows for variance in preferences of the site attributes. These two models are then compared based on model fit. The results indicate that the surveyed anglers do have heterogeneous preferences, as the random parameter logit presents a better fitting model. The results of the random parameter logit are used to estimate marginal willingness to pay for a change in site attributes, as well as the implications of several policy scenarios. This provides valuable insight into a direction for future coarse angling site development.

The effective management and development of Irish coarse angling sites is highly dependent on the correlation between managements' perspective of the sites and users' perspectives. If it is the case that management and users perceive sites differently, then even the most value enhancing development policy may be implemented ineffectively. In chapter 3 a comparison is made between the results of a random parameter logit applied to the users-based data and manager-based data. Comparison between the results are made by testing for statistically significant differences of parameter estimates.

Following this, new data sets are generated from the user-based data. This new data is used to investigate whether the results of the management-based data closely align with any segment of the user-based data.

Analysts of choice data, particularly those using perceived data, often encounter the problem of missing data. The data used throughout the thesis presents with this problem. There are several techniques that can be used to overcome this obstacle. However, a rigorous comparison of these techniques to choice data has yet to be explored. Chapter 4 fills this gap in the literature. Using data with full information, a conditional logit model is applied. The results of this model are used as a benchmark against which the missing data techniques are compared. Data is then generated missing randomly from a subset of the data with full information at three different percentages of missing data. Four missing data techniques are applied to the data sets with missing information; complete case analysis, two forms of mean imputation, and multiple imputations. Following this, conditional logits are applied. Using a host of tests, the results of the conditional logits are then compared to the original results from the data set with full information. Additionally, willingness to pay estimates are generated using each technique to demonstrate the effects of missing data on policy formation.

The quality of fish, both size and quantity, at a coarse angling site is assumed to impact fishing participation, measured in days spent fishing per year. However, for Irish based research, both size and quantity are seldom jointly significant. In many cases, only size or number of specimen fish at the site impacts participation. Chapter 5 explores this idea, through the use of the contingent behaviour method. Anglers are asked how their angling participation would change if the number of specimen fish or the quantity of fish increased at Garadice, the most popular of the site of interest. Unlike the models presented in the earlier chapters both Irish and tourist angler data are used. A traditional travel cost model is also applied to determine the drivers of angler participation at Garadice and to examine if there is a difference in willingness to pay for a day spent fishing for the Irish anglers as opposed to the tourist anglers.

Chapter 1

1. Introduction

1.1 Introduction

Entitled “A treatyse of Fysshynge wyth an Angle”, the first known English language book on angling was published in 1496. It describes the virtues of angling; creating both moral character and reverence for fellow-men and lower creatures. The book’s author, Dame Juliana Berners describes angling as the “meanes and the causes that enduce a man in to a mery spyryte” (Berners 1880, P. 1). Since the book’s publication, many things have changed but angling remains a pastime that continues to be held in the highest regard by all its practitioners.

In Ireland, recreational angling is a popular pastime for many and an important sport for others. It is estimated that a quarter of a million Irish people fish annually with many more overseas anglers travelling for the unique fishing product Ireland has to offer (Tourism development Ireland 2013). These anglers contribute to the Irish economy, providing an additional €755 million per annum (includes multipliers) and support approximately 10,000 jobs (TDI 2013). The quarter of a million Irish anglers and the numerous tourist anglers play an important role in many rural villages around Ireland as they keep them socially and economically vibrant, through fishing festivals, seasonal tourism and local pride.

Anglers are commonly categorised by the type of fish they target. In Ireland, there are five different types of recreational angler; sea, pike, salmon, coarse and trout. However, these categories are not mutually exclusive as an angler can fish for many different species in a year. The focus of this thesis is anglers who target coarse fish. The coarse fish these anglers target are non-game freshwater fish which include; bream, tench, roach, rudd, hybrids, perch, eels, dace and carp.

Coarse anglers make up about 7% of the total number of domestic anglers. However, due to coarse angling having a 12-month fishing season, many anglers will fish for coarse species during the closed season of a more preferred fish species. Consequently, about 19% of all anglers who went on any type day fishing trip and 31% of those that went on an overnight trip, fished for coarse fish (TDI 2013). Coarse angling has also proven to be a big draw for overseas anglers with about 21% of the 118,000 fishing tourists being attributable, in some part, to coarse fishing (IFI 2015). It is estimated that the economic contribution to the Irish economy from coarse angling alone was about 96 million in 2014, supporting 1,380 jobs (IFI 2015).

The protection and management of Irish fisheries is conducted by Inland Fisheries Ireland (IFI). IFI was established in 2010, through the amalgamation of the central fisheries board and seven regional fisheries boards into a unified agency. IFI has jurisdiction over Ireland's 74,000 kilometres of rivers and streams, and 128,000 hectares of lakes, as well as Ireland's sea angling resources. In addition to protection and management, IFI has been charged with the task of re-engaging lapsed anglers, introducing more non-anglers to angling and to increase the number of overseas anglers fishing in Ireland (NSAD 2015). In order to complete this task a strong understanding of the decision-making behaviour of coarse anglers and would be coarse anglers is vital. However, relatively little is known about how coarse anglers make decisions, inhibiting optimal policy implementation. The tools of natural resource economics may be useful in accomplishing these goals.

Natural resources economics aims to provide answers to questions about recreationalists decision-making process; including the value they place on a recreational activity, the trade-offs they are willing to make between site attributes and their behaviour under different policy scenarios; all of which provides useful insights that could be used to build the appropriate policy for the development of Irish coarse angling.

The impetus for the modern method of valuing a natural resource through economic principles can be traced back to Hotelling's letter to the United States National Park Services in which he outlined what would become known as the zonal travel cost model (Hotelling 1947). The travel cost model (TCM) (Parsons 2003) provides important information on a variety of demand characteristics, the most useful of which is an individual's demand in relation to the costs. As stated by Hotelling "the fact that they come [to the park] means that the service of the park is at least worth the cost" (Hotelling 1947, P. 1). This is often the trade-off that provides the most insight to economist and policymakers alike. This information gives a lower bound on the value of the service for an individual and can inform those interested how much additional value an individual receives above the costs incurred. This additional value, also known as consumer surplus, can have quite profound implications for policy development. It suggests that the investment in these amenities could provide economic value to users (and non-users) that is unobservable from market activity alone and indicates that estimates such as the earlier quoted €755 million do not accurately reflect the value of Irish recreational fisheries.

In order to manage sites in a manner that brings the most value to its users, policymakers may also be interested in how one site is chosen over another. To determine this, analysts employ site choice models (McFadden 1973). Site choice models aim to represent the decisions that an individual makes on a given choice occasion; how much of one attribute they are willing to trade off to get an extra unit of another attribute, what cost they are willing to endure to receive an extra unit of some site attribute and how much additional cost they are willing to incur to visit a more preferred site. The estimates from such a model give valuable insights into what the relative importance of site attributes are. This knowledge can lead to policies that aim to increase the most valued of these site attributes or policy that most increases economic value most.

The use of site choice models is dependent on a number of aspects of the data. A fundamental element of the data is its source. Traditionally, data can be collected from one of two sources, an objective source or a subjective source. Economic theory would suggest that individuals make decisions based on their own perception of a good (Puto 1987; Singh 1988; Poor et al., 2001; Artell, Ahtiainen and Pouta, 2013) and, as such, site choice models are most appropriate when applied to data where the respondent is able to rate the sites themselves. From a data collection perspective, this substantially increases many elements of the cost associated with the collection process (Baranzini, Schaerer, and Thalmann, 2010; Artell, Ahtiainen, and Pouta 2013). As it is usually much easier to use objective data, the literature has overwhelming used objective data rather than the theoretically preferred subjective data. However, the use of objective data over subjective data is only an issue if the objective data is a bad approximation of the subjective data, leading to biased results.

Of particular interest to managers of natural resources is a comparison of managerial staff's perspective to the that of the user. It can often be the case that policy, particularly at a local level, is built or implemented based on the managers' understanding of the sites of interest. Even when models, based on user data, accurately determining the relative importance of attributes, less than optimal policy implementation may occur when there is substantial deviation between how management and users view the same site. It is important then to understand if such differences exists.

Although revealed preference methods, like TCM (Parsons 2003) and choice models (Train 2009), are routinely employed in natural resource economics, they are limited by the experiences of the respondents. To estimate the values for the unobserved or unobservable attribute changes, analysts turn to stated preference techniques, where individuals are given a set of choices or scenarios and asked how they would respond in reaction to those events. These stated preference techniques allow for an exploration of the unseen, and, importantly, can provide indications of the appropriateness of proposed projects before implementation.

One such stated preference method is the contingent behaviour method first implemented by Adamowicz (1994) for non-market goods. Using this method individuals are asked about their current trip patterns to a specific site and then asked how they would change their trip frequency under a proposed change. This type of analysis has the benefit of combining both revealed trip frequencies (trips they took in the period of the survey) and stated trip frequencies under the proposed change. The use of revealed trip frequency gives a natural framing for the policy change scenario in which the payment vehicle, travel cost, is explicitly understood.

This thesis applies site choice, travel cost and contingent behaviour models to coarse anglers who fish in Irish waters. The application of these techniques addresses two general requirements; providing information on the decision-making processes of coarse anglers to make more informed policy, and to examine more generally the application of these methods to recreational demand data.

Three key components of demand are analysed within this thesis; how coarse anglers choose where to go fishing, they value the place on a day spent fishing and how they would change their angling participation given a variety of site attribute changes. This demand information allows management to build policy in a manner that will achieve their goals of providing the fishing public with the best experience possible and increasing the number of anglers fishing in Ireland. By understanding what site attributes are important to anglers, appropriate investment and cultivation of sites can be undertaken. Also, by forecasting how specific policies may impact angler participation, the effects of potential policies can be understood in a manner that is consistent with IFI goals. From a more technical perspective, this thesis explores how the trend of using an objective source of data may result in biased parameter estimates and how to address missing data in choice data. Both of these issues are pervasive for choice data and may lead to biased parameter estimates, incorrect policy assessment and misinformed measurement of the economic value of a good or service.

1.2 Research Objectives

There are two primary goals of this thesis. The first is to model the decision-making behaviour of coarse anglers who fish in Irish waters. The second is to examine the methods used to model decision-making behaviour. These goals are addressed in four papers, each looking at a separate topic that will contribute either to the knowledge of coarse angler behaviour or how decision-making behaviour is modelled more generally. These topics are:

1.The determinants of site choice for coarse anglers;

a) What effect do key site attributes have on an individual's probability of choosing that site?
b) What are anglers' willingness to pay for a change in these attributes?
c) How would anglers be affected by policy changes to the sites and does this vary from site to site?

2. Examining whether objective data is an appropriate substitute for subjective data;

a) Are the results of site choice models equivalent using objective data or subjective data?
b) Are welfare estimates equivalent?
c) Would the same policy decisions be made?

3. Comparing techniques for dealing with missing data in choice data;

a) Are the commonly used techniques creating bias in parameter and welfare estimates?
b) How do these missing data techniques compare using several metrics of bias?
c) Does bias increase with an increase in percent missing?
d) Which is the preferred technique?

4. Estimating the value an angler place on a day spent coarse angling;

a) How much is an angler willing to pay for a day spent coarse angler?
b) What level of consumer surplus is received by the average angler?

5. Estimating the effects of various policy scenarios on angler behaviour;

a) Would anglers take more trips if the quantity of fish increased?
b) Would they take more trips if there were more specimen fish?
c) What is the value of an increase in fish quality?
d) Which has more of an impact quantity or number of specimen fish?

1.3 Structure of the Thesis

The remainder of the thesis is structured as follows:

Chapter 2 presents the results of two site choice models, used to determine the attributes that affect the probability of site choice for Irish coarse anglers. These site choices models are applied to a data set of perceived site attributes, in which anglers were asked to rate six key site attributes, on a five-point Likert scale, for five prominent coarse angling sites in the Cavan and Leitrim area. The results of the conditional logit and random parameter logit site choice models are used to estimate the marginal willingness to pay for each site attribute and a number of policy scenarios.

Chapter 3 compares the results of a random parameter logit applied to a data set populated by coarse angler's perception of site attributes with a random parameter logit applied to a data set using managerial perception of the site attributes. The comparisons are used to determine if the parameter estimates are statistically equivalent and to examine if the welfare estimates are comparable. The results allow the researcher to understand if the same policy decisions would be made regardless of the source of the data.

Chapter 4 examines the use of four different techniques for dealing with missing data; complete case analysis, two types of mean imputation, and multiple imputations. These techniques are applied to a data set where true parameter estimates are known. The techniques are then compared using several metrics of bias over a range of percentages of missing data. Comparison is then extended to welfare estimates where, again, measures of bias are employed.

Chapter 5 uses travel cost and contingent behaviour data to estimate the value of a day spent coarse angling in Ireland. The contingent behaviour data examines how angler behaviour would change if the quantity of fish were to change and if the number of specimen fish increased.

Chapter 6 concludes this thesis. An overview of the main findings is presented along with the possible policy implications. The limitations of the thesis are discussed, ending with suggestions for future work.

1.4 Thesis Output

Journal Articles

1) Deely, J., Hynes, S., & Curtis, J. (2018). Coarse angler site choice model with perceived site attributes. *Journal of Outdoor Recreation and Tourism*.
2) DOI: 10.1016/j.jort.2018.07.001
3) Deely, J., Hynes, S., & Curtis, J. (2019). Are objective data an appropriate replacement for subjective data in site choice analysis? *Journal of Environmental Economics and Policy*, *8*(2),159-178.
   DOI: 10.1080/21606544.2018.1528895
4) Deely, J., Hynes, S., & Curtis, J. (2019). Combining actual and contingent behaviour data to estimate the value of coarse fishing in Ireland. *Fisheries Research*, *215*, 53-61. DOI: 10.1016/j.fishres.2019.03.008

Journal Articles Under Review

Deely, J., Hynes, S., & Curtis, J. (2019). Comparing alternative approaches to dealing with missing data in revealed preference site choice models. Journal of Choice Modelling

Presentation

Coarse angler site choice model with perceived site attributes. Inland Fisheries Ireland Annual Economic Conference. Inland Fisheries Ireland, Dublin, Ireland (August, 2017)

Are objective data an appropriate replacement for subjective data in site choice analysis? European Regional Science Association. University College Cork, Cork, Ireland (August, 2018)

Are objective data an appropriate replacement for subjective data in site choice analysis? Inland Fisheries Ireland, Dublin, Ireland (September, 2018)

Modelling the decision-making behaviour of coarse anglers. The Whitaker Institute Research Day, National University of Ireland Galway, Galway, Ireland (April, 2019

Chapter 2

2. Coarse Angler Site Choice Model with Perceived Site Attributes

2.1 Introduction

The value of a day spent fishing is impacted by a number of factors including; site management, choice of site, and duration of stay. Anglers' preference for these factors have been demonstrated to be influenced by a wide range of personal, demographic, and site characteristics including; personal beliefs and attitudes (Arlinghaus and Mehner 2005), race and ethnicity (Hunt et al. 2007a), water quality (Curtis and Stanley 2016), objective site attributes (Curtis and Breen 2017), traditions and weather conditions (Hunt et al. 2007b) and duration since last trip (Provencher et al. 2002). In some cases, preferences for site attributes have been found to be influenced by the species the angler is targeting (Curtis and Breen 2017). This paper focuses on Irish coarse anglers, their preferences and what those preferences suggest in relation to managing Irish coarse fishing sites.

The economic contribution of coarse angling to the Irish economy has been measured by Tourism Development Ireland at €96 million (Inland fisheries Ireland 2015). However, direct expenditure measures, such as this, provide limited information on the value of a day spent fishing to an angler. A number of recent studies have estimated the value of a day spent fishing in Ireland; Hynes et al. (2015) suggested that per trip consumer surpluses for a day spent fishing ranges from €49 to €277, and Curtis and Stanley (2016) estimated a per trip value, at the upper end of Hynes et al.'s (2015) range, of €264. However, these estimates use a single demand function for multiple types of anglers, and as such, it may be inappropriate to attribute these euro values to a single group such as coarse anglers. Curtis and Breen (2017) suggest that a separate demand function for

each type of angler is more appropriate as different angler types have distinct preferences. For example, Curtis and Breen (2017) estimated that the per trip consumer surplus for game fishing in Ireland was more than three times the value of coarse fishing; €787 and €249 respectively.

In addition to welfare estimates, site choice analysis can provide a better understanding as to why certain sites are preferred over others. This, in turn, can facilitate better management and lead to a better-quality site for consumers. Site choice models enable the modeller to create an approximation of the decision-making process undertaken by, in this case, an angler. When an angler is presented with a choice set containing a number of alternative sites, the choice is often informed by the bundle of attributes that each site possesses. These attributes will influence angler's choice in accordance with the angler's preferences. The analysis of angler's choice can reveal the trade-offs that are implicitly made in a given choice occasion. Policy can then be developed based on the analysis of revealed angler preferences.

Site choice models have been widely applied to recreational angling demand (see Hunt (2005) and Johnston et al. (2006) for literature reviews). In a site choice model, random utility style models are often used to determine the probability of site choice and by extension, determine the trade-offs anglers are willing to make to satisfy their preferences. The attributes of a site, used in the model specifications, can be quantified in one of two manners; a perceived rating, like the one used in this paper, or an objective measure, using a scientific measure of the site attributes. An extensive literature has grown debating the merits of using perceived measures over the often more convenient objective measures (Adamowicz et al. 1997; Jeon et al. 2005; Farr et al. 2016). However, the literature has not clearly determined which measures are superior in site choice models. Adamowicz et al. (1997) argued that models solely based on perceived measures slightly outperformed models solely based on objective measures. Elsewhere, Jeon et al. (2005) found that the inclusion of a perceived measure had a significant impact on site choice analysis of recreational anglers.

There have been some attempts, in the recreational angling literature, to extend models based on objective measures of site attributes with the addition of a limited set of perceived measures (Jeon et al. 2005; Artell et al. 2013). In all cases, the use of perceived measures has been limited to one or two attributes. No paper, to the best of our knowledge, within the recreational angling literature, has attempted to use a complete array of subjective site attributes as rated by anglers as the sole method of measuring site quality. In essence, the angling literature has yet to explore how angler's own perception of a site affects site choice. Through the collection of perceived site attributes and revealed preferences data, this paper explores the preferences of Irish recreational anglers using a random parameter logit model. Additionally, this paper adds to the literature by exploring the effects of mean imputation on parameter estimation, which has not previously been dealt with in the recreational site choice literature.

2.2 Materials and Methods

2.2.1 Survey Design[1]

The initial steps of the survey design were to identify a choice set and the attributes that are thought to impact the selection of a coarse angling site. An iterative approach was taken to accomplish this. The earliest draft of the survey was developed through an examination of the relevant literature and discussions with coarse angling experts at Inland Fisheries Ireland (IFI). The literature search incorporated both academic literature and recent National Strategy for Angling Development (NSAD) studies. Once a region was selected, where the survey would be conducted, a focus group of local anglers was assembled. The focus group helped in the selection of specific sites and site attributes, as well as giving valuable insight into their understanding of each element of the survey and how they would answer the survey if they were sampled. Finally, prior to the deployment of the finished survey, a pilot study ran from 28th of July to the 5th of August 2016.

The choice set comprises five angling sites that are thought to be feasible alternatives for coarse anglers, while still incorporating enough variability

[1] The complete survey can be viewed in appendix 2.1

amongst the six site attributes. The six site attributes were defined to the respondents as;

1) Accessibility to the site (this includes parking and ability to reach the location that you fished at),
2) Size of fish at the site (on average does this site provide access to good sized fish),
3) Quantity of fish (on average does this site provide access to a large quantity of fish),
4) Encounters with other anglers,
5) Variety of fish species (are there a large variety of species of fish at this site)
6) Local services (these include pub, shops, accommodation etc…).

The respondents were asked to rate these attributes on a five-point Likert scale for each of the five sites they had visited. To ensure clarity, the lowest and highest points on the Likert scale were defined for the respondent. For example, for the attribute variety of fish species the first point on the Likert scale was marked ‘little to no variety', the highest point was marked ‘lots of variety'.

The sites of interest are located in the Cavan and Leitrim area of Ireland. Both counties are located in the north of the Republic of Ireland and border with Northern Ireland. This geographic location was selected because of its abundance of coarse fishing sites, making it a popular destination for anglers both north and south of the border as well as being home to many local coarse anglers.

The choice set is limited to five sites that could a) feasibly be alternatives b) have a reasonable number of visitors and c) have ostensibly different attribute levels. The sites were purposely selected to maximise the number of sites, from the choice set, each angler may have visited. The five sites selected are Garadice (Leitrim), Killykeen Forest Park (Cavan), Eonish (Cavan), Dernaferst (Cavan) and Church Lake (Cavan). The five sites are situated within 30 kilometres of each other. Both Eonish and Killykeen are fishing sites on the water system Lough Oughter, Church Lake and Dernaferst are on Lough Gowna and Garadice is a lake with multiple access points.

All five sites have some level of road access but vary in the number of access points, accessibility for fishing boats, the proximity of parking to pegs[2], and the number of pegs available. In order to induce variability amongst fish species, the sites were chosen from three different water systems. However, in Ireland, most sites that hold coarse fish will hold similar species. The potential for variability of encounters with other anglers was determined by the perceived popularity of the site and the number of pegs it contains. The potential for variability in size of fish and quantity of fish between sites was informed by expert opinion and focus groups.

Table 2.1 Site Attribute Overview Table

Site	Accessibility	Size	Quantity	Encounters	Variety	Services
Garadice						
Killykeen						
Eonish						
Dernaferst						
Church Lake						

2.2.2 Sampling

Two methods of data collection were used to elicit responses from the coarse angling community; online and intercept surveying. The online survey was accessible via SurveyMonkey from the 6th of August 2016 to January 15th, 2017. Potential participants were contacted through Irish coarse angling Facebook pages, by emailing local coarse angling clubs, contacting local newspapers and through the IFI newsletter. The online survey was completed by 62 individuals, making a total of 4,265 observations.

Intercept surveying began on the 6th of August and ran until the 7th of November. Individuals were approached and invited to complete the survey. Although this has the potential to increase 'length of stay' bias (Lucas, 1963), the alternative of interviewing people at the car park or site entrance, was not

[2] A peg is a cleared designated area an angler can fish from.

feasible as car parking facilities were seldom used as anglers chose to park beside where they fished. Eonish and Church lake were surveyed 15 times each, Killykeen Forest Park and Dernaferst were surveyed 13 times and Garadice was surveyed 12 times. The intercept survey accounts for 43 survey responses and 6,685 observations. In total, the sample is comprised of 105 survey respondents and 10,950 observations.

The respondents were asked a series of questions about their angling experience as well as socio-demographic questions. The respondents were also asked to rate each of the five sites on a set of six attributes. They were also asked to provide information on how many trips they made to each site during the 12-month period prior to completing the survey. The survey design is developed on the assumption that the choice set describes five sites that are possible alternatives for every angler in the sample. Any site that was not previously visited by a respondent was not rated by that respondent. Following Hynes et al. (2008) and Hanley et al. (2001), sites that were not rated by an angler had their attribute levels set equal to the mean of the responses given by all other anglers for those attributes. The impact of replacing the missing values through this mean imputation process is further explored in section 4.4 sensitivity analysis.

The travel cost variable was constructed by calculating the distance from the respondent's home address to each site and is specified as:

$$tc = ((travel\ disance * operating\ cost) + \\ (travel\ time * \ opportunity\ cost\ of\ time)) \ * 2 \qquad (2.1)$$

Here, operating cost equals 0.2475 € per kilometre, which is the operating cost of running a medium sized car according to the Automobile Association of Ireland. This assumes that each trip is a day trip, where the individual travels from their home and back. The opportunity cost of time is taken as 33% of the average hourly wage rate (Parsons, 2003) assuming a 2000 hours' work year. No opportunity cost of time is included to account for time spent on-site.

2.2.3 Considerations for Sampling Bias

Consideration needs to be given to the ability to combine the on-site and online survey responses; in particular, it may be the case that younger anglers are more likely to complete the online survey. The average age of the online cohort is 44 whereas the average age of the on-site cohort is 56. However, the sample as a whole seems to align almost perfectly with the estimated age range of Irish anglers (TDI, 2013). The current sample is comprised of 15% 18-34-year-olds, 54% 35 – 54, and 31% 55 +. The TDI (2013) estimates suggest that 18% of the fishing population is aged between 18 – 34, 51% between 35 – 54, and 30% are older than 55 years of age.

Kolmogorov–Smirnov (KS) tests were also used to assess whether there are substantial differences between the two samples with regard to their perspective on site attributes. The KS test is a non-parametric test for equality of distribution which was first proposed by Kolmogorov (1933) and thought to be more powerful than chi-square test in most situations (Conover 1980). In Kolmogorov style tests the largest vertical distance between two samples is measured and used to test the hypothesis that; all values from, in this case, the online sample come from the same distribution as the on-site sample versus the alternative that at least one value from the online sample falls outside this distribution. Presented in table 2.2 of the appendix, the results suggest that for 24 out of 30 attributes the responses can be considered to be from the same distribution. Additionally, to control for the effects of the sampling methods used, several interaction terms are used to account for differences between the on-site cohort and online cohort.

A further consideration is how the sampling technique may cause biased parameter estimates. Two forms of bias may be present in the sample through the non-random sampling methods employed. The first relates to a sample selection bias. When individuals are intercepted at a particular site, the probability of inclusion in the sample is correlated with site choice, leading to biased parameter estimates (Hindsley et al. 2011). The elimination of this form of bias can be achieved by weighting sampled observations to reflect known population ratios of site choice (Manski and Lerman 1977). Alternatively, the bias in parameter estimates can be confined to a single set of parameters. Manski and Lerman (1977) have demonstrated that by including a full set of alternative specific constants (ASC) the bias can be fully restricted to these dummy variables. In the case of the models presented here, ASCs are included as the requisite population ratios are unknown.

The second form of bias is known as endogenous stratification or avidity bias, where the probability of being sampled is positively correlated with the number of trips the respondent has taken within some time frame (Hindsley et al. 2011). As a result of avidity bias, parameter estimates are more heavily influenced by avid anglers than would otherwise be true. As is common in many recreational site choice models we do not have accurate information on the total number of Irish anglers using the five sites or their associated trip frequencies to the five sites. This means that we cannot generate individual weights that could be used to reduce the influence of the avid anglers in the sample. Similar to other recreational site choice modelling studies such as Hanley et al. (2001) and Scarpa et al. (2005) we, therefore, are unable to correct this issue. However, as pointed out by Hynes and Hanley (2006), the addition of the online respondents to the sample should reduce the number of avid respondents by virtue of the fact that their response is not as a result of being intercepted on-site.

Researcher defined choice set may also lead to biased parameter estimates when respondent choice set and researcher choice set differ (Peters et al. 1995; Parson et al. 1999; Hick and Strand 2000; Li et al. 2015). The literature discusses multiple consequences of inappropriate choice set assumptions[3]. Of particular note for the analysis presented here is how the missing data for unfamiliar sites is treated. The effects of the mean imputation process used to replace this missing data are explored in section 4.4 sensitivity analysis.

Finally, the variable *encounters,* which measures how often a respondent meets or sees other anglers at a particular site, is likely correlated with other angler's site choice leading to endogeneity issues. Reverse causality may

[3] The interested reader may wish to read Li et al. (2015) and Haab and Hicks (1999) for further insight into the multitude of assumptions and considerations that can be employed when trying to recreate an individual's true choice set or consideration set.

indeed be an issue as *encounters* is likely a function of site choice, although not the sole determinate. However, the discussion with both the coarse angling experts and focus group, as well the angling literature on angler site choice (Hunt 2005) suggests that the level of likely encounters with other anglers is an important determinant of site choice.

2.3 Econometric Models

2.3.1 Site Choice Model

McFadden's (1973) random utility model (RUM) asserts that an individual will choose the alternative that will maximize her utility on any given choice occasion. This utility can be written as:

$$u_{in} = V(X_{in}, y_n - p_{in} \mid \theta_n, z_n) + \varepsilon_{in}$$

$$= V_{in} + \varepsilon_{in} \quad (2.2)$$

Where u_{in} is the utility received by individual *n* from choosing site *i*, V is the indirect utility function, X_{in} is a vector of perceived attributes, y_n is individual *n*'s income, p_{in} is the travel cost, θ_n is a vector of individual specific characteristics and z_n are individual specific covariates. ε_{in} is the stochastic error term, by definition, unknown to the modeller and is assumed to be independent and identically distributed (IID) extreme value type 1. The estimated variable parameters are homogenous across individuals and, by implication, each individual has the same taste preferences (Train 1998). The probability of individual *n* choosing site *i* from choice set *J* can then be given as:

$$\Pr(i) = \Pr(V\,(x_{in}, p_{in}\,|\theta_n, z_n) + \varepsilon_{ij}) \geq V\big(x_{jn}, p_{jn}\,|\theta_n, z_n\big) + \varepsilon_{jn}\ \forall\, j \in J \quad (2.3)$$

The probability of individual *n* choosing site *i* is equivalent to the probability that individual *n* will receive more utility from visiting site *i* than any other sites in choice set *J*. As only difference in utility matters when calculating the probability of site choice, individual characteristics like income are differenced away. When the distribution of the error terms are independently and identically drawn from an extreme value distribution, the RUM model takes the form of a conditional logit (CL) (McFadden 1973), where the probability of choosing site i is given as a logit with scale parameters μ which is assumed to be equal to 1 (Boxall and Adamowicz 2002).

$$\Pr(\mathrm{i}) = \frac{\exp(\mu V_{in})}{\sum_{j=1}^{J} \exp(\mu V_{jn})} \tag{2.4}$$

The associated likelihood function is (Train 2009)

$$LL(\beta) = \sum_{n=1}^{N} \sum y_{in} \ln Pr_{in} \tag{2.5}$$

When maximised, the derivative of this function with respect to each parameter is equal to zero.The maximum likelihood estimates are therefore the values of β that statisfy the condition of equalling zero (Train 2009).

2.3.2 Random Parameter Logit

The Random Parameter Logit (RPL) as outlined in Train (2009) overcomes the restrictive quality of the IID error term found in the CL by decomposing the error term into two separate elements. One part is correlated over alternatives and heteroskedastic, the other is IID over alternatives and individuals. In this form utility can be written as:

$$u_{in} = V_{in} + [\eta_{in} + \varepsilon_{in}] \tag{2.6}$$

And

$$V_{in} = \beta_{in} * X$$

Where η_{in} is a random term with zero mean, which may be correlated across individuals and alternatives, ε_{in} remains IID, β_{in} is the coefficient for a variable and X is the corresponding value of that variable. By decomposing the error term, the RPL allows the coefficients of observed variables to vary randomly across individuals. The choice probability remains logit conditional on individual taste. Marginal probabilities across individuals need to be integrated over taste distributions which are specified by the modeller. η_{in} can take on multiple distribution forms (Hensher and Greene 2003). If it is assumed that it takes a multivariate normal form, we can write:

$$\beta_n \sim N(\bar{\beta}, \Omega) \tag{2.7}$$

Where $\boldsymbol{\Omega}$ is the variance-covariance matrix.

The conditional probability for any η_{in} is logit:

$$\Pr(\mathrm{i}) = \frac{\exp(\mu V_{in} + \eta_{in})}{\sum_{j=1}^{j} \exp(\mu V_{jn} + \eta_{jn})} \tag{2.8}$$

This logit is then integrated over all values of η_{in} with appropriate density weightings to form the unconditional choice probability. After accommodating for an unbalanced panel data, the unconditional choice probability becomes:

$$\int \prod_{t=1}^{t=T(n)} \frac{\exp(\mu V_{in} + \eta_{int})}{\sum_{j=1}^{j} \exp(\mu V_{jn} + \eta_{jnt})} \, \varphi(\bar{\beta},)\mathrm{d}\beta_n \tag{2.9}$$

Where T(n) is the revealed preference of each respondent, $\varphi(.)$ denotes the multivariate normal density, $\bar{\beta}$ and $\boldsymbol{\Omega}$ are the mean and variance parameters which are estimated from the sample data.

The simulated log likelihood function for the RPL is (Train 2009):

$$SLL = \sum_{n=1}^{n} \sum_{j=1}^{j} d_{jn} \ln \tilde{p}_{jn} \tag{2.10}$$

Where $\tilde{p}_{jn}$ is the unbiased estimator of the probability of individual *n* selecting site *j* and d_{jn} equals one if *n* chooses *j*. Exact maximum likelihood is not possible for the RPL (Train 2009). The maximum simulated likelihood estimator is the value of β which maximises the SLL, which is achieved through a designated number of draws (500 in this case) over the parameter's distribution.

2.3.3 Welfare Estimates

Willingness to pay (WTP) estimates are calculated following Train (2009):

$$WTP = \frac{\beta_n}{-\beta_{tc}} \tag{2.11}$$

Where β_n represents the coefficient of the attribute of interest for individual n and β_{tc} is the travel cost coefficient. For the RPL, the coefficient of the attribute of interest is estimated through simulation. As WTP is simply a ratio, with the travel cost parameter as the denominator, it is highly influenced by an individual's marginal utility of income. In the models presented in this paper, the travel cost coefficient is fixed to avoid infinite moments of the welfare estimates (Daly et al. 2012). Consequently, all individuals are assumed to have the same marginal utility of income. Fixing the parameter in this way may affect estimation. For example, bias in the parameter estimates may occur due to cofounding with unexplained heterogeneity in the cost coefficient (Daly et al. 2012). However, alternatives, such as normal, truncated normal, uniform and triangular were rejected as it has been demonstrated that these methods can produce infinite moments of welfare estimates (Daly et al. 2012). An alternative is to use a log transformation, which restricts the travel cost parameter to be distributed over a negative range of values. However, this too may be problematic. Changyong et al (2014) report that the results of the transformed data may not be relevant to the original data. Consequently, we used a fixed parameter, acknowledging that further investigation on the topic is warranted but beyond the scope of this paper.

Confidence intervals are constructed using the Krinsky and Robb method (Krinsky and Robb, 1986), which takes a specified number of draws from a multivariate normal distribution. The mean and covariance of this distribution are specified to equal the estimated coefficients and covariance matrix of the site choice model (Hole 2007).

