VALUATION OF ENVIRONMENTAL QUALITY AT MICHIGAN RECREATIONAL FISHING SITES: METHODOLOGICAL ISSUES AND POLICY APPLICATIONS

Prepared by:

Carol Adaire Jones
Yusen D. Sung

EPA Contract No. CR-816247-01-2
FINAL REPORT
September 1993

DISCLAIMER

ACKNOWLEDGEMENTS

We are grateful to the many people who have contributed to the successful completion of this project. In particular we are grateful to Doug Jester of the Fisheries Division, Michigan Department of Natural Resources (MNDR), and to Ted Graham-Tomasi, of the University of Minnesota. Doug initiated the project and, from the beginning, has been a source of creative modeling ideas for the economic analysis and for the Linkages between biology and economics. Doug has provided much of the data used in the analysis, both from environmental resource surveys and from the Michigan angler survey. Almost as important as the data has been his keen appreciation of the sources and limits of the data series. Ted Graham-Tomasi was the original Principal Investigator, starting the project while he was on leave at the School of Natural Resources (SNR), University of Michigan. Due to funding delays, the money came through as Ted was returning to the University of Minnesota, and responsibility for directing the project was passed on to Carol Jones. Ted continued to contribute to the project, as time permitted.

Sharon Nowlen provided valuable assistance as the MDNR Project Manager during the latter portion of the project, facilitating the acquisition of data and smoothing the grant administration process. Anne Wittenberg provided very able research assistance in the early stages for the project. Wendy Silverman provided effective research assistance in sifting information about potential environmental resource measures during the middle of the project.

In its current form, the report is a modified version of Yusen Sung's dissertation, submitted to the Economics Department of the University of Michigan in 1991. Sung has worked on the project from the beginning, originally providing -research and computing assistance, and over time coming to serve as partner in the research. Sung wrote all the computer programs used in the modeling, with the exception of the multinomial logit algorithm, written by C. Manski.

Preliminary results from the research project. have been reported in a series of earlier working papers, cited in the Bibliography. For the most part, revised versions of the earlier work are incorporated in this document.

The Michigan Department of Natural Resources provided generous financial support, in addition to the extensive research support and data noted above. The US Environmental Protection Agency provided support in the form of a Cooperative Agreement: Contract No. CR-816-217-01-2. Resources for the Future (RFF) provided Carol Jones with a Gilbert F. White Fellowship to carry out the research in residence at RFF for a year. and graciously accommodated her during an additional year of leave from the University of Michigan. Other financial support has been provided by the Rackham Graduate School and the Institute for Science and Technology at the University of Michigan.

In addition to the individuals mentioned above! useful comments and collaborative ideas have been offered by Mary Jo Kealy, USEPA and George Parsons, University of Delaware; and by Gary Solon, Joe Swierzbinski, and Al Jensen of the University of Michigan, who served on Yusen Sung's dissertation committee along with Carol Jones. Nonetheless, we alone are responsible for any errors.

In addition, we would like to thank Dean Jim Crow-foot, and Acting Dean Harry Morton, as well as Wayne Say, the Research Director at SNR, for the support that they provided over the years of the project. And finally, we particularly want to thank the team of people in the School of Natural Resources Business Office at the

University of Michigan, including Barb Branscum, Joan Kipfmiller, Barbara Murphy, Tracy Willoughby, Diana Woodworth, and Carole Shadley. Each in her own way has assisted with grace and immeasurable good will with various aspects of the financial and grants administration; all have eased the process of running the grant while on leave elsewhere and have allowed us to concentrate on the research itself.

Addresses for contacting the authors are:

Carol Adaire Jones
National Oceanic and Atmospheric Administration
1305 East West Highway
Silver Spring MD 20910-3281

Yusen D. Sung
Department of Economics
National Taiwan University
Taipei TAIWAN

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF MAPS

LIST OF TABLES

CHAPTER I

INTRODUCTION

In the research described in this report, we have developed a random utility model of demand for recreational fishing in Michigan, covering all water bodies and all species types throughout all counties in the state. The major study sponsor, the Michigan Department of Natural Resources (MDNR), funded the research to produce a model that could be used to improve fisheries resource management and to perform natural resource damage assessments. One out of every two households in Michigan has a fishing license, suggesting that fishing-related benefits will represent a substantial portion of the total benefits of improvements in water and sediment quality.

The travel cost model was designed to value recreational experiences. In a recent state-of-the-art review of recreation models, Bockstael, McConnell and Strand (1991) conclude that the random utility version of the travel cost model is particularly well-suited to valuing changes in quality at one or more recreation sites. The random utility model allows the researcher to model a wide range of substitution possibilities and, consequently, provides a procedure for estimating the value of changes in environmental quality. Nonetheless, many issues remain regarding the correct specification of these models and the sensitivity of welfare estimation to specification errors.

We identified two major research objectives for this project. The first was to address several key methodological issues associated with implementation of the random utility models. The second was to incorporate in the model sufficient data about the environmental attributes of sites in the State to perform the policy analyses of interest.

Below, we outline the model and the policy analysis we perform with the model. With that background, we will then briefly highlight the methodological issues addressed in the report.

Overview of the Model

To implement the random utility framework for modeling recreational trip demand, economists have identified two levels of consumer decisions: (1) How many recreational trips does each individual take during a year or a season? and (2) What attributes do people seek for each recreational trip? The first question pertains to total demand for recreation, the macro decision. The second question pertains to the *micro* decisions associated with an individual trip.

On any given choice occasion in a sport-fishing season, anglers must decide whether or not to take a fishing trip. For participants, we model three levels of choices they make for an individual trip: fishing site, by county; fishing product line, which captures distinctions by macro-species and water-body type; and trip duration. The anglers' decision structure is shown on page 3, along with the options available and the factors hypothesized to influence each decision.

In our context of recreational fishing, the *macro* decision is the total number of fishing trips anglers take during a fishing season. Since anglers may take trips of different lengths: we model separately total demand for different trip- lengths. Consequently, we handle the third-level choice for individual trips, trip duration, within the *macro-level* participation model.

Though it is theoretically possible to model the discrete product-line/site choices

CHOICE STRUCTURE OF SPORTFISHING ANGLERS

Trip Length	Fish Product Line	Destination Site

Alternatives:

Day	Great Lakes Coldwater	83 Michigan counties
Weekend: 2-4 days	Great Lakes Warmwater	
Vacation: 5+ days	Anadromous Runs	
	Inland (Lk+Strm) Coldwater	
	Inland Lakes Warmwater	
	Inland Streams Warmwater	

Factors influencing choice;

Inclusive value of PL	inclusive value of sites	Travel costs
Lodging/food cost	Product line costs	Fish catch rates
Workstatus	Fishing skill/preference	Quantity of resources
Avidness of angler	Demographic attributes	Natural beauty
Household income		Accessibility
Marital status		Contamination
School vacation		

and the total participation decision jointly, the data and computational requirements for the correct treatment of the corner-solutions implied by zero trips of certain categoriest makes an integrated utility-theoretic model practically infeasible. Essentially, researchers face a trade-off: they either implement a utility-theoretic framework that does not properly model the statistics of the corner solutions; or they model the *micro* and *macro* decisions in separate models that may address the corner solution problem but do not form an integrated utility-theoretic framework.

In our analysis we estimate separate models at the *micro* and *macro* levels. We use the nested multinomial logit model (NMNL) to estimate the determinants of site and product line choices on the *micro*- level. Due to severe data limitations at the total participation level, our participation model is somewhat different from the standard treatment in the literature. We do not know the total number of season trips: our *macro* level information is limited to the duration between trips. and this variable is censored because we only observe the duration from last trip to the survey return date, not to the subsequent trip. By incorporating a key result from stochastic renewal theory in our modeling. we are able to estimate the determinants of the between-trip durations with a stochastic renewal model and then to derive the total number of trips in a season from the duration model.

Though necessitated by the data limitations we face. this approach in fact may provide several advantages. The most prominent advantage of the competing risks approach is the capacity for modeling the dependency of choices among trips of different types, which is lacking in most other empirical work with random utility models. Most researchers have limited their analysis to day trips. Another advantage is that we are able to incorporate time-varying covariates to account for changing fishing conditions over the season at individual sites.

Performing Policy Analysis

In order to perform policy analysis with the model, it is important to incorporate appropriate measures of site quality to capture the quality changes associated with the policies. Michigan identified several policies of particular interest. In the resource management area, the key concern was evaluating alternative fish stocking regimes. In the area of natural resource damage assessments, the State wanted the capability to estimate damages from power-plant related fishkills, toxic contamination at state and federal Superfund sites, fishkills from acute toxic episodes, and acid rain contamination.

In order to value these injury scenarios, the determinants of site choice in the model had to include the key measures of environmental quality that change in the scenarios, as they are experienced by anglers. The two key categories of quality change are fish catch rates (to capture the stock effects) and toxic contamination levels. We incorporated detailed information on fish catch rates from the MDNR creel survey for the Great Lakes and anadromous fisheries, and generally found the predicted positive relationships between expected catch rates and anglers' valuation of a site. Due to problems with endogeneity between participation and catch rates for the inland product lines, we were only able to use measures of lake area or stream length, broken down by quality level, for those product lines.

Unfortunately, we were not. able to use a fish consumption advisory measure to capture toxic contamination in the Great Lakes product lines Because fish consumption advisories apply to virtually all of the Great Lakes warmwater and coldwater fisheries (except a few counties with no fish, and a few counties in Lake Superior), the variable lacks the variability required for inclusion in the modeling. We used fish advisory measures for inland product lines, but there were few inland resources with advisories at the time of the angler survey, so there is limited variation in the advisory variable for those product lines also.

Toxic contamination in the Great Lakes product lines is measured by a variable indicating that (selected) water bodies in the county have been designated as part of an Area of Concern by the International Joint Commission. A noteworthy finding in the empirical analysis is that designation of a county as an Area of Concern has a substantial dampening effect on participation, an effect that spills over into water bodies and species (fishing product lines) that are not directly located in the (localized) Area of Concern within the county.

In constructing the model. we estimated how individuals value for fishing at a site varied with the fish catch rates and contamination variables. To carry out a policy analysis with the model, a resource expert must provide the "policy scenario", which specifies how the values of the environmental quality variables will change as a result of the policy.

To illustrate the capabilities of the model for performing policy analysis. we apply the model to two current contexts in which environmental injury is occurring in Michigan. First, we calculate the damages to Michigan-licensed recreational anglers from fish kills due to operation of the largest pumped- storage plant in the US. Second, we calculate the benefits of cleaning up PCB contamination in a river in Michigan, which would allow the State to remove dams currently containing contaminated sediments and to open a substantial reach of the river for anadromous runs. The contamination at this site is sufficient to merit designation of the site as an Area of Concern

Methodological Issues

We identified three key methodological issues raised in implementing the random utility model:

1. modeling total trip participation across the season, given that we have detailed information on a single trip and very limited information about total trip demand;

2. developing a consumer surplus measure that takes into account the changes in predicted number of trips due to policy changes (as well as the change in value per trip); and

3. performing sensitivity analysis of the model to alternative specifications, including alternative treatments of the opportunity costs of time.

Participation modeling

The major methodological challenge is to link a *macro-* level model of total recreational trip demand to the micro-level model of demand for fishing site and fishing product line. Our participation model represents an innovative solution to the estreme limited-data problem we faced. The analytical framework, which develops estimation procedures for a competing risk model with censored duration data and time-varying covariates, has wide applicability beyond the recreational demand contest.

By modeling demand for trips of different durations, we are able to show that two-thirds of the damages in our policy scenarios accrue to anglers taking trips of longer than one day. If we had followed the standard procedure in the literature of analyzing day trips only, we would have seriously underestimated damages.