Haab and McConnell (2002) extend WTP estimates to measure the compensating variation (CV) which is the amount one would be willing to pay to achieve a certain attribute level at one or more sites:

$$CV_{sitechange} = -(\beta_{tc})^{-1}$$

$$\left[\ln\left[\sum_{j=1}^{j} exp\,(\hat{\beta}_n x_n^1)\right] - \ln\left[\sum_{j=1}^{j} exp\,(\hat{\beta}_n x_n^0)\right]\right] \qquad (2.12)$$

where β_{tc} is the marginal utility of income which, here, is the negative reciprocal of the travel cost coefficient, $\hat{\beta}_n$ is a vector of parameters for individual n, x_n^0 is a set of perceived site attributes (or travel cost) and x_n^1 is the same set of attributes after some exogenously imposed change to one or more of the site attributes. When applied to the RPL, WTP estimates for a specific change in a site's attribute follows the same specification but need to be integrated over taste distributions (Train 1998) which is approximated through simulation:

$$\widehat{CV_{sitechange}} = \int WTP_n \varphi(\hat{\beta}, \hat{\Omega})\ d\beta =$$

$$\int\{-(\beta_{tc})^{-1}\left[\ln\left[\ \sum_{j=1}^{j} exp\,(\hat{\beta}_n x_n^1\right)\right] - \ln[\sum_{j=1}^{j} exp\,(\hat{\beta}_n x_n^0)]]\ \varphi(\hat{\beta}, \hat{\Omega})\ d\beta \quad (2.13)$$

2.4 Results

2.4.1 Sample Statistics

Table 2.2 shows the perceived attribute levels as rated by the survey respondents. The within attribute variance is reasonable given that the attributes are rated on a five-point Likert scale with most attribute means lying between two and four on the Likert scale. As the sites were chosen, in part, by the fact that attendance should be reasonably large, it would be unlikely to see many attributes that were rated very low on the Likert scale.

Table 2.2 Mean Perceived Site Attribute Rating

Site	Accessibility	Size	Quantity	Encounters	Variety	Services
Garadice	4.30	3.32	3.15	3.52	3.47	3.22
	(.79)	(.69)	(.74)	(1.05)	(.85)	(1.11)
Killykeen	3.29	2.99	3.23	3.59	3.47	2.97
	(1.13)	(.69)	(.83)	(1.05)	(.85)	(1.03)
Eonish	3.54	3.03	3.17	3.05	3.44	3.08
	(0.93)	(.53)	(.64)	(.92)	(.70)	(.84)
Dernaferst	3.46	3.05	3.29	3.46	3.35	3.50
	(.94)	(.72)	(.73)	(.79)	(.67)	(.80)
Church	3.04	2.93	2.92	2.97	3.98	3.40
Lake	(.80)	(.68)	(.74)	(.81)	(.60)	(.69)

Ratings are rated on a 1- 5 Likert scale. Standard deviation given in parenthesis

As shown in table 2.3, Garadice was visited by the greatest number of anglers as well as having the highest number of mean trips. Garadice's much higher number of total trips is due, in part, to anglers who took more than 50 trips to the site during the survey year, with two reporting to have visited Garadice

100 times. Approximately 40% of the respondents who visited Church Lake only visited it once within the past 12 months. In comparison, less than 20% of respondents who visited Garadice or Killykeen only visited once.

Table 1.3: Mean and Total Trips Per Site

	Number of anglers who have visited each site in the last 12 months	Mean trips	Total Trips
Garadice	71 (67.61%)	15.39	1,093
Killykeen Forest Park	70 (66.67%)	7.06	494
Eonish	45 (42.86%)	4.36	196
Dernaferst	43 (40.95%)	6.02	259
Church Lake	33 (31.4%)	4.48	148

Percentage of sample who visited each site is given in parenthesis

As shown in table 2.4, the majority of respondents stated that angling was their most important pastime. Nearly 95% of the anglers considered their abilities to be intermediate or advanced. The average number of years the sampled anglers have been fishing for was 37 with 80% fishing for more than 20 years.

Table 2.4: Angling Related Experience of Respondents

Items	Frequency	Percentage
Importance of angling as recreation:		
Most important outdoor activity	85	80.95%
Second most important outdoor activity	13	12.38%
Third most important outdoor activity	4	3.81%
One of many outdoor activities	3	2.86%
Ability level:		
Basic	6	5.71%
Intermediate	43	40.95%
Advanced	56	53.33%
Years fishing:		
10 years or less	7	6.67%
11 – 20 years	14	13.33%
21 – 30 years	24	22.86%
31 – 40 years	33	31.43%
41 – 50 years	19	18.10%
51 – 60 years	5	4.76%
61 + years	6	1.20%

As shown in table 2.5, the vast majority of respondents were from the Republic of Ireland with only 13% residing in Northern Ireland at the time of completing the survey. The average sampled angler was 49 years old, and just over half the sample has completed third level education.

Table 2.5: Sociodemographic Characteristics of Respondents

	Mean	Standard deviation
Age	48.6 years	13.39 years
Income	€43,281	€30,258
Education:		
Third level education	50.48%	
Secondary	43.81%	
Primary	5.71%	
Country of residence:		
Ireland	86.67%	
Northern Ireland	13.33%	

2.4.2 Estimation Results

Both conditional logit and random parameter logit models were estimated. The Akaike information criterion statistics suggest that the RPL specification is the preferred model, compared to the CL. On a more fundamental level, the advantages that the RPL provides, by allowing correlations in the decision-making process for individuals across choice occasions, is a much more logical interpretation of how individuals act in a real-life situation. By allowing for unobserved taste heterogeneity, in the manner that the RPL does, a closer approximation of the decision-making process of individual anglers and the sample as a whole is provided. Additionally, the CL model failed the test for independence of irrelevant alternatives (IIA); a problem that is overcome by using the RPL. The discussion on estimated results and welfare estimates will, therefore, focus on the RPL, and will only refer to the CL where explicitly stated.

Table 2.6 shows the results of the econometric analysis. The estimated parameters are not obviously interpretable; however, direction, magnitude, and significance are easily understood. A positive coefficient means a ceteris paribus increase in this variable increases the probability of site selection; the greater the absolute value of the coefficient the larger the absolute increase in this probability. The alternative specific constants are dummy coded with Garadice, the most visited, as the reference site. As discussed in section 2.2.3, these parameters may be biased, and care should be taken when interpreting them.

The first set of parameters, presented in table 2.6, are the variables which are allowed to vary randomly in this application of the RPL. All site attribute coefficients are assumed to be normally distributed, so that negative, as well as positive values, are permitted[4]. The second set of parameters contains the

[4] One could argue that it would make more sense to specify the quantity and size coefficients to be log-normally distributed to ensure only positive estimates. Indeed, alternative specifications of the model were attempted, which included specifying these coefficients as log-normal. However, these models failed to converge. This is not an uncommon result with RPL models; as pointed out by Hynes et al. (2008) non-convergence may result in cases

travel cost parameter and alternative specific constants, which aim to capture the attractiveness of a site that remains unaccounted for by the variables of interest. These were selected to be fixed parameters in order to focus on the site attribute parameters. However, alternative models were run with the ASCs allowed to vary over a normal distribution which showed little variation in welfare estimates between those models and the ones presented here. To capture heterogeneity between the on-site and online groups, four interaction terms are created. Each of the four variables is created by interacting a dummy indicating that the individual completed the survey online interacted with a dummy for each of the alternative specific constants. The aim of these interaction terms is to capture any difference that may exist between the two groups with respect to their site preference.

where restrictions in the choice of distributions are employed when using maximum simulated likelihood because of its reliance on gradient methods to find the maximum.

Table 2.6: Results of Conditional Logit & Random Parameter Logit

	Conditional Logit	**Random Parameters Logit**	
	Mean of Coefficient	Mean of Coefficient	Standard Deviation of Coefficient
Random Parameters			
Access at Site	0.092(0.032)***	0.354 (0.081)***	0.714(0.067)***
Local Services	-0.300 (0.034)***	-0.335 (0.109)***	0.752(0.106)***
Size of Fish	0.053 (0.051)	0.225 (0.113)**	1.589(0.128)***
Quantity of Fish	0.113 (0.042)**	0.024 (0.104)	1.472(0.175)***
Variety of Fish	0.182 (0.053)***	0.331 (0.111)***	1.366(0.143)***
Encounters with other Anglers	-0.008 (0.038)	0.014 (0.075)	1.060(0.138)***
Fixed parameters			
Travel Cost	-0.065 (0.004)***	-0.080(0.009)***	
Killykeen Forest Park	-1.108 (0.097)***	-0.761 (0.251)***	
Eonish	-2.114 (0.118)***	-1.400 (0.224)***	
Dernaferst	-1.189(0.120)***	-0.508 (0.294)*	
Church Lake	-2.336 (0.208)***	-1.267 (0.355)***	
Heterogeneity in mean, parameter:			
Killykeen Forest Park: online	-0.879 (0.135)***	0.087 (0.337)	
Eonish: online	-1.479 (0.170)***	0.561 (0.299)*	
Dernaferst: online	-0.725(0.164)***	0.490 (0.381)	
Church Lake: online	-1.728(0.235)***	0.357 (0.394)	
Log likelihood function	-2551	-2066	
Akaike information criterion	5132	4174	
Bayesian information criterion	5241	4327	
Observations	10950	10950	

*Figures in parenthesis are standard errors. All figures under conditional logit are fixed parameters. *** indicates significant at 1% ** indicates significant at 5% * indicates significant at 10%*

As expected, the *travel cost* coefficient is negative and significant, indicating that an increase in cost will result in a decrease in the probability of site selection. The coefficients of the alternative specific constants are all negative, indicating that, Garadice has some features that draw the sampled anglers to it, as opposed to the other four sites in the choice set. Only one of the four interaction terms are significant, suggesting that the online respondents are more likely to choose Eonish in comparison to their on-site counterparts.

For the sampled anglers an increase in *access* is associated with a higher probability of site selection. *Services*[5], which include; accommodation, pubs, and shops, has a negative and significant impact on site choice. This may indicate that the sampled anglers generally choose sites that are more remote and require few local amenities on their fishing trips. *Variety* plays a positive and significant role in site selection for the sample. The estimated parameter for *encounters* is insignificant suggesting that for the average sampled angler site choice is not correlated with the level of *encounters with other anglers*. The *quantity of the fish* at the site was not a significant driver of site choice amongst the sample. However, the *size of the fish* at a site seems to play an important role in site choice for the sampled anglers as they tend to choose sites with large fish.

Broadly speaking the results of the CL and RPL are similar. Nevertheless, some noteworthy differences do appear. The results of the CL suggest that *size of fish* has an insignificant effect on site choice whereas once preference heterogeneity is controlled for the RPL results suggests that *size of fish* has a mean positive impact. Conversely, the *quantity of fish* plays a significant role in the CL model but does not in the RPL.

[5] The initial sample set used for this analysis included overseas anglers. They were not used in the final estimation due to the absence of travel cost information. However, the results of an RPL model that included the foreign visitors suggested that for Irish anglers, services had a negative and significant impact on site choice but for overseas anglers, local services had a positive and significant impact.

2.4.3 Welfare Estimates

WTP estimates for a marginal change in a site attribute are presented in table 2.7. The confidence intervals for these estimates were computed using the Krinsky-Robb method (Krinsky and Robb, 1986) with 5,000 draws. To do this the Stata post estimation command mixlbeta was used. This command exports the individual specific coefficients for each respondent using the original set of results presented in table 2.6. Equation 2.12 and the individual specific coefficients, were used to estimate state 0 pf this equation. State 1 was estimated using the same coefficients but the level of the attribute of interest was changed as described in method section.

Table 2.7: Willingness To Pay Estimates (€ per person/trip)

Attribute	Conditional Logit	Random Parameter Logit
Access at site	1.42 (0.43 – 2.41)	4.44 (2.49 – 6.79)
Size of fish	0.82 (-0.75 – 2.38)	2.79 (0.10 – 5.77)
Local Services	-4.62 (-5.72 – -3.51)	-4.20 (-6.95 – -1.57)
Quantity of fish	1.74 (0.45 – 3.03)	0.30 (-2.24 – 2.98)
Encounters with Other Anglers	-0.13 (-1.29 – 1.02)	-0.18 (-1.67 – 2.06)
Variety	2.80 (1.13 – 4.6)	4.15 (1.40 – 7.41)

Estimates indicate a euro value WTP per trip for a marginal increase in the perceived value of an attribute. 95% confidence intervals shown in parenthesis.

The sampled anglers have a WTP of €4.44 per trip for a marginal increase in the perceived level of *access*. Estimates suggest that these anglers, on average, would be willing to pay €2.79 per trip for a marginal increase in the perceived *size of the fish* at a site. The estimates suggest that the sampled anglers are willing to pay €4.20 per trip for a marginal decrease in *local services*. After accounting for the observed heterogeneity, the average sampled angler, has a WTP of €4.15 for a marginal increase in perceived *variety*. WTP estimates for both encounters with other anglers and quantity of fish were not statistically different from zero.

As shown in table 2.8, WTP estimates are extended to assess how a variety of changes to a site's attributes would affect the sampled anglers. The first estimate presented investigates how an increase in *access* at each of the five sites would impact the sample on a per choice occasion basis. Then, WTP estimates are presented for an increase in size at Garadice and Killykeen, the two most popular sites, as well as averaged across all five sites. In both cases the exogenously imposed change is a one unit increase on the five-point Likert scale e.g. for those who said that accessibility was three at Garadice, their WTP is calculated as the difference between accessibility being three and four for Garadice. Importantly, attribute ratings are restricted to five on the Likert scale. This will have a significant impact on the WTP estimates for sites that are already highly rated for that attribute. Anglers who rated a site as being five out of five on accessibility will, in essence, be excluded from the WTP calculation (i.e. there is no difference between the status quo and a change in policy for those anglers).

Table 2.8: Compensating Variation for a Change in a Site's Attributes

Site	Attribute	Per person/choice occasion, €	95% confidence intervals
Garadice	Unit increase in access	2.26	2.05 – 2.47
Killykeen	Unit increase in access	3.53	3.29 – 3.78
Eonish	Unit increase in access	3.27	3.05 – 3.49
Dernaferst	Unit increase in access	2.71	2.53 – 2.89
Church Lake	Unit increase in access	3.39	3.18 – 3.61
Garadice	A unit increase in the size of fish	2.39	1.80 – 2.98
Killykeen	A unit increase in the size of fish	1.80	1.21 – 2.39
Average across all five sites	A unit increase in the size of fish	2.17	1.59 – 2.76

The results suggest that the sampled anglers may benefit most from an increase in *access* at Killykeen and least from an increase in *access* at Garadice. The relatively low CV for an increase in *access* at Garadice is, in part, due to the large number of anglers who rated Garadice as having *access* worthy of a five out of five rating.

A TDI (TDI 2013) report suggests that fish quality (both *size* and *quantity*) is the most appealing aspect of Ireland as an angling destination. Consequently, analysis has been extended to demonstrate how a change in *the size of fish* may affect coarse anglers. Simulations were conducted demonstrating how a change in *the size of fish* at Garadice, Killykeen and averaged across all five sites would impact the sampled anglers. This change has been specified to be a one unit increase in the perceived *size of fish* as measured on the five-point Likert scale.

The results of this simulation suggest that the per choice occasion increase in welfare, for a 1-unit Likert scale increase in the *size of fish* is €2.39 at Garadice and €1.80 at Killykeen. Averaged across all five sites the CV is €2.17 per choice occasion. Additionally, it may be true that an increase in the *size of fish* may induce anglers to take more fishing trips during the year, which would have a much greater impact on consumer welfare.

2.4.4 Sensitivity Analysis

Biased parameter estimates are often an issue of concern in economic modelling. Parameter estimates can be biased through numerous mechanisms, some of which have been previously discussed in section 2.2.2. Of particular concern with the methods used here and elsewhere (Hynes et al. 2008 and Hanley et al. 2001) is the practise of replacing unrated site attributes with the mean rating given by all other anglers. Following Hick and Strand (2000), Peters et al. (1995), and Adamowicz et al. (1997) we create two data sets; the first is the full data set with mean imputed missing values, the second is a restricted data set using only familiar sites, in this case, the sites that were rated by the respondents.

As suggested by Parson et al. (1999) the results of the restricted choice set may undervalue the disutility of an individual's travel cost. For example, if an angler is sufficiently satisfied with Garadice he/she may not be willing to travel 45 more minutes to Church Lake, even though they are aware of the site. In this example, the angler has a preference for less travel cost and as a result, Church Lake is not visited by the angler and is, consequently, unrated.

The restricted choice set will not capture this disutility as Church Lake is simply dropped from the individual's choice set. This may result in a travel cost coefficient biased upward towards zero. Of particular interest is the comparison between the site attribute coefficients across the two choice set specifications. If the parameter estimates are different this suggests that the results are sensitive to the method of dealing with missing data.

Table 2.9: Results of Conditional Logit & Random Parameter Logit using Restricted and Full Choice Sets

Attributes	Conditional Logit Full Choice Set	Conditional Logit Restricted Choice Set	Random parameter Logit Full Choice Set	Random parameter Logit Restricted Choice set
Random Parameter				
Access at Site	0.092(0.032)***	0.053(0.034)	0.354(0.081)***	0.225(0.091)**
Standard Deviation			0.639(0.062)***	0.568(0.066)***
Local Services	-0.300(0.034)***	-0.183(0.038)***	-0.335(0.109)***	-0.102(0.103)
Standard Deviation			1.837(0.183)***	0.971(0.120)***
Size of Fish	0.053(0.051)	0.123(0.054)**	0.225(0.113)**	-0.140(0.118)
Standard Deviation			1.432(0.129)***	0.963(0.160)***
Quantity of Fish	0.113(0.042)***	0.154(0.046)***	0.024(0.104)	-0.025(0.085)
Standard Deviation			0.592(0.092)***	0.070(0.069)
Variety of Fish	0.182(0.053)***	-0.089(0.060)	0.331(0.111)***	0.272(0.107)**
Standard Deviation			1.462(0.177)***	0.561(0.089)***
Encounters with other Anglers	-0.008(0.038)	0.035(0.042)	0.014(0.075)	0.168(0.082)**
Standard Deviation			0.0163(0.064)	0.382(0.083)***
Fixed parameters				
Travel cost	-0.065(0.004)***	-0.041(0.005)***	-0.080(0.009)***	-0.062(0.010)***
Killykeen Forest Park	-1.108(0.097)***	-1.105(0.117)***	-0.761(0.251)***	-1.404(0.283)***
Eonish	-2.114(0.118)***	-1.933(0.125)***	-1.400(0.224)***	-1.407(0.213)***
Dernaferst	-1.189(0.120)***	-1.450(0.177)***	-0.508(0.294)*	-1.529(0.376)***
Church Lake	-2.336(0.208)***	-2.27(0.267)***	-1.267(0.355)***	-2.577(0.477)***

Table 2.9 – continued from previous page

Heterogeneity in mean, parameter:				
Killykeen Forest Park: online	0.879(0.135)***	0.967(0.150)***	0.087(0.337)	1.037(0.343)***
Eonish: online	1.497(0.170)***	1.418(0.182)***	0.561(0.299)*	0.808(0.302)***
Dernaferst: online	0.725(0.164)***	1.307(0.215)***	-0.490(0.381)	0.717(0.426)*
Church Lake: online	1.728(0.235)***	1.968(0.291)***	0.357(0.394)	1.808(0.519)***
Log likelihood function	-2551.1284	-1989.4099	-2066.001	-1756.6589
Pseudo R2	0.2762	0.2005		
Akaike information criterion	5132.257	4008.820	4174.001	3555.318
Bayesian information criterion	5241.773	4112.412	4327.324	3700.346
Observations	10950	7377	10950	7377

Table 2.9 shows the results of a CL and RPL applied to the full and restricted choice sets. For the majority of the parameter estimates the sign remains constant throughout. *Access* is positive across all models with overlapping confidence intervals in the CL and in the RPL. However, it is not significant in the restricted choice set model. *Local services* is negative with overlapping confidence intervals in the RPL estimates. The variable *size of fish* is positive and significant in both the restricted choice CL and the full choice set RPL but insignificant in the full choice set CL and the restricted choice set RPL. The variable *quantity of fish* is positive and significant in both CL models and insignificant in both RPL models with overlapping confidence intervals in both cases. *Variety of fish* is positive and significant for all models except the restricted choice set CL. *Encounters with other anglers* was only positive and significant for RPL of the restricted choice set models but insignificant in the full choice set models. As expected, for both estimates of the *travel cost* variable based on the restricted choice set data are lower than the estimates based on the full choice set models. It is also interesting to note that the

pseudo R-squared indicates a better fit for the full choice CL model in comparison to the restricted choice set CL model.

There are a number of differences between the mean imputed data and the restricted data sets. This suggests that these results are sensitive to how the missing data is dealt with and that further analysis may be warranted.

2.5 Discussion

Although many of the results presented in this paper conform to a priori expectations (*access, size of fish,* and *variety of fish species)* some results seem to have counter-intuitive parameter estimates. The parameter estimates for *services* and *quantity of fish* do not suggest, as one would expect, that an angler would prefer a site with greater levels of these attributes. However, these results seem to align with the most relevant literature on the topic.

The parameter estimate for local *services* was negative and significant suggesting that anglers prefer sites away from areas with good local services. The effects of local *services* on angler participation have been relatively unexplored within the Irish recreational angling literature as only one paper (Curtis and Breen 2017) has employed any form of services to determine angler participation. Curtis and Breen (2017) found that the presence of tackle shops had a negative but insignificant role in the determination of trip length for a sample of Irish and overseas coarse angler. For a sample of game anglers, Curtis and Breen (2017) found that accommodation, and a good provision of pubs, dining, and family activities had a insignificant impact on trip duration[6]. However, the presence of a fishing guide was positively correlated with trip duration for game anglers. Although the sample used for our analysis was solely Irish coarse anglers, previous estimation results, that used a sample of both Irish and overseas anglers, suggested that Irish anglers preferred fewer *local services* and overseas anglers preferred more *local services*. In light of these results, it may not be surprising that Curtis and

[6] Curtis and Breen (2017) did not report how accommodation, a good provision of pubs, dining, and family activates, or fishing guides affect coarse anglers.

Breen's sample of Irish and overseas coarse angler would have a insignificant impact on angler participation.

The insignificant parameter estimate for the variable *quantity of fish* is in direct contention with a priori expectations that anglers prefer sites with more fish. However, this result is also supported by the literature; the level of fish stock did not have a significant impact on the number of days spent fishing or the number of trips taken in a year for a sample of Irish game, coarse and sea anglers (Curtis and Stanley 2016). Fish yield was found to be a insignificant determinant of trip length for Irish game anglers (Curtis and Breen 2017) and using, a sample of Irish coarse anglers, Curtis and Breen (2017) found that the ability to catch specimen fish was a positive and significant determinate of trip length but bag weight (total weight of fish caught) was negative and significant. Curtis and Breen (2017) have interpreted their results to mean that anglers spend more days at a site that has larger fish but less overall quantity of fish. Given the results presented here and the literature on coarse angler participation in Ireland further analysis is warranted to determine the importance of fish quantity to Irish coarse anglers.

Although highlighted by Hunt (2005) as an influential variable in angling site choice models *encounters with other anglers* had been unexplored within the Irish recreational angling context. As such, the results presented here are difficult to compare. However, the insignificant parameter estimate is, in some senses to be expected. Although encounters with other anglers is thought to reduce the enjoyment of an angling experience (Martinson and Shelby 1992), there are a number of sampled anglers who will only attend a site during competitions (these include large competitions and weekly local matches). For these anglers, one would expect a positive correlation between site choice and *encounters with other anglers*. Consequently, a insignificant parameter estimate with a relatively large degree of variation (as indicated by the standard deviation of the coefficient) may be a logical result.

In all cases, the standard deviation of the coefficients was statistically significant and relatively larger in comparison to the parameter estimates. This variance suggests that there is a large degree of heterogeneity in the sample (McConnell and Tseng 1999). The application of the RPL in this paper implicitly acknowledges this heterogeneity exists but, suggests that the source is unknown to the researcher (see Hunt (2005) for a detailed explanation of how heterogeneity has been dealt with in the recreational angling site choice model literature). The current study suggests very high levels of heterogeneity, some of which may be of a knowable variety. However, a portion of this may be unknowable to researchers. To investigate this further, a latent class logit was applied. The results tended to suggest that more classes were always favoured over fewer (this was tested for up to 12 classes). This made it difficult to pinpoint whether there was a distinct “lumpiness” to the preferences, which would suggest that a latent class approach is more appropriate.

The results of the sensitivity analysis suggest that bias may have occur due to the method of dealing with missing data. There are many differences between the set of results using mean imputation and the set of results the restricted data sets, demonstrating a sensitivity to how the missing data is dealt with. Without knowing the respondent's true choice set it is difficult to know what the likely true parameter estimates are. However, as noted by others a change in the magnitude of a parameter estimate may be expected (Peters, Adamowicz, & Boxall, 1995; Parsons et al. 1999; Hicks & Strand, 2000). Further research may be needed which could explore alternative methods of replacing missing perceived data, such as alternative means of data imputation like hot deck or multiple imputations. Not fully explored in this paper is what the respondents consideration sites are. The consideration set may, in fact, contain sites that are unrated and potentially not contain sites that have been previously rated. This too may impact parameter estimates even if mean imputation is a good approximation for the missing data.

2.6 Conclusion

Through the application of a site choice model, this paper aimed to develop a better understanding of Irish coarse anglers' preferences. For the first time, in the context of the recreational angling literature, it has been assumed that the perception of multiple site attributes varies across individuals and that these perceived site attributes can be used to model site choice in recreation anglers. A key aim of the NSAD (2016) is to increase the number of domestic anglers that regularly participate in the Irish angling scene. A comprehensive analysis of angler preferences, as was carried out in this paper, may improve management's ability to reach this goal.

To allow for both anglers' preferences and anglers' perception of a site to be heterogeneous between individuals a random parameter logit was applied to a dataset of site attributes constructed from anglers' perception of the site. The estimated parameters are used to constructed willingness to pay estimates that show the value of a marginal increase in site attributes as well as CV for a range of policy changes. Sensitivity analysis was conducted to determine if the practise of replacing unrated sites with the mean rating by those who visited the site may induce biased parameter estimates.

The results suggest that there is a statically significant correlation between anglers' perception of site attributes and their choice of a fishing site. The average size of the fish, the level of access, and the number of different fish species at the site all played a positive and significant role in site selection. The level of local services had a negative and significant impact on site choice, with a WTP of negative €4.20 for a marginal increase. One of the attributes conventionally thought to increase the probability of site selection does not seem to play a dominant role in angler's site choice. In our application, the quantity of fish variable was not statistically significant suggesting, somewhat counter intuitively, that the average angler does not choose a site based on the quantity of fish the site holds.

Two policy scenarios were examined during analysis; the first of which was an increase in access, as it is the most feasible attribute management can develop. The results of this analysis suggest that anglers would not benefit uniformly from an increase in access across all sites, as CV ranged from €2.26 at Garadice to $3.53 at Killykeen. The second explored avenue for development was the average size of fish at the site. This was selected as a TDI (2013) report suggests that fish quality is the most appealing aspect of Ireland as an angling destination. The estimates again varied between sites, with the average CV being estimated as €2.17 per choice occasion.

The sensitivity analysis suggests that the method of dealing with missing data may cause difference in parameters estimates. As it is impossible to know the true parameter estimates, it is impossible to know which method, if not both, cause bias. Further analysis is needed on this topic.

Some care needs to be taken when applying the results of this paper to all Irish coarse anglers as we do not have accurate information on the total number or composition of the entire population of Irish coarse anglers using the chosen sites. Consequently, the results should only be viewed as being representative of the sample. However, they still provide an indication of the likely preferences of Irish coarse anglers and a useful example of how angler attribute perspectives can be incorporated into angler site choice models.

A possible avenue for future research is to further examine how a change in size and quantity might affect the number of trips taken to a particular site. This could be accomplished through contingent behaviour analysis which would extend the estimates presented here to determine how an improvement in the quality of fish at a site would affect trip frequencies. Another area for future research would be to compare the model results here, that used the anglers own subjective ratings of each attribute, to a model that uses expert's/management's objective ratings for the same attributes. Finally, alternative methods of constructing values for unrated sites may be employed and tested; modal and imputations may be possible alternatives. However, this sort of analysis may be more ideally suited to a data set in which respondents have explicitly stated what their consideration sites are.

Chapter 3

3. Are Objective Data an Appropriate Replacement for Subjective Data in Site Choice Analysis?

3.1 Introduction

Economic theory suggests that a rational agent makes decisions based upon their perception of a good (Puto 1987; Singh 1988; Poor et al., 2001; Artell, Ahtiainen, and Pouta, 2013). A rational agent chooses a recreational site based on her perception of a site's attributes (Adamowicz et al. 1997), she buys a house because of the perceived bundle of goods the house possess (Chasco and Gallo 2013), and she decides whether or not to partake in risky behaviour founded on her perception of the risk she will be subjected to (Brewer et al. 2004). One might, therefore, conclude that the econometric analysis of site choice should be based solely on perception-based data, but this is seldom the case. Instead, objective measures are often used.

Objective measures of site characteristics are determined by a source external to the user, whereas subjective measures are based on users' own judgement of site attributes. In general, the literature has favoured objective data over the theoretically preferred subjective data. This predilection for objective data often stems from the comparative ease at which objective data can be collected (Baranzini et al. 2010; Artell et al. 2013) as collecting subjective data is often more time consuming, and costly. Outside of the academic literature subjective data is rarely used as a measurement for the quality of a good as the variance that is present in subjective data can make it more difficult to use in policy formation. As noted by Hynes et al. (2008), policy decisions are typically set in terms of objective measures of attributes indicating that a trade-off exists in what is more useful in terms of predicting recreationists' behaviour and the implementation of environmental policy.

This paper aims to explore the appropriateness of using objective data in place of subjective data when applied to a random parameter logit (RPL) site choice model for coarse anglers. At present, the academic literature is lacking in its exploration of the viability of objective data as a source of recreational site choice attribute levels, with only two papers (Adamowicz et al. 1997; Jeon et al. 2005) tackling this subject. This paper adds to the existing literature; by being the first site choice paper to compare models with identical variables from an objective data source and from a subjective data source and is also the first to compare models using identical attributes and a single choice set. This paper also presents a comparison of parameter estimates, willingness to pay estimates and compensating variation from site choice models applied to the objective and subjective data. This comparison is presented to examine if the objective ratings of site attributes are in line with the subjective ratings of the users of the resource and to determine the impact, if any, of using different sources of data on welfare estimation.

3.2 Literature review

The relative convenience of objective data has meant that the theoretically grounded subjective data (Baranzini et al. 2010; Artell et al. 2013) are seldom used in large-scale revealed preference choice-based analysis. In response, literature has developed assessing the relationship between subjective and objective data, and the appropriateness of the use of one source over another. This literature has been varied and spans across an assortment of models and applications. Hedonic modelling, for instance, has been used to determine the effect of air quality, water quality and noise pollution on house prices using both subjective and scientifically measured attribute levels (Poor et al., 2001; Chasco and Gallo, 2013; Baranzini et al. 2010). Site choice models have been developed using both managerial perception and users' perception of site attributes (Adamowicz et al., 1997), as well as site choice models comparing the scientific measure of water quality and users' perception of water quality (Jeon et al., 2005). Kappa statistics were used by Ma and Dill (2016) to test 'mismatch' between perceptions of neighbourhood bike-ability and objective data. Elsewhere, Farr et al. (2016) compared the extent to which objective or subjective perceptions of water quality affected willingness to pay estimates for an improvement in water quality at the Great Barrier Reef. Across these papers, the unified aim was to test the merits of using a single source of data, determining if there is value in collecting the more time-consuming subjective data when objective data are available.

The literature has taken two approaches to determine the need, or use, of incorporating subjective data into economic models. The first is a comparative method; researchers test if one source of data is superior to another in terms of predicting the dependent variable. Adamowicz et al. (1997) used several site choice models, half of which were applied to data comprised of users' perception of sites and the other half were applied to data of expert opinion. Models using users' perception performed better indicating that, given Adamowicz et al.'s (1997) data, users' perception of sites is a better indicator of site choice. Adamowicz et al. (1997) also demonstrated that the compensating variation estimate differed between the data sources.

The second objective of the literature is to determine if subjective data adds explanatory power to a model. This was examined by Baranzini et al. (2010), who saw no improvement in their hedonic price model through the addition of perceived levels of noise. It was determined that scientifically measured noise pollution sufficiently captured the effects of noise on house prices. Baranzini et al. (2010) note that there was a convergence between the subjective data and scientific data; this may be an indication as to why no improvement was found.

Additionally, the literature has also taken more explicit steps to test convergence between subjective and objective data. The literature has used correlation coefficients (Baranzini et al. 2010), and Kappa statistics (Ma and Dill, 2016; Kirtland et al., 2003), while others (Artell et al. 2013) have tried to establish the factors that are correlated with systematic divergence between the two data sources. Using bivariate probit and multinomial models Artell et al. (2013) investigated the factors correlated with the divergence between a subjective and an objective measure of water quality. They found that water body type, the level of objective classification, and distance to the site were all correlated with a difference between subjective and objective measures of water quality. This reveals that, in some cases, perceptions may be altered by objectively measurable variables.

Much of the literature looking at this issue has dealt with non-identical attributes. The levels of precision of subjective measures are usually much less than scientific measures. Scientific measures can, for instance, determine the exact decibel level of a source of noise pollution whereas subjective measures are often limited to a Likert scale. Additionally, a scientific measure can be extended to attributes that are unknown to users. In these cases, a researcher can restrict the scientific data to an aspect of water quality that is known to the users. Jeon et al. (2005) followed this protocol by restricting their comparison to water clarity. They compare the users' perception of depth visibility to scientifically measured water clarity. They found that user's perception deviated from scientific data but models including both scientific data and subjective data outperformed models using either one separately. Jeon et al. (2005) report that subjective measures of water clarity, as measured using their method, did not sufficiently describe the impact of water quality on site selection. An alternative to restricting the scientific data is to make a composite variable. This method was employed by Chasco and Gallo (2013) who made a composite index for both air quality and noise pollution to compare subjective and scientific data sources. They found that the subjective hedonic price model was preferred, with the objective model presenting counterintuitive signs for pollutants.

The literature has, in general, seemed to favour models based on subjective data. The inclusion of subjective data has been found to improve model fit (Jeon et al., 2005). Models solely using subjective data generally outperformed models using objective data (Adamowicz et al., 1997; Chasco and Gallo, 2013), or, in some cases, objective variables were found to have no statistically significant impact on the dependent variable (Farr et al., 2016; Lee et al., 2017). However, there are some examples where objective data outperforms subjective data (Poor et al., 2001; Baranzini et al. 2010). With respect to site choice models, only two papers exist where subjective data are compared to objective data (Adamowicz et al., 1997; Jeon et al., 2005) and only Adamowicz et al. (1997) uses multiple subjective site attributes. Although Adamowicz et al. (1997) collected identical attributes from the two data sources the authors use different choice sets for the objective and subjective models as well as including different variables in each model. Jeon et al. (2005) use the same choice set for both objective and subjective models but does not have identical attributes from both data source, and, consequently cannot have identical variables included in both models.

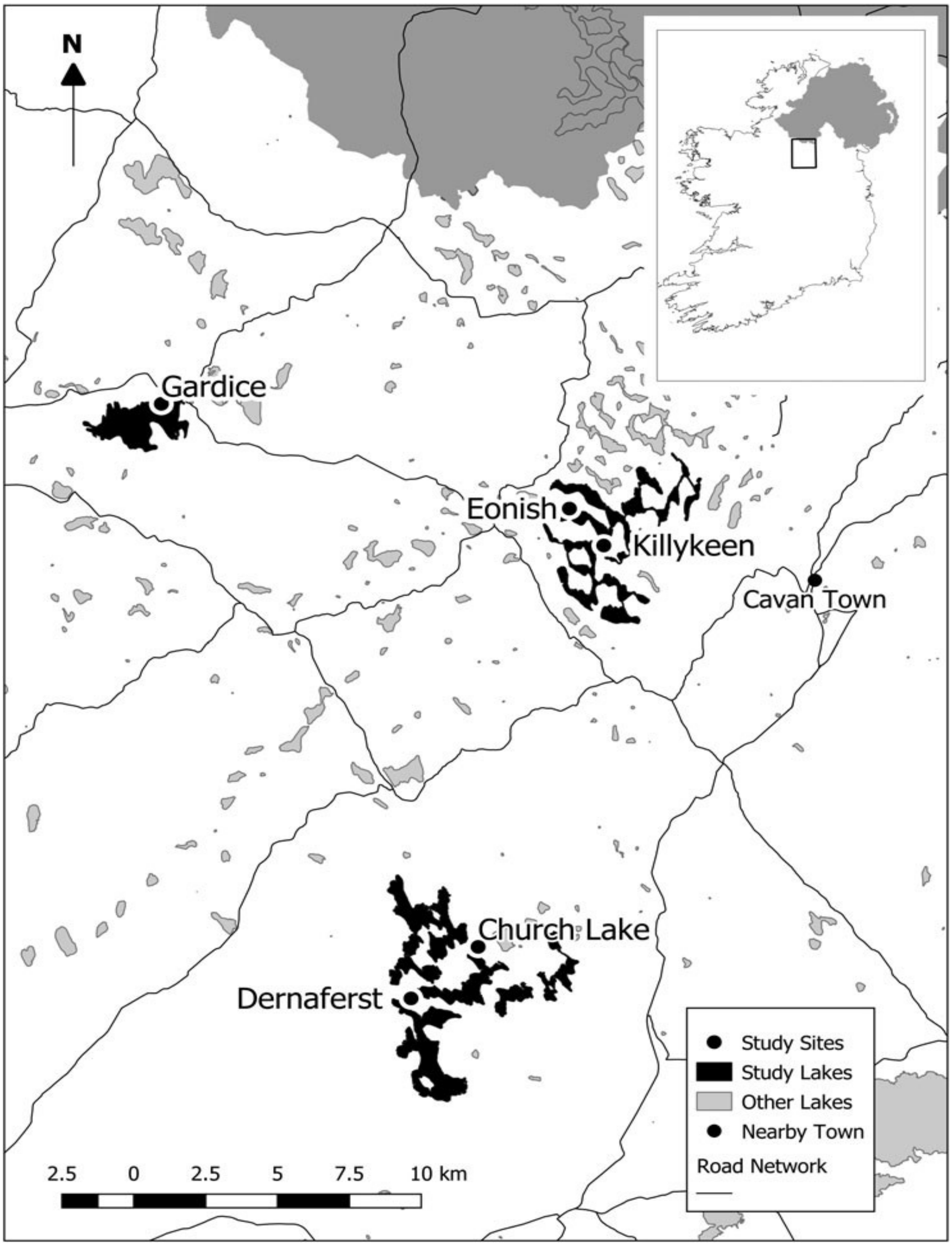
To the authors' knowledge, the comparison between identical measures of site attributes from objective and subjective data while incorporating preference heterogeneity in a site choice model or the use of identical measures and a single choice set has not been made in the literature. This paper adds to the existing literature by making both comparisons; using a RPL to account for preference heterogeneity. It also adds to the literature in a more general sense through further analysis of objective and subjective data.

3.3 Data

3.3.1 The Sites

The sites that comprise the researcher defined choice set (presented in Figure 3.1) are; Garadice, Killykeen Forest Park (referred to as Killykeen throughout this thesis), Eonish, Dernaferst, and Church Lake. All sites are situated within the counties Cavan and Leitrim, in the Republic of Ireland; both counties border Northern Ireland. Cavan and Leitrim were selected as this area is renowned for its coarse fishing due to the number of quality fishing sites available. As a result, there are multiple large fishing competitions held in both Cavan and Leitrim throughout the year. This area also has numerous fishing clubs indicating a strong contingent of enthusiastic anglers.

Figure 3.1: Map of Sites of Interest



Map display the study area and the five sites of interest

Garadice is a 3.9km lake situated in County Leitrim with multiple access points for boats and cars as well as parking beside fishing pegs. Due to the layout of the lake, anglers can choose a fishing point that best suits the weather conditions on a given day, making it a popular year-round destination. Garadice is a popular site with both recreational anglers and match anglers, hosting large annual competitions and smaller club competitions year-round. Garadice is also the best developed of the sites, providing multiple toilet and showering facilities.

Killykeen, like the remaining three sites, is located in County Cavan. Killykeen provides a beautiful scenic area for anglers to fish from which is enveloped by a forest park with trails that draw non-anglers to the site. For the most part, anglers must walk from the car park to the fishing pegs. Although this is a short distance this may be inhibiting to the less firm or fit anglers, particularly given the large amounts of gear coarse anglers travel with. There are two main access points to Killykeen. These access points lead to either side of a reasonably narrow fishing stretch. However, simply due to the road network, it would take approximately 20 minutes to drive from one bank to the other. It is assumed, for analysis, that the respondent chooses the access point closest to their home. Fishing quality was known to be particularly good at Killykeen as the coarse fish were drawn to the site by runoff from local chalets. Recently, these chalets have been shut down which may have impacted fishing quality.

Like Killykeen, Eonish is part of the Oughter water system. The fishing pegs on Eonish are all accessed by one road, that allows parking beside each peg. Eonish is one of the quieter sites as there is no park (Killykeen), play area (Dernaferst) or numerous recreational walkers (Garadice). Eonish also provides boat access and is in the closet proximity to accommodation of any of these sites with numerous lodges only meters from fishing pegs.

Dernaferst is a fishing site on the Gowna water system. It has two access points and a large parking area. A sizeable portion of the recreational fishing at Dernaferst takes place on the large boat ramp, allowing anglers to park a few meters from where they fish. Shore fishing can also be found a short walk away but requires the angler to carry their equipment through a field for a short distance. Dernaferst also provides a picnic area, children’s park and toilet facilities.

Like Dernaferst, Church Lake is part of the Gowna water system. Church Lake has some of the poorest access of all the sites, with anglers having to climb over a step gate to reach the fishing pegs. Until recently, Church Lake was renowned for its fishing. However, there seems to have been a downturn in recent years. Church Lake also has some of the deepest shore fishing of all the sites of interest. As coarse fishing is a year-round activity this may make Church Lake a much better winter fishing site than the other sites.

3.3.2 Subjective Data

Data were collected from 105 coarse anglers who fished in at least one of the five sites and was limited to residents of the Island of Ireland. Intercept surveying began on the 5th of August and ran until the 7th of November 2016 garnering 43 responses. Each of the five sites was visited multiple times during surveying, including both weekends and weekdays. The remainder of the surveys were completed online, which ran from the 6th of August to January 15th, 2017. The potential online participants were contacted through Irish coarse angling Facebook pages, by emailing local coarse angling clubs, and through the Inland Fisheries Ireland (IFI) newsletter. In order to increase the number of anglers participating in the survey who fish less frequently, local newspapers printed the details of the survey and how individuals could complete the survey online.

Due to the sampling procedures employed the data is likely to over represent anglers who fish frequently, in comparison to a random sampling framework sample. Anglers who frequent one or many of the five sites often have a higher probability of being sampled than their less avid counterpart. Although methods do exist to correct this avidity bias (Hindsley et al. 2011), like other recreational site choice models (Hanley et al. 2011; Scarpa and Thiene 2005; Deely et al. 2018) the requisite information is not available for the sites of interest and, as such, is uncorrected for. As a result, due care may need to be taken when interpreting the results and considerations may need to be given to the fact that the perceived data may be more representative of experienced anglers.

The respondents were asked to rate all the sites they had attended on a one to five-point Likert scale for six different attributes. An example of the rating system used for each site, containing the attributes and levels, is presented in table 3.1. These attributes were chosen based on a review of the relevant literature (Curtis and Stanley 2016; Hynes et al. 2015; NSAD, 2015), expert opinion, and focus groups[7]. In particular, the attributes were chosen so that the respondent's task of rating the sites would closely resemble the product criteria evaluation carried out by the National Strategy for Angling Development (NSAD, 2015) without being too cognitively difficult for the respondents to complete.

[7] Three focus groups were organised to improve the quality of the survey. The first group was comprised of environmental economists who gave insight into previous surveys they had undertaken which informed the overall formatting of the survey. The second group were employees of IFI who have expert knowledge of coarse angling and the Irish product. They had a large impact on both attribute levels, wording and the site choice. The final group was of local anglers. These individuals provided insight into their perception of the importance of the attributes, their ability to complete the survey and proposed new wording for some attributes. The focus groups were followed by a pilot study.

Table 3.1: Example Site Attribute Rating Table

Factor	Score/level of Factor				
Accessibility to the site (this includes parking and ability to reach the location that you fished at.) Score from 1 = very difficult to access to 5 = easily accessed	1 Difficult to access	2	3	4	5 Easy to access
Size of fish at the site (On average does this site provide access to good sized fish) Score from 1 = small fish to 5 = large fish	1 Small fish	2	3	4	5 Large fish
Quantity of fish (on average does this site provide access to a large quantity of fish) Score from 1 = low quantity to 5 = high quantity	1 Low quantity	2	3	4	5 High quantity
Encounters with other anglers Score from 1= none to 5 = frequent	1 No encounters	2	3	4	5 Frequent encounters
Variety of fish species (are there a large variety of species of fish at this site) score from 1 = low level of variety of fish to 5 = high level of variety of fish	1 Little to no variety	2	3	4	5 Lots of variety
Local services (these include pub, shops, accommodation etc…) Score from 1 = low level of local services to 5 = high level of services	1 Lacks local Services	2	3	4	5 Plenty of local services

This site attribute rating table was repeated for each of the five sites. The users were asked to rate each site they had ever attended. The managers were asked to give their managerial opinion on all sites.

The product evaluation criteria was chosen as it has previously been used as 'objective data' (Curtis and Breen 2017) and represents a good source of information on the important attributes of an Irish coarse angling site. The six site attributes selected were; *accessibility* (how easy it is to get to the location the angler will fish from), average *size* of fish caught at the site, average *quantity* of fish caught at the site, *encounters* (how often do they meet or see other anglers at the site), *variety* of fish species and *local services* (including shops, pubs, restaurants and accommodation). The five sites of interest, the chosen attributes and their means are presented in table 3.2.

Table 3.2: Mean Subjective and Objective Site Attribute Rating

Site	Access at Site	Size of Fish	Quantity of Fish	Local Services	Encounters with Other Anglers	Variety Of Fish
			Subjective	Data		
Site	Accessibility	Size	Quantity	Encounters	Variety	Services
Garadice	4.30	3.32	3.15	3.52	3.47	3.22
	(.79)	(.69)	(.74)	(1.05)	(.85)	(1.11)
Killykeen	3.29	2.99	3.23	3.59	3.47	2.97
	(1.13)	(.69)	(.83)	(1.05)	(.85)	(1.03)
Eonish	3.54	3.03	3.17	3.05	3.44	3.08
	(0.93)	(.53)	(.64)	(.92)	(.70)	(.84)
Dernaferst	3.46	3.05	3.29	3.46	3.35	3.50
	(.94)	(.72)	(.73)	(.79)	(.67)	(.80)
Church Lake	3.04	2.93	2.92	2.97	3.98	3.40
	(.80)	(.68)	(.74)	(.81)	(.60)	(.69)
			Objective	Data		
Garadice	5	4.5	4.5	4	4	3.5
	(0.0)	(0.5)	(0.5)	(1.0)	(0.0)	(0.5)
Killykeen	4.5	4	4.5	3.5	3	4.5
	(0.5)	(0.0)	(0.5)	(0.5)	(1.0)	(0.5)
Eonish	4.5	4	4	4	2.5	3.5
	(0.5)	(0.0)	(0.0)	(1.0)	(0.5)	(0.5)
Dernaferst	5	3.5	3.5	2.5	2.5	3.5
	(0.0)	(0.5)	(0.5)	(0.5)	(0.5)	(0.5)
Church Lake	2.5	3	3	3	2	3.5
	(0.5)	(0.0)	(0.0)	(1.0)	(0.0)	(0.5)

Ratings are on a 1- 5-point Likert scale. Standard deviation given in parenthesis.

When filling out the survey the respondents were asked to report the number of trips they had taken to each of the five sites of interest in the 12-month period prior to completing the survey. Angling experience and demographic questions were also asked, including the hometown or village where each respondent lived. The travel cost variable is calculated using

$$tc = ((travel\ distance * operating\ cost) + (travel\ time * \ opportunity\ cost\ of\ time)) * 2 \quad (3.1)$$

Where travel distance is the distance from the individual's home to the fishing site, operating cost is equal to 0.2475 cent per kilometre[8] and opportunity cost is calculated as 33% of the individual's hourly wage (Parsons, 2003). Following Hynes, Hanley, and Scarpa. (2008) and Hanley et al. (2001), if an angler did not rate a site on any of the site attributes the missing value was set equal to the mean response of other anglers for that attribute.

3.3.3 Objective Data

The objective data were collected in the same manner as the angling product evaluation criteria (NSAD, 2015) using some of the same respondents and, coincidentally, the same method as employed by Adamowicz et al. (1997). The two fisheries managers for the area containing the sites of interest were asked to rate the sites using an identical questionnaire as the one presented to the angler respondents. To ensure that the management perspective is as objective as possible a number of tactics were employed; firstly, to provide consistency between the NSAD measurements and the objective data used in this study, the managers who completed the NSAD survey, for the study area, were also asked to complete the present survey. Secondly, the managers were informed of the aim of the study and that comparison would be made between the management perspective, provided by them, and the users' perspective. This step was taken so that the managers understood that it is their perspective as a manager that was of concern to this study.

[8] This is running cost of operating a medium sized vehicle according to Automobile Association of Ireland.

The decision not to use the product evaluation criteria already available from IFI was multifaceted. Firstly, the product evaluation criteria evaluated much larger areas than the sites the user respondents were expected to rate. These areas often encompassed multiple fishing sites, which would complicate estimates of travel distance as these areas frequently included sites miles apart. Secondly, the cost of getting management to rate the sites was negligible compared to the possible drop in respondent retention through the necessary expansion of the user survey to match the product evaluation criteria. And, finally, by getting management to rate the sites, an exact comparison between subjective and objective rating can be made[9]. Mean objective ratings are presented in the lower portion of table 2.

In order to test if the objective data can be considered to be from the same distribution as the subjective data, a Kolmogorov-Smirnov test was conducted. The test results are found in table 3.3. The results indicate that for the attributes *size*, *access* and *quantity* the objective data cannot be considered to be from the same distribution as the subjective data. A visual inspection of the objective and subjective ratings also reveals that objective ratings tend to be higher. In many cases, the objective ratings are above the mean of the subjective ratings by more than one standard deviation of the subjective data. This may explain why these variables failed the Kolmogorov-Smirnov test of equal distribution. The disparity between the two data sources is particularly evident for the variables rating the average *size* and *quantity* of fish at each site. In six of the ten cases the objective rating for *size* or *quantity* was more than one standard deviation above the mean of the subjective rating.

[9] Unfortunately, it is not possible to assess how closely the management rating and the NSAD product evaluation coincide; the NSAD areas include so many sites that it is impossible to know the contribution of any one site in order to compare them.

Table 3.3: Results of Kolmogorov–Smirnov Tests

	Kolmogorov-Smirnov test statistic
Access at Site	0.027*
Size of Fish	0.000*
Quantity of Fish	0.012*
Local Services	0.568
Encounters with other Anglers	0.113
Variety of Fish	0.104

*Note: P values reported, *indicates significance at the 5% level suggesting that for these attributes the expert opinion does not come from the same distribution as user's opinion.*

The variable *variety* has almost no variability suggesting that management view all the sites as having the same *variety* of fish species. This may render the objective variable *variety* unsuitable for predicting site choice. It is also unclear if management and users rated the variable *variety* using the same criteria. It may be the case that management rated each site based on the presence of different species or their abundance. This lack of *variety* could be due to management considering only the presence of the species. For the users, abundance might play a much more vital role in their rating, particularly for anglers who seldom fish at the sites of interest. One would expect that an angler would rate the *variety* of fish species based on the fish they have caught or heard of others catching, in this case, the abundance of each species could play a pivotal role in each anglers rating. In relation to the subjective data, a lack of variety between respondents is not a problem and as such models applied to this data set can support all site variables.

To test how attribute ratings move between sites, for a given attribute simple correlation tests are employed, the results of which can be seen in table 3.4. A correlation coefficient of less than one indicates that a unit change in the objective rating of an attribute is not met with an equal change in the subjective rating. However, as the subjective rating varies between people, the expectation is that none of the variables will present with a coefficient of one, although a positive coefficient is expected for all variables.

Table 3.4: Correlation Statistics between Subjective and Objective Variables

	Pearson's correlation	Spearman's correlation
Access at Site	0.3175	0.3267
	(0.000)*	(0.000)*
Size of Fish	0.1804	0.3027
	(0.000)*	(0.000)*
Quantity of Fish	0.0662	0.1418
	(0.000)*	(0.000)*
Local Services	-0.1867	-0.2276
	(0.000)*	(0.000)*
Encounters	0.2501	0.3220
	(0.000)*	(0.000)*
Variety	0.1008	0.1141
	(0.003)*	(0.000)*

*Note: * indicates significance at 5% level. P values given in parenthesis.*

The correlation coefficients indicate that there is a consensus between the subjective and objective data on the direction of the ratings but no coefficient is close to one, meaning that the rates of change between sites vary. However, in the case of *services*, there is a negative and significant relationship demonstrating that users' perception of the quality of *services* near a site runs in the opposite direction to the objective data. Due to the difference in how these variables change between sites, as measured by the correlation statistics, there is an expectation that parameter estimates will vary between data sources.

3.3.4 Trip Frequencies

In total 2190 trip observations were taken to the five sites of interest. The mean and total number of trips taken to each site as well as the number of respondents who visited them can be seen in table 3.5. Garadice was the most popular site with almost as many trips taken there as the other four sites combined. Although seeing fewer trips, Killykeen was visited by the second most anglers just one less than Garadice. At the other end of the spectrum, Church Lake was visited by the fewest anglers and had the lowest total number of trips. In general, those who went fishing at one of the sites of interest once a month, or once a week tended to spread out their site choice in a similar manner. However, for those who went fishing more than once a week, there is a strong preference for Garadice. In part, this may be due to local intraclub matches being held there, but also Garadice's ability to provide different fishing points that are distinct enough to make fishing more hospitable during any weather condition.

Table 3.5: Mean and Total Trips Per Site

	Number of anglers who have visited each site in the last 12 months	Mean trips	Total Trips
Garadice	71 (67.61%)	15.39	1,093
Killykeen Forest Park	70 (66.67%)	7.06	494
Eonish	45 (42.86%)	4.36	196
Dernaferst	43 (40.95%)	6.02	259
Church Lake	33 (31.4%)	4.48	148

Note: Percentage of sample who visited each site is given in parenthesis. Mean number of trips refers to the average number of trips taken by anglers who visited at least once.

In order to test whether the perspective of the on-site cohort and online cohort were similar Kolmogorov–Smirnov tests were used to determine if they came from the same distribution. The test shows that the two cohorts' responses can be considered to be from the same distribution for all but 6 of the 30 attributes[10]. These attributes are *access* at Garadice and Eonish, *services* at Garadice, *encounters* at Garadice and Killykeen and *variety* in Eonish. To account for this difference an interaction term is added to the analysis.

3.4 Methods

To test the suitability of objective data to accurately represent the sites as perceived by site users a number of procedures are undertaken. A RPL is applied to both the subjective data and objective data, measures of fit are

[10] This table can be viewed in table 2.2 of the appendix

compared between the models, as well as the number of correct predictions made by each model. Then, the magnitude and direction of the coefficients are compared between models to assess differences. Willingness to pay (WTP) estimates are used to compare welfare effects. Compensating variation is used to demonstrate, under the two different sources of data, the welfare loss to anglers from the closure of each of the five sites of interest. Finally, a further two datasets are created in order to examine if the perspective of the management is representative of the average perspective of the users.

3.4.1 Model

McFadden (1973) stated, through the use of a random utility model (RUM), that an individual will select the site that maximises her utility on a given choice occasion. This utility can be written as:

$$u_{in} = V(X_{in}, y_n - p_{in} \mid \theta_n, z_n) + \varepsilon_{in}$$

$$= V_{in} + \varepsilon_{in} \tag{3.2}$$

Where u_{in} is the utility that individual n receives from choosing site i, V is the indirect utility function, X_{in} is either a vector of subjective attributes or a vector of objective attributes, y_n is the income of individual n, p_{in} is the travel cost, θ_n is a vector of individual n's characteristics and z_n are individual n's covariates and ε_{in} is the stochastic error term and is unknown to the modeller. It is assumed that the error term is independent and identically distributed (IID) extreme value type 1. The resulting estimated parameters are homogenous across individuals; implying that every individual sampled has the same taste preferences (Train 1998). The RUM model takes the form of a conditional logit (CL) (McFadden 1973) when the error terms are independently and identically drawn from an extreme value distribution.

As noted by Train (2009), by decomposing the error term the restrictive IID quality of the CL is overcome. The decomposed error term has two distinct elements, the first is correlated over alternatives and is heteroskedastic, the second is IID over alternatives and individuals. The resulting model is the RPL. The utility equation with a decomposed error term can be written as:

$$u_{in} = V_{in} + [\eta_{in} + \varepsilon_{in}] \tag{3.3}$$

and

$$V_{in} = \beta_{in} * X$$

Where η_{in} is a zero mean random term, which may be correlated across alternatives, and individuals, ε_{in} remains IID. The decomposition of the error term allows the parameter estimates to vary randomly across individuals but remain homogenous across choice occasions for an individual. The probability of individual *n* selecting site *i* is logit and can be written as:

$$\Pr\,(\mathrm{i}) = \frac{\exp\,(\mu V_{in} + \eta_{in})}{\sum_{k=1}^{j} \exp\,(\mu V_{jn} + \eta_{jn})} \tag{3.4}$$

where μ is a scale parameter and η_{in} can take on a number of distributional forms (Hensher and Greene 2003), which must be specified by the modeller. Assuming that η_{in} takes a multivariate normal distribution, it can be written that:

$$\beta_n \sim N(\bar{\beta}, \Omega)$$

where $\bar{\beta}$ is the mean of the parameter and $\mathbf{\Omega}$ is the variance-covariance matrix.

Accommodating for an unbalanced panel data the logit is integrated across all values of η_{in}, with appropriate density weightings. This forms the unconditional choice probability and can be written as:

$$\int \prod_{t=1}^{t=T(n)} \frac{\exp\,(\beta_{in} + \eta_{int})}{\sum_{j=1}^{j} \exp\,(\beta_{jn} + \eta_{jnt})}\ \varphi(\bar{\beta},)\mathrm{d}\beta_n \tag{3.5}$$

Where T(n) is each respondent's revealed preference, $\varphi(.)$ is the multivariate normal density, $\bar{\beta}$ and $\mathbf{\Omega}$, the mean and variance parameters, are estimated from the sample data.

3.4.2 Welfare estimates

Two methodological approaches to estimating welfare are employed in this paper. The first is willingness to pay estimates (WTP) and the second is compensating variation (CV). WTP estimates measure marginal value. WTP estimates are calculated following Train (2009):

$$WTP = \frac{\beta_n}{-\beta_{tc}} \tag{3.6}$$

Where β_n is the coefficient of the attribute of interest for individual n and -β_{tc} is the negative of the travel cost coefficient, which, here, represents the marginal utility of income. In the context of this paper, WTP estimates have an added advantage. WTP estimates are standardised into a monetary value. This standardisation allows for a meaningful comparison across models.

The second method used is CV. CV determines the amount of money an individual would have to pay or receive for their utility to be unchanged after a change to a site in their choice set. Following Hanemann (1982) CV is calculated as:

$$CV_n = -(\beta_{tc})^{-1} \left[\ln\left[\, \sum \exp(\hat{\beta}_n x_n^1\right) - \ln[\sum \exp(\hat{\beta}_n x_n^0)]]\right] \tag{3.7}$$

The negative of the travel cost coefficient β_{tc} represents the marginal utility of income, which in the models presented in this paper is fixed across all individuals. $\hat{\beta}_n$ is a vector of parameters for individual n. x_n^0 is either a vector of subjective site attributes or objective site attributes and x_n^1 is the same vector after some exogenous change to the site. For RPLs, CV must be integrated over simulated taste distributions (Train 1998):

$$\widehat{CV} = \int CV_n \varphi\left(\hat{\beta}, \hat{\Omega}\right)\, d\beta$$

$$= \int\{-(\beta_{tc})^{-1} \left[\ln\left[\, \sum \exp(\hat{\beta}_n x_n^1\right) - \ln[\sum \exp(\hat{\beta}_n x_n^0)]]\right]\, \varphi(\hat{\beta}, \hat{\Omega})\, d\beta \tag{3.8}$$

CV, in this paper, focuses on the closure of each of the five sites individually is presented as the average per person per choice occasion. State zero is the value of all five sites to an individual n and state one is the value of four of the sites to the same individual. Although it is conceivable that a site could be estimated to have a negative value for any one individual, it is assumed that an individual cannot be made better off by the closure of a site and as such all negative values are set equal to zero.[11] Additionally, although the researcher defined choice set is comprised of five sites, each individual's choice set can, and more than likely do, contain more sites.

3.4.3 Model Comparison Procedure

Parameters are estimated for three models; the first model is applied to the objective data, the second model is applied to the subjective data based on the same set of attributes as in the first model, the final model uses an extended set of parameters that could not be used in the model applied to the objective data. Comparison is made both between the estimated parameters of each model, and between their corresponding welfare estimates. The second stage of comparison is to determine if the subjective data can replicate the findings of the objective data, through a number of logical contractions of the subjective data. The aim of this comparison is to determine if the managers and the users are rating the site attributes using the same criteria. If this is the case then a strong argument can be made that the added variability of the subjective data, assuming a better fitting model, allows for more precise estimation of real-world preferences. For this comparison, two adjustments to the dataset were used to create new attribute levels with accompanying site choice models. The first was a simple averaging of the subjective data. Through the use of this averaged subjective data hypothesis tests are conducted to determine if the coefficients of the objective data align with the coefficients from the site choice model applied to the average ratings of the attributes. A second and maybe more plausible consideration is that management perspective is more closely aligned with the anglers who fish at these sites most often. To test this hypothesis the observations in the dataset are reweighted by the number of trips an angler has taken to each of the five sites, in essence, the more often an angler went fishing the heavier their weight. As the survey was not conducted using a random sampling framework the sample is composed of more avid anglers than the national average. This

[11] Negative cases range from 1 at Killykeen using the extended subjective model to 32 at Church Lake using the objective model. Nearly 50% of the sites across all three models had less than 10 cases where the value was less than zero.

combined with the weighted mean system employed could result in a data set that is much closer to the views of the more avid angler than would be expected from a national survey. Consequently, due care should be taken when interpreting the results.

3.5 Results

The first model (column 1 in table 3.6) is a RPL applied to the objective data. Two of the site attributes have been excluded from this analysis; *variety* because it lacked variance across the sites and *encounters* due to collinearity issues. In the case of a site choice model a dummy variable, indicating whether the angler completed the survey online, cannot be fit directly to the model as there would be no variance between sites for an individual. Consequently, the interaction term *access: online*, is used to capture differences between the online cohort and the onsite cohort. It is constructed by multiplying a dummy variable indicating that the survey was completed online with the variable *access*. This interaction term shows heterogeneity in the mean, indicating that, in the event of a significant coefficient, the average of the online cohort has a statistically different preference to the onsite cohort for *access*. The second model is a replication of the first model applied to the subjective data. This allows for a direct comparison between the two models. The third model is the extended model given the subjective data. The subjective data set has much more variability than the objective data; this allows for the inclusion of all the site attributes thought to impact site choice as well as alternative specific constants for each site.

Table 3.6: Results of Random Parameter Logits

	Objective Model Mean of Coefficient	Reduced Subjective Mean of Coefficient	Extended Subjective Mean of Coefficient
Random Parameters			
Access at Site	0.657 (0.277)***	0.828(0.075)***	0.569(0.092)***
Standard Deviation	1.372(0.420)***	1.14(0.080)***	1.091(0.096)***
Size of Fish	1.372(0.420)***	0.399(0.108)***	0.307(0.116)***
Standard Deviation	1.70(0.155)***	-2.226(0.292)***	1.763(0.204)***
Quantity of Fish	0.847(0.277)***	0.263(0.094)***	-0.004(0.092)
Standard Deviation	3.659(0.263)***	1.635(0.130)***	0.752(0.157)***
Local Services	-1.025 (0.241)***	-0.394(0.094)***	-0.510(0.095)***
Standard Deviation	2.264(0.189)***	1.822(0.163)***	1.141(0.169)***
Encounters with other Anglers			0.165(0.080)**
Standard Deviation			0.544(0.717)***
Variety of Fish			0.287(0.117)**
Standard Deviation			1.302(0.254)***
Fixed parameters			
Travel Cost	-0.092 (0.008)***	-0.074(0.007)***	-0.066(0.009)***
Killykeen			-0.472(0.172)***
Eonish			-0.882(0.163)***
Dernaferst			-0.804(0.164)***
Church Lake			-0.797(0.197)***
Heterogeneity in mean, parameter:			
Access: Online	-1.056 (0.226)***	-0.379(0.159)**	-0.321(0.122)***
Model fit			
Log likelihood function	-2131.05	-2110.03	-2062.77
AIC	4282.09	4240.07	4161.53
BIC	4355.105	5303.829	4292.95
Correct Predictions	27%	30%	32%
Observations	10950	10950	10950

*Notes: Figures in parenthesis are standard errors. *** indicates significant at 1%, ** indicates significant at 5% and * indicates significant at 10%*

The best fitting model is the extended subjective model as it has the log-likelihood function that is closest to zero. This is to be expected in some respects; it should be the case that the extended subjective model should outperform the reduced subjective model as it has additional parameters. In this case, the Akaike information criteria (AIC) (Akaike 1998) and the Bayesian information criteria (BIC) (Schwarz 1972) may be a more appropriate measure of fit as both add penalties for the number of parameters estimated. Both the BIC and AIC also indicate that the best fitting model is the extended subjective model. The extended subjective model also predicts the correct site choice in the largest percentage of cases, predicting the right site choice in 32% of choice occasions. This is 2% more often than the reduced subjective model, and 5% more often than the objective model

A comparison of the three models shows that in all cases but one direction is identical across parameter estimates. The variable *Quantity of fish* is significant in both reduced models but not in the extended subjective model. For all three models the *travel cost* parameter is negative and significant, suggesting that, all else being equal, anglers will choose to visit the site with the lowest travel cost. The results of all three models also indicate that *access* plays a significant role in site choice. However, the significance of the interaction term *Access: online* indicates that the online cohort have a statistically different preference for *access* than the onsite cohort. There are some reasonable explanations as to why this may be. *Access* may be correlated with general activity at the fishing site. Good *access* may be correlated with high volumes of recreational activities other than fishing; examples of recreational activities that occur at some of the sites of interest are dog walking, cycling, kayaking, and picnics. This level of activity may be a deterrent for some of the sampled anglers. It may also be the case that the scenery or atmosphere of the fishing site is detracted from in some way by the development of *access*. These factors could make a site less appealing for certain anglers.

Across all models *local services* play a negative and significant role in site choice indicating that anglers tend to pick sites that are away from good local services. This may suggest that the more appealing sites are more remote and further away for bigger towns or villages. The *size of fish* variable had a positive and significant role in all three models, indicating that anglers prefer sites with bigger fish. The effect that average *quantity of fish* played on site choice differed between the models; it was positive and significant for the objective model and the reduced subjective model but insignificant for the extended subjective model. The significance of the standard deviation suggests that there are some individuals that prefer sites with larger quantities of fish while others prefer sites with smaller quantities of fish. However, with this variable, and all others used in this analysis, a somewhat strong assumption is that all attributes are considered when choosing a fishing site. It may well be the case that for some anglers, or even just some choice occasions for individual anglers, the *quantity of fish* did not play a role in their decision on where to go fishing resulting in them choosing a site that has, by their own estimation, a lower *quantity of fish* than other sites. This could result in a coincidental correlation between site choice and low rated *quantity of fish* rather than a purposeful decision to choose a site where they have a lower chance of catching a fish.

The extended subjective model contains a number of variables not contained in either the objective or reduced subjective models. Both *variety of fish* species and *encounters with other anglers* are included in the model. *Variety* seems to play a positive role in site selection, indicating that anglers prefer sites with more species of fish. *Encounters with other anglers* has a positive effect indicating that anglers tend to pick sites where there is a good chance of meeting other anglers. It is worth considering that there may be an endogeneity issue as there is likely correlation between *encounters* (or more accurately number of anglers at a site) and being sampled. The extended model also contains four alternative specific constants. The ASCs are all negative and significant implying that these sites possess attributes that negatively affected site selection in comparison to Garadice, the base case, which are unaccounted for by the other variables presented in the model. Conversely, Garadice may contain positive attributes that the other sites do not. Garadice seems to hold certain attributes that were not quantifiably measured that may have induced this result. For instance, as there are multiple points to fish around Garadice an angler can be assured some level of shelter from the weather regardless of wind direction. It was also a popular spot for local angling clubs, often booking pegs for regular intraclub matches.

In order to test the similarity of the estimated parameters across models, simple hypothesis tests are employed. Following Clogg (1995), hypothesis testing was conducted using the formula …

$$Z = \frac{\beta_1 - \beta_2}{\sqrt{SE\beta_1{}^2 + SE\beta_2{}^2}} \qquad (3.9)$$

Where β_1 and β_2 are parameter estimates of the same variable from two different models and $SE\beta_1{}^2$and $SE\beta_2{}^2$ are the respective coefficient variances.