In order to validate the participation estimates from the model, we compared the estimated trip-days derived from our model against estimated trip-days based on analysis of the MDNR creel survey. Because the procedures and criteria for counting trips and trip-days are different in the two datasets, the comparison is not suited to statistical testing. Though the differences between the surveys limit our ability to compare the estimates, we conclude that the similarity of predicted participation between the model and the annual diary data provides some evidence corroborating the participation model.

Several possible avenues exist for improving model specification. We have not explicitly addressed the “corner solution” problem, as Bockstael, Hanemann, and Strand have labelled it. We need to test to see whether non- participants should be treated differently from participants. Resolution of this issue is more complicated in our dataset than in a more typical survey, where total trips are measured for a fixed time period across all individuals. In our dataset, we observe “no trip” outcomes over very different time periods, ranging from one to fourteen months. To model "no-participation'', we must confront the question, over what length of time must a licensed angler not participate to be considered a different type of person?

Consumer Surplus Measure

Linked with the *macro* modeling issue is the correct specification of the consumer surplus measure. The standard measure employed for discrete choice models is based on the assumption that total trips do not change with policy changes. This measure will result in an under- or over-estimate of “true” consumer surplus, depending upon whether total trips increase or decrease. We develop a consumer surplus measure that incorporates the change in trips predicted by the participation model. Additional complexity is added to the measure with a nested multinomial logit model (NMNL), when the choice occasion income is not observed and the marginal utility of income is not constrained to be constant across alternatives due to the computational complexity of such a procedure. We propose a simplifying procedure that makes the calculation tractable under these circumstances.

Model Specification Issues

Finally, we analyze the sensitivity of model estimates to alternative treatments of the time constraints faced by anglers in making their trip choices. Extensive exploration in conventional (continuous demand) travel cost models has shown that consumer welfare measures are extremely sensitive to the treatment of time, though no consensus has emerged on the appropriate method for valuing time. Discrete choice models have not been subjected to comparable exploration. In this study, we develop a careful accounting of household allocation of time; the accounting highlights the fact that different treatments of the time constraints imply different choice sets of feasible sites: as well as different treatments of the opportunity costs of time in the modeling.

Outline of the Report

The report is organized as follows. Chapter II reviews the literature on random utility models of recreation demand. The emphasis is on highlighting the methodological issues associated with implementing the random utility model. Chapters III through V specify the theoretical framework for modeling the PL-site choice, for modeling total trips in a season, and for calculating the exact seasonal consumer surplus. Chapter VI is a description of the data sources. Chapters VII and VIII present estimation results of the multinomial logit and the participation models, respectively. Chapters IX and X apply the model to two natural resource damage scenarios in Michigan fisheries, one relating to fishkills and the other to toxic contamination, and calculate the loss in consumer value as a result of the injuries. In the Appendix, we report the sensitivity analysis of site choice model estimates with alternative treatments of the value of time.

CHAPTER II

DISCRETE CHOICE MODELS OF RECREATION DEMAND: A BRIEF REVIEW

The purpose of this chapter is to provide a brief overview of random utility models (RUM) of recreation demand, highlighting some key methodological issues that remain in model design and implementation. First used by Luce (1959) to model psychological choice behavior. RUM was shown by McFadden (1974, 1978) to be consistent with underlying consumer utility maximization behavior.[1]

An individual, upon deciding to take a trip on a choice occasion, is assumed to choose the site among the available alternatives that offers him/her the highest utility. The utilities that can be derived from visiting different sites are usually considered deterministic to the individuals, but stochastic to the outside investigators due to unobserved personal/site characteristics, data measurement errors, or simply random elements in human decision-making process.

By assuming *weak complementarity* which posits that a consumer will not care about marginal improvements of a commodity if he/she consumes none of it,[2] i.e.,

$$\frac{\partial u(x, \ldots)}{\partial x} = 0, \text{ if } x = 0,$$

---

[1] See McFadden (1976, 1961, 1982, 1984), Amemiya (1981), Hensher and Johnson (1981), or Maddala (1983) for surveys and discussions of qualitative response models.

[2] This in effect rules out the non-use value of the commodity. See Maler (1974. p. 134) or Feenberg and Mills (1980. p. 64).

the utility function; conditional on site j being chosen: of individual i can be specified as

$$u_{ij} = v_{ij}(q_j, \bar{y}_i - p_{ij})$$

where q_j is the characteristics vector of site j, p_{ij} is the cost of i travelling to site j, and $\bar{y}_i$ is the budget allocated to the trip duration in question. All the characteristics vectors pertaining to unchosen sites are excluded as a result of the weak complementarity assumption. Note that individual-specific variables can also be omitted if v_{ij} is linear in its parameters since they have the same values across all alternatives and thus will not affect the utility ranking of the feasible sites.[3]

Since the conditional utility appears stochastic to researchers: a disturbance term must be added to form the random utility

$$u_{ij} = v_{ij}(q_j, \bar{y}_i - p_{ij}, \varepsilon_{ij})$$

An individual i will then choose k among a set of feasible sites C_i if

$$u_{ik} > u_{ij}, \ \forall j \neq k, \ j \in C_i. \tag{II.1}$$

By strategically choosing a utility function u and defining the joint probability distribution for ε to make the mathematics tractable, we can calculate the probability of an individual i going to site k. given i's decision of participation:

$$\pi_{ik} = \text{Prob}\{u_{ik} > u_{ij}, \ \forall j \neq k, \ j \in C_i\}.$$

The most widely adopted multinomial response model in the literature is the multinomial logit (MNL) model.[4] because it yields a simple form of π_{ik} as well as other computational advantages. In the MNL model, the random terms ε are assumed to

[3] This is in fact the result of adopting an additively separable utility form usually assumed for estimation convenience.

[4] See Train (1986), Ben-Akiva and Lerman (1985), McFadden (1974, 1976, 1984) and Maddala (1983) for model specification.

be i.i.d. type I extreme value distributed [5] The probability of an individual i choosing site k among a collection Ω_i of sites can then be shown to be

$$\pi_{ik} = \frac{e^{u_{ik}}}{\sum_{j \in \Omega_i} e^{u_{ij}}}.$$

A restrictive feature of the MNL model is the *Independence from Irrelevant Alternatives* (IIA) property. which states that the probability ratio of two sites being chosen will stay the same regardless of the addition or deletion of other sites (or their properties).[6] This can be easily verified since the probability ratio

$$\frac{\pi_{ij}}{\pi_{ik}} = \frac{e^{u_{ij}}}{e^{u_{ik}}},$$

depends only on variables in u_{ij} and u_{ik}. Given the weak complementarity assumption. u_{ij} and u_{ik} consist solely of the quality variables of sites j and k, respectively.

While the multinomial logit models have the IIA property which is not very desirable in many situations, researchers car, circumvent this problem by using the more flexible *generalized extreme valve* (GEV) model,[7] which embodies the correlation among sites within its joint distribution structure of the error terms. The most commonly employed GEV model is the nested multinomial logit (NMNL);[8] which captures the inter-site correlation in the coefficient of the inclusive value index. Derivation of both MNL and NMNL from GEV can be found in Ben-Akiva and Lerman (1985).[9] The NMNL model is particularly useful when the number of

---

[5] Ben-Akiva and Lerman use the *Gumbel* distribution, which is a slightly more general structure than the type I extreme value distribution.

[6] See Maddala (1963, pp. 61- 62) or Amemiya (1985, p. 298). Ben-Bkiva and Lerman (1985. p. 109) point out that any model assuming the independence of all the disturbances would necessarily yield the IIA property.

[7] Introduced by McFadden (1978, 1961).

[8] Some examples of empirical NMNL studies are Carson and Hanemann (1967) and Bockscael et al. (1988)

[9] Pages 127 and 304, respectively.

alternatives is very large but the decision process itself can be properly described by a tree structure to reduce computational complexity. [10]

Like other discrete choice models, the RUM is used to explain the choice of site to visit and possibly other characteristics for a specific trip, which is referred to as the *micro* decision. As discussed below, the total number of trips taken during a season, the *macro* decision, is generally estimated by other means.

Many researchers have estimated models based on the random utility discrete choice approach to explain trip allocation decisions and to measure the welfare effects from environmental quality changes, including Hanemann (1978, 1982, 1984. 1985), Binkley and Hanemann (1978), Feenberg and Mills (1980), Caulkins (1982), Caulkins, Bishop and Bouwes (1986), Rowe, Morey, Ross and Shaw (1985), Bockstael, McConnell and Strand (1988), Morey et al. (1991, 1989). Jones et al. (1988, 1989, 1990), Parsons and Kealy (1990). Smith and Kaoru (1990), and Carson, Hanemann, Gum, and Mitchell (1987).

The multinomial logit model is attractive not only because it can avoid some of the problems of conventional travel cost methods, but also due to its computational tractability and feasibility when the number of alternatives gets large. In a recent state-of-the-art review of recreation models, Bockstael, McConnell, and Strand (1991) conclude that the random utility version of the travel cost model is particularly well-suited to valuing changes in quality at one or more recreation sites. The random utility model allows the researcher to model a wide range of substitution possibilities and, consequently, provides a procedure for estimating the value of changes in environmental quality.

Simulations have been run to show the advantages the random utility method has over other approaches. Kling (1986, 1988) uses Monte Carlo methods to generate var-

[10] Conditions to be met for the employment of a nested analysis are explained in Ben-Akiva and Lerman (1985, pp. 291-93) for a three- dimensional case.

ious data sets for a Stone-Geary utility function and compares the welfare estimates of different models with actually known measures. In their review, Bockstael, McConnell, and Strand conclude that Kling's "stylized simulation experiments . . . give preliminary support to the notion that discrete choice models produce better benefit estimates in problems characterized by much substitution among sites, especially when a large portion of the sample is observed to choose more than one site to visit in a season." (p. 256)

Nonetheless several fundamental methodological issues remain. Perhaps the most thorny is to integrate the *micro* and *macro* levels of the modeling, with the correct statistical treatment of the corner-solutions implied by zero trips of certain categories, (otherwise known as the 'corner-solution' problem.) We consider this issue in some detail in Chapter IV.

In this section, we discuss specification issues associated with specifying time constraints and choice sets in the random utility models. One important issue that has not been explored in the random utility context is the valuation of the opportunity costs of time. As pointed out by Bockstael et al. (1987), recreationists often cite time as more constraining than money in their recreation consumption. So the time spent on recreation consumption is, in many cases, an important determinant of the demand.

It has been recognized, since the early period of recreation demand modeling, that the omission of time costs (i.e., the opportunity costs of on-site and travel) in conventional travel cost models biases the parameter estimates and understates the final welfare measures.[11] The time-valuation literature since has focused on the context of conventional travel cost demand models.[12] In the multinomial logit models of recre-

---

[11] See Clawson and Knetsch (1966) or Cesario and Knetsch (1970).

[12] E.g.. Cesario (1976), Smith et al. (1983), Kealy and Bishop (1986). Bockstael et al. (1987), and McConnell (1990).

ational demand reported in the literature, the treatment of travel time apparently has varied substantially. However. our literature review revealed that authors frequently did not explain how they defined the travel cost/time variables, rarely explained how they defined a choice occasion, and never explained how an individual's choice set of feasible sites related to the time constraint for the choice occasion.

1. In the studies conducted by Bockstael et al. (1986, p. 213; 1987), all we know is that they have a "trip cost" variable. No details are provided.

2. In their MNL model of southcentral Alaska sport fishing, Carson, Hanemann, Gum, and Mitchell (1987) include only a round-trip distance cost[13] variable, computed as round-trip distance multiplied by the individua1 respondent's reported motor vehicle cost per mile. No time cost is included.

3. Morey et al. (1991. p. 4) state only that they have the "cost of a trip to site *j* mode *m* " in their model. No explanation is given as to how this variable is calculated.

4. Bockstael et al. (1988) calculate their travel cost variable as $.10 per mile plus 80 percent of the wage rate for individuals who worked for a wage and could vary their time. A separate travel time variable is used for anglers who cannot vary their work time.

5. McConnell et al. (1990) assume that anglers spend a fixed amount of time fishing at the site, whatever site is chosen. Both the distance cost and the cost of travel time thus enter the angler's site decision. For anglers who work flexible hours, the cost of travel time, valued at the wage rate, is included as part of the total travel cost. For anglers without such discretion. travel time enters the utility function directly.

6. Parsons (1990) allows the recreation period to be longer than the trip duration, and includes the individual's opportunity cost of time, distance cost, as well as other expenses in the price of taking a trip.

7. In their Wisconsin lake recreation study, Parsons and Kealy (1990) measure the travel cost as the sum of transit costs and opportunity cost of time. The transit cost is assumed to be $.10 (1978 dollars) per mile. For the time cost, they assume that all individuals stay on site for a fixed four-hour period. Each

---

[13] We use the term distance cost to refer to the cost of motor vehicle operation for the trip. The term *travel cost* is meant to be the all inclusive measure, which consists of the distance cost and the opportunity time cost of travelling

individual is then assumed to value an hour at one third of his/her wage rate for the travel time and on-site time [14]

8. The travel cost measure in Smith and Kaoru (1990) is the sum of a distance cost plus the opportunity cost of travel time. The former consists of the vehicle operating costs measured as round-trip mileage times $.20 per mile; the opportunity costs of travel is measured as the predicted wage per hour for employed respondents and the minimum wage for non-working individuals times travel time. The travel time is estimated from the round-trip mileage by assuming an average speed of 40 miles per hour.

The multinomial logit literature on recreational demand has not focused on the question of how an individual's choice set of feasible sites is defined. When is a site too far for an individual to reach on a choice occasion? What are the time constraints used for defining the choice sets? None of the papers mentioned above provide enough information to answer these questions.[15] However, as we will show, these specification choices indeed have a large impact on the MNL estimates.[16]

We know of only one study that has analyzed explicitly the sensitivity of model estimates to alternative definitions of choice sets. Smith and Kaoru (1990) consider the geographical resolution of site definitions, evaluating the specification error from increasing levels of aggregation across heterogeneous sites. They observe that site definition does have important implications for specification of the nesting structure of the model and for the benefits measurements associated with quality changes. However: they conclude that their findings provide "rather strong support for using

---

[14] The wage rate is calculated as annual income divided by 2080, the average number of hours worked in a year of their sample.

[15] Parsons and Kealy (1990) only mention that they include "lakes within a day's drive from an individual's home" in the individual opportunity set. In our case, though all the 83 Michigan counties form the units of our site MNL analysis. not every county is in the choice set of an angler on a certain choice occasion. This is especially true for the day anglers. Some counties are simply beyond reach for a day trip. Some counties may be within reach, but heavy driving may make trips to them infeasible. For example, it is unlikely that people are willing to drive ten hours each way to a distant site in a single day.

[16] As Smith and Kopp (1980) point out, there are spatial limits to the legitimate use of travel cost methods.

random utility models to [estimate] the effects of quality attributes on people's decision to use different recreation sites." They further note that their study "strongly reinforces the Bockstael, McConnell and Strand (1991) conclusions supporting the RUM framework even in cases where the site definition and specification of the set of alternatives is unclear." (p. 27)

CHAPTER III

INDIVIDUAL MICRO-LEVEL CHOICE MODELING

In this chapter we specify a utility-theoretic model to analyze the PL-site choices of recreational anglers. In the following chapter we present the model of the macro-level demand for total recreational trips per season.

Consumer Preferences and Behavior

Consider a consumer *i* who derives utility from two kinds of activity: consuming market goods and taking fishing trips Let Z = $(\mathbf{Z}^1, \mathbf{Z}^2, \ldots, \mathbf{Z}^T)$ denote the numeraire composite market good consumed by *i* in the *T* periods of a fishing season. The periods are determined in such a way that the individual *i* can take no more than one fishing trip in each period $t = 1, 2, \ldots, T$. In each period, individual *i* will decide whether to take a fishing trip for one of the total *M* product lines. The number of feasible sites for product line *m* is J_m which varies with product line choice and individuals. The attributes of all PL-site combinations in all periods are denoted by $\mathbf{Q} = \{q^t_{(l,j)}, \forall\, t,\ l,\ j\}$, where *t, l* and *j* are the indices for periods, PLs and sites, respectively. Also, denote the costs[1] individual *i* has to incur to fish for all possible PL-site choices *(l,j)* in all periods *t* as P $= \{p^t_{(l,j)}, \forall\, t,\ l,\ j\}$. The participation and

[1] The costs of recreational activities generally include license fees, site entrance fees, travel cost. gear purchases. etc.

PL-site decisions made by i are $\delta = \{\delta^t_{(l,j)}, \forall t, l, j\}$, where $\delta^t_{(l,j)}$ has a value of 1 if i decides to visit site j for PL l during period t, and a value of 0 otherwise. The indicator variable $\delta^t_{(l,j)}$ will be zero for all (l, j) alternatives if the individual i does not take a trip in period t. Individual i is assumed to maximize his or her utility, given annual income y. The maximization problem facing i is thus

$$\begin{aligned} \text{maximize} \qquad & U_i(Z, \delta Q) = U(Z, \delta Q, S_i) \qquad \text{(III.2)} \\ \text{subject to} \qquad & \delta P + Z = y \\ & \delta^t_{(l,j)} \cdot \delta^t_{(m,k)} = 0, \ \forall (l,j) \neq (m,k), \ \forall t \\ & \delta^t_{(l,j)} = 0 \text{ or } 1, \ \forall (l,j), \ \forall t \\ & Z \geq 0 \end{aligned}$$

where S_i is the vector of socioeconomic attributes of individual i. The first constraint is the budget constraint, while the other constraints force the corner solution in which the consumer i can only buy at most one of the quality-differentiated fishing trips. Note also that the parameter (δQ) in the utility function U_i embodies the weak complementarity assumption, which asserts that i will only obtain utility from quality attributes Q through realized trips. The indirect utility function can then be derived as $V_i = V_i(P, Q, y)$ or $V_i = V(P, Q, y, S_i)$.

Since the decision indicators $\delta^t_{(l,j)}$ can only take on integer values. a solution to the above maximization problem can only be found by first comparing the utility levels yielded by all possible trip choices over all T periods and then selecting the one that generates the highest utility. The procedure to solve this problem is described in both Kling (1986) and Bockstael et al. (1986). Since solving the problem (III.2) is computationally infeasible, simplification of the model is necessary.

A common practice is to impose further structure on the utility function. A useful and reasonable strategy is to assume that individuals adopt a *two-stage budgeting* process. Individuals, seeking to maximize their utility, are hypothesized first to optimally allocate their season budget y among all T time periods, and then to determine the actual consumption pattern in each period with the period budget y^t.[2]

An implication of the two-stage budgeting process is that the utility function is characterized by weak separability across budget categories, (such as recreation, housing, food etc), where weak separability is defined as follows:[3]

Definition III.1 For a utility function $u = u(q_1, q_2, \cdots, q_K)$ where q_k is the vector of commodities in kth category, the *weak separability* assumption requires that the utility function u be expressible as $u = f(v_1(q_1), v_2(q_2), \ldots, v_K(q_K))$, while strong (or additive) separability further implies the simpler form of $u = f(v_1(q_1) + v_2(q_2) + \cdots + v_K(q_K))$. ∎

Because our dataset (as with most datasets in recreation demand studies) has data only on the most recent fishing trip, we must further assume weak separability across choice occasions within the recreational fishing budget branch. With this restriction, an angler's ranking of possible fishing trips on a particular choice occasion does not depend on how many fishing trips of different types he/she has already taken or will take later in the season.

We further assume weak separability across site choices: the quality of a site only affects an individual's utility if the site is chosen. The recreational fishing sub-utility function is defined over the vector of market goods associated with recreation Z^t and the vector of site characteristics for the chosen site, for each choice occasion t. Therefore, i's season utility function can be written in the form

$$U_i = U\left(u(Z^1, \delta^1 Q^1, S_i), u(Z^2, \delta^2 Q^2, S_i), \ldots, u(Z^T, \delta^T Q^T, S_i) \right).$$

When the allocation of y to each period t is done, the utility maximization problem

---

[2] This assumption can be justified by observing that people frequently form a general price aggregate about market prices in the near future and allocate their long-term income to different time periods accordingly.

[3] See Deaton and Muellbauer (1983, Part 2, Chapter 5) or Morey (1984) for a thorough treatment of this topic. Discussion on two-stage budgeting can also be found in Varian (1984, pp. 146-49).

can be attacked by solving the following maximization problem for each period t

$$\begin{aligned} \text{maximize} \quad & u^t = u(Z^t,\ \delta^t Q^t,\ S_i) \\ \text{subject to} \quad & \delta^t P^t + Z^t = y^t \\ & \delta^t_{(l,j)} \cdot \delta^t_{(m,k)} = 0,\ \forall (l,j) \neq (m,k) \\ & \delta^t_{(l,j)} = 0 \text{ or } 1,\ \forall (l,j). \\ & Z^t \geq 0 \end{aligned} \tag{III.3}$$

By substituting the budget constraint $Z^t = y^t - \delta^t P^t$ into the utility function u^t, we can reformulate the problem as

$$\begin{aligned} \text{maximize} \quad & u = u(y^t - \delta^t P^t,\ \delta^t Q^t,\ S_i) \\ \text{subject to} \quad & \delta^t_{(l,j)} \cdot \delta^t_{(m,k)} = 0,\ \forall (l,j) \neq (m,k) \\ & \delta^t_{(l,j)} = 0 \text{ or } 1,\ \forall (l,j) \end{aligned}$$

Consider period t where a micro PL-site choice decision has to be made by individual i. Let Ω_i be i's choice set of available PL-site (l,j) alternatives. Upon choosing not to take a trip in period t, individual i will obtain the no-trip utility

$$u^t_0 = u_0(y^t,\ S_i).$$

Otherwise, if the PL-site (l,j) combination is chosen, he or she will, by weak complementarity, receive the following conditional utility

$$\begin{aligned} u^t_{(l,j)} &= u_{(l,j)}\left(Z^t,\ q^t_{(l,j)},\ S_i\right) \\ &= u_{(l,j)}\left(y^t - p^t_{(l,j)},\ q^t_{(l,j)},\ S_i\right) \end{aligned}$$

We can also define the unconditional utility function as

$$V_i(P^t,\ Q^t,\ y^t) \equiv \max\left\{u^t_{(l,j)},\ \forall (l,j) \in \Omega_i\right\}$$

One way to incorporate the participation decision in this framework is to compare the no-trip utility u_0^t with the unconditional utility V_i. A no-participation decision will consequently be made if

$$u_0^t > V_i(P^t, Q^t, y^t),$$

and hence $\delta^t_{(l,j)} = 0, \forall (l,j) \in \Omega_i$. On the other hand, the PL-site (m,k) will be chosen if

$$u^t_{(m,k)} = \max\left\{u^t_{(l,j)},\ \forall (l,j) \in \Omega_i\right\} > u_0^t,$$

giving us $\delta^t_{(m,k)} = 1$. and $\delta^t_{(l,j)} = 0$ for all $(l,j) \neq (m,k)$.