Table 3.7 shows the results of the hypothesis tests. A P-value of less than 0.05 signifies that the null, $\beta_1 = \beta_2$, can be rejected. The results indicate that the coefficient estimates from the objective model are statistically different to the estimates provide by the subjective models for almost all variables; four of the six estimated coefficients are different when comparing the results of the objective model against the reduced subjective model and five of the estimated coefficient are different from the objective model to the extended subjective model. In comparison, only two of the estimated coefficients are statistically different from the reduced subjective model to the extended subjective model. These results seem to suggest that, for our samples, parameter estimates do vary based on the source of the data.

Table 3.7: Equality of Coefficient Hypothesis Testing

Variable	Objective versus reduced	Objective Versus extended	Reduced versus extended
Access at Site	0.549	0.762	0.029*
Size of Fish	0.025*	0.015*	0.562
Quantity of Fish	0.046*	0.004*	0.042*
Local Services	0.014*	0.047*	0.385
Travel Cost	0.091	0.031*	0.483
Access: Online	0.014*	0.014*	0.773

Note: P-value reported, * denotes significance at 5% level.

3.5.2 Welfare Estimates

WTP estimates are presented in table 3.8. WTP estimates were calculated using the WTP command in Stata. This follows equation 3.6, calculating WTP as the ratio of the estimated parameter of interest over the travel cost parameter. Confidence intervals were computed using the Krinsky-Robb method (Krinsky and Robb 1986) with 5,000 draws. Estimates for interaction terms follow the approach used by Nahuelhual et al. (2004):

$$WTP = \frac{\beta_r + \sum_i \frac{\beta_{r*it\, i}}{n}}{\beta_{tc}} \tag{3.10}$$

Where β_r is the random coefficient (i.e. *access*), $\beta_{r*it\, i}$ is the interaction term for individual *i,* that is associated with that random coefficient (i.e. *access: online*), n denotes the sample size and β_{tc} remains the marginal utility of income.

Table 3.8: Willingness to Pay Estimates (€ per choice occasion)

Attribute	Objective Model	Reduced Subjective Model	Extended Subjective Model
Access at Site	7.11(1.22 , 13.33)	11.17(8.68 , 7.83)	8.61(5.25 , 13.87)
Size of Fish	14.84(5.94 , 24.66)	5.38(2.52 , 8.57)	4.64(1.22 , 9.02)
Quantity of Fish	9.16(3.33 , 15.82)	3.55(1.03 , 6.35)	-0.50(-2.79 , 2.94)
Local Services	-11.08(-16.49 , -5.97)	-5.31(-8.13 , -2.74)	-7.71(-10.94 , - 5.04)
Encounters with Other Anglers			2.49(0.16 , 5.45)
Variety of Fish			4.33(0.72, 8.81)
Access: online	-4.31(-10.21 , -1.91)	6.06(3.57 , 9.28)	3.74(0.39 , 9.00)

Note: 95% confidence intervals in parenthesis

For all but two variables the estimates are of the same sign. The WTP estimates for *quantity of fish* is negative but insignificant from the subjective extended model results but positive and significant from the two smaller models. This ranged from -€0.50 in the extended subjective model to €9.16 in the objective model. *Access: online* has a negative WTP estimate from the objective model but a positive WTP estimate from the models applied to the subjective data. The estimates from the objective data indicate that anglers who completed the survey online have a negative WTP of €4.31 for an increase in *access*, whereas the estimates based on subjective data results in WTP of €6.06 and €3.74 for the reduced and extended models, respectively. Although a negative WTP for an increase in *access* may seem counter-intuitive this may suggest a more complex relationship between the use of the site by non-anglers and the sites desirability for anglers.

WTP estimates for *Local Services* were negative for all three models, indicating in each case, that anglers are willing to pay for a reduction in *local services*, although it is more likely that this may be acting as a proxy for remoteness. Across all three models, anglers have a positive WTP for an increase in the *size of fish* at a site. This ranges from €4.64 in the subjective model to €14.84 in the objective model.

The remaining estimates for the extended subjective model show that there is a positive WTP for both *variety* and *encounters*. As indicated by the significant standard deviation estimate of the RPL model, for some anglers the level of *encounters* may be a deterrent and a positive draw for others. In particular, it would be expected that anglers who regularly fish in competitions or are members of local clubs will have a strong correlation between *encounters* and site choice. The remaining alternative specific constants are not shown here but mirror the results of the RPL model. All four have negative WTP, suggesting that an angler would have to be compensated to pick one of these sites over Garadice.

Table 3.9: Compensating Variation for Site Closure (per person per choice occasion, €)

Site Closure	Objective Model	Reduced subjective Model	Extended subjective model
Garadice	57.54 (45.08, 70.00)	48.78 (38.39, 59.17)	52.39 (42.25, 62.53)
Killykeen	64.50 (55.27, 73.73)	49.03 (41.15, 56.91)	51.88 (44.33, 59.43)
Eonish	45.56 (37.41, 55.71)	34.92 (27.29, 42.55)	32.19 (25.56, 38.82)
Dernaferst	33.82 (27.18, 40.46)	27.47 (19.84, 35.10)	27.70 (21.50, 33.90)
Church Lake	14.65 (9.96, 19.34)	16.59 (12.38, 20.90)	13.87 (9.91, 17.83)

Note: 95% confidence intervals given in parenthesis.

The average, per person, per choice occasion, welfare loss from a site closure, displayed in table 3.9, is similar for each site across all models. The ranked order of the sites, in terms of CV, is almost identical across the three models with the exception that; Killykeen has the largest CV, and Garadice has the second largest CV in objective and reduced subjective models, whereas the reverse is true for the extended subjective model. In all cases but one, the value of CV for a site closure is larger for the objective model than either of subjective models. CV for the closure of Church Lake is larger in the reduced subjective model than the objective model. However, the difference between these estimates is not large with the greatest difference being between the CVs for Killykeen from the reduced subjective model to the objective model. In this case the objective model estimate is 31% larger than the reduced subjective model.

The top two most visited sites, Garadice and Killykeen, were also the sites that would need the greatest compensation for their closure. Compensation for a closure of Garadice ranges from €48.78 to €57.54 per trip and compensation for Killykeen ranges from €49.03 to €64.50. The two sites that would cost the least, in terms of CV, if the sites were closed are; Dernaferst and Church Lake. These were also the two sites visited by the lowest number of surveyed anglers; however, Eonish received a lower number of total trips than Dernaferst. In part, Dernaferst and Church Lake had the lowest CV because the average sampled angler had to travel the furthest to reach these sites.

Compensation for the closure of the sites to the sampled anglers for the survey year ranged from €30,375 for the closure of Church Lake based on the results of the extended model to €141,255 based on the results of the objective model. However, these results are based on a sample that could be overrepresented by the most eager anglers and, as such, due care should be taken when interpreting these results.

3.5.3 Management Perspective and the Average Angler

Table 3.10 shows the results of the reduced objective model (repeated from table 6), the unweighted mean model, and the weighted mean model. The results of the data set comprised of anglers' mean perception seem unlike any other model presented in this paper. Although the direction of the parameter estimates is the same as all the previously presented models the magnitudes differ greatly. The most striking is the parameter estimate for the variable *Size of Fish;* it is estimated to be over six times greater than either the objective or weighted mean models. This may be the result of the averaging process reducing variability across the sites; the difference between the site with the largest fish and the smallest fish is 0.35 on the five-point Likert scale. Consequently, if an angler chooses one site over another, based on *size*, they are making a decision based on a small change in average *size*, which in turn produces a relatively large coefficient for a one-unit change in *size*.

Table 3.10: Results of *RPL applied to Mean and Weighted Mean data sets*

	Objective Model Mean of Coefficient	Mean Model Mean of Coefficient	Weighted Mean Model Mean of Coefficient
Random Parameters			
Access at Site	0.657 (0.277)***	0.323(0.573)	0.182(0.200)
Standard Deviation	1.372(0.420)***	3.171(0.373)***	1.273(0.183)***
Size of Fish	1.372(0.420)***	8.308(1.903)***	1.264(0.697)***
Standard Deviation	1.70(0.155)***	-5.318(0.570)***	-3.74(0.523)***
Quantity of Fish	0.847(0.277)***	2.028(0.601)***	2.849(0.717)***
Standard Deviation	3.659(0.263)***	3.507(0.570)***	-5.934(0.450)***
Local Services	-1.025 (0.241)***	-2.60(0.414)***	-1.208(0.280)***
Standard Deviation	2.264(0.189)***	-.3.390(0.251)***	2.506(0.217)***
Fixed parameters			
Travel Cost	-0.092 (0.008)***	-0.130(0.0145)***	-0.088(0.014)***
Heterogeneity in mean, parameter:			
Access: Online	-1.056 (0.226)***	-0.800(0.292)***	-0.952(0.256)***
Model fit			
Log likelihood function	-2131.05	-2084.17	-2073.81
Akaike information criterion	4282.09	4188.35	4167.61
Bayesian information criterion	4355.105	4261.36	4240.62
Observations	10950	10950	10950

*Notes: Figures in parenthesis are standard errors. *** indicates significant at 1%, ** indicates significant at 5% and * indicates significant at 10%*

The weighted mean model produces results that are similar to the objective data, with all but one variable having overlapping confidence intervals. In the case of the variable whose confidence intervals do not overlap, *Quantity of Fish,* the WTP, although not presented here, do overlap. This level of similarity is not found between any two other models, even the reduced subjective and extended subjective models do not share overlapping confidence intervals for two of their variables. This may reveal that management perspective is more closely aligned with anglers who spend a lot of time fishing these waters. It may, in fact, be the case that the less experienced anglers, who make up a small but not negligible portion of the weighted mean sample, may pull the results away from the objective model results. Although alternative specification of contracting the subjective data set could have been attempted, like removing all anglers who have only been fishing for a certain period of years, or taken less than a certain amount of trips, these tests were not conducted as cut-offs would be arbitrary and not informed by any *a priori* assumptions.

Hypothesis tests are applied to the results of the objective model and the models of the newly created samples and presented in table 3.11. Comparisons are made between the objective model results and the mean sample model results, as well as between the objective model results and the weight mean model results. The null hypothesis that $\beta_1 = \beta_2$, where β_1 is the estimated parameter of a particular variable from one model and β_2 is the estimated parameter of the same variable estimated from a different model, can be rejected if P is less than 0.05. In the comparison between the objective and mean model result, the hypothesis tests indicate that $\beta_1 = \beta_2$ can be rejected for half of the variables. This result is slightly better than the earlier comparison between the objective model results and two subjective model results. Although it should be noted that the absolute difference between the estimated parameters is much larger when comparing the objective model results against the mean model results as opposed to the objective model results against either of the subjective model results. The results of the comparison between the coefficients of the objective and weighted mean models reveal that for all but one variable we fail to reject the null hypothesis. This result indicates a level of similarity that is not found between any other two models estimated within this paper and may suggest that the results of models applied to samples of objective data may be similar to the results of models applied to samples of data giving heavier weight to frequent users.

Table 3.11: Equality of Coefficient Hypothesis Testing

Variable	Objective versus Mean	Objective Versus Weighted mean
Access at Site	0.599	0.162
Size of Fish	0.000*	0.894
Quantity of Fish	0.074	0.009*
Local Services	0.001*	0.620
Travel Cost	0.022*	0.804
Access Online	0.488	0.760

P-value reported, * denotes significance at 5% level

3.6 Discussion and Conclusion

The prevalence of objective data used in the recreational, environmental and hedonics literature could lead to biased estimates. It has been argued (Puto 1987; Singh 1988; Poor et al., 2001; Artell et al. 2013) that economic agents act based on the perception of the bundle of attributes a good (site) possess. Additionally, the use of objective data instead of subjective data may lead to poor policy development and implementation if there is dissonance between objective measure and users' opinion. It is then worth assessing if the objective data used are a reasonable substitute for the perceptions it is believed decisions are based on. This paper has compared two contrasting sources of data for revealed preference discrete choice analysis; objective and subjective site attribute ratings, to determine if objective data is indeed a reasonable substitute when subjective data is unavailable.

RPL models were applied to both sources of data resulting in three different models; an objective model, a comparable 'reduced' subjective model, and an extended subjective model. The reduced subjective model is a direct replication of the objective model in terms of variables. The extended subjective model incorporates the variables excluded from the reduced subjective model and alternative specific constants. Parameter estimates were used to compute willingness to pay for an increase in site attributes, as well as compensating variation for the closure of each site.

The results reveal that both subjective models outperform the model based on objective data; a finding in accordance with Adamowicz et al. (1997). For all but one variable the direction of estimated parameters is the same across all three models. However, the magnitude of the parameters differs substantially, with hypothesis testing demonstrating that most coefficient estimates are not statistically equivalent.

The confidence intervals of the willingness to pay estimates overlapped in only one of the variables found across all three models. However, some differences are to be expected. The extended subjective model includes variables that are not present in the other two models. In many cases this should reduce omitted variable bias, which in turn affects parameter estimates and therefore WTP estimates. The compensating variation estimates demonstrate that the rating of the sites, in terms of how much it would cost to compensate an angler for a site's closure, remains similar regardless of the source of the data. In contrast to Adamowicz et al. (1997)'s finding, the CV estimates for the objective data indicate that a greater compensation would need to be paid for site closure. Adamowicz et al. (1997) state that, in the case of their data, the higher CV for a site closure is due to the fact that, on average, the subjective data had a higher rating. In the instance of the current data, the objective data had the higher rating and the higher CV.

Comparison between the results of the objective data and the mean and weighted mean models seem to demonstrate that the objective data, based on management perspective, is most closely aligned to the anglers who fished the sites most often. This has an intuitive appeal as one would expect the management to have a similar view of the sites as those who frequent it most often. It may also suggest that the two data sets are fundamentally using the same criteria to value the sites. It could be the case that this type of objective data is an appropriate substitute for avid angler data but may be less suitable for data of those anglers who have spent less time at each site.

For the purpose of practical application, we find that the welfare estimates presented from the results of the objective model are in many respects similar to the results of the subjective data. The direction and significance of most parameter estimates are the same across the models, as are the willingness to pay estimates for a marginal change of a site attribute. Additionally, the ranked order of the CV for a site closure was almost identical across the three main models. The real difference between the objective and the subjective was the magnitude of the parameter estimates which in turn dictate the magnitude of the welfare estimates. In most cases the objective estimates were higher than the subjective estimates; often to a degree that meant the objective parameter estimates could not be considered to be statistically similar to the subjective estimates. The consequence of this on policy may be nuanced but there is a consistency between the objective and subjective results that could result in the similar policies being implemented; both data sets suggest the same attributes are positive or negative and the ranked order of site values in terms of CV are the same. However, estimates based on objective measure, as used here, could result in an overly generous estimate of the value placed on coarse angling within Ireland. It is also important to reiterate that, because of the sampling techniques used, the sample may over represent the keenest anglers and as such these results may not be representative of the national view.

The benefits of the subjective data are not to be overstated; Hynes et al. (2008) cautioned that while they used subjective ratings in their site choice analysis doing so meant that "there could be a potential trade-off between possible bias (if the use of subjective measures leads to endogeneity) and a loss of efficiency (if the loss of information from moving from the individual to some sort of average or objective measure is important)" (Hynes et al. 2008, P. 1016). The authors suggest that the direction of the possible bias will depend on whether the respondent overestimates or underestimates the true value of the quality of the site attribute. This bias may be low in cases where respondents are very familiar with the good.

Chapter 4

Comparing Alternative Approaches to Dealing with Missing Data in Revealed Preference Site Choice Models

4.1 Introduction

Missing data are a common occurrence in recreational revealed preference site choice surveys. There are numerous techniques for dealing with these missing data. As such, analysts are often faced with the challenge of selecting the most appropriate method for their data. This selection is a non-trivial matter. The literature comparing missing data techniques has, so far, demonstrated that all techniques can cause some level of bias, but some techniques will cause substantial bias (Downey and King 1998; Ali et al. 2011; Zhu 2014; Nakai et al. 2014 etc).

When choosing a missing data technique, numerous aspects of the analysist's data must be considered. As pointed out by Ali et al. (2011), the size of the sample, the proportion of missing data, the number of modelled variables, the correlation between the missing and observed variables, and the association between the dependent variable and all other pertinent variables can impact the estimates of a missing data method.

The research to date seems to point to some method of multiple imputations being the most appropriate for dealing with missing data (Shrive et al. 2006; Ali et al. 2011; Zhu 2014; Nakai et al. 2014). However, there is limited research on techniques for choice data, with none using multiple bias measures comparing results against known parameters, as has been done for other data types.

In this article, four techniques for analysing recreational site choice data with missing attribute ratings are compared using a variety of bias measures. The four techniques compared are complete case (CC) analysis, where only observations with full information are used, per person mean substitution (PP) where the average rating of all respondents for a particular attribute is used as a substitute for the missing data, per observation mean substitution (PO), where the average of all observations[12] are used to replace the missing values and multiple imputations (MI) where observed data is used to predict values for the missing data over a specified number of imputations. These techniques have been chosen as they have been previously used in recreational choice data under the implied assumption that they do not create bias.

To the best of the authors' knowledge no paper has attempted to compare the commonly used techniques of CC analysis, mean imputation and MI for choice data. There is also no paper that has made a rigorous attempt to compare any methods using known parameters and used a range of bias metrics, as have been done for many other types of data. As a number of studies using choice-based data have employed CC (Whitehead et al. 1998; Deely et al. 2018), mean imputation (Hynes et al. 2008; Hanley et al. 2001; Deely et al. 2019) and multiple imputations (Steimetz and Brownstone 2005), a comparison of these techniques may be of benefit to analysts of this type. More relevantly to the recreational choice literature, we also compare the welfare estimates produced by each of these techniques. The results of these comparisons shed light on the best practice analysts should follow in the likely event of missing data and highlights the drawbacks that are associated with commonly used techniques for dealing with missing data in revealed preference site choice modelling exercises.

The remainder of this paper is structured as follows, section two describes some of the terminology of missing data mechanisms and how they work when applied to choice data. Section three discusses the previous literature comparing missing data techniques. Section four describes the data, and the missing data generation strategy. Section five present the tests of comparison. The results are presented in section six. Section seven discuss the results and their impact on policy formation. Finally, section eight gives a brief conclusion and some direction for possible future works.

[12] Each respondent may have different numbers of observations, this will depend on the number of recreational trips they took during the survey period.

4.2 Missing Data Mechanisms and Methods

Missing data is generally organized into one of three categories depending on the mechanism of its missingness. In theory, these categories should play a fundamental role in the selection of a method for dealing with the missing data. In practice, the mechanism of the missing data is largely untestable and, often, the mechanism is not the same for all individuals or even all missing variables for the same individual. The three categories, missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR), can, in some sense, be thought of as on a continuum, going from the category with the strongest assumption MCAR to the category with the weakest assumption MNAR. As researchers may be unsure where their data fit on this spectrum, they may be unsure how to deal with their missing data.

The mechanism of missingness is simply why the data is missing. For MCAR data the reason for the missingness must not be correlated with either observed data or unobserved data. It is important to differentiate between the reason for the missing value and the value itself. The mechanism for the missingness may be random, say an individual simply forgetting to fill in an answer, but the missing value itself is not random and may be correlated with other answers given by the respondent. The missing value may be imputable from the observed data. Following Allison (2001) the MCAR assumption can be described as:

$$\Pr\,(R_z = 1 |\, X, Z) = \Pr\,(R_z = 1) \tag{4.1}$$

Where R_z is a binary indicator, which has the value of 1 if data is missing on Z. X is a vector of observed variables and Z is the missing value.

Given that MCAR is predicated, in part, on the assumption that the missing value is not correlated with the missingness mechanism and the obvious fact that the missing value is missing, the MCAR assumption cannot be tested for.

The MAR assumption is somewhat less restrictive; for MAR to hold the missingness mechanism must not be dependent on the missing value itself but can be correlated with the observed data. This can be formally stated as (Allison, 2001):

$$\Pr(R_z = 1|\ X, Z) = \Pr(R_z = 1|X) \tag{4.2}$$

For the same reasons as the MCAR assumption, MAR cannot be tested for. However, under the assumption that the data is either MCAR or MAR it is feasible to test which category the missing data falls into. This may play a fundamental role in the technique applied to data with missing values. Under the MCAR assumption, both the CC analysis and MI can provide unbiased parameter estimates (Schafer and Graham 2002). However, if MCAR does not hold but MAR does then CC analysis will result in biased parameter estimates, whereas MI will not.

The final category, NMAR, assumes that the missingness is, at least, a function of the missing value. A commonly given example is income. Individuals with extremely high or low levels of income will have a higher probability of not answering questions on income. In this case, the missingness is, to some degree, predictable based on the missing value. NMAR are the most problematic for analysis as most standard techniques are thought to produce biased parameter estimates in its presence. Under the NMAR assumption, the missing data mechanism itself may be modelled as part of the estimation process (Allison 2001).

4.3 Literature Review

The literature comparing missing data techniques is extensive, spanning numerous disciplines including health research (Ali et al. 2011), transport (Li et al. 2013), psychology (Hawthorne et al. 2005 and Shrive et al. 2006), recreational economics (Whitehead 1994), methodological issues like percentage of missing data (Downey and King 1998; Nakai et al. 2014), proof of concept (Washington et al. 2014) and, routinely, the missing data mechanism (Zhu 2014; Ali et al. 2011)). Generally, one of two types of data have been used to compare missing data techniques, simulated (Zhu 2014; Nakai et al. 2014) or real data (Downey and King 1998; Ali et al. 2011). When simulated data is employed, a series of data sets are generated where the parameter coefficients are prespecified. The analyst then deletes some portion of the data, according to the set of criteria they wish to explore, e.g. percentage of missing data and/or missing data mechanism. Missing data methods like mean imputation, CC or MI are applied to the data sets with the missing values followed by the statistical models. The results of these models are then compared, usually using a variety of bias measures, to the prespecified parameters.

A recent example of the simulated data method is Zhu (2014), who use simulated longitudinal data in order to compare four methods of handling missing data; CC, mean substitution, MI and last observation carried forward (LOCF), where the missing value is replaced by the last corresponding observed value. The simulated data have five time-points, 100 subjects and were randomly generated 1000 times. These datasets were replicated, and the missing data were randomly generated with a set of criteria which included percent missing, values of the slope and the missing data mechanism. Zhu (2014) reports that no one method performs best under all situations. CC performs best when the missing data mechanism is MCAR, whereas MI performs best when it is MAR. The author strongly recommends the discontinuation of the practice of LOCF due to the likelihood of this method producing biased parameter estimates.

Similarly to Zhu (2014), Nakai et al. (2014) compared CC, mean imputation, MI and LOCF using 1000 simulated data sets. However, they focus on varying the percentage of missing data from 5% to 50% under a MCAR missingness mechanism. The authors write that the MI method is most effective; performing well even under large percentages of missing data. Like Zhu (2014), Nakai et al. (2014) demonstrate that LOCF can produce biased parameter estimates.

The alternative approach to using simulated data is to use real data. In this case, real data is collected and all responses with missing data are removed, leaving a full or "complete" data set. Statistical models are applied to the "complete" data set to get "true" parameter estimates. The analyst then deletes some portion of the data, according to the set of criteria they wish to explore. Missing data techniques are applied followed by statistical models. The results are then compared to the “true” parameters.

This method was conducted by Downey and King (1998), to compare two forms of mean imputation, in relation to their ability to generate precise measures of attitude using the Likert method. The Authors created 49 different datasets with varying percentage of respondents (5% to 35%) having varying percentages of missing items[13] (10% to 70%). The paper compares person mean substitution, where the mean of the person’s completed responses is used as a substitute for the missing items, to an item mean substitution, where the mean of the person’s items is used. When the percentage of respondents with missing data was 20% or below and the percentage of missing items were also 20% or below both techniques provided good results. As the percentage of missing items increased, item mean substitution performed much better than person mean substitution.

Ali et al. (2011) also used real data to compare CC, mean substitution and two forms of MI, the first was MI without using the dependent variable and the second was MI with the dependent variable. Ali et al. (2011) used a sample of 5,443 cases with full survival time data from breast cancer. One hundred datasets with missing data were created under the MAR assumption and a further 100 under the MCAR assumption. The authors found that all four techniques were reasonably robust. The confidence intervals under CC are larger than for the other techniques but there was no systematic over or underestimation of the variables of interest. The confidence intervals for both the mean substitution and MI without the outcome variable were smaller than the CC but they tended to over or underestimate parameters. Ali et al. (2011) report that MI with the dependent variable performs better than the other three techniques used. However, the authors state that it was not possible to definitively rank the techniques as their appropriateness depends on a multitude of factors including missingness mechanism, percent of missing data and correlation between variables.

[13] An item is a variable for which data has been collected.

Both the use of real or simulated data have their benefits and drawbacks when comparing missing data techniques. Simulated data allows the analyst to compare the ability of missing data techniques to deal with differing parameter slopes with relative ease. The use of simulated data also means that an entirely new data set is created each time, which enables comparison over a more varied range of data sets. However, as simulated data is often much less complicated than real data, it may demonstrate comparison in an optimal situation which may not be as informative as real data applications for those who wish to apply it to their own data.

Comparison between missing data techniques for choice data has been somewhat limited. There have only been a few studies that have compared techniques on choice data, but these comparisons have been somewhat limited in either, the number of techniques being compared, the ways bias has been tested or, have been focused on very specific variables. For instance, Sanko et al. (2014) look at the impact of missing income values on estimation results from choice models. Using two case studies they compare sample mean imputation, in which they estimated separate values for coefficients for those with and without missing values against a single imputation and against a latent income variable approach. Interestingly, income values were imputed for those with full information and those with missing information. This allowed imputation of precise amounts of income as opposed to the collected categorical data. These four techniques were applied to a sample of intra-mode commuters who completed a stated preference questionnaire and car owners who completed a revealed preference questionnaire; the former having 12.5% missing data and the latter having 36.8% missing data. The comparison metric measured goodness of fit via the log-likelihoods. The authors found that mean imputation and separate analysis performed best for the reported income data, but single imputation and using a latent income variable performed best for the unreported income variable. Sanko et al. (2014) note that for their case studies, little was gained from the latent variable approach, which is much more computationally expensive, in comparison to single imputation. In the case of Sanko et al. (2014), no "true" parameter estimates were established, and bias was not measured for any of the techniques. This makes it difficult to assess how well the techniques did in generating unbiased coefficients.

Steimetz and Brownstone (2005) used data on Southern Californian motorists who could choose between two free-flowing toll lanes with dynamic pricing or the main lanes to measure motorist value of time. Missing data, related to time savings, were present in the data set. To address this, imputations were used. The first was a single imputation technique and the second was a multiple imputation technique using 200 imputations. The comparison shows "sharp" (as measured by smaller t-stats) results for the single imputation technique suggesting that the single imputation technique does not include enough uncertainty about the estimated parameter following imputation. Interestingly, the value of time estimates from the single imputation technique lies in the 25th percentile of the multiple imputation technique. This demonstrates how far a single imputation result may be from the mean of the imputation distribution. Of course, the single imputation estimation could lie on the very extremes of this distribution without the knowledge of the analyst.

Other papers have compared augmented or novel imputation techniques with more traditional techniques. In one such paper, Washington et al. (2014) propose a Bayesian imputation technique for missing data of non-chosen attributes values. The Bayesian priors are formed using the complete data. Applying this technique to a revealed preference household travel survey, the authors compare the Bayesian procedure to network skims which are generated based on the "cheapest" route between two points of interest. The results show that it is feasible to construct the missing values using their Bayesian method but using skim values to calibrate the model may be less than optimal.

Using simulated multiple-choice questions Wolkowitz and Skorupski (2013) describe and implement a multiple imputation technique based on estimates from a multiple-choice method model. Missing values are simulated missing at a rate of 16.5% using a MCAR, MAR and MNAR structures, creating three datasets, one for each missing data mechanism. Using a single measure of bias for CC and MI (with 100 imputed data sets), the levels of bias are calculated by comparison to the "true" values. For the data sets based on MCAR and MAR, CC and MI both produced little bias. However, under the MNAR condition, MI performed best.

Although many analysts have compared missing data techniques across a wide variety of data types, to the best of the authors' knowledge no paper has attempted to compare the commonly used techniques of complete case analysis, mean imputation and multiple imputation for choice data. There is also a gap in the literature with respect to a rigorous comparison of any methods using known "true" parameters and a range of bias metrics for choice data, as have been done for many other types of data. As a number of studies using choice-based data have employed CC (Whitehead et al. 1998; Deely et al. 2018), mean imputation (Hynes et al. 2008; Hanley et al. 2001; Deely et al. 2019) and multiple imputations (Steimetz and Brownstone 2005), a comparison of these techniques may be of benefit to analysts of this type. This paper produces such a comparison filling the gap in the literature and providing a resource for choice analysts who face the almost ubiquitous problem of missing data.

4.4 Data and Simulated Data

4.4.1 Survey Design and Data Collection

The revealed preference data used in this paper comes from a survey of coarse anglers that attempts to determine the site choice preference of the respondents based on a series of site related attributes. The initial stage of the survey design examined which site attributes were important to coarse anglers. This process involved exploring the relevant literature, focus groups with coarse anglers and discussions with the coarse angling experts at Inland Fisheries Ireland, the state agency managing Ireland's recreational fisheries. The attributes of interest are; average size of fish (*Size),* average quantity of fish (*Quantity),* encounters with other anglers (*Encounters),* level of services at nearest town/village (*Services),* variety of fish species (*Variety),* and accessibility to the point from which the respondent fished (*Access).*

The next step was to create the researcher defined choice set. The choice set comprises five coarse angling sites located in the Cavan and Leitrim area of the Republic of Ireland; Garadice (Leitrim), Killykeen Forest Park (Cavan), Eonish (Cavan), Dernaferst (Cavan) and Church Lake (Cavan). The sites that compose the choice set were selected because they are thought to be feasible alternatives, have ostensibly different levels for each of the attributes of interest and all had a reasonable chance of being visited by each respondent. A pilot study ran from the 28th of July to the 5th of August 2016. After which, the onsite collection ran from the 6th of August to the 7th of November and the online survey ran from the 6th of August to January 15th, 2017.

Each respondent was asked how many times they had visited each of the five sites in the 12 months prior to completing the survey. They were then asked to rate each of the sites on the attributes of interest using a five-point Likert scale. 105 individuals took 2,190 trips to one of the five sites in the 12 months prior to completing the survey. The 2,190 trips equate to 10,950 site choice observations, five for each choice occasion. However, not every individual rated all the attributes for each site. Some individuals were unable or unwilling to rate a site/sites on any of the site attributes, suggesting they had never attended the site/sites. This creates about 25% missing attribute values. Other individuals were unable or unwilling to rate some of the site attributes for a site they stated they had visited previously, creating about 1.5% additional missing attribute values.

This type of missing data is common amongst the literature (Hynes et al. 2008; Hanley et al. 2001; Deely et al. 2018; Deely et al. 2019) and is the impetus for the current paper. It is impossible to know what the true parameter estimates would have been had the data been complete. As such, comparison of the missing data methods alone, without the benchmark of "true" parameter estimates, may not provide much value to the interested analysist. Consequently, missing data were created from the responses where complete information was available.

4.4.2 Simulating Missing Data

The analysis begins with the observations where a complete ranking of the sites is available. Each respondent has between two and five sites in their choice set, with a complete rating of all attributes. This "full" data set is comprised of 75 individuals, 1864 choice occasions and 7377 observations. A conditional logit model is applied to this data set to get the "true" parameter estimates. From here, a subset of the data is randomly generated missing.

The subset of the data used in the missing data generation process are the sites that an individual did not visit during the survey period but had rated the site attributes. This comprise 1308 observation or 18% of the data set. This subset was chosen under the assumption that real world missing data is generated by individuals who know of the site but choose not to visit it. As such, the data where individuals choose not to go to the site but have rated the attributes would be most similar.

Three percentages of missing data were created; 6%, 12% and 18%[14]. For the 6% and 12%, random numbers were generated for every site an individual had not visited. The random numbers were sorted and ranked. For the 6% missing data set, the data associated with the top 6% of the random number were set to missing, for the 12% missing, the top 12% were set to missing. As the random numbers were generated per site per individual and, as each individual may have different numbers of observations, the number of missing observations could differ between each simulated missing data set. For example, if an individual had five observations in the data set because they took one trip to the sites of interest during the survey period and one of their sites was selected to be eliminated from the data set. The data set lose one observation. However, if an individual made 20 trips to the sites of interest and one of their sites was selected the data set would lose 20 observations.

The missing data generation process was repeated 100 times each for the 6% and 12% missing data. The 18% missing data is at the extreme, where all rated

[14] These values represent 6%, 12% and 18% of the "complete" data set but are also 33%, 66% and 100% of the subset of the data set that the missing data was generated from.

but unvisited sites are changed to missing data. In this case, CC, PO, PP will only generate one set of results. However, as MI is generated through simulation each set of results from the same data set may be different. Consequently, the MI process was conducted 100 times on the 18% missing data set as well.

In reference to the missing data mechanism underlying this newly created missing data, it is impossible to say with certainty which category these data fall into. Although random number generators are used to produce the missing data they are taken from a select subsample of observations. There is a reason, personal to each respondent, why certain sites have not been visited. It may be the case that the cost of visiting these sites is too great regardless of quality improvements which may suggest, but not confirm, that the missing data are MAR. The respondent may have some special access needs or may only be concerned with a subset of the attributes of interest, in which case some attributes will be NMAR while others could be MAR. Or, it could be the case that the site was not visited because of a combination of the travel cost and the perceived value of the attributes of interest, signifying that all attributes for that site lie somewhere between MAR and NMAR.

4.4.3 Site Choice Model

McFadden's (1973) random utility model (RUM) states that, on any given choice occasion, an individual will choose the alternative that maximises her utility. This can be written as:

$$u_{in} = V(X_{in}, y_n - p_{in} \mid \theta_n, z_n) + \varepsilon_{in}$$

$$= V_{in} + \varepsilon_{in} \tag{4.3}$$

Where u_{in} is the utility received by individual *n* from choosing site *i*, V is the indirect utility function, X_{in} is a vector of perceived attributes, y_n is individual n's income, p_{in} is the travel cost, θ_n is a vector of individual-specific characteristics and z_n are individual specific covariates. The stochastic error term ε_{in} is unknown to the modeller and assumed to be independent and identically distributed (IID) extreme value type 1. The probability of an individual choosing a site, form choice set J, can be written as:

$$Pr(i) = Pr(V\ (x_{in}, p_{in}\ |\theta_n, z_n) + \varepsilon_{ij}) \geq V(x_{jn}, p_{jn}\ |\theta_n, z_n) + \varepsilon_{jn})\ \forall\ j\ \in\ J \quad (4.4)$$

Showing that, the probability of choosing site *i*, for individual n, is equivalent to the probability that site *i* will offer individual *n* greater utility than any other site in choice set *J*.