Since there exist some unobserved factors affecting PL-site and participation decisions, the utilities u_0^t and $u^t_{(l,j)}$ are random from the analyst's perspective. A PL-site specific disturbance is, hence, introduced into the various utility functions to form the random utility functions

$$\begin{aligned}
\tilde{u}_0^t &= u_0(y^t, S_i) + \epsilon_0 \\
\tilde{u}^t_{(l,j)} &= u_{(l,j)}\left(y^t - p^t_{(l,j)},\ q^t_{(l,j)},\ S_i\right) + \epsilon_{(l,j)} \\
\tilde{V}_i(P^t, Q^t, y^t) &= \max\{\tilde{u}^t_{(l,j)},\ \forall (l,j) \in \Omega_i\}.
\end{aligned} \tag{III.4}$$

The Micro-level Product Line/Site Decision

This section presents a model of the micro PL-site choice given that an individual i has decided to take a trip. A rational individual i will prefer PL-site combination (m,k) to (l,j) if

$$\begin{aligned}
& \tilde{u}_{(m,k)} > \tilde{u}_{(l,j)} \\
\Longleftrightarrow\quad & u_{(m,k)} + \epsilon_{(m,k)} > u_{(l,j)} + \epsilon_{(l,j)} \\
\Longleftrightarrow\quad & \epsilon_{(l,j)} < \epsilon_{(m,k)} + u_{(m,k)} - u_{(l,j)}.
\end{aligned}$$

Conditional on participation, consumer i will choose PL-site (m, k) from his or her feasible set of alternatives Ω_i if and only if

$$\tilde{u}_{(m,k)} > \tilde{u}_{(l,j)}, \ \forall (l, j) \neq (m, k), \ (l, j) \in \Omega_i,$$

or equivalently

$$\epsilon_{(l,j)} < \epsilon_{(m,k)} + u_{(m,k)} - u_{(l,j)}, \ \forall (l, j) \neq (m, k), \ (l, j) \in \Omega_i.$$

The probability $\pi_{(m,k)}$ of individual i choosing PL-site (m, k) is then

$$\begin{aligned} \pi_{(m,k)} &= \text{Prob}\left\{\epsilon_{(l,j)} < \epsilon_{(m,k)} + u_{(m,k)} - u_{(l,j)}, \ \forall (l, j) \neq (m, k), \ (l, j) \in \Omega_i\right\} \\ &= \int_{-\infty}^{\infty} \left\{ \left[\prod_{(l,j)\neq(m,k)} F(\epsilon_{(m,k)} + u_{(m,k)} - u_{(l,j)}) \right] f(\epsilon_{(m,k)}) \right\} d\epsilon_{(m,k)} \end{aligned}$$

where $F(\cdot)$ and $f(\cdot)$ are the cumulative distribution function (CDF) and probability density function (PDF), respectively, of the residuals ϵ.

What matters here is the difference $u_{(m,k)} - u_{(l,j)}$ between utility levels offered by (m, k) and (l, j), not their absolute magnitudes. Therefore, if the conditional utility u is additively separable between the choice-specific and non-choice-specific attributes, leading to the following form

$$\tilde{u} = v\left(y^t - p^t, \ q^t\right) - h(S_i) + \epsilon,$$

the non-choice-specific personal attributes will drop out of the micro choice decision because they are constant across all PL-site alternatives.

The Multinomial Logit Model

If the residuals ϵ are independently and identically distributed with *type I extreme value* distribution for which the CDF is

$$F(\epsilon) = \exp(-e^{-\epsilon})$$

and PDF is

$$f(\epsilon) = \exp(-\epsilon - e^{-\epsilon}).$$

then it can be shown that

$$\pi_{(m,k)} = \frac{e^{u_{(m,k)}}}{\sum_{(l,j)\in\Omega_1} e^{u_{(l,j)}}} = \frac{e^{u_{(m,k)}}}{\sum_{l=1}^{M}\sum_{j=1}^{J_l} e^{u_{(l,j)}}}$$

which is the multinomial logit model.[4]

The type I extreme value distribution is in fact a special case of the Gumbel distribution[5] that has the CDF

$$F(\epsilon) = \exp\left(-\exp[-\mu(\epsilon - \eta)]\right),\ \mu > 0$$

and PDF

$$f(\epsilon) = \mu \exp[-\mu(\epsilon - \eta) - \exp(-\mu(\epsilon - \eta))]$$

where η is a location parameter and μ is a scale parameter. The type I extreme value distribution simply assumes that $\eta = 0$ and $\mu = 1$. The Gumbel distributed residuals ϵ all have the same mean

$$\eta + \frac{\gamma}{\mu}$$

and variance

$$\frac{\pi^2}{6\mu^2}$$

where $\gamma(\approx 0.5772)$ is the Euler constant, and result in the probability of PL m-site k alternative:

$$\pi_{(m,k)} = \frac{\exp(\mu\, u_{(m,k)})}{\sum_{(l,j)\in\Omega_1}\exp(\mu\, u_{(l,j)})}.$$

Since the parameter μ is not econometrically identifiable, it is common practice to set it arbitrarily to 1, yielding the same probabilities as the type I extreme value distribution. As pointed out by Ben-Akiva and Lerman (1985, p. 104), the assumption

[4] It is called conditional logit by McFadden (1974).

[5] See Ben-Akiva and Lerman (1985, pp. 104-107) for a discussion.

of a constant η for all alternatives is not restrictive as long as each systematic utility has a constant term. Though the Gumbel distribution is used for analytic convenience, its choice can be defended as an approximation to the normal density.

Note that the probability (III.5) can also be expressed as the product of a conditional probability and a marginal probability

$$\pi_{(m,k)} = \pi_{k|m} \cdot \pi_m \tag{III.6}$$

where

$$\pi_{k m} = \frac{e^{u_{(m,k)}}}{e^{I_m}} \quad \text{and} \quad \pi_m = \frac{e^{I_m}}{\sum_{l=1}^{M} e^{I_l}} \tag{III.7}$$

and

$$I_m \equiv \log\left(\sum_{j=1}^{J_m} e^{u_{(m,j)}}\right).$$

With the Gumbel assumption, it can be shown that

$$E\left[\max_j\{\tilde{u}_{(m,j)}\}\right] = \log\left(\sum_{j=1}^{J_m} e^{u_{(m,j)}}\right) + \gamma = I_m + \gamma.$$

Hence the inclusive value index I_m reflects weighted information about the alternatives in PL m and is a measure of the expected maximum utility one can get from choosing PL m.[6]

We assume that the systematic part u of the random utility $\tilde{u}$ can be separated into the part that varies only with PLs and the part that varies with both PLs and sites as follows:

$$u_{(m,k)} = \gamma Z_m + \beta X_{(m,k)}.$$

In this case the probabilities (III.7) become

$$\pi_{k|m} = \frac{e^{\beta X_{(m,k)}}}{\sum_j e^{\beta X_{(m,j)}}} \quad \text{and} \quad \pi_m = \frac{e^{\gamma Z_m + I_m}}{\sum_{l=1}^{M} e^{\gamma Z_l + I_l}}.$$

---

[6] A discussion of the Gumbel properties is in Ben-Akiva and Lerman (1985, p. 105).

Now the inclusive value index for product line m becomes:

$$I_m = \log\left(\sum_{j=1}^{J_m} e^{\beta X_{(m,j)}}\right)$$

where the "inclusive value" is the expected utility of an individual for the site-specific attributes X, net of the integrating constant γ.

Estimation can thus be carried out by sequentially applying MNL to each PL m, calculating I_m for all m PLs, and then calculating $\pi_{(m,k)}$ using formula (III.6). Obtaining the maximum likelihood estimates of a multinomial logit model in general poses no computational difficulty since it has been proved by McFadden (1974) that the log likelihood function

$$LL = \sum_{i=1}^{N} \log P^{i}_{(m,k)} = \sum_{i=1}^{N}\left(u_{i,(m,k)} - \log \sum_{(l,j)\in\Omega_i} e^{u_{(l,j)}}\right)$$

is globally concave under relatively weak conditions. The Newton-Raphson algorithm will therefore always converge within finite steps, often in just a few iterations, to a unique solution.

The way a simple MNL models the PL-site decision is to treat each PL-site combination as a feasible choice. Given that we have M product lines and J_m potential sites for each product line m (= 1, 2, . . . , M), the total number of alternatives one faces is $\mathcal{J} = \sum_{m=1}^{M} J_m$, as illustrated in figure III.1. A restrictive feature of this modeling approach is the aforementioned *Independence from Irrelevant Alternatives.* It is highly implausible that the odds ratio $\frac{\pi_{(m,k)}}{\pi_{(l,j)}}$ of any two PL-site choices (m,k) and (l,j) will be independent of the conditions of other available alternatives, as implied by the MNL specification where

$$\frac{\pi_{(m,k)}}{\pi_{(l,j)}} = \frac{\exp(u_{(m,k)})}{\exp(u_{(l,j)})}.$$

Consider a situation where an individual i can only choose between site *A* for PL 1 and site *B* for PL 2. With the addition of a site *C* for PL 2 that has exactly

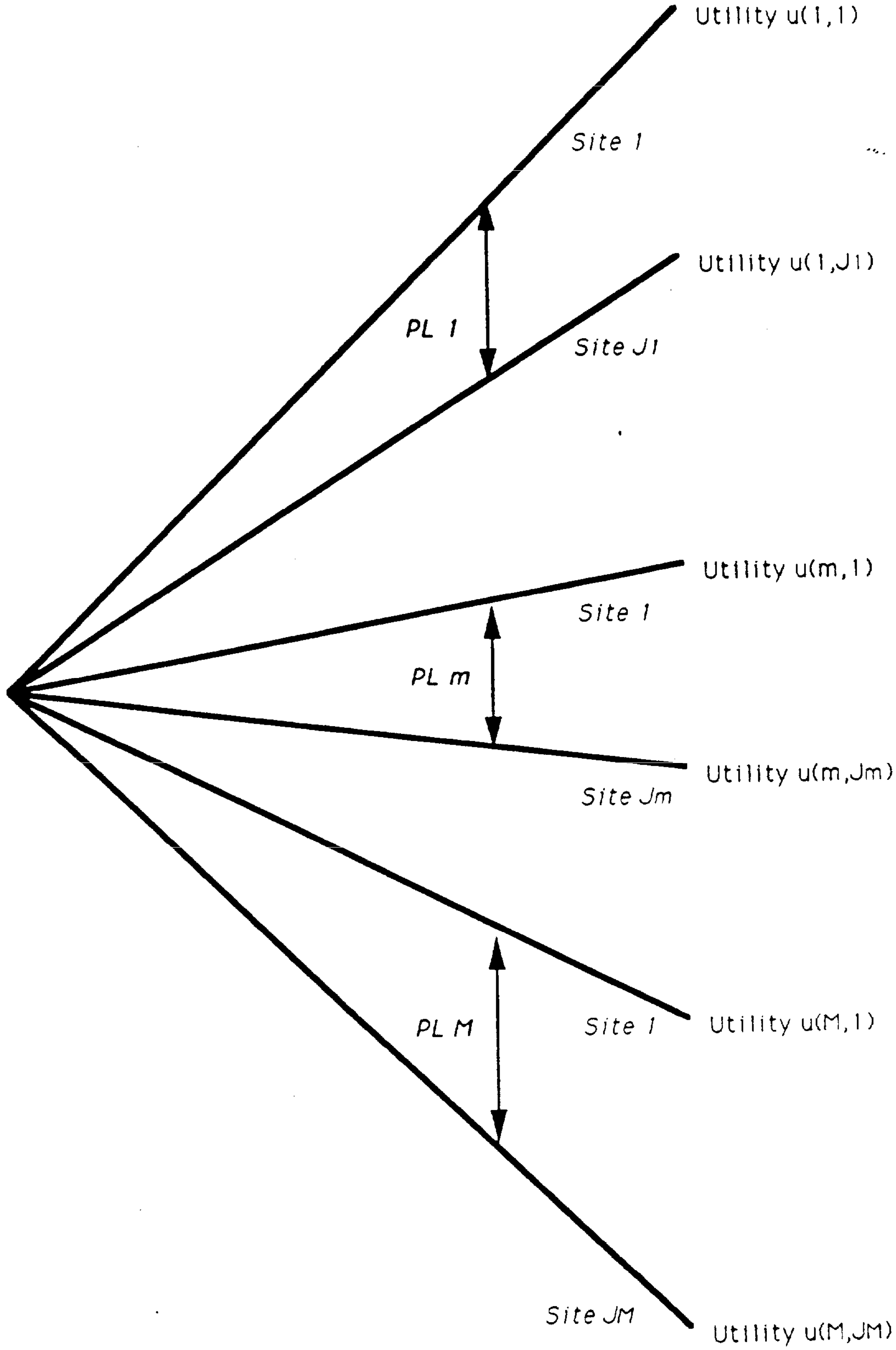


Figure III.1: The flat micro PL-site decision structure.

the same attributes [7] as site *B*, the probability $\pi_{(2,B)}$ would probably be only half its original level, while $\pi_{(1,A)}$ will, most likely, not change. This is just an analog of the famous red-bus/blue-bus problem in the transportation literature. Therefore, where there are obvious differences in patterns of substitution and complementarity across alternatives, the IIA assumption, and hence the MNL, is not appropriate.

Nested Multinomial Logit Model

To avoid the IIA restriction, the *nested multinomial logit* (NMNL) model is better suited for our study. Individuals are hypothesized to adopt a two-level tree-like decision process on any choice occasion. They first determine the target product line, and then choose a site conditional upon the product line decision. This is illustrated in figure III.2. The result is that the IIA property is imposed on sites within a product line, but not across product lines.

Assume that the random utility $\bar{u}$ *(mj),* an individual can receive from first choosing PL *l* arid then site *j* is

$$\tilde{u}_{(m,j)} = \alpha Z_m - \beta_m X_{(m,j)} + \epsilon_{(m,j)}$$

where the attribute vector *X(m,j)* and random terms $\epsilon_{(m,j)}$ are specific to the PL-site choice *(m,j),* while variables in vector Z_m vary only with PLs. The PL characteristics Z_m are shared by all sites available to PL *m.* Also assume that the random terms ϵ follow the generalized extreme value (GEV) distribution defined below.

Definition III.2 The generalized extreme value distribution is defined as

$$F(\epsilon_1, \epsilon_2, \ldots, \epsilon_N) = \exp\left[-G(e^{-\epsilon_1}, e^{-\epsilon_2}, \ldots, e^{-\epsilon_N})\right]$$

where $G(y_1, y_2, \ldots, y_N)$ satisfies the following conditions

1. G is a nonnegative function of $y_i \geq 0$, $i = 1, 2, \ldots, N$..

[7] And consequently C is a perfect substitute for B.

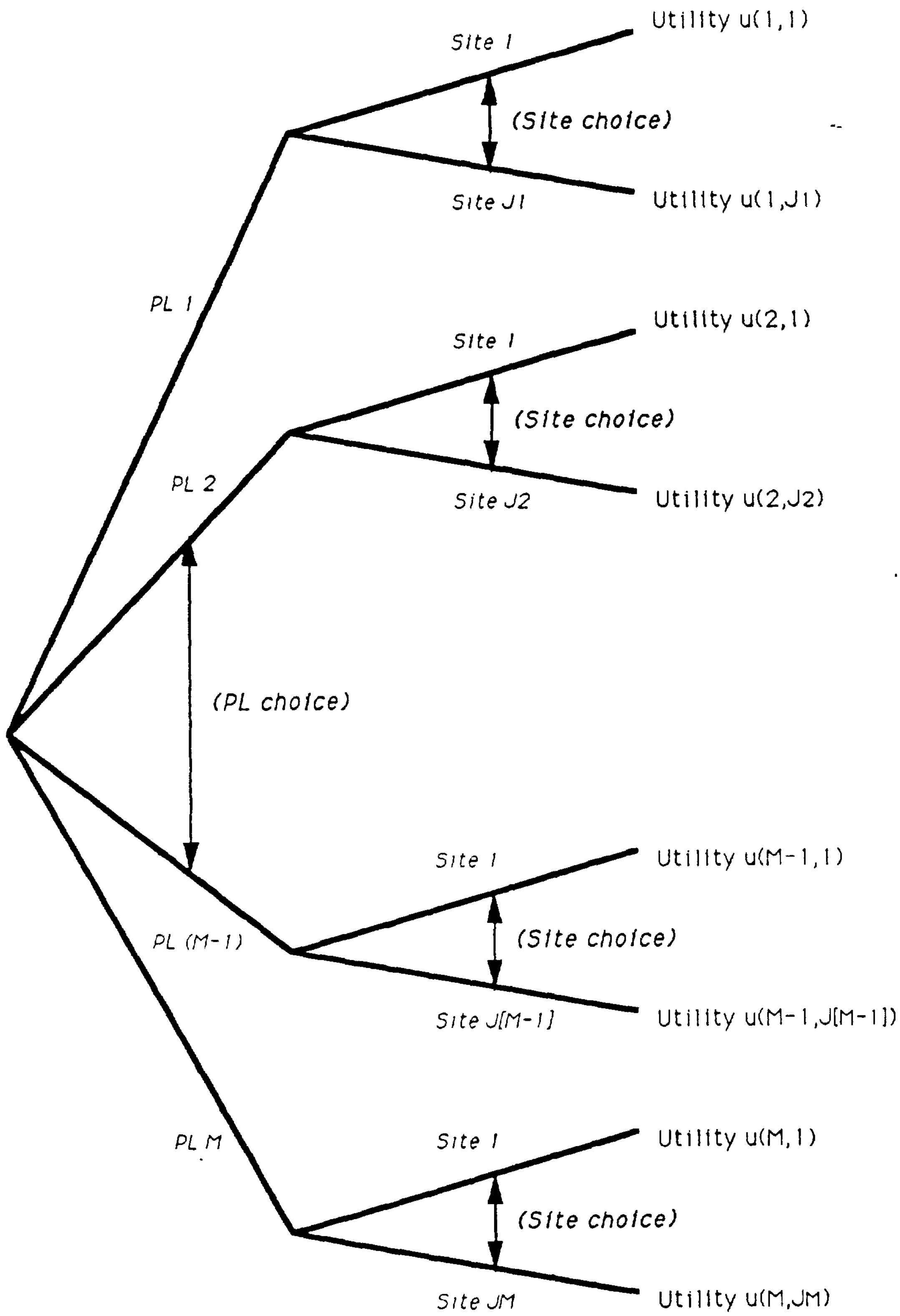


Figure III.2: The two-stage micro PL-site decision structure.

2. G is homogeneous of degree $\mu > 0$
3. $\lim_{y_i \to \infty} G(y_1, y_2, \ldots, y_N) = +\infty$ for $i = 1, 2, \ldots, N$
4. The sth derivative of G with respect to any combination of s distinct y_i's, $i = 1, 2, \ldots, N$, is non-negative if s is odd, and non-positive if s is even. ∎

McFadden (1978) proves the GEV distribution implies that the probabilistic-choice model consistent with utility maximization gives choice probability of the form

$$\pi_i = \frac{e^{u_i} G_i(e^{u_1}, e^{u_2}, \ldots, e^{u_N})}{\mu\, G(e^{u_1}, e^{u_2}, \ldots, e^{u_N})}$$

where G_i is the first derivative of G with respect to y_i.

Now assume that the function G is homogeneous of degree 1 and has the form

$$G(y_{(m,j)},\ \forall (m,j)) = \sum_m \left[\sum_j \left(y_{(m,j)} \right)^{1/\theta} \right]^\theta .$$

Therefore, the disturbances ϵ have the joint distribution

$$F(\epsilon_{(m,j)},\ \forall (m,j) \in \Omega_i) = \exp\left(- \sum_{m=1}^{M} \left[\sum_{j=1}^{J_m} \exp\left(\frac{-\epsilon_{(m,j)}}{\theta} \right) \right]^\theta \right). \qquad \text{(III.8)}$$

The probability of (m,k) being chosen is then

$$\begin{aligned}
\pi_{(m,k)} &= \frac{\exp(\beta_m X_{(m,k)}/\theta)}{\sum_{j=1}^{J_m} \exp(\beta_m X_{(m,j)}/\theta)} \cdot \frac{\exp(\alpha Z_m + \theta I_m)}{\sum_{l=1}^{M} \exp(\alpha Z_l + \theta I_l)} && \text{(III.9)} \\
&= \frac{\exp(\beta_m X_{(m,k)}/\theta)}{\exp(I_m)} \cdot \frac{\exp(\alpha Z_m + \theta I_m)}{\sum_{l=1}^{M} \exp(\alpha Z_l + \theta I_l)} && \text{(III.10)} \\
&= \pi_{k|m} \cdot \pi_m && \text{(III.11)}
\end{aligned}$$

where

$$I_m \equiv \log\left(\sum_{j=1}^{J_m} \exp(\beta_m X_{(m,j)}/\theta) \right) = E\left[\max_j \{ \tilde{u}_{j|m} \} \right] - \text{constant} \qquad \text{(III.12)}$$

is the inclusive value of the sites in PL m. Note that

$$\frac{\pi_{(m,k1)}}{\pi_{(m,k2)}} = \frac{\exp(\beta_m X_{(m,k1)}/\theta)}{\exp(\beta_m X_{(m,k2)}/\theta)}.$$

Consequently, the IIA property continues to hold for sites in the same product line, but not across product lines.

The inclusive value I_m is an index of the overall quality of fishing opportunities of the sites in PL m, or the expected maximum utility the site, of PL m can offer, excluding the utility one can get from PL attributes Z that do not vary across sites. Similarly we can calculate the inclusive value

$$I^* = \log\left(\sum_{m=1}^{M} e^{\alpha Z_m + \theta I_m}\right), \tag{III.13}$$

as an index for the desirability of participation in recreation. The value I^* represents the expected utility of taking a fishing trip[8] and will be used in the participation decision modeling.

Estimates of the parameters can be obtained by employing a two-step procedure: First, the estimates for β'_m ($\equiv \beta_m/\theta$) are obtained by repeatedly applying MNL to each product line m. The inclusive values I_m can then be calculated and used, along with the PL-specific variables Z_m, in the second-stage MNL estimation of α and θ. The original parameters β can then be recovered as $\beta_m = \theta\,\beta'_m$. Note that in the simple logit setup (III.7). the parameter θ is exogenously set to 1, thus excluding the case where different PLs have inherently different utility effects.

The way the NMNL avoids the IIA property is to allow a general pattern of dependence among the choices. This is embodied by the GEV distribution assumption, as opposed to the independent residuals assumption of the simple logit model. This can be more intuitively seen from an alternative derivation of the NMNL by Ben-Akiva and Lerman (1985, pp. 285-91) or Cardell and Steinberg (1988).

They start by splitting the PL-site alternative residuals into two parts and assuming that

$$\tilde{u}_{(m,k)} = u_m + u_{(m,k)} + \epsilon_m + \theta\,\epsilon_{(m,k)}$$

where ϵ_m is the random component attributable to the product line m and common to all sites in PL group m. This generates a correlated structure for the errors across

[8] Net of the constant terms from both the site and product line levels of analysis.

alternatives in the same PL group When $\epsilon_{(m,k)}$ is Weibull distributed, it can be shown that there exist some $\theta \in [0, 1]$ and a random variable ϵ_m, independent of $\epsilon_{(m,k)}$, such that $[\epsilon_m + \theta\, \epsilon_{(m,k)}]$ is also Weibull distributed. The probability statements derived from this specification are exactly identical to (III.9) and (III.11).

The parameter θ is called the *dissimilarity index* because it indicates the share of the common components in the error variance. The smaller θ is, the more similar the sites under PL m are.[9] Therefore, as θ approaches 0, the nested MNL becomes more appropriate. The flat MNL can only be justified when θ is close to 1.