When the distribution of the error terms is IID from an extreme value distribution, the RUM model takes the form of a conditional logit (CL) (McFadden 1973), where the probability of choosing site *i* is given as a logit with scale parameters μ.

$$\text{Pr (i)} = \frac{\exp\ (\mu V_{in})}{\sum_{j=1}^{j} \exp\ (\mu V_{jn})} \quad (4.5)$$

Although more sophisticated models that allow for heterogeneity in preferences are preferred; for the purpose of testing for the impact of missing attribute data the CL is employed.

Willingness to pay (WTP) estimates are calculated for each of the attributes of interest. These estimates demonstrate how much the average respondent is willing to pay for a one unit increase in a site attribute. Following Train (2009) the WTP estimates are calculated using the formula:

$$WTP = \frac{\hat{\beta}}{\widehat{-\beta_{tc}}} \quad (4.6)$$

Where $\hat{\beta}$ denotes the coefficient of one of the attributes of interest and $\widehat{\beta_{tc}}$ is the travel cost coefficient which is assumed to be equivalent to the marginal utility of income. Confidence intervals are calculated using the Krinsky-Robb method with 5,000 draws (Krinsky and Robb, 1986).

4.5 Missing Data Techniques

Four techniques for dealing with missing data are compared within this paper. The first is Complete Case (CC) analysis (also known as listwise deletion).

Traditionally CC restricts the modelling data to individuals who have no missing data. The major drawback with such a technique is the loss of data. It is often the case, in the presence of missing data, only a fraction of the variables of interest are missing for any one individual. As most analysts would prefer to keep the data that is available to them, CC may represent an unnecessary deletion of otherwise important information.

CC analysis for site choice models is somewhat different to many other data sets. Site choice data is stacked. For each choice occasion in this dataset, an individual has five observations, one for each of the sites in the choice set. CC analysis, when applied to site choice models, eliminates observations rather than individuals. If an individual did not rate any attribute for a site, the site is removed from their choice set. This can lead to a large reduction in the number of observations but will only reduce the number of survey respondents if an individual has only rated one site. However, this restriction of the choice set by deleting observations may have an impact on the size of the travel cost parameter (Deely et al. 2018, Peters et al. 1995) which, in turn, may make it much less efficient for welfare estimates even if it has a relatively small bias for other parameter estimates.

CC analysis has the somewhat obvious benefit of not imputing data. There is generally little ambiguity over the validity of the data as no external source or technique has been used to replace the missing data. CC has also been demonstrated to not produce biased estimates under MCAR and, under certain conditions, MAR (Schafer and Graham, 2002).

When sites are deleted and CC is employed there is a reduction in the modelled choice set. This may cause some level of bias which we may be able to observer in the comparison that follows.

The second technique is mean imputation. Mean imputation is a process where the missing data are replaced with the average of the observed data. Two forms of mean imputation are employed, the first uses the average of all individuals and the second uses the average of all observations. In many respects mean imputation is the simplest method for dealing with missing data; it preserves all the observed data and because the data set is "complete” all traditional methods of modelling data are applied as easily as when no missing data are present. However, traditionally no indicator is associated with the replacement data. As such, the mean imputed data is treated as if it were part of the original data set. This may disregard the inherent uncertainty associated with any missing data techniques.

The final technique for dealing with missing data used in this paper is multiple imputations by chained equations (MICE). MI, first proposed by Rubin (1987), is a method by which the observed data is used to predict values for the missing data over a specified number of imputations. In the case of the current data, ordinal regressions are used to predict the values of the missing data. Random draws are then taken from the simulated error distribution. The random errors are then added to each individual's predicted values. Parameter estimates, based on these imputed data sets, are approximately unbiased (Allison 2001). However, standard errors (SEs) will be underestimated. In order to address this issue, the data generation process is repeated a number of times, each time creating a new "full" data set. The parameter estimates are then a simple average of the results from models applied to each "full" data set. The SEs, on the other hand, use the variance within each data set and the variance between each data set. The formula for calculating these SEs is (Allison 2001):

$$\sqrt{\frac{1}{M}\sum_{k=1}^{M} s_k^2 + \left(1+\frac{1}{M}\right)\left(\frac{1}{M-1}\right)\sum_{k=1}^{M}(\widehat{\beta}_k - \bar{\beta}_k)^2} \tag{4.7}$$

where *M* is the number of data sets, s_k is the standard error of β_k in the *kth* data set and β_k is the parameter estimated in the *kth* data set.

The chained equation method using the MICE algorithm performs the imputation in a slightly different manner. However, the theory and method for combining the results are the same. Initially, the missing values are filled at random, then the variable with the least missing data is regressed on the other variables in the imputation model using only the observations with observed data for that variable. The missing values of the first variable are then replaced from draws from the posterior predicted distribution of this variable. The process then begins with the second variable with missing data.

This time, however, the imputed values of the first variable are included in the estimation process. This process is repeated for multiple cycles before one "full" data set is created. It is also worth noting that although the analysis presents the variables, *Size, Quantity,* etc the variables being imputed were *Size at Garadice*, *Size at Killykeen,* etc, in much the same way as longitudinal data would impute variables at different time points. This allows for imputation based on a more complete knowledge of how individuals perceived the same attributes at other sites.

During the imputation process a problem with one of the variables was encountered. The variable *Access at Garadice* would not update[15], meaning it was stuck at stage one of the imputation process; filled by a random number between one and five[16]. As the imputed data of other variables is, in part, based on the imputations of this variable this posed a problem, potentially for all the variables. Although not presented here, three different methods were compared, over 25 iterations for all percentages of missing data. The first was to leave the variable as it was originally imputed, the second was to impute the variable *Access at Garadice* with a separate imputation model and replace it in the first imputed data set. The final method was to leave out *Access at Garadice* from the imputation model, generate all other variables and generate *Access at Garadice* separately and then combine both data sets. The benefit of the first two methods is that there is more information to impute the other variables from, particularly seen as *Access at Garadice* was one of the variables with the lowest percent of missing data. However, there is reason to believe that the non-updated random number could induce bias in the imputation of other variables.

[15] Royston (2004) suggest using the persist option in Stata for “difficult” variables such as these. However, when the variable will not update over any iteration, Royston suggests dropping the variable from the model. No resolution seems to be available when the variable is integral to the analysis model.

[16] The precise reason for the problem imputing Access at Garadice was not fully established. A process of elimination suggests that this problem arises due to the interaction between the variables Access at Garadice and Quantity at Garadice. The imputation model runs without fault if either of these variables are omitted. Correlation tests were run to see if these two variables were highly correlated or looked different from the other variables in the imputation model. However, no obvious difference could be found.

The results indicate that the method that produced the least amount of bias over all percentages of missing data was the second method, where the variable *Access at Garadice* is in the imputation model but is later replaced by *Access at Garadice* generated from a different imputation model. However, this may not always be the case. *Access at Garadice* has a small portion of the missing values. If a variable had large percentages of missing data, it may cause larger amounts of bias in the other imputed variables.

A fourth method, not attempted here, is to impute the missing variable once, using some single imputation method, replace it in the data set that will be used for imputation, treat it as having full information and impute the data set as many times as the analyst deems appropriate, then impute this missing value that will not update separately. This method may reduce the bias caused by the random numbers while keeping all the original data.

4.5.2 Measures of Bias

Five metrics are used to determine which method of imputation performs best: bias, root mean squared error, hypothesis testing, type one errors and type two errors.

Bias is the difference between the average estimated parameter and the “true” estimated parameter ($\hat{\beta}_k - \beta_k$). The root mean squared error (RSME) is an extension of the bias measure in which the bias measures are squared, which eliminates any cancelling out that might be found from some results being greater than the “true” parameter and some results being less than the “true” parameter. It can be calculated as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i-1}^{n}(\hat{\beta}_{ki} - \beta_{ki})^2} \tag{4.8}$$

Where *n* is the number of simulated data set, *i* denotes a specific data set and $\hat{\beta}_{ki}$ is the estimated parameter for variable *k* and β_{ki} is the true estimate.

Hypothesis tests are applied to test for significant difference between estimated and real parameter estimates. Following Clogg, (1995) the test statistic is

$$Z = \frac{\beta_k - \hat{\beta}_{ki}}{\sqrt{SE\beta_k{}^2 + SE\hat{\beta}_{ki}{}^2}} \tag{4.9}$$

$SE\beta_k$ is the standard error for the "true" parameter estimate for variable k and $SE\hat{\beta}_{ki}$ is the standard error for the estimated parameter for variable k in simulated data set i.

This test has an advantage over the bias or RMSE measures when comparing methods for dealing with missing data. Missing data techniques should, as a general rule, avoid overly precise parameter estimates i.e. small confidence intervals. This is simply due to the inherent uncertainty associated with missing data. The hypothesis test uses the standard error of the coefficient when calculating the Z score, as such variables that have small standard errors, are more likely to reject the null hypothesis of statistically equivalent parameters.

Substantial importance is often placed on the significance of the variables. Consequently, type one and type two errors are reported. In relation to this analysis, a type one error is when a parameter is not statistically significant at the 5% level in the "true" parameter but is in the estimated model. A type two error is when a parameter is statistically significant for the "true" model but not significant in the estimated model.

4.6 Results

The results of conditional logit models using four different techniques for dealing with missing data are presented in table 4.1A for 6% missing, 4.1B for 12% missing and 4.1C for 18% missing. The second column is the "true" parameter estimates, the third column shows the average results of the CC analysis, the fourth shows the average results of the PP mean substitution, the fifth shows the PO mean imputation and the final column shows the results of the MI analysis. For simplicity, the model has been confined to the parameters of interest and alternative specific constant. For each method it is relatively simple to create a model that includes interactions terms that analysts may find appropriate, however, this should be considered when creating the imputation model.

Table 4.1A: Average Parameter Estimates From Conditional Logit Model at 6% Missing

Variable	"True" Parameter estimate	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Travel cost	-0.063	-0.057	-0.064	-0.064	-0.062
	(0.004)***	(0.005)***	(0.005)***	(0.005)***	(0.005)***
Access	0.108	0.064	0.072	0.081	0.075
	(0.029)***	(0.029)**	(0.030)**	(0.021)***	(0.033)**
Size	0.046	-0.019	0.042	0.056	0.072
	(0.050)	(0.053)	(0.052)	(0.052)	(0.062)
Quantity	0.156	0.154	0.159	0.167	0.138
	(0.045)***	(0.045)***	(0.046)***	(0.046)***	(0.056)***
Variety	-0.100	-0.076	-0.100	-0.156	-0.069
	(0.058)*	(0.059)	(0.058)*	(0.057)***	(0.076)
Services	-0.206	-0.174	-0.196	-0.205	-0.167
	(0.036)***	(0.037)***	(0.037)***	(0.037)***	(0.043)***
Encounters	0.051	0.047	0.086	0.049	0.047
	(0.041)	(0.041)	(0.041)**	(0.041)	(0.049)
Alternative specific constants					
Killykeen	-0.664	-0.737	-0.711	-0.706	-0.696
	(0.079)***	(0.080)***	(0.079)***	(0.079)***	(0.083)***
Dernaferst	-0.757	-0.779	-0.792	-0.785	-0.798
	(0.101)***	(0.102)***	(0.102)***	(0.103)***	(0.105)***
Church Lake	-0.978	-0.931	-1.044	-1.037	-1.036
	(0.115)***	(0.117)***	(0.114)***	(0.115)***	(0.123)***
Eonish	-1.330	-1.319	-1.324	-1.340	-1.339
	(0.090)***	(0.092)***	(0.092)***	(0.091)***	(0.094)***
Log likelihood	-2045	-1962	-2046	-2043	N/A
AIC	4111	3946	4114	4108	N/A
Observations	7377	6840	7377	7377	44262

Standard deviation given in parenthesis, 1% significance denoted by ***, 5% by ** and 10% by *

Table 4.1B: Average Parameter Estimates From Conditional Logit Model at 12% Missing

	"True" Parameter estimate	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Travel cost	-0.063	-0.052	-0.064	-0.065	-0.061
	(0.004)***	(0.005)***	(0.005)***	(0.005)***	(0.005)***
Access	0.108	0.026	0.044	0.058	0.048
	(0.029)***	(0.030)	(0.030)	(0.030)*	(0.036)
Size	0.046	0.01	0.004	0.069	0.091
	(0.050)	(0.055)	(0.053)	(0.053)	(0.067)
Quantity	0.156	0.153	0.163	0.177	0.146
	(0.045)***	(0.045)***	(0.048)***	(0.048)***	(0.061)***
Variety	-0.100	-0.58	-0.106	-0.200	-0.063
	(0.058)*	(0.059)	(0.057)*	(0.056)***	(0.081)
Services	-0.206	-0.139	-0.188	-0.201	-0.154
	(0.036)***	(0.038)***	(0.037)***	(0.037)***	(0.046)***
Encounters	0.051	0.041	0.106	0.044	0.047
	(0.041)	(0.042)	(0.042)**	(0.042)	(0.052)
Alternative specific constants					
Killykeen	-0.664	-0.802	-0.745	-0.736	-0.718
	(0.079)***	(0.081)***	(0.079)***	(0.080)***	(0.085)***
Dernaferst	-0.757	-0.810	-0.819	-0.811	-0.827
	(0.101)***	(0.103)***	(0.103)***	(0.104)***	(0.109)***
Church Lake	-0.978	-0.877	-1.093	-1.085	-1.066
	(0.115)***	(0.118)***	(0.113)***	(0.115)***	(0.129)***
Eonish	-1.330	-1.342	-1.329	-1.357	-1.348
	(0.090)***	(0.094)***	(0.093)***	(0.092)***	(0.096)***
Log likelihood	-2045	-1813	-2047	-2039	N/A
AIC	4111	3649	4117	4100	N/A
Observations	7377	6069	7377	7377	44262

Standard deviation given in parenthesis, 1% significance denoted by ***, 5% by ** and 10% by *

Table 4.1 C: Average Parameter Estimates From Conditional Logit Model at 18% Missing

	"True" Parameter estimate	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Travel cost	-0.063	-0.042	-0.065	-0.067	-0.061
	(0.004)***	(0.005)***	(0.005)***	(0.005)***	(0.005)***
Access	0.108	-0.039	0.004	0.023	0.027
	(0.029)***	(0.030)	(0.031)	(0.031)	(0.038)
Size	0.046	-0.004	0.046	0.086	0.096
	(0.050)	(0.058)	(0.055)	(0.056)	(0.084)
Quantity	0.156	0.157	0.175	0.200	0.174
	(0.045)***	(0.047)***	(0.046)***	(0.049)***	(0.068)***
Variety	-0.100	-0.015	-0.105	-0.244	-0.050
	(0.058)*	(0.060)	(0.057)*	(0.055)***	(0.091)
Services	-0.206	-0.074	-0.178	-0.195	-0.147
	(0.036)***	(0.040)*	(0.037)***	(0.037)***	(0.050)***
Encounters	0.051	0.037	0.132	0.038	0.074
	(0.041)	(0.043)	(0.042)***	(0.043)	(0.061)
Alternative specific constants					
Killykeen	-0.664	-0.933	-0.791	-0.779	-0.760
	(0.079)***	(0.083)***	(0.079)***	(0.080)***	(0.089)***
Dernaferst	-0.757	-0.864	-0.862	-0.857	-0.868
	(0.101)***	(0.105)***	(0.105)***	(0.106)***	(0.113)***
Church Lake	-0.978	-0.793	-1.151	-1.153	-1.085
	(0.115)***	(0.120)***	(0.113)***	(0.116)***	(0.136)***
Eonish	-1.330	-1.410	-1.337	-1.385	-1.351
	(0.090)***	(0.098)***	(0.093)***	(0.094)***	(0.100)***
Log likelihood	-2045	-1900	-2047	-2041	N/A
AIC	4111	3822	4116	4105	N/A
Observations	7377	6491	7377	7377	44262

Standard deviation given in parenthesis, 1% significance denoted by ***, 5% by ** and 10% by *

The estimated parameters averaged over the simulated missing data sets show some interesting results in comparison to the "true" parameter estimates. There are, in many cases, substantial difference between the true estimate and the estimate from the comparison methods. The travel cost variable is estimated with good precision by all three imputation models, across all percentages of missing data. However, the CC analysis estimate is a good deal closer to zero than the true estimate and becomes more biased as the percentage of missing data increases. This bias has been noted elsewhere (Deely et al. 2018) and is of importance in non-market valuation studies. The bias is derived from the type of sites being removed from the analysis when the CC method is employed on data of this type. On average, the sites that are not visited are further away than sites that are visited, in part because it is costlier to visit them. When these sites are removed important information on how an individual reacts to cost is removed; in essence, truncating the information to closer sites, on average.

All four methods erroneously produced insignificant estimates and performed poorly at estimating the variable *Access* for all but the lowest amount of missing data. However, only PO produced significant average estimates when the "true" parameter estimate was insignificant for any of the variables. It is also interesting to observe that the alternative specific constants, for which there is full information, are biased using each method.

A dominant concern with missing data is the measure of SEs. Tables 4.1A, 4.1B and 4.1C indicate reasonably good approximations of the SEs for all variables. Mean imputation, in particular, is known to bias downward SE which does not appear to be the case here. MI, almost by construct, have the largest SE. One would expect this may influence tests, which fail to reject the null hypothesis less often when larger SEs exist such as the hypothesis test.

Table 4.2A to 4.6C show the results of the bias, RMSE, hypothesis testing, type one errors and type two errors tests. Column two shows the results for CC analysis, column three is the PP, column four is the PO and column five is the MI results.

Table 4.2A: Average Bias at 6% Missing

	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Travel cost	0.006	-0.001	-0.002	0.001
Access	-0.044	-0.036	-0.028	-0.033
Size	-0.027	-0.004	-0.010	0.025
Quantity	0.002	0.004	0.012	-0.017
Variety	0.024	0.000	-0.056	0.031
Services	0.032	0.009	-0.001	0.039
Encounters	-0.003	0.035	-0.002	0.000
Alternative specific constants				
Killykeen	-0.074	-0.048	-0.042	-0.032
Dernaferst	-0.022	-0.035	-0.028	-0.041
Church Lake	0.047	-0.065	-0.059	-0.057
Eonish	-0.010	-0.006	0.010	-0.009
Absolute Average bias	0.291	0.243	0.250	0.285

Table 4.2B: Average Bias at 12% Missing

	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Travel cost	0.011	-0.002	-0.003	0.001
Access	-0.082	-0.064	-0.051	-0.055
Size	-0.045	-0.002	0.022	0.045
Quantity	-0.002	0.007	0.022	-0.010
Table 4.2B – continued from previous page				
Variety	0.042	-0.006	-0.100	0.037
Services	0.067	0.018	-0.005	0.052
Encounters	-0.010	0.055	0.007	-0.004
Alternative specific constants				
Killykeen	-0.139	-0.082	-0.073	-0.054
Dernaferst	-0.053	-0.062	-0.054	-0.070
Church Lake	0.102	-0.114	-0.106	-0.087
Eonish	0.013	0.001	0.028	0.018
Absolute Average bias	0.566	0.413	0.471	0.433

Table 4.2C: Average Bias at 18% Missing

	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Travel cost	0.021	-0.003	-0.004	0.002
Access	-0.147	-0.104	-0.085	-0.081
Size	-0.082	-0.000	0.040	0.049
Quantity	0.002	0.019	0.044	0.018
Variety	0.085	-0.006	-0.145	0.050
Services	0.132	0.028	-0.011	0.059
Encounters	-0.014	0.082	-0.013	0.023
Alternative specific constants				
Killykeen	-0.270	-0.127	-0.115	-0.096
Dernaferst	-0.107	-0.104	-0.099	-0.111
Church Lake	0.186	-0.172	-0.173	-0.106
Eonish	0.081	0.007	0.055	0.021
Absolute Average bias	1.127	0.652	0.784	0.616

The three tables concerning bias, presented in tables 4.2A to 4.2C, demonstrate that as the percentage of missing data increases so does the amount of bias for each technique. CC constantly produces the largest amount of bias with the other three producing similar amounts. PP performs best for 6% and 12% missing data whereas MI performs best for the largest percentage of missing data. Additionally, the increase in bias from one percentage of missing data to the next, is lowest for MI. This may suggest a trend which may have shown even more pronounced results had the percentage of missing data increased further.

Table 4.3A: Average RMSE at 6% Missing

	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Travel cost	0.065	0.001	0.002	0.001
Access	0.052	0.042	0.034	0.040
Size	0.056	0.035	0.035	0.049
Quantity	0.039	0.034	0.037	0.037
Variety	0.048	0.050	0.077	0.054
Services	0.049	0.032	0.029	0.050
Encounters	0.032	0.047	0.035	0.043
Alternative specific constants				
Killykeen	0.098	0.055	0.050	0.038
Dernaferst	0.080	0.050	0.046	0.052
Church Lake	0.105	0.081	0.072	0.069
Eonish	0.070	0.021	0.024	0.026
Total RMSE	0.694	0.448	0.441	0.459

Table 4.3B: Average RMSE at 12% Missing

	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Travel cost	0.012	0.002	0.003	0.002
Access	0.087	0.067	0.054	0.059
Size	0.070	0.034	0.039	0.059
Quantity	0.041	0.036	0.043	0.039
Variety	0.060	0.045	0.110	0.057
Services	0.079	0.033	0.027	0.060
Encounters	0.038	0.064	0.038	0.041
Alternative specific constants				
Killykeen	0.152	0.085	0.077	0.058
Dernaferst	0.096	0.071	0.065	0.076
Church Lake	0.146	0.120	0.112	0.094
Eonish	0.085	0.020	0.037	0.031
Total RMSE	0.866	0.577	0.605	0.576

Table 4.3C: Average RMSE at 18% Missing

	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Travel cost	0.021	0.003	0.004	0.003
Access	0.147	0.104	0.085	0.082
Size	0.082	0.000	0.040	0.056
Quantity	0.002	0.019	0.044	0.030
Variety	0.085	0.006	0.145	0.061
Services	0.132	0.028	0.011	0.061
Encounters	0.014	0.082	0.013	0.030
Alternative specific constants				
Killykeen	0.270	0.127	0.115	0.098
Dernaferst	0.107	0.104	0.099	0.113
Church Lake	0.186	0.172	0.173	0.112
Eonish	0.081	0.007	0.055	0.027
Absolute Average bias	1.127	0.652	0.784	0.673

Note: results of CC, PP and PO are the same as in equivalent bias table as these methods were only used once for complete missing data case

The RMSE, presented in tables 4.3A to 4.3C, produces slightly different results to the bias test with a different imputation method performing best for each percentage of missingness. PP and MI performed almost identically across the all percentages of missing data. Although the PO method performed best at the 6% level, the increase in bias was a lot more pronounced than for the other two methods of imputation. Again, CC performed the poorest at each percentage.

Table 4.4A: Hypothesis Testing at 6% Missing

	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Travel cost	4	0	0	0
Access	0	0	0	0
Size	0	7	14	29
Quantity	0	0	0	0
Variety	31	11	4	30
Services	8	0	0	4
Encounters	7	43	14	8
Alternative specific constants				
Killykeen	0	0	0	0
Dernaferst	0	0	0	0
Church Lake	0	0	0	0
Eonish	0	0	0	0
Total	50	61	32	71

Values show the percentage of times a significantly different estimate is calculated

Table 4.4B: Hypothesis Testing at 12% Missing

	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Travel cost	61	0	0	0
Access	0	0	0	0
Size	1	6	24	40
Quantity	0	0	0	0
Variety	37	9	0	36
Services	30	0	0	7
Encounters	7	63	8	10
Alternative specific constants				
Killykeen	0	0	0	0
Dernaferst	0	0	0	0
Church Lake	0	0	0	0
Eonish	0	0	0	0
Total	136	78	32	93

Values show the percentage of times a significantly different estimate is calculated

Table 4.4C: Hypothesis Testing at 18% Missing

	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Travel cost	100	0	0	0
Access	0	0	0	0
Size	0	0	0	23
Quantity	0	0	0	0
Variety	100	0	0	44
Services	100	0	0	4
Encounters	0	100	0	0
Alternative specific constants				
Killykeen	0	0	0	0
Dernaferst	0	0	0	0
Church Lake	0	0	0	0
Eonish	0	0	0	0
Total	300	100	0	71

Values show the percentage of times a significantly different estimate is calculated

Tables 4.4A, 4.4B and 4.4C show the percentage of times a statistically significantly different estimate is calculated by each of the methods across the 100 simulated data sets. The results show that it is not uncommon for parameters to be estimated statistically different than the "true" value.

Overall, PO seems to perform best based on the hypothesis test metric, having the lowest total percentage for each percentage of missing data. MI performs poorly by this metric. This may indicate that there may be some questionability on the reliability of this technique over different generated complete data sets.

Failure to reject the null hypotheses across metrics concentrate on the same variables, with the variables *Size, Variety* and *Encounters* resulting in numerous instances of statically different estimates. These three variables are also the only insignificant variables (at the 95% level) based on the "true" parameter estimates, with the largest ratio of standard errors to coefficients. It may be the case that there is more heterogeneity in relation to these variables (the two largest standard errors are also amongst these three variables). This may suggest that if data are missing for variables with relatively large standard errors it may have a far great influence on estimates, then those with small standard errors.

Table 4.5A: Type One Errors at 6% Missing

	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Travel cost	0	0	0	0
Access	0	0	0	0
Size	0	1	8	16
Quantity	0	0	0	0
Variety	19	42	82	8
Services	0	0	0	0
Encounters	12	55	18	12
Alternative specific constants				
Killykeen	0	0	0	0
Dernaferst	0	0	0	0
Church Lake	0	0	0	0
Eonish	0	0	0	0
Total	31	98	108	36

Values denote the percentage of times a method committed a type one error

Table 4.5B: Type One Errors at 12% Missing

	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Travel cost	0	0	0	0
Access	0	0	0	0
Size	3	2	16	22
Quantity	0	0	0	0
Variety	10	43	98	5
Services	0	0	0	0
Encounters	13	75	19	8
Alternative specific constants				
Killykeen	0	0	0	0
Dernaferst	0	0	0	0
Church Lake	0	0	0	0
Eonish	0	0	0	0
Total	26	120	133	35

Values denote the percentage of times a method committed a type one error

Table 4.5C: Type One Errors at 18% Missing

	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Travel cost	0	0	0	0
Access	0	0	0	0
Size	0	0	0	3
Quantity	0	0	0	0
Variety	0	0	100	0
Services	0	0	0	0
Encounters	0	100	0	8
Alternative specific constants				
Killykeen	0	0	0	0
Dernaferst	0	0	0	0
Church Lake	0	0	0	0
Eonish	0	0	0	0
Total	0	100	100	11

Values denote the percentage of times a method committed a type one error

Tables 4.5A to 4.5C show the percentage of times a method produced a type one error. Considering that only three variables; *Variety, Encounters* and *Size* had insignificant "true" estimates all methods performed poorly. The PO method performed the worst. CC, which has performed poorly using all other metrics, performed the best using the type one metric with MI performing best of the imputation methods. Surprisingly, the variable *Encounters* committed a type one error nearly as often as the parameter estimates for *Variety. A priori* expectations would be that the variable *Variety*, which was significant at the 90% level would be estimated incorrectly often but *Encounters*, whose coefficient was not close to being significant, would rarely commit a type one error.

Table 4.6A: Type Two Errors at 6% Missing

	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Travel cost	0	0	0	0
Access	30	24	16	29
Size	0	0	0	0
Quantity	2	0	0	22
Variety	0		0	0
Services	1	0	0	0
Encounters	0	0	0	0
Alternative specific constants				
Killykeen	0	0	0	0
Dernaferst	0	0	0	0
Church Lake	0	0	0	0
Eonish	0	0	0	0
Total	33	24	16	51

Values denote the percentage of times a method committed a type two error

Table 4.6B: Type Two Errors at 12% Missing

	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Travel cost	0	0	0	0
Access	87	73	54	72
Size	0	0	0	0
Quantity	5	1	1	31
Variety	0	0	0	0
Services	8	0	0	2
Encounters	0	0	0	0
Alternative specific constants				
Killykeen	0	0	0	0
Dernaferst	0	0	0	0
Church Lake	0	0	0	0
Eonish	0	0	0	0
Total	100	74	55	105

Values denote the percentage of times a method committed a type two error

Table 4.6C: Type Two Errors at 18% Missing

	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Travel cost	0	0	0	0
Access	100	100	100	100
Size	0	0	0	0
Quantity	0	0	0	10
Variety	0		0	0
Services	100	0	0	4
Encounters	0	0	0	0
Alternative specific constants				
Killykeen	0	0	0	0
Dernaferst	0	0	0	0
Church Lake	0	0	0	0
Eonish	0	0	0	0
Total	200	100	100	114

Values denote the percentage of times a method committed a type two error

Table 4.6A, 4.6B and 4.6C show the percentage of times that each method does not correctly produce a significant variable. Although all variables are presented, a type two error can only occur when the "true" parameter estimate is significant. The results show that nearly all type two errors occur for the same variable, *Access*. The reoccurrence of type two errors for the same variable may suggest that certain qualities of a coefficient may make them more susceptible to this type of error.

MI performed poorest using this metric with a large amount of type two errors for *Quantity* as well as *Access*. PO performed best with the lowest amount of Type two errors.

Table 4.7A: Willingness to Pay at 6% Missing, given in €

	"True" Parameter estimate	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Access	1.73	1.12	1.14	1.26	1.21
	(0.75 – 2.80)*	(0.05 – 2.20)*	(0.17 – 2.11)*	(0.29 – 2.22)*	(0.14 – 2.36)*
Size	0.74	0.35	0.65	0.87	1.16
	(-0.80 – 2.34)	(-1.50 – 1.84)	(-0.96 – 2.26)	(-0.74 – 2.46)	(-0.81 – 3.23)
Quantity	2.48	2.71	2.51	2.60	2.24
	(1.13 – 3.93)*	(1.14 – 4.43)*	(1.08 – 3.98)*	(1.18 – 4.05)*	(0.41 – 4.05)*
Variety	-1.59	-1.32	-1.57	-2.43	-1.12
	(-3.44 – 0.24)	(-3.36 – 0.72)	(-3.34 – 0.022)	(-4.17 – -0.68)*	(-3.46 – 1.37)
Services	-3.29	-3.05	-3.10	-3.18	-2.70
	(-4.45 – -2.18)*	(-4.36 – -1.78)*	(-4.31 - - 2.06)*	(-4.31 - - 2.06)*	(-4.05 – -1.35)*
Encounters	0.81	0.84	1.35	0.76	0.77
	(-0.46 – 2.17)	(-0.59– 2.32)*	(0.07 – 2.68)*	(-0.50 – 2.05)	(-0.75 – 2.37)

Confidence intervals given in parenthesis, 5% significance given by *

Table 4.7B: Willingness to Pay at 12% Missing, given in €

	"True" Parameter estimate	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Access	1.73	0.50	0.69	0.88	0.86
	(0.75 – 2.80)*	(-0.62 – 1.73)	(-0.22 – 1.68)	(-0.02 – 1.87)	(-0.28 – 2.10)
Size	0.74	0.052	0.68	1.05	1.49
	(-0.80 – 2.34)	(-2.05 – 2.23)	(-0.94 – 2.37)	(-0.56 – 2.72)	(-0.62 – 3.78)
Quantity	2.48	2.97	2.54	2.71	2.38
	(1.13 – 3.93)*	(1.19 – 4.85)*	(1.06 - 4.04)*	(1.25 – 4.19)*	(0.45 – 4.38)*
Variety	-1.59	-1.10	-1.65	-3.05	-1.02
	(-3.44 – 0.24)	(-3.32 – 1.27)	(-3.36 – 0.13)	(-4.75 – -1.36)*	(-3.60 – 1.58)
Services	-3.29	-2.68	-2.93	-3.06	-2.50
	(-4.45 – -2.18)*	(-4.17 – -1.23)*	(-4.10 - - 1.80)*	(-4.21 - - 1.97)*	(-3.95 – -1.08)*
Encounters	0.81	0.78	1.66	0.67	0.76
	(-0.46 – 2.17)	(-0.79 – 2.47)	(0.43 – 3.01)*	(-0.55 – 1.97)	(-0.90 – 2.50)

Confidence intervals given in parenthesis, 5% significance given by *

Table 4.7C: Willingness to Pay at 18% Missing, given in €

	"True" Parameter estimate	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Access	1.73	-0.94	0.08	0.39	0.44
	(0.75 – 2.80)*	(-1.16 – 3.10)	(-0.86 – 1.04)	(-0.54 – 1.34)	(-0.79 – 1.74)
Size	0.74	-0.85	-0.12	1.32	1.64
	(-0.80 – 2.34)	(-3.63– 1.98)	(-1.36 – 1.64)	(-0.11– 2.83)	(-1.07 – 4.59)
Quantity	2.48	3.80	2.83	2.97	2.90
	(1.13 – 3.93)*	(1.63 – 6.33)*	(1.38 - 4.34)*	(1.59 – 4.45)*	(0.66 – 5.25)*
Variety	-1.59	-0.36	-1.42	-3.58	-1.94
	(-3.44 – 0.24)	(-3.13 – 2.64)	(-3.07 – 0.25)	(-.5.22 – -2.01)*	(-3.81 – 1.99)
Services	-3.29	-1.78	-2.74	-3.01	-2.45
	(-4.45 – -2.18)*	(-3.63 – 0.07)	(-3.88 - - 1.63)*	(-4.14 - - 1.94)*	(-4.15 – -0.76)*
Encounters	0.81	0.88	2.15	0.54	1..25
	(-0.46 – 2.17)	(-1.16 – 3.10)	(0.90 – 3.52)*	(-0.71 – 1.83)	(-0.74 – 3.39)

Confidence intervals are given in parenthesis, 5% significance given by *

Tables 4.7A to 4.7C present the average results of the WTP estimates for the various percentages of missing data. On average, all methods did reasonably well at producing estimates that were significant when the "true" estimate was significant and producing insignificant estimates when the "true" estimates were not significant. However, the PP method routinely predicted a significant estimate for the variable *Encounters* when it was not appropriate. All methods again struggled with the variable *Access*, producing erroneous insignificant estimates.

Given the upward bias of the travel cost coefficient for the CC method, the expectation was that the CC method would overestimate the WTP for each variable as the travel cost coefficient is the denominator in the WTP formula. This was not the case. Nor was it the case that CC estimated larger WTP estimates than the other methods. It would seem that the bias in the parameter estimates played a greater role than the bias of the travel cost coefficient.