McFadden (1978) shows that a sufficient condition for a NMNL model to be consistent with random utility maximization is that the coefficient θ of the inclusive value I lies in the [0, 1] unit interval. An estimated θ outside the unit interval range hence raises questions about a potential mis-specification of the model.

Estimation of the Sequential Multinomial Logit

Due to the nonlinearity and complexity of the model (III.9), estimation by maximum likelihood is practically infeasible. Instead, the sequential estimation method described above is employed.

One complication arising from this stepwise procedure needs to be addressed. Because the inclusive value index variables used in all stages above the lowest level are in fact estimated from the lower stages, not actually observed variables, the covariance matrices calculated by MNL will be biased and have to be corrected.[10]

A more serious problem with the sequential estimation method that can not be

---

[9] Many authors use $\rho = 1 - \theta$ instead of θ as a measure to indicate the correlation among alternatives in the same group. For example, Maddala (1983), Bockstael et al. (1986, 1988), and Greene (1989).

[10] The process of correcting the estimated covariance matrix to generate consistently estimated standard errors is described in the Appendix in McFadden (1981). Schmalensee and Joskow (1986) also discuss techniques of using estimated parameters as independent variables.

so easily corrected is the loss of efficiency in the estimation process. Note that since parameters β and θ appear in both $\pi_{k|m}$ and π_m of the probability statement (III.11), the full information maximum likelihood (FIML) estimates can only be obtained by taking derivatives of the complete log likelihood function

$$LL = \sum_{i=1}^{N} \log \pi^{i}_{(m,k)} = \sum_{i=1}^{N} \log \pi^{i}_{k|m} + \sum_{i=1}^{N} \log \pi^{i}_{m}$$

with respect to the parameters β, θ and α and setting the first derivatives to zero.

Alternatively, the multi-stage procedure estimates $\beta' = \beta/\theta$ in the first stage by maximizing only

$$LL_1 = \sum_{i=1}^{N} \log \pi^{i}_{k|m},$$

and then estimates α and θ in the second stage by maximizing

$$LL_2 = \sum_{i=1}^{N} \log \pi^{i}_{m}$$

These estimates are thus only limited information maximum likelihood. (LIML) estimates since not all available information in the data is utilized.

The Nested Multinomial Logit Specification

We adopt a linear utility function for this study, which implies there are no income effects from quality changes. Consequently, the compensating variation and equivalent variations for a quality change will be equal.[11] The utility an individual i receives from choosing PL m and site k is assumed to be

$$\tilde{u}^{i}_{(m,k)} = \alpha_i + [\xi_m + \gamma Z_m] + [\lambda_{(m,k)} + \beta_m X_{(m,k)}] + \epsilon_{(m,k)} \tag{III.14}$$

---

[11] Most empirical MNL studies. from the early mathematical psychology work of Luce (1959) to recent recreation demand studies of Bockstael et al. (1988) and Morey et al. (1988), adopt a linear form for the conditional utility function. Feenberg and Mills (1980) argue that, though linearity can hardly be literally true, it “may be a good approximation if the utility function is smooth and the sample variances of the parameters are not too large” (p. 112). The most important reason for our adopting a linear utility function, however, is the ease of consumer surplus computation. When the conditional utility is nonlinear formula for the consumer surplus per choice occasion is very hard to obtain.

where α_i is the individual constant, ξ_m is the PL constant, $\lambda_{(m,k)}$ is the PL-site constant, Z_m is the characteristics vector that varies only with PLs, and $X_{(m,k)}$ are the attributes specific to both PLs and sites. The random elements ϵ are assumed to follow the GEV distribution defined by (III.8).

Given the PL choice m, individual i will select a site k that offers the highest "site" utility

$$\tilde{v}_{(m,k)} = \lambda_{(m,k)} + \beta_m X_{(m,k)} + \epsilon_{(m,k)}.$$

The variables X we use in the estimation include site quality Q and the choice occasion income net of travel costs to site $k\,(Y_i - P_{ik})$, which is available to spend on consumption of market goods, where P_{ik} is the travel cost of an angler i visiting site k. Therefore, the conditional utility function $\tilde{v}_{(m,k)}$ becomes

$$\tilde{v}_{(m,k)} = \lambda_{(m,k)} - \eta\,[Y_i - P_{ik}] + \beta_m Q_{(m,k)} + \epsilon_{(m,k)}.$$

As explained above, we cannot obtain estimates for λ, η and β at this stage; instead we get only $\lambda' = \lambda/\theta, \eta' = \eta/\theta$, and $\beta'_m = \beta_m/\theta$ for each PL m using the sample of individuals who we observed choosing PL m.

In our analysis (as is generally the case) we do not have data on choice occasion income. These missing data are not a problem in the site-choice level of analysis, because the income is constant across sites and so drops out in the estimation (which employs differences between the conditional utility functions.) However, when the marginal utility of income is not constant across alternatives in a nested MNL model, the lack of choice occasion income will affect the higher-level estimation – in our analysis, the choice of product line – and the welfare calculations.

Consequently, we derive in some detail below the value of the lower-level inclusive value index for sites in product line m, I_m, and of the higher-level inclusive value index across product lines. The goal is to explicitly identify the role of the choice occasion income variable. All inclusive value calculations are performed separately for each trip duration category.

The inclusive value I_m for sites in product line m, as defined in (III.12), is

$$
\begin{aligned}
I_m &= \log\left(\sum_j \exp(v_{(m,j)}/\theta)\right) \\
&= \log\left(\sum_j \exp(\lambda'_{(m,j)} + \eta'_m[Y_i - P_{ij}] + \beta'_m Q_{(m,j)})\right) \\
&= \log\left(\sum_j \left[\exp(\eta'_m Y_i)\ \exp(\lambda'_{(m,j)} - \eta'_m P_{ij} + \beta'_m Q_{(m,j)})\right]\right) \\
&= \log\left(\exp(\eta'_m, Y_i) \sum_j \exp\left[\lambda'_{(m,j)} - \eta'_m P_{ij} + \beta'_m Q_{(m,j)}\right]\right) \\
&= \eta'_m Y_i + \log\left(\sum_j \exp\left[\lambda'_{(m,j)} - \eta'_m P_{ij} + \beta'_m Q_{(m,j)}\right]\right) \\
&\equiv \eta'_m Y_i + \log\left(\sum_j \exp(v'_{(m,j)})\right) \\
&\equiv \eta'_m Y_i + \bar{I}_m
\end{aligned}
$$

The estimated parameter η of the travel cost variable P_{ij} is individual i's constant marginal utility of income for product line m. Because of missing data on choice occasion income Y_i, we can only calculate the *pseudo*-inclusive value $\bar{I}_m$ from the estimates λ', η'_m and β'_m.

In the upper-level PL-choice modeling in the NMNL we estimate the parameters of the PL conditional indirect utility function

$$
\begin{aligned}
v^i_m &= \alpha_i + \xi_m + \theta I_m + \gamma Z_m \\
&= \alpha_i + \xi_m + \theta[\eta'_m Y_i + \bar{I}_m] + \gamma Z_m \\
&= \alpha_i + \xi_m + \eta_m Y_i + \theta \bar{I}_m + \gamma Z_m
\end{aligned}
$$

where I_m is the inclusive value index calculated above from the lower-level site-choice MNL estimation. As defined by (III.13), the inclusive value of taking a trip is

$$
I_i^* = \log\left(\sum_m \exp(v_m)\right)
$$

$$= \log\left(\sum_m \exp(\alpha_i + \xi_m + \theta\, I_m + \gamma\, Z_m)\right)$$

If $\eta_m = \eta$ for all m, then we can further simplify the formula:

$$\begin{aligned}
I_i^* &= \log\left(\sum_m \left[\exp(\alpha_i + \eta\, Y_i^*)\, \exp(\xi_m + \theta\, \bar{I}_m + \gamma\, Z_m)\right]\right) \\
&= \log\left(\exp(\alpha_i + \eta\, Y_i^*) \sum_m \exp(\xi_m + \theta\, \bar{I}_m + \gamma\, Z_m)\right) \\
&= \alpha_i + \eta\, Y_i^* + \log\left(\sum_m \exp(\xi_m + \theta\, \bar{I}_m + \gamma\, Z_m)\right) \\
&= \alpha_i + \eta\, Y_i^* + \log\left(\sum_m \exp(v'_m)\right) \\
&\equiv \alpha_i + \eta\, Y_i^* + \bar{I}_i^* \qquad \text{(III.15)}
\end{aligned}$$

The individual-specific constants α_i are not identifiable, and as stated above, we do not know the choice occasion income Y_i^*. Therefore, we cannot calculate the real level I_i^*, and so will use the pseudo-inclusive value $\bar{I}_i^*$ in chapter V for the calculation of consumer surplus.

Because $\alpha_i + \eta Y_i^*$ is constant across product lines (assuming $\eta_m = \eta$), estimation with $\bar{I}_i^*$ is equivalent to estimation with I_i^*. In Chapter V, we further show that if the marginal utility of income is constant across alternatives, the lack of choice occasion income does not pose problems for the welfare analysis. However, if MUI is *not* constant across product lines, then we cannot calculate I_i^* as the three separable components in (III.15): the choice occasion income term remains an integral component of the calculation. As a consequence, estimation using $\bar{I}_i^*$ in place of I_i^* yields a mis-specification.

In our NMNL estimation procedure below, we do not impose the constraint of constant MUI across product lines within a duration group,[12] due to the computational

[12] We would not expect constant MUI across duration groups, but this issue does not pose any difficulties because we model trip-duration choice within the participation model.

complexity it would cause, though such a restriction seems conceptually appropriate. We have developed a procedure for handling the problems posed by not constraining the MUI to be constant. In the process of specifying the correct consumer surplus formulas for the case of varying marginal utility of income, we derive in Chapter V a “weighted” marginal utility of income, where the weights represent the ex-post probabilities of choosing the alternative. The weighted MUI serves the same role in the consumer surplus formula that the (constant) MUI serves in the simpler context. We will substitute the weighted MUI in the calculation of I_i^*, in lieu of performing the NMNL estimation with cross-estimation constraints. Employing this conceptual framework, use of $\bar{I}_i^*$ in the estimation procedure is not a mis-specification.

The Valuation of Time

The Analysis Framework

A critical component of the conditional indirect utility function for site alternatives specified above in equation III.14 is the travel cost P_{ik} per trip by individual i to site k. Conceptually. travel cost consists of two components – distance costs and time costs. As discussed above in Chapter II the time cost component is controversial. We derive in some detail several models which support several different treatments of the opportunity costs of trip time in the discrete choice literature. We show how the models imply not only different measures of travel cost but also different definitions of the choice set of feasible sites.

Anglers are assumed to maximize utility subject to a full-income constraint for the choice occasion, where full income refers to the money budget plus the value of time, following the household production function literature. Note that, among the sites within the angler’s choice set, we only observe the amount of time the individual allocated for a trip to the *chosen* site. We must make assumptions about how much time an individual would allocate for trips to other sites.

The first two frameworks are based on an assumption that *total trip time* will be the same for all sites (as for the chosen site); we label this the "exogenous total trip time" framework. The two variants we develop employ alternative measures of trip time. The third framework is based on the assumption that *on-site time* will be the same for all sites (as for the chosen site); we label this the "exogenous on-site-time" framework. We show below in the Appendix the potentially large effect these differences may have on model results.

Variable Definitions

We first define the following variables for our discussion.