Table 4.8A: Willingness to Pay bias at 6% Missing, given in €

	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Access	-0.60	-0.58	-0.47	-0.53
Size	-0.39	-0.08	-0.13	0.10
Quantity	0.22	0.03	0.12	-0.24
Variety	0.27	0.02	-0.83	0.58
Services	0.23	0.19	0.11	0.61
Encounters	0.29	0.54	-0.05	0.03
Absolute Average bias	2.00	1.44	1.71	2.08

Table 4.8B: Willingness to Pay bias at 12% Missing, given in €

	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Access	-1.23	-1.03	-0.84	-0.86
Size	-0.67	-0.05	0.31	0.75
Quantity	0.48	0.06	0.23	-0.10
Variety	0.49	-0.06	-1.46	0.57
Services	0.61	0.36	0.22	0.78
Encounters	-0.03	0.85	0.14	-0.05
Absolute Average bias	3.51	2.41	3.20	3.11

Table 4.8C: Willingness to Pay bias at 18% Missing, given in €

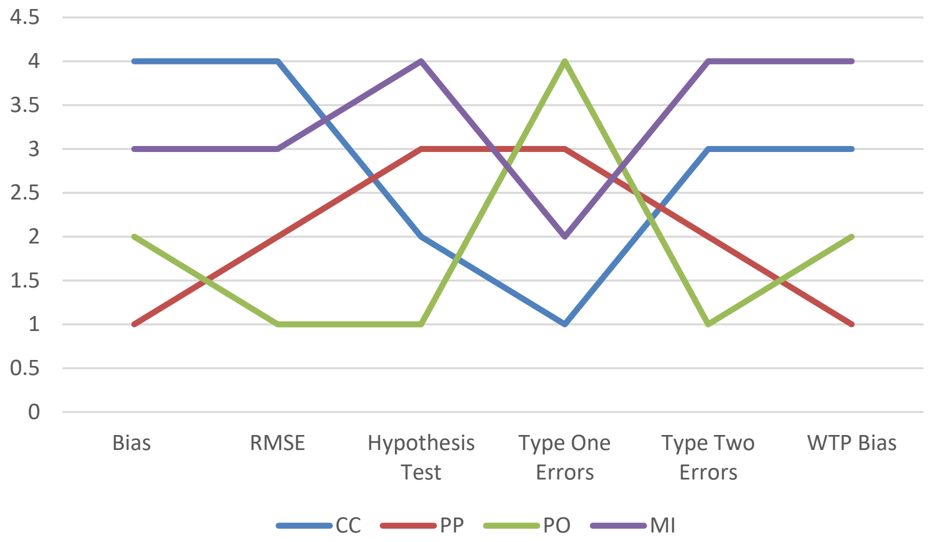
	Complete Case Analysis	Per person mean imputation	Per observation Mean Imputation	Multiple Imputations
Access	-2.67	-1.65	-1.34	-1.28
Size	-1.59	-0.86	0.58	0.85
Quantity	1.32	0.35	0.49	0.40
Variety	-1.23	0.17	1.99	0.76
Services	1.51	0.55	0.28	0.85
Encounters	0.07	1.34	-0.27	-0.41
Absolute Average bias	8.39	4.92	4.95	4.55

Finally, the bias in relation to the WTP estimates in comparison to the "true" estimates are shown in tables 4.8A, 4.8B and 4.8C. As the percentage of missing data increased so did the level of bias. CC performed poorest overall. PP, PO and MI performed similarly with PO performing best at the lowest percentages of missing data and MI performing best at the maximum percentage of missing data. Again, there seems to be a trend in the increase in bias as the percentage of missingness increases. Indicating that MI is superior at higher percentages of missing data. Somewhere before 18% in the case of this data.

4.7 Discussion

Of the compared missing data techniques, PP performed best at the two lowest percentages of missing data and MI performed best at the highest percentage. PP had the lowest level of bias for the two lowest percentages of missing data, but the bias associated with MI increased by the least amount from one percentage to another resulting in MI performing best at the highest percentage of missing data. For RMSE, the best performer changed with each level of missing data; cumulatively, PP performed best. Under the hypothesis test metric, PO performed much better than any other technique with PP coming second. For type one errors MI performed best with PP again doing second best and, for type two errors, PO performed best. Although PP had the lowest overall bias for WTP the bias increased at a much quicker rate as the percentage of missing data increased in comparison to any of the other methods. MI produced much less bias for the largest percentage of missing data. PP was also the only method to erroneously estimate a significant WTP for any of the variables.

Graph 4.1A: Plot of relative performance of each missing data method for 6% missing data



Graph 4.1B: Plot of relative performance of each missing data method for 12% missing data

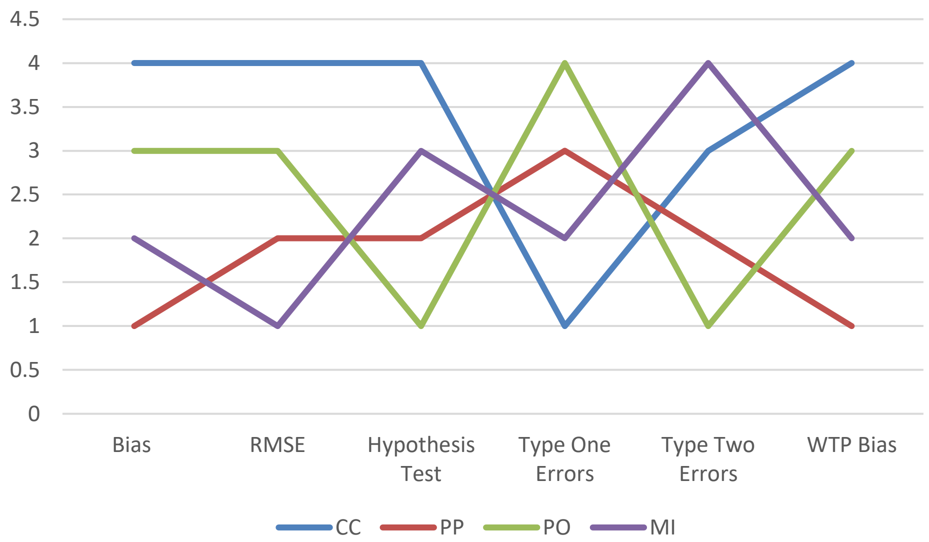
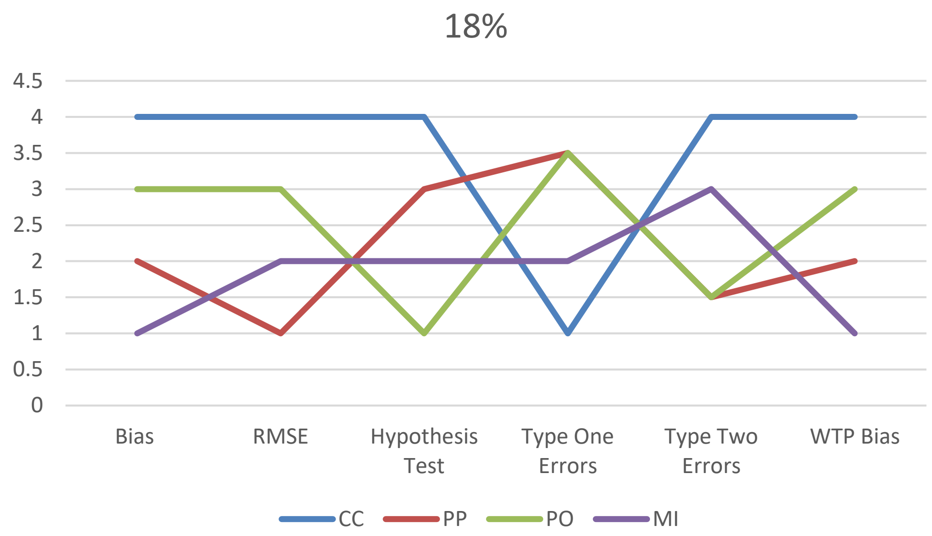


Table 4.1C: Plot of relative performance of each missing data method for 18% missing data



Graphs 4.1A, B and C plot how well each method did in comparison to the other methods. The values along the left-hand side display position i.e. first, second, third and fourth. Along the bottom of the graphs the colour coding for each method is displayed as well as the bias metric.

In general, the best course of action for low percentages of missing data is to use PP mean imputation. When employed, it produces reasonable results and most types of models can be easily applied to a data set with mean imputed data. For large amounts of missing data, the answer seems to be MI. As well as performing best at the 18% missing data the trend seemed to suggest that as the percentage of missing data increased MI would have performed even better in comparison to the other methods.

It is also worth noting that the MI application used here may be suboptimal. MI performs best if the dependent variable is a component of the imputation model. This is not possible in this type of analysis. By construction, all the missing data will have a zero for the dependent variable. Therefore, no information can be gained from including it in the imputation model. Although this is sub-optimal it will be the challenge that most analysts of this type of data will face. Most individual who have visited a site will be able to rate the site attributes and, as such, will not have this type of missing data.

The current analysis used five imputations, but others have suggested that MI would benefit from using many more imputations (Graham et al. 2007). The analysis omits a comparison of different numbers of imputations because of the time costs but this is an avenue for future research.

Even if MI had performed better than the other methods, analysts would still have to consider the extra time and effort that is needed to apply MI. analysts may also find it difficult to find commercial software packages that support MI in conjunction with more advanced specifications such as the mixed logit or latent class logit models (the authors are unaware of such a package). The results are, however, clear that CC should not be used with missing data of this type.

In comparison to other papers that have used similar techniques to the ones employed here, the results of this paper are not as favourable to MI. MI performed best in several other papers (Nakai et al. 2014; Zhu 2014) CC, which performed well in other papers (Ali et al. 2011; Zhu 2014) performed the worst on almost all metrics used in this paper. As highlighted earlier in this paper, this was to be expected given the type of data. As such, the authors would advise against its further use for data of this sort.

Although this paper was methodological in construct, it is worth noting how missing data may lead to miss valuation of natural resources, as can be seen in the amount of bias produced by the WTP estimates. As well as increasing the likelihood of poor policy formation. For example, had the results of most of these analyses been used to develop a policy aimed at improving the quality of coarse fishing sites in Ireland, a policy maker may think that *Access* was unimportant. When, in fact, it can be seen from the "true" parameter estimates that *Access* plays a significant and positive role on site choice and, likely, the perceived quality of the site. With this in mind, it is important that analysts understand that missing data can impact their results even after applying missing data techniques.

4.8 Conclusion

Missing data are a common problem in most empirical non-experimental work. The methods for dealing with such problems are numerous. However, the decision, over which technique best suits the data, can be an onerous one. This paper has explored four of these missing data techniques, complete case analysis, per person mean imputation, per observation mean imputation and multiple imputations as applied to site choice data. The results indicate that no one method performed best across all metrics and all percentages of missing data. Overall, PO performed best at the lowest percentages of missing data, but not across all metrics. For the highest percentage of missing data MI performed best.

There are many possible avenues for further research in this area. As outlined previously, a comparison of different numbers of imputations would greatly benefit researchers looking to apply MI. As the literature has shifted towards techniques that focus more on individual or group heterogeneity, a comparison of different missing data techniques using models such as the mixlogit and latent class logit would be fruitful for further research. Finally, a replication of the metrics used here to similar data is important to test the generalization of our findings to more than just the current data set.

Chapter 5

5. Combining Actual and Contingent Behaviour Data to Estimate the Value of Coarse Fishing in Ireland

5.1 Introduction

Fishing quality, indicated by both the size and quantity of a fish stock, is generally considered to be an important determinant of angler participation. This assertion has been tested in a multitude of ways, from the use of site attribute variables such as size, quantity, catch or simply quality of angling experience in site choice models (see Hunt 2005 for a review of the application of site choice model to fishing sites) and travel cost applications (Bilgic and Florkowski*,* 2007; Shrestha et al. 2002; Du Preez and Hosking, 2011) to explicit scenarios in contingent valuation (Rolfe and Prayaga, 2007) and contingent behaviour models (Alberini et al. 2007; Prayaga et al. 2010). However, these studies have produced mixed results, with some indicating that size and/or catch are positive corollaries of participation (Bilgic and Florkowski*,* 2007; Shrestha et al. 2002; Du Preez and Hosking, 2011) while others have not found a statistically significant relationship between size and/or catch and participation (Alberini et al., 2007; Prayaga et al. 2010).

Even if it assumed that an increase in fishing quality does have a positive impact on angler participation the distinction between the size of fish being the driver of increased participation or quantity being the primary reason for an increase in fishing days may have important implications for the angling community. Fisheries managers implement policies and practices based on the needs of the angling community. Legislation in many countries dictates how an angler can fish to preserve or improve some aspect of fishing quality. For example, it is common practice that countries will legislate how many fish can be kept and what size these fish can be. The primary goal of these laws is to maintain or improve fish numbers and to allow the larger fish to spawn or the juveniles a chance to reach reproductive age. However, these laws may come at a cost to the anglers who may wish to keep more, or larger fish than is permitted. It may then be of interest to policymakers to determine if both size and quantity of fish are important to anglers or if only one or neither is. Further to this, sufficient changes in both welfare and angler participation as a direct consequence of changes to fishing quality may point to the economic viability of stocking practices.

The principal aim of this paper is to determine if angler participation would change with an increase in either the number of specimen fish or the quantity of fish and what additional consumer surpluses (CS) are associated with these changes. Specimen fish is a commonly used term for coarse anglers. It refers to a fish over a certain size for its species. The quantity of fish simply refers to the number of fish regardless of size. We aim to compare a change in the number of these specimen fish to the same percentage increase in the fish quantity regardless of size. However, given the relatively low number of specimen fish compared to all other fish, a percentage increase in specimen fish is a much smaller number than the same percentage increase in the total fish quantity.

To make this comparison, the contingent behaviour (CB) method is employed. The focus of this study is coarse anglers who fish in Lake Garadice, a fishing site located in the north of the Republic of Ireland. Substantial benefits may arise from a comparison between a change in fish quantity to fish size in an Irish context as a report presented by Tourism Development Ireland stated that fish quality (both size and quantity) was the most important aspect of angling in Ireland (TDI, 2013). However, no distinction in the relative importance of size and quantity was made.

The remainder of this paper is structured as follows, section 2 describes the relevant previous literature on recreational angling. Section 3 then presents the study area, discusses the sampling method and data, and describes the contingent behaviour questions presented to the respondents. Section 4 describes the CB model and its application within this paper. Section 5 presents the results of the travel cost models (TCM) using the revealed trip preferences of Irish and overseas anglers who fish in Lake Garadice. This is followed by a set of contingent behaviour models that examine changes to the quantity of fish and the number of specimen fish. Finally, a discussion of the results within the context of the literature is offered with concluding remarks in section 6.

5.2 Previous research

The first application of the panel data approach to CB data was presented by Englin and Cameron (1996) in their assessment of the effects of changing prices on recreational fishing demand in Nevada. Since then, CB analysis has been widely applied to recreational demand. For example, Hanley et al. (2003) examined the effects of improvements in coastal water quality on Scottish seaside bathers, Christie et al. (2007) used a series of CB scenarios to determine the value of a number of changes to Great British forest and woodland areas for a variety of recreational users, Barry et al. (2011) estimated welfare changes from a proposed trail along the Irish coastline and, more recently, the CB method was used by Filippini et al. (2017) in their assessment of welfare changes caused by the provision of an alpine centre that would provide services that aim to reduce risk of both injury and death. The literature has also explored some of the more methodological issues of CB analysis; tests for validity have been undertaken within several papers (Grijalva et al. 2002; Lienhoop and Ansmann, 2011; Hoyos and Riera, 2013), survey non-response has been explored (Cameron et al. 1996), and methods of incorporating preference heterogeneity have been presented by Hynes and Greene (2013, 2016).

Numerous studies have applied the CB method to data on recreational angling. In addition to Englin and Cameron (1996), the application of the CB method to recreational fishing has included but is not limited to; examining how changes to water levels would affect the users of a drying Nevada lake (Eiswerth et al. 2000), exploring how price changes effect stated trip frequencies (Egan and Herriges 2003), examining the welfare impact from changes in water clarity for a Wisconsin lake (Eiswerth et al. 2008), while Cameron et al. (1996) used the CB method to "break near perfect multicollinearities among water levels at some waters".

The CB method, in application to fish quality, has been applied less often. Prayaga et al. (2010) used an onsite sample to look at how four levels of change in catch rate, ranging from a decrease of 25% to an increase of 25%, would affect anglers who fish in Australia's Capricorn Coast. Their results indicate that none of these changes had a statistically significant impact on trip frequency. Prayaga et al. (2010) also found that a 50% increase in the probability of catching a legal sized red emperor, a 30% increase in crowding or an increase in the length of algae blooms did not have a statistically significant impact on user behaviour.

Alberini et al. (2007) used a mail survey of users of the Lagoon of Venice to examine the effects of a 50% increase in catch. They report that respondents revealed angling participation was positively correlated with a respondent's historical experience of catch rates but the estimated effects of a hypothetical change in catch rate were insignificant. This non-significance remained consistent across numerous model variations containing variables that interact respondent characteristics with the hypothetical change dummy.

The results of Prayaga et al. (2010) and Alberini et al. (2007) may demonstrate several important differences between the revealed and stated data for fish quality. In both cases, individuals demonstrated a preference for higher catch rates, but this did not result in statistically significant results for hypothetical changes in catch. Based on the results of Prayaga et al. (2010) this is understandable; although the parameter associated with historical catch rate was significant it was small, and one may not expect that the level of hypothetical changes would induce more trips. However, the results of Alberini et al. (2007) reveal that for an extra kilogram of fish historically caught, respondents would take an extra 10 trips per year. With an average catch of about 3kg, the hypothetical increase of 50% should have resulted in about 15 more trips per year, instead of a insignificant negative value. It may be the case that for both the Prayaga et al. (2010) and Alberini et al. (2007)

samples, the revealed data do not perform well in predicting future behaviour, or that their respondents may have reached some threshold on the amount of time they would like to spend fishing at the respective sites.

Poor and Breece (2006) examined how a water quality change would affect anglers who fish in the Chesapeake Bay through CB analysis. However, the results are difficult to interpret. The CB dummy was positive and significant, but the framing of the question makes it difficult to know why the respondents are willing to take more trips. Respondents were informed, before the CB question, that due to poor water quality both fish size and population had been affected. The CB question specially stated that an improvement in water quality would result in larger rockfish. It is difficult then to parse why an individual would take more trips; more fish, larger fish, larger rockfish, better water quality or a combination of all these factors.

Duffield et al. (2001) took a two-stage approach to CB modelling fish quality changes to five stocked Alaskan rivers. Using a mail survey of registered sport anglers respondents were asked to first rate a series of stocking practices which included stocking fewer but larger fish or greater numbers of smaller fish. For the CB scenario question, the respondents were asked how they would change their trip pattern if their preferred stocking method was implemented. However, by only asking respondents about their preferred stocking practise there is no way of knowing the impact of any scenario for an individual who ranked the stocking practice anywhere lower than first. Ultimately, Duffield et al. (2001) did not produce the results of the CB model but suggested that the raw data indicated that the respondents would increase the number of trips taken if any of the proposed stocking practises were employed.

No application of the CB method has been applied to Irish angling. However, travel cost models have been extensively applied to estimate CS for a day spent fishing (Curtis and Breen, 2017; Hynes et al. 2015, 2017; Grilli et al. 2017). The exploration of angling behaviour in Ireland has also extended beyond the application of TCM. Anglers' own perception of site attributes have been used to determine how coarse anglers choose where they go fishing

(Deely et al. 2019), and how the use of objective or subjective data may impact parameter estimates (Deely et al. 2018). Also, Curtis and Stanley (2016) found that fish stock was positively correlated with the number of days spent on a trip but was not with the number of days spent fishing in a year or the number of trips taken in the year for a sample of game and coarse anglers.

At present the Irish angling literature seems to suggest that the quantity of fish at a site, or the catch rate, plays limited to no role in angler participation. Conversely, the number of specimen fish or size of fish has been shown to be a significant determinant of angling demand (Curtis and Stanley, 2016; Curtis and Breen, 2017; Deely et al. 2018, Deely et al. 2019). The international literature, on the other hand, has presented positive results for both size and quantity of fish (Bilgic and Florkowski, 2007; Shrestha et al. 2002; Du Preez and Hosking, 2011) using the TCM.

This paper explores the relationship between angler participation and the number of specimen fish and quantity of fish available at an Irish fishing site. To the best of the authors' knowledge, no attempt has been made to compare an increase in specimen fish to an increase in the quantity of fish using the CB method. This comparison may be of interest to managers and legislators as policy current dictates both the size and quantity of fish that can be kept. This paper is also, to the best of the authors' knowledge, the first recreational angling CB analysis to combine an onsite survey with an online survey as well as being the first to incorporate non-users into an analysis of either a change in fish size or quantity which should give a better estimate of the effect of the CB scenarios on a more diverse group of anglers.

5.3 Data

Coarse angler data were collected in relation to the respondent's use of Lake Garadice. Garadice is a 3.9 km^2 lake located in County Leitrim, Ireland, which is a border county to Northern Ireland. Garadice provides year-round freshwater fishing, with road access to a large number of pegs distributed around the lake. The site also contains boat access at two designated points as well as showering and toilet facilities. Numerous fishing competitions are held at Garadice every year. These competitions include regular local intraclub matches and international competitions, the latter providing an important source of revenue for the local communities.

Data were collected through a survey constructed with the assistance of experts in the field of Irish coarse angling, a focus group of anglers who fish in Garadice and a pilot study. The data collection process took two forms, an online survey and intercept sampling. The online survey ran from the 6th of August 2016 to the 15th of January 2017. Potential participants were contacted through Irish coarse angling Facebook pages, by emailing local coarse angling clubs, through the Inland Fisheries Ireland newsletter, and, to contact less avid anglers, local institutes of learning and local newspapers assisted in disseminating the survey link. In total, 45 respondents completed the contingent behaviour questions through the online survey, only two of which were from an overseas country. Intercept sampling also began on the 6th of August 2016 and finished on the 7th of November 2016, garnering 78 respondents.

Although the respondents were only asked CB questions related to Garadice, data were collected at four other sites as a portion of the survey was constructed for a site choice model. Before surveying began Garadice was chosen, for the CB portion of the survey, over the other four sites as it is the largest, most popular, and best known of these sites. By collecting data at other sites as well as the site of interest, the correlation between the number of days spent at Garadice and the probability of being surveyed is reduced for the sample, in comparison to traditional onsite sampling. This should reduce avidity bias.

During the survey, respondents were asked a series of questions pertaining to their angling experience, the number of days spent fishing at Garadice and all other angling sites in Ireland, expenditure on angling within the last year and demographic questions including the location of their home. The anglers were also asked a series of contingent behaviour questions.

The contingent behaviour questions (Table 5.1) explore how respondents' number of fishing days would change in response to a change in two characteristics of fishing quality; the number of specimen fish and quantity of fish. The respondents were first asked how many days they had spent fishing at Garadice in the 12-month period prior to completing the survey. They were then asked how many days they intended on spending at Garadice next year if conditions remained the same. Following this, the respondents were asked how many more days they would spend next year under each of the four contingent behaviour scenarios; a 25% increase in the number of specimen sized fish, a 50% increase in the number of specimen sized fish, a 25% increase in the quantity of fish, or a 50% increase in the quantity of fish. The CB questions and levels were constructed during meetings with Irish fisheries managers but were not chosen as definitive changes that may result from stocking practice or changes to management policy. The combination of revealed preference and stated preference data makes a sample of 738 observations, 6 for each individual. However, it should be noted that each CB scenario is modelled and presented separately in the results section, so a panel of 3 observations is used in each case.

Table 5.1: Contingent Behaviour Questions Posed to the Respondents

How many days have you spent fishing at Garadice in the previous 12 months?
How many days do you intend on spending fishing at Garadice in the next 12 months?
If there were a 25% increase in the quantity of fish at Garadice how many days would you spend fishing there?
If there were a 50% increase in the quantity of fish at Garadice how many days would you spend fishing there?
If there were a 25% increase in the number of specimen sized fish at Garadice how many days would you spend fishing there?
If there were a 50% increase in the number of specimen sized fish at Garadice how many days would you spend fishing there?

5.3.1 Summary Statistics

The survey was completed by 123 coarse anglers. They differed with respect to country of origin, income, angler type, self-reported ability level, and many other aspects of both angler and personal characteristics. Every respondent is, by construction, an adult over the age of 18. Only two respondents were female.

Approximately 38% of the sample are overseas anglers (Table 5.2); this is larger than the TDI (2013) estimate. TDI (2013) estimated that 28% of the anglers who fish in Irish waters come from overseas. A possible reason for the apparent oversampling of overseas anglers may be Leitrim's (the county Garadice is situated in) proximity to Northern Ireland. The majority (61%) of the overseas anglers came from Great Britain. The single currency between Great Britain and Northern Ireland may result in Northern Irish counties being visited more often by British anglers, and by extension, nearby fishing sites in counties such as Leitrim may have a higher representation of British anglers than the national average. Additionally, the TDI (2013) study includes all angler types and is not necessarily reflective of coarse anglers. Curtis and Breen (2017) state that 37% of the game and coarse anglers from the same TDI sample are from overseas, which is almost identical to the present sample.

Table 5.2: Sample Descriptive Statistics

	Irish and Overseas Respondents (n =123)	Irish Respondents (n =76)	Overseas Respondents (n = 47)
Variable	Mean	Mean	Mean
Average one-way distance (km)	287.45 (293.55)	81.19 (70.83)	620.98 (190.84)
Average expenditure per day (€)	76.75 (71.40)	65.01 (56.32)	95.74 (88.07)
Mean number of days spent at Garadice	10.21 (18.88)	14.28 (22.37)	3.64 (7.69)
Years fishing	37.75 (14.55)	34.80 (14.00)	34.80 (14.00)
Number of species Targeted	1.71 (1.13)	1.99 (1.29)	1.26 (0.57)
Age	52.88 (11.73)	49.11 (12.60)	59.32 (6.25)
Income (€)	44,684 (33,579)	43,975 (29,073)	43,214 (37,302)
Survey online	0.37 (0.48)	0.57 (0.50)	0.04 (0.20)
Overseas	0.38 (0.49)	N/A	N/A
Match	0.42 (0.50)	0.43 (0.50)	0.40(0.50)
Specimen	0.13 (0.34)	0.20 (0.40)	0.02 (0.15)
Pleasure	0.45 (0.50)	0.37 (0.49)	0.57 (0.50)
Ability level	2.34 (0.66)	2.46 (0.64)	2.15 (0.66)
Retired	0.24 (0.43)	0.22 (0.42)	0.28 (0.45)

Note: Standard deviation presented in parenthesis.

The sample average one-way distance to Garadice is 287.45 kilometres and the average per day expenditure is €76.75. Per day expenditure[17] is composed of three components; angling expenditure[18], accommodation, and transport cost. The transport cost element of the total travel cost is computed differently for the overseas anglers as opposed to the Irish anglers. For the Irish anglers, transport cost was computed as the operating cost of running a medium-sized vehicle from the respondent's home to the fishing site and back. This accounts

[17] All expenditure except transport cost for Irish anglers is self-reported for the 12-month period prior to filling out the survey. For overseas anglers all expenditure is self-reported. This expenditure refers only to money spent by the individual in Ireland.

[18] Angling expenditure only includes items that are purchased for each trip such as bait and excludes investment items such as gear or tackle.

for 61% of the total travel cost of the average Irish angler. For the overseas anglers, their self-reported cost of fuel and vehicle rental spent was divided by the total number of days spent fishing in Ireland. For the average overseas angler, transport cost accounts for just 23% of their total per day expenditure.

Although it is unusual in the literature to calculate any element of the travel cost variable differently for two segments of the sample, the Irish and overseas are pooled to form one sample for the TCM and CB models. While not reported here we also ran models with spilt samples for domestic and overseas visitors and compared the results to the pooled sample model. The results were similar with respect to almost all the variables of interest and, the consumer surplus calculated from the Irish only model fell into the confidence interval from the overseas model. There remains some difference in magnitude between the estimated travel cost coefficients of the Irish model and the overseas model. Consequently, the pooled model contains an interaction term between the dummy overseas and the travel cost variable.

The average number of days spent fishing at Garadice is 10.21. However, the trip frequency is not truncated at one for all respondents as some (17) of these respondents took zero trips to Garadice within the previous 12 months before completing the survey. For the users that spent at least one day at Garadice, the mean cost per day is €68.41, the average one-way distance is 273.15 kilometres, and the average active angler spent 11.85 days fishing at Garadice.

The average respondent has been fishing for almost 38 years and is 53 years old with nearly a quarter being retired. The mean angler targets 1.7 species of fish[19]. This suggests that the sampled anglers may attend more than one type of site during the year and may have further implications on their total financial investment in fishing as targeting certain types of fish may require specialised equipment. For the current sample, those who targeted 3 or more fish species spent about 60% more on angling related expenses in the twelve-month period before completing the survey in comparison to those who fished for one or two species. Ten income brackets were provided to the respondents.

[19]For the purpose of this metric coarse fish are considered to be one species.

In order to ascertain the respondent's income: the minimum was less than €15,000 and the maximum was €150,000 plus. Average income was €44684 per annum.

The surveyed anglers were also asked what type of angler they considered themselves to be; a match angler (someone who participates regularly in fishing competitions), a specimen angler (someone who aims to catch large fish of a breed) or a pleasure angler. The largest proportion of respondents considered themselves to be pleasure anglers, followed closely by match anglers, with only 14% declaring they were specimen anglers, and only one overseas angler was a specimen angler. The respondent's self-reported ability level could take one of three categories; basic, intermediate or advanced. Only 10% of the surveyed anglers considered themselves to be of a basic level, while 45% believed they were of an intermediate level, and a further 45% believed they were an advanced level angler.

5.4 Methodology

Within the travel cost modelling framework, the economic value of a non-market good is estimated through the uses of revealed preference data, individual characteristics, site characteristics, and an estimated price. The estimated price usually includes the cost of reaching the site of interest and may also include the cost of all the necessary equipment needed to take the trip, accommodation and opportunity cost of time (Parsons, 2003). The demand function for the travel cost model can be written as:

$$Y_i = f(X_i) \tag{5.1}$$

where Y_i, the dependent variable, is the number of trips taken by individual i, and X_i is the vector of variables that impact the individual's decision to take these trips e.g. price, site characteristics, and individual characteristics. The dependent variable, number of trips taken, is a non-negative integer value by definition. Consequently, count data estimation approaches, such as the Poisson and Negative Binomial models, are routinely employed to determine the probability of the number of trips taken by an individual equalling some integer value.

Contingent behaviour models can be thought of as an extension to the traditional travel cost model, in which revealed trip frequencies are combined with stated trip frequencies to form a panel data set. Like the TCM, the dependent variable in the CB framework is also the number of trips, consequently, panel count models are used. The most common of these count data models are the Poisson and negative binomial. Following Hausman et al. (1984) the panel Poisson probability function can be specified as:

$$pr\ (y_{ij}) = \frac{e^{-\lambda_{ij}}\lambda^{y_{ij}}}{y_{ij}!}; y_i = 0,1,.,., \infty \tag{5.2}$$

where $y_{ij}$is the number of trips taken or intended to be taken per year, i denotes individual i, j is one of the six CB scenarios, and λ, the rate parameter, is equal to the mean and variance of y_{ij} which can be expressed as an exponential function (Hausman et al. 1984):

$$\lambda_{ij} = \exp(X_{ij}\beta) = E(Y_{ij}|X_{ij}\beta) \tag{5.3}$$

where β is a vector of parameters to be estimated and x_{ij} is a vector of explanatory variables.

An assumption of the Poisson model is equidispersion, where the mean and variance are equivalent. In the case where data are over dispersed a negative binomial distribution may be more suitable. Following Hynes et al. (2016)

$$Pr(Y_i) = \frac{\Gamma(Y_i + 1/\alpha)}{\Gamma(Y_i + 1)\Gamma\ (1/\alpha)}(\alpha\lambda_i)^{y_i}(1 + \alpha\lambda_i)^{-(Y_i+1/\alpha)} \tag{5.4}$$

Where Γ denotes a gamma function and the scalar α is a nuisance parameter to be estimated. α is a measure of the ratio of the mean to the variance, with larger values corresponding with greater amounts of over-dispersion.

When revealed and stated preference data are combined, as they are in CB analysis, assumptions over the correlations of the error terms play a fundamental role in model specification. Correlated error terms imply that the individual's responses are correlated with unobservable tastes and characteristics which remain consistent across their responses. Data with correlated error terms are treated using a random or fixed effects approach. Uncorrelated error terms are assumed to be independent and identically distributed across all observations. In the presence of uncorrelated error terms, standard statistical models can be used as the pooled data observations do not require models that address this issue. Although not presented here, both pooled and panel methods were employed on the data set used for this study. However, chow tests indicated that a panel approach is appropriate for the data.

Welfare estimates are often a primary concern in non-market valuation papers, for the TCM consumer surpluses are estimated and, for the CB method, the marginal effect of the proposed change is estimated. CS, in this case, is the value of a day spent fishing to an individual above the money spent. It has been demonstrated, by Hellerstein et al. (1993), to be the sum of the values under the demand function of a TCM over the relevant price range. The per-day CS can be conveniently calculated as:

$$CS = -1/\beta_{tc} \quad (5.5)$$

whereβ_{tc} is the estimated parameter of the travel cost variable.

The marginal effect is the additional value that is associated with the CB change. Following Prayaga et al. 2010, it can be calculated using the formula:

$$ME = \beta_{cb} * [\frac{-1}{\beta_{tc}}] \quad (5.6)$$

Where β_{cb} is the coefficient for the contingent behaviour scenario.

An important consideration when employing CB analysis is biased parameter estimates as a direct consequence of the data collection procedure. Two common forms of bias in CB analysis are endogenous stratification and zero truncation. Endogenous stratification has been well documented in applications of the TCM (Shaw, 1988; Englin and Shonkwiler, 1995) as well as for CB analysis (Egan and Herriges, 2006; Hynes and Greene, 2016). Endogenous stratification is a consequence of the respondent's probability of attending a site being correlated with their probability of being sampled. As a result, onsite samples are often overpopulated with respondents who have a predilection for the site of interest, which is not representative of the population. By combining intercept samples from several sites and an internet sample, the entanglement of the correlated probabilities is only attributable to a reasonably small portion of our sample i.e. those sampled at Garadice. As such, no correction for endogenous stratification is applied to the data presented in this paper.