X = market goods (set price=1 as numeriare

D = number of days in trip

w = post-tax wage rate

c = vehicle operating cost per mile

s = driving speed (miles per hour)

q_j = quality of site j, where j = 1, . . . , J

D_j = round-trip distance in miles to site j

$P_j = c\,D_j$ = round trip travel cost to site j

$R_j = D_j/s$ = round trip travel time in hours to site j

S_j = time spent on site j in hours

$T_j = R_j + S_j$ = total trip time to site j in hours

C = choice occasion time in hours

y = money allocated to the recreational choice occasion

Y = full income allocated to the choice occasion

Exogenous Trip Time

Within the exogenous trip-time framework, we impose the assumption that an individual allocates all of the choice occasion time C either to visiting a site j, thereby incurring round-trip travel time R_j plus on- site time S_j, or to other activities (working, other recreation). We can write the generalized time budget, conditional upon participation, for the choice occasion:

$$\sum_j \delta_j (R_j + S_j) = \sum_j \delta_j T_j = C.$$

As defined previously, the indicator δ_j = 1 if site j is chosen for the visit, and δ_j = 0, otherwise. If δ_k = 1, then δ_j = 0 for all $j \neq k$.

The *sources* of full income Y^* include wC, time during the choice occasion valued at the post-tax wage rate w,[13] and y, the money income allocated for expenditure during the period:

$$Y^* = wC + y \tag{III.16}$$

The *uses* of full income during the period are expenditures during the trip on market

---

[13] Recognizing that recreational activities take up time, much of the recreation demand literature relates the opportunity time costs to the wage income forgone when a trip is taken. However, the labor supply literature now recognizes that work, time may not be a continuous choice variable over which individuals can freely trade-off income and recreation at the wage rate. Only individuals with flexible work hours can adjust their marginal rate of substitution between work and recreation and make it equal to their marginal wage rates. These individuals are said to be at interior solutions. Others, who either have to work fixed hours or do not work at all, are at corner solutions, and their wage rates cannot serve as the value of their leisure time.

Some authors in the recreation demand literature have also adopted this view and have treated interior solutions and corner solutions differently. For people at interior solutions, work time is at their discretion, and trip time can be traded for income at their marginal wage rate. For others without such freedom, no opportunity, wage cost exists since they cannot increase their work effort even if no trip is taken.

Unfortunately no wage rate was directly measured in our data, so the above treatment is impossible. The budget frontier therefore has to be assumed a straight line, and people are assumed to be at interior solutions. We assume that people value their time at their wage rate, calculated as annual personal income divided by projected working hours per year.

goods X plus distance costs and time costs to the chosen site:

$$Y = X + \sum_j \delta_j [cD_j + wR_j + wS_j]$$

Setting *sources* equal to *uses,* we have

$$wC + y = X + \sum_j \delta_j [P_j + wR_j + wS_j]. \tag{III.17}$$

Conditional upon participation in recreation at site j (i.e., $\delta_j = 1$), we can solve for $X = y - P_j$. The indirect utility function conditional upon participation at site j is

$$V_j = V(y - P_j, q_j).$$

Assuming a linear functional form, the conditional indirect utility function becomes

$$V_j = \beta_1 [y - P_j] + \beta_2 q_j.$$

The important point to note is that time costs have completely dropped out of the travel cost measure for site choice, because the amount of time allocated to "producing" recreation (the choice occasion) equals the trip duration. The use of the standard travel cost variable incorporating both time and distance costs cannot be supported in this framework. [14] With a conditional direct utility function of the form

---

[14] Besides monetary vehicle costs P_j, the amount of driving could conceivably have at least two other effects on an angler's utility: a reduction in available fishing time S_j and the (dis)utility of driving R_j itself. A more elaborate and complete utility function will, therefore, be

$$\begin{aligned} V_j &= V(y - P_j, R_j, S_j, q_j) \\ &= V(y - cD_j, R_j, T_j - R_j, q_j) \end{aligned}$$

The linear estimating function is then

$$\begin{aligned} V_j &= -\beta_1 [cD_j] + \beta_2 R_j + \beta_3 [T_j - R_j] + \beta_4 q_j \\ &= -\beta_1 [cD_j] + \beta_2 [D_j / s] + \beta_3 [T_j - D_j / s] + \beta_4 q_j \\ &= \left(-\beta_1 c + \frac{\beta_2 - \beta_3}{s} \right) D_j + \beta_3 T_j + \beta_4 q_j \\ &= \left(-\beta_1 + \frac{\beta_2 - \beta_3}{cs} \right) P_j + \beta_3 T_j + \beta_4 q_j \end{aligned}$$

$V_j(X, q_j)$, the time costs affect angler decision-making only at the higher level where the participation choices for each trip duration are made.

We suggest two alternative methods for implementing the exogenous trip duration model: the first measures trip duration in hours, based on the self-reported hours (and days) for the beginning and end of the trip; the second measures trip duration based on the number of days the individual reported being away. Though the conditional indirect utility specification is the same, the definition of the choice set for each individual, the consistency checks for selecting individuals into the sample, and the value of the time cost variable in the participation model differ.

Hypothesis 1: Exogenous Trip Time (Using Self-Reported Trip Hours)

In this case, total trip time T is based on self-reported trip duration. The amount of income allocated to the choice occasion is calculated as (III.16). but has no practical significance in the site choice analysis.

Let constant h be the presumed number of hours people are awake and active during a day. Its complement h' $(\equiv 24 - h)$ is then the time in hours people rest each day. For a trip of D days, people necessarily must rest for $(D - 1)$ nights, a total of $(D - 1)\,h'$ hours, at their fishing site or on the road. Therefore. we impose two time constraints for a D-day trip to any potential site j:

$$\textbf{A1}: T_j \leq h^* \equiv D\,h + (D-1)\,h'$$

$$\textbf{A2}: S_j = T_j - R_j \geq (D-1)\,h'$$

The real marginal utility of income $\eta = \beta_1$ cannot be identified. The marginal utility of income measure we obtain without including the R_j and S_j terms in V_j is $\eta' = \beta_1 - \frac{\beta_2 - \beta_3}{cs}$, which is actually a combination of the various driving effects. Anticipating negative utility from driving and reduced fishing time, we have $\beta_2 < 0$ and $\beta_3 > 0$. Therefore,

$$\eta' > \eta > 0.$$

The resulting consumer surplus we derive using η' will consequently be an under-estimate of its real value.

Constraint A1 asserts that the total trip duration cannot exceed the limit of *h'* hours, which is the sum of active and resting time during the day. Constraint A2 enforces that people still have time left after accounting for driving and resting to fish and enjoy site amenities. Combined: they imply that $R_j \leq D\, h$. In words; this posits that round-trip driving time (R_j) cannot take up all the time people are awake during the trip.

Any site *j* that violates either A1 or A2 is considered infeasible for a visit on the choice occasion in question. The time constraints A1 and A2 thus define the individual choice set of feasible sites. People whose reported destination sites of the observed trips violate these time constraints are thus treated as outliers and deleted from the MNL sample.[15] The value selected for *h* hence plays a critical role in the site choice analysis.[16] When a larger *h* is used: more people will be included in the sample, and more sites will be included in each individual's choice set.

To be included in the MNL sample: each angler's destination site *k* must satisfy time constraints A1 and A2 specified above. That is,

$$\mathbf{S_1}: \begin{cases} T_k & \leq h^* \\ S_k = T_k - R_k & \geq (D-1)\, h' \end{cases}$$

For an individual in the sample S_1 who chooses site *k* (and hence has a pre-determined total trip duration T_k). site *j* will be included in his or her choice set if and only if

$$\mathbf{C_1}: \begin{cases} T_j \equiv T_k & \leq h^* \\ S_j = T_k - R_j & \geq (D-1)\, h' \end{cases}$$

[15] In the actual implementation of the model, an individual can be excluded from entering the MNL sample due to three reasons: (1) the trip duration data T is missing, (2) the observed chosen site itself violates the time constraint A1 and/or A2, or (3) the chosen site is the only feasible site, so no other site is contained in his or her choice set. People who are left out from the site MNL estimation for the third reason may be included in the upper level PL MNL estimation because the inclusive value can still be calculated even though there is only one feasible site for the chosen PL.

[16] The choice set definition in turn has direct effects on the MNL and total trip estimation, as well as the final welfare calculation.

The satisfaction of the first constraint of C_1 is guaranteed since it is exactly the same as the first restriction of S_1 for inclusion in the sample. The restrictions of S_1 are applied to the selected sites, while the restrictions of C_1 are applied to all other sites.

Hypothesis 2: Exogenous Trip Time (Using Total Trip-Days)

In this case, we eschew using self-reported measures of total trip time and alternatively impose the assumption that the total trip duration T equals the maximum number of trip hours allowed in a D-*day* trip, h^*. Therefore, $T = h' = Dh - (D - 1) h'$. This procedure eliminates the measurement error and missing data problems posed by the first procedure: but incorporates probably greater measurement error in the trip time (cost) variable by imposing the assumption that each trip uses all waking hours of the day.[17]

Under this hypothesis, the conditions A1 and A2 combined are equivalent to

$$\mathbf{S_2} : \; R_k \leq D\,h.$$

The choice set of an individual can also be computed by including in it any site j that satisfies

$$\mathbf{C_2} : \; R_j \leq D\,h.$$

Hypothesis 3: Exogenous On-Site Time

The final hypothesis imposes an alternative assumption that on-site time is fixed across site choices: rather than trip duration. To ensure an exogenously defined choice occasion, we must include in the model the possibility of 'slack time' during the choice occasion Even using this device. there is some question for day trips as to whether

[17] We can calculate the difference from the site-hours variable, but we do not know how much measurement error there is in site-hours.

one of the conditions for welfare analysis necessarily holds: that only one trip could be taken during the choice occasion [18] We rewrite the time-budget from above to acknowledge that choice occasion time C and trip time T are no longer assumed to be the same duration:

$$\sum_j [R_j + S_j + \theta_j] = \sum_j [T_j + \theta_j] = C,$$

where θ_j is slack, and $C = h^* = Dh - (D - 1)\, h'$.

The solution to equation (III.17) in which we set sources equal to uses is:

$$X = [y - P_j] + w[C - R_j - S_j].$$

By assumption, y, C, and S_j are constant across sites, and so are not relevant to the site choice decision-making. The measure of travel cost, however, is $(-P, -w\dot{R}_j)$: the time cost of travel is included along with the distance cost! unlike for the models above. in which only the distance cost ($-P_j$) is included.

The conditions an individual must meet to get into sample S_3 are the same constraints A1 and A2 that define S_1 above. Therefore, $S_3 = S_1$. For an individual i in the sample S_3 whose actual destination is site k, the constant on-site time is calculated to be

$$S^* \equiv S_k = T_k - R_k.$$

The choice set C_3 of i is defined by including any site j that satisfies

$$\mathbf{C_3} : \begin{cases} T_j = S_k + R_j & \leq h^* \\ S_j \equiv S_k & \geq (D - 1)\, h' \end{cases}$$

Now the second constraint of C_3 is always satisfied.

Though the two samples of anglers S_1 and S_3, defined respectively for the exogenous trip duration and the exogenous on-site time hypotheses, are identical, there is

---

[18] This hypothesis is adopted by McConnell et al. (1990) and Parsons and Kealy (1990). The latter assumes that all anglers spend a fixed amount of four hours on site.

no relationship between the choice sets C_1 and C_3 defined for an individual under the alternative hypotheses.[19]

It is also obvious that

$$S_1 \subseteq S_2 \text{ and } S_3 \subseteq S_2$$

since for S_2 (1) the deletions necessitated by S_1 and S_3 due to to missing data on self-reported trip-lengths are avoided, and (2) the trip durations in S_2 are assumed to be their maximum value h^*. It can further be shown that, for people in both samples S_1 and S_2,

$$C_1 \subseteq C_2$$

and, for people in both samples S_3 and S_2,

$$C_3 \subseteq C_2.$$

We will show the sensitivity of the model to the different trip time hypotheses in the Appendix.