Although endogenous stratification is not corrected for, the probability of inclusion in the sample is still correlated with the avidity of the respondent, which may imply that the sample is more reflective of avid coarse anglers than a random sampling of coarse anglers. The implications of a sample being relatively overpopulated with avid anglers are nuanced and may impact the contingent behaviour analysis in numerous ways. However, these implications are beyond the scope of this paper.

The second, and equally well documented, form of bias is zero truncation (Shaw, 1988; Grogger and Carson, 1991; Hynes and Greene, 2013; Egan and Herriges 2006). This problem arises due to the simple fact that, for an angler to be surveyed at a site they must have taken at least one trip; excluding individuals who took zero trips from the analysis. Parsons (2003) states that this may lead to less accurate estimates of the choke price and, consequently, less accurate estimates of CS. Fletcher (1990) suggests that zero truncation can lead to an upward bias of CS estimates. Additionally, zero truncation may undervalue improvements to a site as the allure of the proposed changes to anglers with zero trips is not explored. As the data collection process of the current study employed an online sampling component as well as sampling at other sites than the site of interest, the sample is not truncated. This allows for a more precise estimate of CS and a more effective examination of how the CB changes may impact anglers, including those who are currently non-users of Garadice.

Although, neither endogenous stratification nor zero truncation is corrected for in the current study, the sample has been restricted to anglers who have some knowledge of the site. In particular, the respondents needed to know where the site is situated and have a general knowledge of both the number of specimen sized fish and the quantity of the fish at the site even if they have stated that they have not visited the site in last 12 months. This, in turn, may result in a sample of more 'enthusiastic' users of Garadice than would be expected from a random sampling framework. However, the requisite statistics to transform the results presented here into more nationally representative results do not exist. Consequently, no transformation is made and as such the results found in this paper should be seen as representative of the sample only.

5.5 Results

5.5.1 Travel Cost Model

The results of a negative binomial TCM estimated on the revealed trip frequency observed in the data are reported in Table 5.3. A Poisson model was also fitted but are not presented as the data are overdispersed with regard to its mean and variance, which can be seen in the significant overdispersion parameter of the negative binomial model.

Table 5.3: Travel Cost Model with Negative Binomial Specification

Variables	Coefficients
Travel cost	-0.011 (0.002)***
Travel cost * Overseas	0.007 (0.003)**
Years fishing	0.015 (0.009)*
Ability level	-0.094 (0.176)
Online	-1.628 (0.288)***
Retired	0.139 (0.314)
Match	0.525 (0.225)**
Specimen	0.971 (0.436)**
Income	0.085 (0.057)
Targets 3 or more fish species	1.267 (0.315)***
Overseas	-2.154 (0.326)***
Constant	0.266 (0.591)
Alpha	0.920 (0.142)***
Observations	123
Pseudo R2	0.124
Log-likelihood	-345
AIC	717
BIC	753

*Note: Figures in parenthesis are standard errors. *** indicates significant at 1%, ** indicates significant at 5% and * indicates significant at 10%.*

The results of the model indicate that, as expected, travel cost has a negative and significant impact on the number of days spent fishing at Garadice. The statistically significant interaction term *travel cost overseas* suggests that the negative effect from increasing travel cost is less for the overseas anglers. This may be reflective of several overseas anglers receiving a higher level of utility from a day spent at Garadice. Alternatively, overseas angler may be receiving utility from other activities that may justify their lower disutility for travel costs.

A priori expectations were that the number of years the respondent has been fishing is related to their devotion to the sport, experience, and their age and as such the variable *Years Fishing* would have a positive effect on angling demand. The *Years Fishing* variable conformed to expectations as the estimated coefficient was positive and significant. This implies that the predicted number of trips to Garadice increases with every year of fishing experience the respondent obtains.

Self-reported *ability level* does not seem to have a statistically significant impact on the number of trips an angler took to Garadice. This suggests that respondents who consider themselves to be an advanced level angler take as many trips to Garadice as a novice angler, all else being equal. The dummy variable *online* indicates where the respondent completed the survey, either online or at one of the five survey sites. This is negative and significant, indicating that the onsite cohort takes more trips to Garadice than their online counterparts. This may be due, in part, to a portion of the onsite sample being collected at Garadice, resulting in oversampling of avid users of Garadice.

Unlike other analysis of angler participation (Curtis and Breen, 2017; Curtis and Stanley, 2016; Grilli et al. 2017) whether or not an angler is retired (often proxied by being 65 years or older) does not seems to play a significant role in their decision of how many days they spend fishing at Garadice. However, the analysis used elsewhere (Curtis and Breen, 2017; Curtis and Stanley, 2016; Grilli et al. 2017) did not include an age or years fishing variable, which one would expect is correlated with retirement.

Both the dummies *match* and *specimen* are positive and significant indicating that these anglers tend to spend more days fishing at Garadice than the pleasure anglers. *Income* was not statistically significant for any of the model results; a not uncommon result in travel cost modelling (Hynes et al. 2015, 2017; Curtis and Breen 2016). The variable *targets 3 or more species* is positive and significant. It may be the case that, as targeting multiple species of fish is a much more costly endeavour than just targeting one species, these individuals are willing to spend more on fishing experiences more generally. This may result in more days spent at Garadice. The final variable, *Overseas,* is a dummy indicating whether the respondent is not a resident of the island of Ireland. This is positive and significant showing that, as expected, Irish anglers take more trips to Garadice than their overseas counterparts.

The per-day CS for the Irish anglers is estimated at €93, with 95% confidence intervals ranging from €50.75 to €135.25 (Table 5.4). The estimated average consumer surplus for the overseas anglers is much larger at €296.42 per day. A number of factors may have influenced the large difference between the mean CS of the Irish anglers in comparison to the overseas anglers. The Irish angling product may offer a unique experience to the overseas angler as fishing in Ireland is known for its natural state, scenery and fishing quality, all of which are ranked as important factors for overseas anglers visiting Ireland (TDI, 2013). There may also be an overestimation of the overseas CS if some of the visits to Ireland were for multiple purposes (Kuosmanen et al. 2004).

Table 5.4: Estimated Consumer Surplus Per Day

	Value
Consumer surplus €	93.00
Confidence Intervals €	50.75 – 135.25
Total value[20] €	158.02
Predicted Number of Days	11.52

The predicted number of days spent at Garadice is somewhat larger than the actual frequency of 10.21. The estimate predicts 11.52 days would be taken by the sample over a year long period, 1.31 days more than the actual number of trips. For the Irish anglers this estimate is 16.76 days, 2.49 days more than the actual figure but for the overseas anglers, the estimate of 3.05 days is lower than the actual trips of 3.64 days.

[20] Total value is the average consumer surplus plus the average expenditure per day.

5.5.2 Contingent Behaviour Analysis

Using Stata 15's xtnberg[21] command, a random effects panel negative binomial model was estimated for the sample of coarse anglers (Table 5.5). For each model presented, every respondent contributes three observations; current trips, future trips and trips taken after the specified contingent behaviour change. The current trips are the number of trips they took in the 12 months prior to completing the survey, the future trips are the trips they intend to take in the next 12 months under status quo conditions, and the final observation is the stated number of trips the respondent would take if one of the contingent scenarios were to occur.

[21] As pointed out by one of the anonymous referees of this paper, the xtnbreg command, in Stata, estimates random effects with respect to the dispersion parameter, not to Xβ, so that over-dispersion is assumed to follow a beta distribution B(r,s), where r and s are the rate and scale of the beta distribution respectively. By using this command, one has a distribution of the dispersion parameter across respondents, which is very similar to what a correction for endogenous stratification does in a cross-section negative binomial model.

Table 5.5: Contingent Behaviour Models

	25% increase in quantity fish	50% increase in quantity fish	25% increase in specimen fish	50% increase in specimen fish
Travel cost	-0.006 (0.002)***	-0.005 (0.002)***	-0.005 (0.002)***	-0.005 (0.002)**
Travel cost * overseas	0.005 (0.003)*	0.004 (0.003)*	0.004 (0.003)	0.004 (0.003)
Dummy current	-0.222 (0.071)***	-0.225 (0.077)***	-0.221 (0.073)***	-0.218 (0.079)***
25% increase in quantity fish	0.292 (0.062)***			
50% increase in quantity fish		0.474 (0.065)***		
25% increase in specimen fish			0.292 (0.064)***	
50% increase in specimen fish				0.384 (0.068)***
Years Fishing	0.016 (0.007)**	0.012 (0.006)*	0.014 (0.007)**	0.014 (0.007)**
Ability Level	-0.117 (0.147)	-0.087 (0.139)	-0.089 (0.149)	-0.073 (0.144)
Online	-1.134 (0.261)***	-1.050 (0.249)***	-1.074 (0.264)***	-1.035 (0.257)***
Retired	0.073 (0.245)	0.106 (0.230)	0.076 (0.246)	0.051 (0.236)
Match	0.466 (0.188)**	0.471 (0.177)***	0.423 (0.189)**	0.382 (0.182)**
Specimen	-0.169 (0.334)	-0.256 (0.316)	-0.060 (0.332)	-0.073 (0.322)
Income	0.075 (0.047)	0.061 (0.044)	0.079 (0.047)*	0.072 (0.045)
Target 3 or more fish species	0.374 (0.253)	0.370 (0.237)	0.319 (0.253)	0.348 (0.242)
Overseas	-1.477 (0.290)***	-1.359 (0.276)***	-1.427 (0.297)***	-1.358 (0.287)***
Constant	0.460 (0.472)	0.384 (0.445)	0.378 (0.478)	0.219 (0.459)
r	2.470 (0.404)	2.461 (0.396)	2.280 (0.3720	2.191 (0.352)
s	2.466 (0.487)	3.087 (0.637)	2.391 (0.485)	2.850 (0.600)
Observations	369	369	369	369
Log-likelihood	-1043	-1075	-1048	-1077
AIC	2118	2183	2128	2185
BIC	2180	2245	2191	2248

*Note: Figures in parenthesis are standard errors. *** indicates significant at 1%, ** indicates significant at 5% and * indicates significant at 10%.*

For the CB model, the travel cost coefficient is negative and significant, the interaction term *travel cost * overseas* is positive, but only significant for a change in the quantity of fish, the dummy for current trips is negative and significant indicating that the respondents, on average, took fewer trips in the current period than they intend to take in the next 12 months without any change to Garadice. Finally, all four contingent behaviour dummies are positive and significant indicating that, on average, anglers declared that they would take more trips next year, under each of the proposed scenarios, than they had otherwise intended to. In both cases, the magnitude of the CB dummies reveals that a 50% increase in either the number of specimen fish or the quantity of fish results in more trips than a 25% increase.

Table 5.6 shows the marginal effect of each of the CB scenarios. The 95% confidence intervals were calculated using the Krinsky-Robb method (1986) with 5,000 draws. For the marginal change in days the Stata command margins was used.

Table 5.6: Marginal Effect of Contingent Behaviour Scenarios

	25% increase in quantity fish	50% increase in quantity fish	25% increase in specimen fish	50% increase in specimen fish
Marginal Effect, €, per day Irish anglers	50.86	89.01	54.97	81.09
95% confidence intervals	24.25 – 163.87	48.28- 300.72	24.72 – 204.08	37.96 – 349.13
Marginal Effect, €, per day overseas anglers	296.32	212.36	253.62	440.810
95% confidence intervals	-1257.47 – 1850.11	-548.86 – 973.57	-888.97 – 1396.21	-1491.94 – 2373.54
Marginal change in days fishing	3.8	6.56	3.78	5.25
Change for Irish anglers	6.53	10.69	6.34	8.59
Change for overseas anglers	1.49	2.75	1.52	2.21
Change for specimen anglers	2.47	3.92	2.8	3.96
Change for Match anglers	4.66	8.12	4.55	6.25
Change for pleasure angler	2.93	5.07	2.98	4.26

All the CB scenarios add additional value when compared to the status quo. The associated confidence intervals indicate that for the Irish anglers all estimated MEs are statistically different from zero. However, for the overseas anglers MEs is not statistically different from zero. The marginal change in days spent at Garadice ranges from 3.78 extra days for a 25% increase in the number of specimen fish to 6.56 days for a 50% increase in the quantity of fish. As would be expected, the Irish anglers are much more likely to increase the number of fishing days at Garadice as a result of one the CB changes, than the overseas anglers. The Irish anglers would increase the number of days fishing at Garadice by approximately four times that of the overseas anglers, after one of the CB changes.

Table 5.6 also displays the marginal effect, in terms of increased fishing days, associated with the CB changes for specimen anglers, match anglers and pleasure anglers. A priori expectations are that for a change in the number of specimen fish the marginal effect will be largest for specimen anglers, whereas a change in the quantity of fish will have the greatest impact on match anglers. However, the results do not fully meet expectations.

A 50% increase in the quantity of fish results in the match anglers' spending approximately 8 more days fishing; many more than either the pleasure or specimen anglers. Although the estimated change for the specimen anglers is larger for an increase in specimen fish rather than an increase in quantity of fish, it is a comparatively small value; smaller than for match or pleasure anglers. The estimates imply that match anglers are the most responsive to a change in fish quality followed by pleasure anglers, whereas specimen anglers do not react strongly to any of the CB scenarios. However, specimen anglers are the least well represented amongst the sample which may make the sample less representative of specimen anglers than either match or pleasure anglers.

5.6 Discussion and Conclusion

Fish quality is considered to be the most important aspect of Irish recreational fishing (TDI, 2013). However, little is known about how anglers respond to an increase in either the number of specimen fish or the quantity of fish. The literature has given somewhat conflicting results but, generally, suggests that size or number of specimen fish is an important aspect of angler participation whereas quantity seems to play a less dominant role. However, no attempt has been made to directly estimate how changes to fish quality would affect angler participation in Ireland.

The current study employed contingent behaviour analysis to estimate the changes associated with a 25% and 50% increase in either the number of specimen fish or the quantity of fish at Lake Garadice. The results indicate that both the number of specimen fish and the quantity of fish play a significant role in angler participation. The results also indicate that anglers are almost equally well off from an increase in the number of specimen fish as an increase in the quantity of fish. The difference between the mean estimates is approximately €4 for a 25% and €8 for a 50% increase. The marginal change in the number of days spent fishing is also similar; a 25% increase in either the number of specimen fish or the quantity of fish results in approximately 3.80 more days fishing at Garadice and a 50% increase in quantity of fish results in 6.56 more days, whereas a 50% increase in the number of specimen fish would induce anglers to take 5.25 more days fishing at Garadice.

These results imply that recreational users of Irish coarse fishing sites may benefit from stocking practices as an increase in fish stock may result in Irish anglers spending more days angling during the year while also enticing overseas anglers to holiday in Ireland for longer periods of time. However, stocking for coarse fish is seldom practised in Ireland, particularly for natural waters (IFI 2015). In part, this may be because Irish waters are seen as natural and wild, and its fish robust and challenging. Hatchery fish, as reported by Inland Fisheries Ireland (IFI), the state agency responsible for the protection, management and conservation of inland fisheries, are genetically inferior and provide less of a challenge to catch (IFI 2015). This, in itself, poses a question on the trade-off anglers are willing to make for an increase in fish stock. Anglers may want more fish but be unwilling to substitute natural fish for hatchery fish.

At present, IFI has undertaken a policy of trying to keep and restore Irish recreational fishing waters in a natural state. This includes the improvement and preservation of water quality, protection of nurseries, and legislation over the size and quantity of fish that are allowed to be kept. It may be possible to extend these practices to increase the number of specimen fish and the quantity of fish at Garadice. By altering the legislation on what fish can be kept an increase in fish abundance and possibly size may be achieved without the need to stock lakes. However, this may be a delicate balancing act, as anglers may want an increase in fish quality but may not be willing to release more caught fish than is currently legislated. It could, in fact, be the case that anglers may be, on average, worse off from an increase in fish quality if it is brought about by a change in legislation. Careful consideration would need to be given not only to the trade-off between fish quality and the amount of fish an angler can keep but also to the time horizon over which a change in fish quality might occur and to anglers' discounting of future utilities.

The study also highlighted several characteristics of angler participation that may be of interest to managers of fisheries outside of Ireland. Firstly, it is clear that tourist anglers are impacted differently by travel cost than Irish anglers. This may indicate higher WTP for a day spent coarse angling, or alternatively, that the trips are multipurpose. Consequently, models from any region that encompasses both Irish and tourist anglers should consider the differences highlighted in this paper. Secondly, the number of extra days that tourist anglers are willing to spend fishing due to any of the purposed changes was about a quarter of the size of the Irish anglers. It may be the case that if managers aim to increase tourist angling by increasing fish quality the change may be much less dramatic than for local anglers. Finally, as a consequence of the previous two points, analysts of data such as this may need to consider the demographics of the sample as data sets overrepresented by tourist anglers may undervalue CB changes in terms of extra days spent fishing but overvalue associated CS.

With respect to the international literature, reviewed in section 2, there are a number of reasons that this study may result in different estimates from previous papers. The first and most obvious being the populations that the samples are drawn from are different. Secondly, as the source of the extra days spent at Garadice is uncertain, it may be the case that anglers will substitute fishing days at other sites for days at Garadice. If this were the case and if respondents of other analysis did not have this opportunity one may expect a larger estimate for this analysis than others as substituting one fishing day for another may be easier than non-fishing days for fishing days. Also, this paper is the first, with respect to size and quantity changes in CB analysis, to incorporate current non-users of the site of interest into the sample as well as being the first to combine an onsite and online sample which may result in different estimate than previous studies. As the normal issue of endogenous stratification and avidity bias should have been greatly reduced, the results may be more generalizable to the population of interest than previous estimates and, consequently, different CB estimates.

This paper also presented a traditional travel cost model. The estimates imply that the CS is €93 per day spent fishing in Garadice. However, no accurate information exists on the total number of visitors or the compositions of the users of Garadice. As such, the values cannot be extrapolated to all user and potential users of Garadice but reveal important information on the CS associated with a day spent coarse angling.

With respect to both the TCM and CB, the different treatment of the Irish anglers' and the overseas anglers' travel expenditure also merits some discussion. It is assumed that the travel expenditure of overseas anglers, who are on multiday visit to Ireland, can be conveniently parcelled into a number of single day costs. This implies that all the expenditure the overseas anglers spent in Ireland can be attributable to fishing and that each fishing day is of equal value. This may not be the case. Some anglers may see these multi-day trips as multipurpose and, as such, a portion of the expenditure may rightly be attributable to other activities. This may, in fact, account for some of the difference between the Irish and overseas anglers with respect to their disutility from additional costs. There is a growing literature on the incorporation of multipurpose/multidestination trips into travel cost models (Hill et al. 2014; Saengavut, 2018). However, the required information was not available to address this issue here. Although it is common amongst the recreational angling literature to combine overseas and native anglers into one travel cost model due care must be taken when extrapolating result from this paper given this uncertainty.

Chapter 6

6. Conclusion

6.1 Main Findings and Policy Implications

Inland Fisheries Ireland is currently aiming to increase the number of fishing days spent in Irish waters. The ability to do so requires the implementation of strategies, that will motivate non-users to start fishing and will create an environment that encourages users to spend more time fishing. There are many complex factors that impact individuals' decisions to fish in Ireland. Amongst the decision-making factors are the quality of the fishing sites and the fish quality at these sites. This has been a primary focus of this thesis.

Economic theory states that individuals make a decision based on the bundle of attributes that a good possesses (Lancaster 1966). In order to improve a site, one must know how the attributes it possesses contribute to the probability that individuals will fish at that site. The first paper 'Coarse angler site choice model with perceived site attributes' (Chapter 2) is motivated to answer this question. Conditional and random parameter logit models are applied to a data set of perceived coarse angling site attributes, to determine the interaction between angler behaviour and key site attributes. The second paper 'Are objective data an appropriate replacement for subjective data in site choice analysis?' (Chapter 3) follows on from this idea in two respects. Firstly, by looking at the methodological issue of the appropriateness of using objective data instead of perceived data to answer questions about choice behaviour. Secondly, from a more practical perspective, the paper examines how well mangers' assessment of site attributes align with users'. This has some obvious implication for the usefulness of the results of the first paper. If it were the case that manager and users' perspectives were entirely unconnected then the results of the first paper may be of little use at a national scale, as managers may not be able to determine which sites are appropriate for development. The third paper 'Are objective data an appropriate replacement for subjective data in site choice analysis '(Chapter 4) compares missing data techniques commonly used for site choice data. It highlights, with some caveats, how one should approach the nearly ubiquitous problem of missing data. The final paper 'Combining actual and contingent behaviour data to estimate the value of coarse fishing in Ireland' (Chapter 5) looks specifically at one of the sites of interest to determine the consumer surpluses associated with a day spent coarse angling. Contingent behaviour methods are also used to see if anglers would take more trips, given some policy change. The amount of additional benefit the surveyed anglers would receive from each contingent behaviour scenario was also estimated.

Focusing on Chapter 2, Conditional and random parameter logit models were applied to choice data, in which individuals were asked to rate five of the keys coarse fishing sites in the Cavan and Leitrim area on a set of six attributes. The findings indicate that accounting for individuals' preferences, which is achieved through the use of a random parameter logit, results in a better fitting model, as measured by the log-likelihood. However, the results were similar in all respects but one. For the conditional logit *accessibility*, *variety* and *quantity* of fish are significant positive determinants of site selection but for the random parameter logit model *size* was a positive and significant determinant but *quantity* was not. However, as both *size* and *quantity* were positive this change is one of magnitude and not direction. *Local services* has a negative impact on site selection, whereas the level of encounters with other anglers does not play a significant role in both models.

The results of both models were used to estimate willingness to pay for site attributes and a series of policy scenarios. WTP estimates, for a one-unit change in a site attribute, range from negative €4.62 for a unit increase in *services*, estimated from the results of the conditional logit model, to positive €4.44 for a unit increase in access, from the random parameter logit. The policy scenarios estimated how WTP estimates varied between sites for a change in access and size. These welfare estimates were estimated using the results of the random parameter logit model. The policy change scenario would increase access and size up by one unit on the five point Likert scale used to rate the sites. These results demonstrate, particularly for access, that the benefit of a change in site attributes is dependent both on the sites current state and its popularity. For access, a change to the second most popular site, Killykeen, would result in the largest economic increase. Conversely, the same change at the most popular site, Garadice, would add the least additional value. This is largely as a result of Garadice's already high rating for site *access* (mean of 4.27).

In comparison to similar studies, these results are complementary however there are few relevant studies to compare many of the variables of interest. Although previous research has stated that *encounters with other anglers* is often a determinant of angler participation (Hunt 2005), no comparable study has examined this attribute. Only one other Irish study has looked at access as a determinant of fishing participation. Curtis and Breen (2017) found that both having a carpark and disabled access increased the number of trips taken per year, which is in line with our results. Contrary to the finding of the paper presented in chapter 2, local services have been found to have a significant impact on days spent fishing (Curtis and Been 2017). This difference may be attributable to the differences in samples. Curtis and Breen's (2017) work contained overseas anglers, which in previous iteration of our work had a positive preference for local services. It may well be the case that if we used both Irish and overseas anglers it would lead to the same insignificant impact for local service as found in their research. The insignificant parameter estimate for the variable quantity of fish was also found in papers by Curtis and Stanley (2016) and Curtis and Breen (2017). Curtis and Breen (2017) also found the number of specimen fish to be a positive determinant of angling participation, similar to our findings that the *size of fish* and *quantity of fish* was not significant.

In relation to the goals of IFI, the most noteworthy findings from the first paper are the importance of attributes for site choice and the fact that not all sites will benefit equally from improvements. However, for this research to be of use a consideration must be explored; do the management of Irish fisheries view the site attributes in a similar manner to the users. If not, then management may not be able to use their own perception to choose sites that would most benefit from improvements. This is explored in the second paper, presented in chapter 3.

For chapter 3, a second data set was collected to contrast the original data set of users' perceptions of site attributes. For this data set, the two managers of the sites of interest were asked to rate the site as the users had done before them. Random parameter logit models were then applied to the original user data set and the manager data set, both of which use users' trip frequencies. Parameter estimates were then compared.

The results of the models from the two data sets were similar with respect to direction and significance. However, due to magnitude difference, very few of the parameter estimates could be considered statistically similar. Compensating variation for site closure lead to a similar ranked order of the sites. However, the manager-based results valued the sites higher than user-based results. Further analysis was conducted to determine if the manager-based results were similar to any particular user group in the dataset. This comparison indicates that the manager-based results are most closely aligned to the results based on anglers who visit the sites most often.

The results from the existing literature have not been unanimous with some results indicating that models solely using subjective data generally outperformed models using objective data (Adamowicz et al., 1997; Chasco and Gallo, 2013), or, that, objective variables had no statistically significant impact on the dependent variable (Farr et al., 2016; Lee et al., 2017). However, there are some examples where objective data outperforms subjective data (Poor et al., 2001; Baranzini et al., 2010). Only two papers compared subjective and objective sources of data for site choice models (Adamowicz et al., 1997; Joen et al., 2005). However, only Adamowicz et al. (1997) compared identical site attributes from the two sources. Their finding are very similar to those found within the paper presented in Chapter 4. Adamowicz et al. (1997) also found that the models applied to the subjective data outperformed those applied to the models applied to the objective data.

With regard to the applicability of this research to policy development, the results of this paper have two clear implications. Firstly, if the management ratings are similar enough to the user's, such that there are only small magnitude differences, it may be the case that future data collection for site choice data need only ask individuals for their trip frequencies. This would require much less effort on the part of both the respondent and the data collector. This reduced effort may also allow for more sites in the choice set, something that may have enhanced the current survey. Secondly, it would seem that policy implemented based on management's perspective of the sites would be in line with one based on the collective opinion of the users, removing the need to collect the labour-intensive data used for the comparison model.

The third paper departs somewhat from the obviously policy orientated analysis to the more methodological. Most data sets are afflicted by missing data of some sort. Choice data is no different. However, to date, no extensive comparison has been made between the various methods of dealing with this pervasive problem. This paper aims to deal with this problem by comparing four techniques, complete case analysis, per person mean imputation, per observation mean imputation and multiple imputations. Data were constructed so that "true" parameter estimates could be generated, which over varying percentages of missingness, could be compared to the estimates generated from the four missing data techniques.

For low percentages of missingness (less than 18%) per person mean imputation seemed to fare best but not across all metrics. As the percentage of missingness increased a trend developed with multiple imputations performing best at the greatest amount of missingness. In contrast, it is clear that complete case analysis led to the worst result, by some margin, over most metrics. This has serious implication for the question asked during the data collection process for data of this kind and similar data using objective measures. It would seem to suggest that serious bias can be created by removing choices from the choice set of respondents. It may then be important to get a full understanding of the choice set an individual possess before analysis beings.

In comparison to other papers that have used similar techniques, the results presented in this chapter are not as favourable to MI. MI performed best in several other papers (Nakai et al. 2014; Zhu 2014) CC, which performed well in other papers (Ali et al. 2011; Zhu 2014) performed the worst on almost all metrics used in this chapter. However, the poor performance of CC was, in some respects, to be expected given the type of data used for analysis.

The fourth and final paper returns to an empirical exploration of coarse angler behaviour. Using contingent behaviour data this paper examines how individuals would change their angling participation given a change in either the quantity of fish or the number of specimen fish at Garadice, the most popular of the five sites, for both Irish and overseas anglers. Marginal changes are estimated for groups within the sample to determine how responsive they would be to changes in fish quality. Additionally, a travel cost model was employed to determine the predictors of the number of fishing days. Consumer surpluses estimates were calculated to measure the additional value of a day spent coarse fishing in Garadice.

The results suggest that, in the case of both a change in the quantity of fish and the number of specimen fish would result in anglers being willing to take more days fishing, ranging from 3.78 days for a 25% increase in the number of specimen fish to 6.56 days for a 50% increase in the quantity of fish. Under status quo conditions, for the sampled anglers, the consumer surplus was €93 per day, and years fishing, being a match or specimen angler, targeting 3 or more fish species, whether or not the respondent filled out the survey online or at a site, as well as being Irish, all contributed to the likelihood of taking more days fishing.

This study aimed to examine the recurring but somewhat surprising result that suggests that the quantity of fish does not have a positive impact on coarse angler participation but the size of fish does (Curtis and Stanley 2016; Curtis and Breen 2017; Deely et a., 2018). As such there are no comparable Irish contingent behaviour models. However, the results do differ from the international literature (Poor and Breece 2006; Alberini et al., 2007; Prayaga et al., 2010). This could be due to a host of reasons such as the difference in fish species being targeted, difference in samples or difference in the percentages of change used for the CB question.

Combined, these papers demonstrate which attributes are important in site selection, that objective management based data leads to similar policy as users based data, that individuals do receive additional benefits beyond the cost of their trip and that the average angler would be willing to spend more days fishing if fish quality improved. From a policy perspective, management may wish to improve sites that are popular but not extremely well developed. Improving access at sites that have a variety of coarse fish, but that is situated away from areas with large town or villages may attract the largest number of anglers. As the quantity and number of specimen fish had comparable returns in terms of extra days spent at Garadice if a choice had to be made over a policy that would increase the number of specimen fish or quantity it would seem sensible to employ the method that is least burdensome for the current anglers.

From a methodological perspective, this thesis has added to the current literature on the use of objective or subjective data, finding that, in this case, it is a reasonable replacement. It has also been the first to truly compare the use of missing data techniques as applied to site choice data. The results seem to suggest that for low percentages of missingness (less than 18%) per person mean imputation works best but for larger percentages of missingness multiple imputations should be used.

6.2 Limitations and Further Research

Some of the limitations of this research should be highlighted. The data itself, had a reasonably small response rate, in terms of the number of respondents, making it difficult to extrapolate to a national level. It also seems to be the case that the respondents spend many more days fishing than the average angler. As a consequence, the results may be skewed towards the perspective of the more avid angler. The overseas anglers also had to be dropped from the analysis of papers 1, 2 and 3 due to missing information for this group from the survey. These individuals were not asked where they stayed during their trip to Ireland meaning that it was impossible to know, on any given choice occasion, the travel cost associated with one site as opposed to the others.

As examined by Hicks et al. (1999) and Peters et al. (1995), the consideration site can have a large impact on parameter estimates. Within this thesis, it was assumed that all respondents had the five sites of interest in their choice sets. It may be the case that due to differences in locations and accessibility, as well as preferences, anglers may restrict their consideration set to a subset of the five possible sites. The Independent Availability Logit (IAL) model (Habib et al. 2013) is a modelling approach that could test this hypothesis and is interesting area for future research. Although not suggested by focus groups, pilot studies, fisheries expert or the reviewers of paper 4, it could be argued that there is a logical inconsistency in the contingent behaviour question. It states that “If there were a 25% increase in the quantity of fish at Garadice how many days would you spend fishing there?” and also asks “If there were a 25% increase in the number of specimen sized fish at Garadice how many days would you spend fishing there?”. It could be argued, from the first question that if the quantity of fish increased so too would the number of specimen fish. If everyone thought of the question in this manner, then the increase in days, due to an increase in quantity, could be attributable, nearly entirely, to an increase in the number of specimen fish given the proximity of the two sets of results. Given that this was not pointed during any of the testing stages and the fact that a 25% increase in the number of specimen fish induced more fishing days than a 25% increase in quantity (this should not logically happen if the above logic was used by respondents), the questions seem not to have been interpreted in this way by most people.

A concern with the results of the contingent behaviour analysis is that there is no information on where these extra fishing days are coming from. If the anglers are substituting days fishing at one site for another it may be of little use in the effort to increase the number of fishing days spent in Ireland. In fact, it may be the case that if a national project was taken which resulted in better fish quality at all sites in Ireland no more days fishing may be taken if anglers were unwilling to substitute other activities for fishing days. However, it would still be the case that some additional consumer surplus would be created. It is also worth noting that, for the contingent behaviour analysis, trips made by overseas anglers have been treated as single purpose when this may not be the case.

Individuals who did not rate some of the site attributes but attended the site may be revealing something about the contribution of this/these variable/s to their decision-making process. It would seem that if an individual would pick a site without the knowledge of some attribute then any correlations between this attribute and site choice are probably coincidental. Additionally, just because an individual is able to rate an attribute it does not mean it affects site choice. It may be of use, in future works, to see how individuals rate the importance of site attributes in comparison to a ranked order of the absolute value of the parameter estimates. Large differences may indicate coincidental correlation, or that the attribute is correlated with something outside of the "attributes of interest" rather than meaningful decision making.

The comparison between the objective and subjective model results also had limitations. The starting point of this analysis was getting the users to rate the sites, this can be a laborious task for the respondents and as such the choice set was constructed using few sites. This, in turn, meant that the objective data was limited to five sites. As a consequence, there was a lack of variability between sites according to the objective ratings. One of the benefits of using objective data is the ease at which the data can be collected for numerous sites. A complete comparison using a larger data set for the objective data, which would be more consistent with other published works, and a smaller choice set for the subjective data, may have given analyst a better understanding of the trade-offs they are making by choosing one set of data over the other. To this point, although the results of paper 3 are stated as being similar, two of the six variables of interest are missing from the objective results, which, in all reality may have a huge impact on policy formation. Future work may want to compare choice sets of unequal dimensions to determine how much accuracy is been given up using one method over another.

Future works may also want to consider the difference between an information finding trip and a trip made based on knowledge of the site. In a case where an individual chooses one site over another, it is assumed that the individual prefers that site. However, if an individual has never visited the site before, they may not be aware of the attributes the site possesses and whether they prefer it to other sites. To remedy this, individuals could be asked, if they had only been once to a site, if they would return. If they would not return, then the analysts may consider removing this choice occasion from the data.

Bibliography:

Adamowicz, W., Swait, J., Boxall, P., Louviere, J., & Williams, M. (1997). Perceptions versus objective measures of environmental quality in combined revealed and stated preference models of environmental valuation. *Journal of environmental economics and management*, *32*(1), 65-84.

Akaike, H. (1998). Information theory and an extension of the maximum likelihood principle. In *Selected papers of hirotugu akaike* (pp. 199-213). Springer, New York, NY..

Alberini, A., Zanatta, V., & Rosato, P. (2007). Combining actual and contingent behavior to estimate the value of sports fishing in the Lagoon of Venice. *Ecological Economics*, *61*(2-3), 530-541.

Ali, A. M. G., Dawson, S. J., Blows, F. M., Provenzano, E., Ellis, I. O., Baglietto, L., ... & Pharoah, P. D. (2011). Comparison of methods for handling missing data on immunohistochemical markers in survival analysis of breast cancer. *British journal of cancer*, *104*(4), 693.

Allison, P. D. (2001). *Missing data* (Vol. 136). Sage publications. Chapter 4

Arlinghaus, R., & Mehner, T. (2005). Determinants of management preferences of recreational anglers in Germany: habitat management versus fish stocking. *Limnologica*, *35*(1-2), 2-17.

Artell, J., Ahtiainen, H., & Pouta, E. (2013). Subjective vs. objective measures in the valuation of water quality. *Journal of environmental management*, *130*, 288-296.

Baranzini, A., Schaerer, C., & Thalmann, P. (2010). Using measured instead of perceived noise in hedonic models. *Transportation Research Part D: Transport and Environment*, *15*(8), 473-482.

Barry, L., van Rensburg, T. M., & Hynes, S. (2011). Improving the recreational value of Ireland's coastal resources: A contingent behavioural application. *Marine Policy*, *35*(6), 764-771.

Berners, J. (1880). *A Treatyse of Fysshynge wyth an Angle*. E. Stock.

Bilgic, A., & Florkowski, W. J. (2007). Application of a hurdle negative binomial count data model to demand for bass fishing in the southeastern United States. *Journal of Environmental Management*, 83(4), 478-490.

Bousquet, A., & Ivaldi, M. (1998). An individual choice model of energy mix. *Resource and Energy Economics*, *20*(3), 263-286.

Boxall, P. C., & Adamowicz, W. L. (2002). Understanding heterogeneous preferences in random utility models: a latent class approach. *Environmental and resource economics*, *23*(4), 421-446.

Brewer, N. T., Weinstein, N. D., Cuite, C. L., & Herrington, J. E. (2004). Risk perceptions and their relation to risk behavior. *Annals of behavioral medicine*, *27*(2), 125-130.

Brownstone, D., Golob, T. F., & Kazimi, C. (2002). Modelling non-ignorable attrition and measurement error in panel surveys: an application to travel demand modeling.

Cameron, A. C., & Trivedi, P. K. (1986). Econometric models based on count data. Comparisons and applications of some estimators and tests. *Journal of applied econometrics*, *1*(1), 29-53.

Cameron, T. A., Shaw, W. D., Ragland, S. E., Mac Callaway, J., & Keefe, S. (1996). Using actual and contingent behavior data with differing levels of time aggregation to model recreation demand. *Journal of Agricultural and Resource Economics*, 130-149.

Changyong, F. E. N. G., Hongyue, W. A. N. G., Naiji, L. U., Tian, C. H. E. N., Hua, H. E., & Ying, L. U. (2014). Log-transformation and its implications for data analysis. *Shanghai archives of psychiatry*, *26*(2), 105.

Chasco, C., & Gallo, J. L. (2013). The impact of objective and subjective measures of air quality and noise on house prices: a multilevel approach for downtown Madrid. *Economic Geography*, *89*(2), 127-148.

Christie, M., Hanley, N., & Hynes, S. (2007). Valuing enhancements to forest recreation using choice experiment and contingent behaviour methods. *Journal of Forest Economics*, *13*(2-3), 75-102.

Clogg, C. C., Petkova, E., & Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology*, *100*(5), 1261-1293.

Conover, W. J., & Conover, W. J. (1980). Practical nonparametric statistics.

Curtis, J., & Breen, B. (2017). Irish coarse and game anglers' preferences for fishing site attributes. *Fisheries Research*, *190*, 103-112.

Curtis, J., & Stanley, B. (2016). Water quality and recreational angling demand in Ireland. *Journal of Outdoor Recreation and Tourism*, *14*, 27-34.

Curtis, J., & Breen, B. (2017). Irish coarse and game anglers' preferences for fishing site attributes. *Fisheries Research*, *190*, 103-112.

Daly, A., Hess, S., & Train, K. (2012). Assuring finite moments for willingness to pay in random coefficient models. *Transportation*, *39*(1), 19-31.

Deely, J., Hynes, S., & Curtis, J. (2018). Coarse angler site choice model with perceived site attributes. *Journal of Outdoor Recreation and Tourism*. https://doi.org/10.1016/j.jort.2018.07.001

Deely, J., Hynes, S., & Curtis, J. (2019). Are objective data an appropriate replacement for subjective data in site choice analysis?. *Journal of Environmental Economics and Policy*, *8*(2), 159-178.

Downey, R. G., & King, C. V. (1998). Missing data in Likert ratings: A comparison of replacement methods. *The Journal of general psychology*, *125*(2), 175-191.

du Preez, M., & Hosking, S. G. (2011). The value of the trout fishery at Rhodes, North Eastern Cape, South Africa: a travel cost analysis using count data models. *Journal of Environmental Planning and Management*, *54*(2), 267-282.

Duffield, J. W., Neher, C. J., & Merritt, M. F. (2001). Alaska angler survey: Use and valuation estimates for 1995, with a focus on Tanana Valley major stocked waters. Alaska Department of Fish and Game, Division of Sport Fish, Research and Technical Services

Egan, K., & Herriges, J. (2006). Multivariate count data regression models with individual panel data from an on-site sample. *Journal of Environmental Economics and Management*, 52(2), 567-581.

Egan, K., & Herriges, J. A. (2003). Mixed Poisson regression models with individual panel data from an on-site sample. *Unpublished manuscript. Department of Economics, Iowa State University*.

Eiswerth, M. E., Englin, J., Fadali, E., & Shaw, W. D. (2000). The value of water levels in water-based recreation: A pooled revealed preference/contingent behavior model. *Water Resources Research*, *36*(4), 1079-1086.

Eiswerth, M. E., Kashian, R. D., & Skidmore, M. (2008). Examining angler behavior using contingent behavior modeling: A case study of water quality change at a Wisconsin lake. *Water Resources Research*, *44*(11).

Englin, J., & Cameron, T. A. (1996). Augmenting travel cost models with contingent behavior data. *Environmental and Resource Economics*, *7*(2), 133-147.

Englin, J., & Shonkwiler, J. S. (1995). Estimating social welfare using count data models: an application to long-run recreation demand under conditions of endogenous stratification and truncation. *The Review of Economics and Statistics*, 104-112.

Erdem, T., Keane, M. P., & Sun, B. (1998). Missing price and coupon availability data in scanner panels: Correcting for the self-selection bias in choice model parameters. *Journal of Econometrics*, *89*(1-2), 177-196.

Farr, M., Stoeckl, N., Esparon, M., Larson, S., & Jarvis, D. (2016). The importance of water clarity to Great Barrier Reef tourists and their willingness to pay to improve it. *Tourism Economics*, *22*(2), 331-352.

Feather, P., & Shaw, W. D. (1999). Estimating the cost of leisure time for recreation demand models. *Journal of Environmental Economics and Management*, *38*(1), 49-65.

Filippini, M., Greene, W., & Martinez-Cruz, A. L. (2018). Non-market value of winter outdoor recreation in the Swiss Alps: the case of Val Bedretto. *Environmental and resource economics*, *71*(3), 729-754.

Fletcher, J. J., Adamowicz, W. L., & Graham-Tomasi, T. (1990). The travel cost model of recreation demand: theoretical and empirical issues. *Leisure Sciences*, *12*(1), 119-147.

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention science*, *8*(3), 206-213.

Grijalva, T. C., Berrens, R. P., Bohara, A. K., & Shaw, W. D. (2002). Testing the validity of contingent behavior trip responses. *American Journal of Agricultural Economics*, *84*(2), 401-414.

Grilli, G., Curtis, J., Hynes, S., & O'Reilly, P. (2018). Sea bass angling in Ireland: A structural equation model of catch and effort. *Ecological Economics*, *149*, 285-293.

Grogger, J. T., & Carson, R. T. (1991). Models for truncated counts. *Journal of Applied Econometrics*, *6*(3), 225-238.

Haab, T. C., & Hicks, R. L. (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 McConnell, K.E. (2002) Valuing Environmental and Natural Resources: The Econometrics of Non-Market Valuation. Edward Elgar Publishing, Northampton.

Habib, K. N., Morency, C., Trépanier, M., & Salem, S. (2013). Application of an independent availability logit model (IAL) for route choice modelling: Considering bridge choice as a key determinant of selected routes for commuting in Montreal. *Journal of choice modelling*, *9*, 14-26.

Hanemann, W. M. (1982). Applied welfare analysis with qualitative response models. *Department of Agricultural & Resource Economics, UCB*.

Hanley, N., Bell, D., & Alvarez-Farizo, B. (2003). Valuing the benefits of coastal water quality improvements using contingent and real behaviour. *Environmental and Resource Economics*, *24*(3), 273-285.

Hanley, N., Koop, G., Alvarez-Farizo, B., Wright, R. E., & Nevin, C. (2001). Go climb a mountain: an application of recreation demand modelling to rock climbing in Scotland. *Journal of Agricultural Economics*, *52*(1), 36-52.

Hausman, J. A., Hall, B. H., & Griliches, Z. (1984). Econometric models for count data with an application to the patents-R&D relationship.

Hawthorne, G., Hawthorne, G., & Elliott, P. (2005). Imputing cross-sectional missing data: comparison of common techniques. *Australian & New Zealand Journal of Psychiatry*, *39*(7), 583-590.

Hellerstein, D., & Mendelsohn, R. (1993). A theoretical foundation for count data models. *American Journal of Agricultural Economics*, *75*(3), 604-611.

Hensher, D. A., & Greene, W. H. (2003). The mixed logit model: the state of practice. *Transportation*, *30*(2), 133-176.

Hicks, J. R. (1942). Consumers' surplus and index-numbers. *The Review of Economic Studies*, *9*(2), 126-137.

Hicks, R. L., & Strand, I. E. (2000). The extent of information: its relevance for random utility models. Land Economics, 374-385.

Hill, R., Loomis, J., Thilmany, D., & Sullins, M. (2014). Economic values of agritourism to visitors: a multi-destination hurdle travel cost model of demand. *Tourism Economics*, *20*(5), 1047-1065.

Hindsley, P., Landry, C. E., & Gentner, B. (2011). Addressing onsite sampling in recreation site choice models. *Journal of Environmental Economics and Management*, *62*(1), 95-110.

Hole, A. R. (2007). A comparison of approaches to estimating confidence intervals for willingness to pay measures. *Health economics*, *16*(8), 827-840.

Hotelling, H. (1947). Letter to the national park service. *An Economic Study of the Monetary Evaluation of Recreation in the National Parks (US Department of the Interior, National Park Service and Recreational Planning Division, 1949)*.

Hoyos, D., & Riera, P. (2013). Convergent validity between revealed and stated recreation demand data: Some empirical evidence from the Basque Country, Spain. *Journal of forest economics*, *19*(3), 234-248.

Hunt, K. M., Floyd, M. F., & Ditton, R. B. (2007) African-American and Anglo anglers' attitudes toward the catch-related aspects of fishing. *Human Dimensions of Wildlife*, *12*(4), 227-239.

Hunt, L. M. (2005). Recreational fishing site choice models: insights and future opportunities. *Human Dimensions of Wildlife*, *10*(3), 153-172.

Hunt, L. M., Boots, B. N., & Boxall, P. C. (2007). Predicting fishing participation and site choice while accounting for spatial substitution, trip timing, and trip context. *North American Journal of Fisheries Management*, *27*(3), 832-847.

Hynes, S., & Hanley, N. (2006). Preservation versus development on Irish rivers: whitewater kayaking and hydro-power in Ireland. *Land Use Policy*, *23*(2), 170-180.

Hynes, S., & Greene, W. (2013). A panel travel cost model accounting for endogenous stratification and truncation: a latent class approach. *Land Economics*, *89*(1), 177-192.

Hynes, S., & Greene, W. (2016). Preference Heterogeneity in Contingent Behaviour Travel Cost Models with On-site Samples: A Random Parameter vs. a Latent Class Approach. *Journal of Agricultural Economics*, *67*(2), 348-367.

Hynes, S., Gaeven, R., & O'Reilly, P. (2017). Estimating a total demand function for sea angling pursuits. *Ecological Economics*, *134*, 73-81.

Hynes, S., Hanley, N., & Scarpa, R. (2008). Effects on welfare measures of alternative means of accounting for preference heterogeneity in recreational demand models. *American journal of agricultural economics*, *90*(4), 1011-1027.

Hynes, S., O'Reilly, P., & Corless, R. (2015). An on-site versus a household survey approach to modelling the demand for recreational angling: Do welfare estimates differ?. *Ecosystem Services*, *16*, 136-145.

Hynes, S., Hanley, N., & O'Donoghue, C. (2009). Alternative treatments of the cost of time in recreational demand models: an application to whitewater kayaking in Ireland. *Journal of Environmental Management*, *90*(2), 1014-1021.

Hynes, S., Hanley, N., & Scarpa, R. (2008). Effects on welfare measures of alternative means of accounting for preference heterogeneity in recreational demand models. *American journal of agricultural economics*, *90*(4), 1011-1027.

Inland fisheries Ireland (2015). Fish stocking Guidance Document, IFI Publication, Dublin https://www.fisheriesireland.ie/documents/620-ifi-fish-stocking-guidance-document/file.html Accessed 1.11.2018

Inland Fisheries Ireland, 2015. The Economic Contribution of Coarse Angling in Ireland. IFI publication, Dublin. https://www.fisheriesireland.ie/national-strategy-for-angling-development/socio-economics/529-the-economic-contribution-of-coarse-angling-in-ireland/file.html Accessed 07.06.2019

Jeon, Y., Herriges, J. A., Kling, C. L., & Downing, J. (2005). The role of water quality perceptions in modeling lake recreation demand. *The international handbook on non-market environmental valuation*, 74-101.

Johnson, N. S., & Adams, R. M. (1989). On the marginal value of a fish: some evidence from a steelhead fishery. *Marine Resource Economics*, *6*(1), 43-55.

Johnston, R. J., Ranson, M. H., Besedin, E. Y., & Helm, E. C. (2006). What determines willingness to pay per fish? A meta-analysis of recreational fishing values. *Marine Resource Economics*, *21*(1), 1-32.

Kirtland, K. A., Porter, D. E., Addy, C. L., Neet, M. J., Williams, J. E., Sharpe, P. A., ... & Ainsworth, B. E. (2003). Environmental measures of physical activity supports: perception versus reality. *American journal of preventive medicine*, *24*(4), 323-331.

Kolmogorov, A. (1933). Sulla determinazione empirica di una lgge di distribuzione. *Inst. Ital. Attuari, Giorn.*, *4*, 83-91.

Krinsky, I., & Robb, A. L. (1986). On approximating the statistical properties of elasticities. *The Review of Economics and Statistics*, 715-719.

Kuosmanen, T., Nillesen, E., & Wesseler, J. (2004). Does ignoring multidestination trips in the travel cost method cause a systematic bias?. *Australian Journal of Agricultural and Resource Economics*, *48*(4), 629-651.

Kuriyama, K., Hilger, J., & Hanemann, M. (2013). A random parameter model with onsite sampling for recreation site choice: an application to southern California shoreline sportfishing. *Environmental and Resource Economics*, *56*(4), 481-497.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of political economy*, *74*(2), 132-157.

Lee, S. M., Conway, T. L., Frank, L. D., Saelens, B. E., Cain, K. L., & Sallis, J. F. (2017). The relation of perceived and objective environment attributes to neighborhood satisfaction. *Environment and behavior*, *49*(2), 136-160.

Li, L., Adamowicz, W., & Swait, J. (2015). The effect of choice set misspecification on welfare measures in random utility models. Resource and Energy Economics, 42, 71-92.

Li, Y., Li, Z., & Li, L. (2014). Missing traffic data: comparison of imputation methods. *IET Intelligent Transport Systems*, *8*(1), 51-57.

Lienhoop, N., & Ansmann, T. (2011). Valuing water level changes in reservoirs using two stated preference approaches: An exploration of validity. *Ecological economics*, *70*(7), 1250-1258.

Liu, F. C. (2014). Using Multiple Imputation for Vote Choice Data: A Comparison across Multiple Imputation Tools. *Open Journal of Political Science*, *4*(02), 39.

Loomis, J., Yorizane, S., & Larson, D. (2000). Testing significance of multi-destination and multi-purpose trip effects in a travel cost method demand model for whale watching trips. *Agricultural and Resource Economics Review*, *29*(2), 183-191.

Lucas, R. C. (1963). Bias in estimating recreationists' length of stay from sample interviews. *Journal of Forestry*, *61*(12), 912-914.

Ma, L., & Dill, J. (2016). Do people's perceptions of neighborhood bikeability match "Reality"?. *Journal of Transport and Land Use*, *10*(1).

Manski, C. F., & Lerman, S. R. (1977). The estimation of choice probabilities from choice based samples. *Econometrica: Journal of the Econometric Society*, 1977-1988.

Martínez-Espiñeira, R., & Amoako-Tuffour, J. (2009). Multi-destination and multi-purpose trip effects in the analysis of the demand for trips to a remote recreational site. *Environmental Management*, *43*(6), 1146-1161.

Martinson, K. S., & Shelby, B. (1992). Encounter and proximity norms for salmon anglers in California and New Zealand. *North American Journal of Fisheries Management*, *12*(3), 559-567.

McConnell, K. E., & Tseng, W. C. (1999). Some preliminary evidence on sampling of alternatives with the random parameters logit. *Marine Resource Economics*, *14*(4), 317-332.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.

Nahuelhual, L., Loureiro, M. L., & Loomis, J. (2004). Using random parameters to account for heterogeneous preferences in contingent valuation of public open space. *Journal of Agricultural and Resource Economics*, 537-552.

Nakai, M., Chen, D. G., Nishimura, K., & Miyamoto, Y. (2014). Comparative study of four methods in missing value imputations under missing completely at random mechanism. *Open Journal of Statistics*, *4*(01), 27.

NSAD (2015). National strategy for angling development: Angling product evaluation criteria. http://www.fisheriesireland.ie/550-angling-product-evaluation-criteria/file Accessed 07.09.2017

NSAD (2016). National strategy for angling development: Policy document for coarse angling. http://www.ncffi.ie/new/wp-content/uploads/2016/01/NSAD-Submission-Jan-2016.pdf Accessed 07.09.2017

Parsons, G. R. (2003). The travel cost model. In A primer on nonmarket valuation (pp. 269-329). Springer, Dordrecht.

Parsons, G. R., Massey, D. M., & Tomasi, T. (1999). Familiar and favorite sites in a random utility model of beach recreation. *Marine Resource Economics*, *14*(4), 299-315.

Peters, T., Adamowicz, W. L., & Boxall, P. C. (1995). Influence of choice set considerations in modelling the benefits from improved water quality. *Water Resources Research*, *31*(7), 1781-1787.

Poor, P. J., & Breece, M. (2006). The contingent behavior of charter fishing participants on the Chesapeake Bay: welfare estimates associated with water quality improvements. *Journal of Environmental Planning and Management*, *49*(2), 265-278.

Poor, P. J., Boyle, K. J., Taylor, L. O., & Bouchard, R. (2001). Objective versus subjective measures of water clarity in hedonic property value models. *Land Economics*, *77*(4), 482-493.

Prayaga, P., Rolfe, J., & Stoeckl, N. (2010). The value of recreational fishing in the Great Barrier Reef, Australia: a pooled revealed preference and contingent behaviour model. *Marine Policy*, *34*(2), 244-251.

Provencher, B., Baerenklau, K. A., & Bishop, R. C. (2002). A finite mixture logit model of recreational angling with serially correlated random utility. *American Journal of Agricultural Economics*, *84*(4), 1066-1075.

Puto, C. P. (1987). The framing of buying decisions. *Journal of consumer research*, *14*(3), 301-315.

Raguragavan, J., Hailu, A., & Burton, M. (2013). Economic valuation of recreational fishing in Western Australia: statewide random utility modelling of fishing site choice behaviour. *Australian Journal of Agricultural and Resource Economics*, *57*(4), 539-558.

Rolfe, J., & Prayaga, P. (2007). Estimating values for recreational fishing at freshwater dams in Queensland. *Australian Journal of Agricultural and Resource Economics*, *51*(2), 157-174.

Royston, P., 2004. Multiple imputation by the MICE system of chained equations. *MRC Clinical Trials Unit, London*, *35*.

Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. Handbook of Statistical Modeling for the Social and Behavioral Sciences.

Saengavut, V. (2018). The Effect of Preference for Nature-Based Recreations: Application of a Multi-Destination Travel Cost Method. *Chiang Mai University Journal of Economics*, *22*(1), 1-16.

Sanko, N., Hess, S., Dumont, J., & Daly, A. (2014). Contrasting imputation with a latent variable approach to dealing with missing income in choice models. *Journal of choice modelling*, *12*, 47-57.

Scarpa, R., & Thiene, M. (2005). Destination choice models for rock climbing in the Northeastern Alps: a latent-class approach based on intensity of preferences. *Land economics*, *81*(3), 426-444.

Samuelson, P. A. (1948). Consumption theory in terms of revealed preference. *Economica*, *15*(60), 243-253.

Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods*, *7*(2), 147.

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461-464.

Shaw, D. (1988). On-site samples' regression: Problems of non-negative integers, truncation, and endogenous stratification. *Journal of Econometrics*, *37*(2), 211-223.

Shaw, W. D., & Ozog, M. T. (1999). Modelling overnight recreation trip choice: application of a repeated nested multinomial logit model. *Environmental and Resource Economics*, *13*(4), 397-414.

Shrestha, R. K., Seidl, A. F., & Moraes, A. S. (2002). Value of recreational fishing in the Brazilian Pantanal: a travel cost analysis using count data models. *Ecological Economics*, *42*(1-2), 289-299.

Shrive, F. M., Stuart, H., Quan, H., & Ghali, W. A. (2006). Dealing with missing data in a multi-question depression scale: a comparison of imputation methods. *BMC medical research methodology*, *6*(1), 57.

Simões, P., Barata, E., & Cruz, L. (2013). Joint estimation using revealed and stated preference data: an application using a national forest. *Journal of Forest Economics*, *19*(3), 249-266.

Singh, H. (1988). Investigating the compatibility of econometric forecasts and subjective expectations: A suggested framework. *Journal of economic psychology*, *9*(2), 233-249.

Steimetz, S. S., & Brownstone, D. (2005). Estimating commuters'"value of time" with noisy data: a multiple imputation approach. *Transportation Research Part B: Methodological*, *39*(10), 865-889.

TDI (2013). TDI Socio-Economic Study of Recreational Angling in Ireland
http://www.fisheriesireland.ie/media/tdistudyonrecreationalangling.pdf
Accessed 1.11.2018

Train, K. E. (1998). Over People. *Land economics*, *74*(2), 230-39.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Washington, S., Ravulaparthy, S., Rose, J. M., Hensher, D., & Pendyala, R. (2014). Bayesian imputation of non-chosen attribute values in revealed preference surveys. *Journal of Advanced Transportation*, *48*(1), 48-65.

Whitehead, J. C. (1994). Item nonresponse in contingent valuation: should CV researchers impute values for missing independent variables?. *Journal of Leisure Research*, *26*(3), 296-303.

Whitehead, J. C., Huang, J. C., Blomquist, G. C., & Ready, R. C. (1998). Construct validity of dichotomous and polychotomous choice contingent valuation questions. *Environmental and Resource Economics*, *11*(1), 107-116.

Wolkowitz, A. A., & Skorupski, W. P. (2013). A method for imputing response options for missing data on multiple-choice assessments. *Educational and Psychological Measurement*, *73*(6), 1036-1053.

Yeh, C. Y., Haab, T. C., & Sohngen, B. L. (2006). Modeling multiple-objective recreation trips with choices over trip duration and alternative sites. *Environmental and resource economics*, *34*(2), 189-209.

Zhu, X. (2014). Comparison of four methods for handing missing data in longitudinal data analysis through a simulation study. *Open Journal of Statistics*, *4*(11), 933.

Appendices

Appendix 2.1 Data Collection Survey

The Following is the survey used for data collection.

The Economic Value of Recreational Angling in Ireland: Questionnaire

Introduction and Aims of the Study

The Department of Economics at the National University of Ireland, Galway is currently carrying out a study on the economic value of recreational angling in Ireland.

The key aims of this study are as follows:

⇒ To gain insight into which factors/site characteristics influence coarse anglers in their choice of fishing destination.
⇒ To assess how these factors differ amongst angler groups.
⇒ To determine the value of a day spent fishing.
⇒ To gather information that will facilitate better management of Ireland's fisheries.

The following questionnaire has been designed in an attempt to satisfy these aims. Your cooperation in answering the questions below, as accurately as possible, would be greatly appreciated. All responses will be treated in confidence.

Definition of Coarse angling

In this questionnaire we are concerned with the fishing habits and preferences of coarse anglers. Coarse angling is defined by the species of fish the angler is targeting. Coarse fish include the following species: bream, tench, roach, rudd, hybrids, perch, eels, dace and carp.

Part 1: Angling Experience

1.1. Compared to your other outdoor recreational activities (such as walking, cycling, swimming etc.) how would you comparatively rate angling?

1. Your most important outdoor activity
2. Your second most important outdoor activity
3. Your third most important outdoor activity
4. Only one of many outdoor activities

1.2. Would you describe your proficiency (ability) level at angling as:

1. Basic 2. Intermediate 3. Advanced

1.3. Would you describe yourself as a:

1. Match angler 2. Specimen angler 3. Pleasure angler

1.4. How many years have you been fishing for?

_____________ YEARS

1.5. In the last 12 months have you fished abroad? (for those not living in Ireland please include trips to Ireland)

Yes No

1.6. If you have fished abroad in the past 12 months, what countries have you fished in?

Country	No. of Visits
1.	
2.	
3.	
4.	

1.7. Do you generally fish for:

1.Pike 2. Other coarse fish

1.8. Besides coarse fish, what other type of fish do you fish for?

1. Salmon
2. Sea trout
3. Brown trout
4. Pike
5. Bass
6. Other sea species
7. Other please specify
8. Only coarse fish

Part 2: Angling Activity and Choice of Fishing Sites over the past 12 Months

2.1.Which of the following fishing sites have you visited in the past 12 months?

- Please indicate, to the best of your recollection, how many days you have fished at each location.

Angling Site	Total Number of Days	Number of Days that were Competitions or Festivals
1.Killykeen Forest Park (Lough Oughter)		
2.Eonish (Lough Oughter)		
3.Dernaferst (Gowna)		
4.Church Lake (Gowna)		
5.Garadice Lough		
6.Days at other sites in Ireland		

Part 3: Evaluation of fisheries in Ireland

3.1. How would you describe each of the fisheries below, in regard to the following attributes?

- You may evaluate (rate) the fisheries you have visited at any stage throughout your angling experience, but please do not comment on sites you have never visited

Site 1 Killykeen Forest Park (Lough Oughter), County Cavan

Factor	Score/level of Factor				
Accessibility to the site (this includes parking and ability to reach the location that you fished at.) Score from 1 = very difficult to access to 5 = easily accessed	1 Difficult to access	2	3	4	5 Easy to access
Size of fish at the site (On average does this site provide access to good sized fish) Score from 1 = small fish to 5 = large fish	1 Small fish	2	3	4	5 Large fish
Quantity of fish (on average does this site provide access to a large quantity of fish) Score from 1 = low quantity to 5 = high quantity	1 Low quantity	2	3	4	5 High quantity
Encounters with other anglers Score from 1= none to 5 = frequent	1 No encounters	2	3	4	5 Frequent encounters
Variety of fish species (are there a large variety of species of fish at this site) score from 1 = low level of variety of fish to 5 = high level of variety of fish	1 Little to no variety	2	3	4	5 Lots of variety
Local services (these include pub, shops, accommodation etc…) Score from 1 = low level of local services to 5 = high level of services	1 Lacks local Services	2	3	4	5 Plenty of local services

Site 2: Eonish (Lough Oughter), County Cavan

Factor	Score/level of Factor				
Accessibility to the site (this includes parking and ability to reach the location that you fished at.) Score from 1 = very difficult to access to 5 = easily accessed	1 Difficult to access	2	3	4	5 Easy to access
Size of fish at the site (On average does this site provide access to good sized fish) Score from 1 = small fish to 5 = large fish	1 Small fish	2	3	4	5 Large fish
Quantity of fish (on average does this site provide access to a large quantity of fish) Score from 1 = low quantity to 5 = high quantity	1 Low quantity	2	3	4	5 High quantity
Encounters with other anglers Score from 1= none to 5 = frequent	1 No encounters	2	3	4	5 Frequent encounters
Variety of fish species (are there a large variety of species of fish at this site) score from 1 = low level of variety of fish to 5 = high level of variety of fish	1 Little to no variety	2	3	4	5 Lots of variety
Local services (these include pub, shops, accommodation etc…) Score from 1 = low level of local services to 5 = high level of services	1 Lacks local Services	2	3	4	5 Plenty of local services

Site 3: Dernaferst (Gowna), County Cavan

Factor	Score/level of Factor				
Accessibility to the site (this includes parking and ability to reach the location that you fished at.) Score from 1 = very difficult to access to 5 = easily accessed	1 Difficult to access	2	3	4	5 Easy to access
Size of fish at the site (On average does this site provide access to good sized fish) Score from 1 = small fish to 5 = large fish	1 Small fish	2	3	4	5 Large fish
Quantity of fish (on average does this site provide access to a large quantity of fish) Score from 1 = low quantity to 5 = high quantity	1 Low quantity	2	3	4	5 High quantity
Encounters with other anglers Score from 1= none to 5 = frequent	1 No encounters	2	3	4	5 Frequent encounters
Variety of fish species (are there a large variety of species of fish at this site) score from 1 = low level of variety of fish to 5 = high level of variety of fish	1 Little to no variety	2	3	4	5 Lots of variety
Local services (these include pub, shops, accommodation etc…) Score from 1 = low level of local services to 5 = high level of services	1 Lacks local Services	2	3	4	5 Plenty of local services

Site 4: Church Lake (Gowna), County Leitrim

Factor	**Score/level of Factor**				
Accessibility to the site (this includes parking and ability to reach the location that you fished at.) Score from 1 = very difficult to access to 5 = easily accessed	1 Difficult to access	2	3	4	5 Easy to access
Size of fish at the site (On average does this site provide access to good sized fish) Score from 1 = small fish to 5 = large fish	1 Small fish	2	3	4	5 Large fish
Quantity of fish (on average does this site provide access to a large quantity of fish) Score from 1 = low quantity to 5 = high quantity	1 Low quantity	2	3	4	5 High quantity
Encounters with other anglers Score from 1= none to 5 = frequent	1 No encounters	2	3	4	5 Frequent encounters
Variety of fish species (are there a large variety of species of fish at this site) score from 1 = low level of variety of fish to 5 = high level of variety of fish	1 Little to no variety	2	3	4	5 Lots of variety
Local services (these include pub, shops, accommodation etc…) Score from 1 = low level of local services to 5 = high level of services	1 Lacks local Services	2	3	4	5 Plenty of local services

Site 5: Garadice Lough, County Leitrim

Factor	**Score/level of Factor**				
Accessibility to the site (this includes parking and ability to reach the location that you fished at.) Score from 1 = very difficult to access to 5 = easily accessed	1 Difficult to access	2	3	4	5 Easy to access
Size of fish at the site (On average does this site provide access to good sized fish) Score from 1 = small fish to 5 = large fish	1 Small fish	2	3	4	5 Large fish
Quantity of fish (on average does this site provide access to a large quantity of fish) Score from 1 = low quantity to 5 = high quantity	1 Low quantity	2	3	4	5 High quantity
Encounters with other anglers Score from 1= none to 5 = frequent	1 No encounters	2	3	4	5 Frequent encounters
Variety of fish species (are there a large variety of species of fish at this site) score from 1 = low level of variety of fish to 5 = high level of variety of fish	1 Little to no variety	2	3	4	5 Lots of variety
Local services (these include pub, shops, accommodation etc…) Score from 1 = low level of local services to 5 = high level of services	1 Lacks local Services	2	3	4	5 Plenty of local services

Part 4: Contingent Behaviour

In this section we would like to explore if Irish coarse anglers would change the number of days fishing they take per year, to Garadice Lough, if changes are made to certain site attributes.

Please indicate how many days you plan to take **next** year to Garadice Lough.

_______________Days

Please indicate in the table below, how many extra or fewer days you would make if each of the following changes were made e.g. 2 = two more days -1 = one fewer day.

Change at Garadice lough next year	**Change in number of days**
25% increase in the quantity of fish at each site	
50% increase in the quantity of fish at each site	
25% increase in the number of specimen sized fish at each site	
50% increase in the number of specimen sized fish at each site	
25% increase in the quantity of bream	
50% increase in the quantity of bream	

Part 5: Personal Expenditure on Angling

5.1. What has been your approximate spend on angling over **all** the days fishing you have taken in **Ireland** in past 12 months, in the following categories? Please outline your responses in the table below:

Category of Spend	**€ Spend over the last 12 months**	**% of each category spent in the location where you were fishing at**
Travel to sites (e.g. petrol/diesel, car hire etc)		
Flights and ferries		
Food and drink		
Accommodation		
Angling equipment excluding fishing tackle		
Fishing tackle		
Bait		
Angling licences/permits		
Magazines/guides/ books/ maps etc.		
Entry to competitions		
Other angling related expenditure		

Part 6: Classification Questions

6.1. What is your nationality?_______________

6.2. What country do you live in? _______________

6.3. What is the nearest town or village to where you live? _______________

6.4. What age are you? ___________ Years old

6.5. Are you: Male ______ Female ______

6.6. Marital Status:

Single ____ Married ____

Single with children ____ Married with children ____

Partnership ____ Divorced ____

Widowed ____

6.7. Level of Education (Please tick the highest level of education that you have obtained):

1. Third level education: degree/certificate/diploma or post graduate qualification
2. Secondary
3. Primary
4. Still in education

6.8. Which of the following categories best describes your employment status:

Employed, working full-time ______ Employed, working part-time ____

Seeking employment ______ Student ____

Retired ______ Currently not able to work ____

6.10. What is or was your occupation?

____

6.11. What is your approximate total income, before taxes? (Please highlight one)

1. Less than €15,000

2. €15,000 - €29,999

3. €30,000 - €44,999

4. €45,000 - €59,999

5. €60,000 - €74,999

6. €75,000 - €89,999

8. €90,000 - €119,999

9. €120,000 - €149,999

10. €150,000 or above

11. I would rather not say

Appendix 2.2: Results of Kolmogorov–Smirnov tests

	Garadice n = 64	Killykeen n = 75	Dernaferst n = 49	Eonish n = 58	Church Lake n = 43
Access	0.002*	0.068	0.782	0.000*	0.226
Size	0.697	1.000	0.946	0.997	0.950
Quantity	0.845	0.459	0.914	0.647	0.688
Services	0.024*	0.813	0.338	0.850	0.515
Encounters	0.002*	0.006*	0.164	0.143	0.518
Variety	0.332	0.227	0.846	0.034*	1.000

*P-values reported. * denote significance at the 5% level, suggesting that for these attributes at a particular site the online cohort perceived the site differently to the on-site cohort. Critical values for the two sample K-S test were calculated using the sample size presented in the table at a significance level of 5%.*