---

[19] It is easy to show this with examples. Suppose h = 15 and D = 1. For an angler with T_k = 5 and R_k = 2, a site j with R_j = 6 will be in set C_3, whereas not in C_1. But if T_k = 14. R_k = 1 and R, = 4, constraints for C_3 will be violated, while those for C_1 are satisfied. The first example is a short day trip (e.g., an afternoon trip) to a nearby site. In this case faraway sites might be included in C_3 and excluded from C_1. The second example is a long day trip (e.g.. a whole-day fishing excursion) to a nearby site. In this case farther sites will probably be left out from C_3, but still incorporated in C_1. Conceivably this will mainly happen to people taking day trips, as most sites will be available to all anglers taking long trips.

CHAPTER IV

THE MACRO-LEVEL PARTICIPATION MODELING

After modeling the *micro* PL-site choice decision, the next step is to model the determinants of the total number of trips a licensed angler takes during a season.. It is theoretically possible to model jointly the discrete product-line/site choices and the total participation decision; however, the data and computational requirements for the correct treatment of the corner- solutions implied by zero trips of certain categories makes an integrated utility-theoretic model practically infeasible. Essentially, researchers appear to face a trade-off: they either implement a "utility-theoretic" framework that does not properly model the statistics of the corner solutions; or they model the *micro* and *macro* decisions in separate models that may address the corner-solution problem but do not form an integrated utility-theoretic framework.

In this chapter, we first discuss variants of the former approach, in which total participation is modeled as the sum of independent participation decisions made on each choice occasion throughout the season. We then summarize the Bockstael, Hanemann and Strand (1986) critique of this approach and their alternative proposal to model directly the corner solution. Finally, we develop our own model, which is in the spirit of the second approach.

Due to severe data limitations at the total participation level, our model is substantialiy different from the standard treatment in the literature. We do not know the total number of season trips: our *macro-level* information is limited to the duration between trips, and this variable is censored because we only observe the duration from last trip to the survey return date, not to the subsequent trip. By incorporating a key result from stochastic renewal theory in our modeling, we are able to estimate the determinants of the between-trip durations with a stochastic renewal model and then to derive the total number of trips in a season from the duration model. To accommodate the different trip durations, we develop a competing risks model; to allow for variations in site quality throughout the open-water season: we incorporate time-varying covariates in the model.

Participation as the Sum of Independent Trip Decisions

To integrate the participation decision with the PL-site decision in one framework: the utility level u_0 associated with not taking a trip on the current choice occasion has to be specified. Individuals are hypothesized to determine whether to take a trip by comparing u_0 with the expected maximum utility of taking a trip. A trip will consequently be taken if and only if

$$I^* = max\{u_{m,j};\ \forall(m,j)\} > u_0.$$

For empirical estimation, the relationship between the random element ϵ_0 associated with the no-trip utility u_0 and the other random terms $\epsilon_{(l,j)}$ has to be specified. Bockstael et. al. (1986) derive the repeated NMNL model by extending the generalized extreme value (GEV) distribution (III.8) employed for modeling of PL-site choices in the previous chapter to the joint distribution

$$F(\epsilon_0;\ \epsilon_{(m,j)};\ \forall(m,j)) = \exp\left(-e^{-\epsilon_0} - \left[\sum_m \left(\sum_j \exp(-\epsilon_{(m,j)}/\theta)\right)^{\theta/\sigma}\right]^{\sigma}\right). \tag{IV.18}$$

The parameter θ is still the common index of correlation of the random terms *(m, j)* for sites under PL m.[1] The participation decision is illustrated in figure IV.1

The probability that an individual will take a trip in period *t* with the given GEV distribution (IV.18) can be shown to be

$$\pi_*^t = \frac{\exp(\sigma I^*)}{\exp(u_0) + \exp(\sigma I^*)}, \tag{IV. 19}$$

while the probability of no participation is

$$\pi_0^t = 1 - \pi_*^t = \frac{\exp(u_0)}{\exp(u_0) + \exp(\sigma I^*)}. \tag{IV.20}$$

Because the micro-level decisions regarding the trips are nested within the participate/ no-participation decision. the participation choice is not characterized by the IIA restriction. With this model. the *micro* PL-site choices and *macro* participation decision can be estimated simultaneously if the participation and PL-site choice data are available for all periods. However: the framework is one of *repeated* choices, where the decision on any choice occasion is independent of the choices on all other choice occasions

The Alaska fisheries study by Carson: Hanemann, Gum and Mitchell (1987) is the only one to our knowledge that estimates a repeated nested logit model with complete trip information throughout a season [2] In their model, a sport fishing angler can take up to a maximum of three trips in a single week. Let v_{it}^k denote the utility an individual *i*. can receive from taking *k* trips during week *t*. Then the participation probability for having *m* (= 0, 1,2,3) fishing trips is

$$\pi_{it}^m = \frac{\exp(v_{it}^m)}{\sum_{k=0}^3 \exp(v_{it}^k)}.$$

They then maximize the likelihood function

$$\mathcal{L} = \prod_i \prod_t \pi_{it}^{k*},$$

[1] When $\sigma = 1$. the no-trip option is treated as just another alternative, and the model degenerates to a standard MNL (i.e., it is not nested over the participate/no-participate decision.)

[2] Stating that fishing opportunities in Alaska change dramatically over a season. Carson et. al. incorporate weekly choices in the mode! and allow the covariates to vary from week to week.

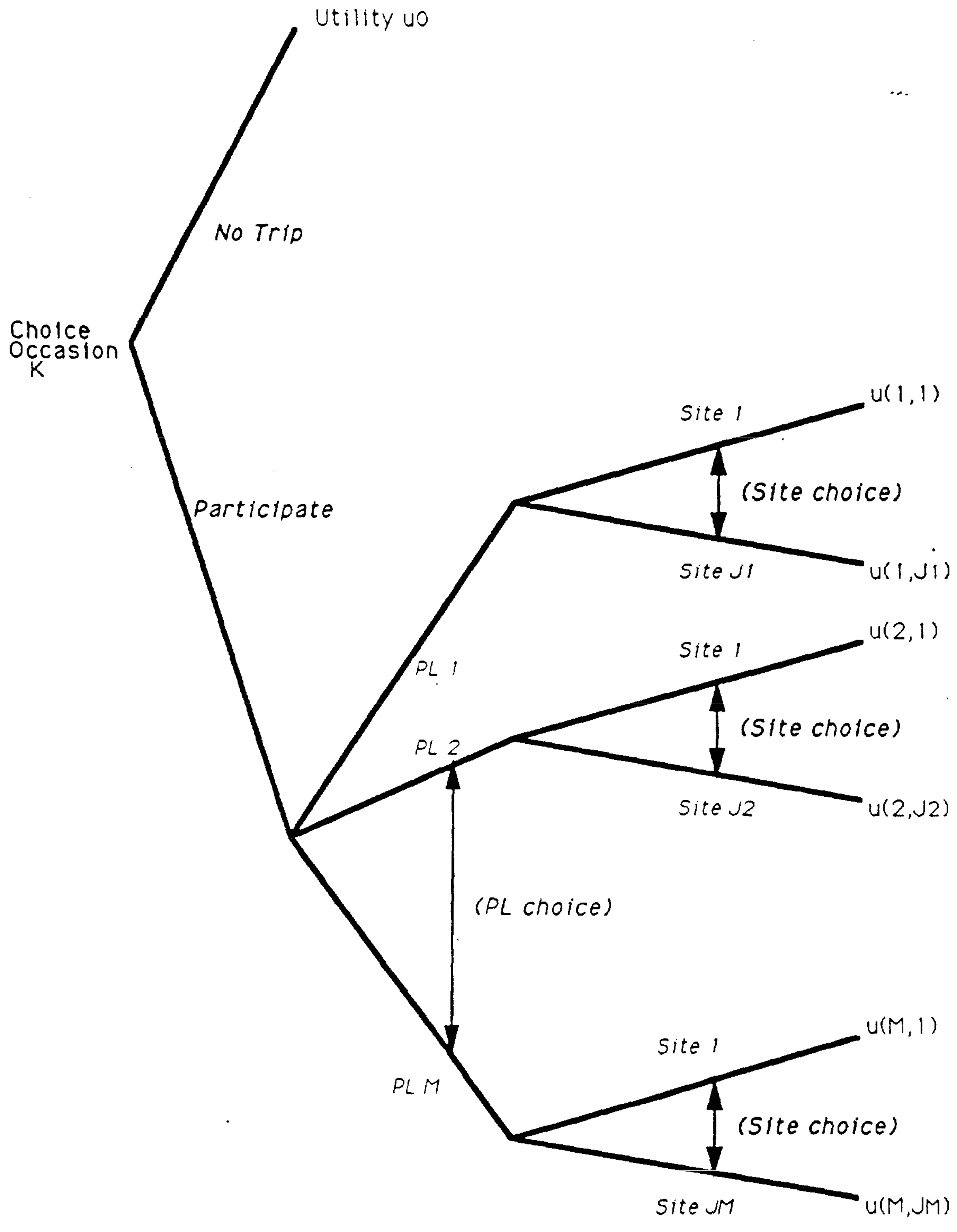


Figure IV.1: The choice occasion participation decision where k^* is the number of trips taken by i in week t. Because the mean number of trips during a week taken by those with more than two trips was 3.63, the expected number of seasonal trips is calculated as

$$Q_i = \sum_t \left[\pi_{it}^1 + 2\pi_{it}^2 + 3.63\,\pi_{it}^3\right].$$

More typically, researchers know the total number of trips in a season: but only have detailed trip information about one trip. The recent paper by Morey et al. (1991) provides a good example of a model designed for such data.[3] With the available site choice information J. the probability density for the micro decision can be formulated as

$$f(J) = \prod_i \pi_i^*$$

where π_i^* is the probability that i would choose his or her actual destination J, in the MNL setup. The total number of trips K is then used to derive the combinatorial participation probability density

$$g(K) = \prod_i \left\{ \left[\frac{T!}{K_i!\,(T-K_i)!}\right] (\pi_i^0)^{T-K_i}(1-\pi_i^0)^{K_i} \right\}$$

where π_i^0 is the probability of i not taking a trip during the period that the site decision is known, K_i is the number of total trips taken by i, and T is the number of choice occasions in the whole season. Morey et al. maximize the complete likelihood function

$$\mathcal{L}(J,K) = f(J)\cdot g(K).$$

The expected total number of trips an individual i would take when there are T choice periods in a year is simply $[T \cdot (1 - \pi_i^0)]$.

---

[3] Their model allows for different distributions for the participation and site choice decisions. However, as they note, it is neither a repeated standard MNL nor a repeated nested MNL because it does not incorporate a stochastic component in the indirect utility function conditional upon no participation.

The Critique and an Alternative Proposal

Bockstael et al. (1986, pp. 185-86 and 1987, p. 13) critique. the class of participation models highlighted above; on the grounds that they characterize total participation simply as a sum of independent decisions on each choice occasion. In particular, they criticize the models because the occurrence of a season with no trips happens merely by accident: the probability of no participation throughout the season is simply the product of the probabilities of no participation on each choice occasion.

When individuals no longer choose “interior solutions” to the utility maximization problem. then the well-behaved, continuous properties of neoclassical demand theory no longer hold. One must instead model the probability statements with Kuhn-Tucker conditions. The problem with the models developed above is that they do not incorporate the discontinuity of the indirect utility functions as individuals switch among different consumption regimes.

A switching regressions model is appropriate to capture statistically the different regimes. Unfortunately the dimensionality of the problem is generally one less than the number of commodities not consumed. Given the level of detail in the random utility models and the many expected corner solutions for most individuals, it appears practically infeasible to integrate over the number of cumulative distribution functions that would be required with either the direct or indirect Kuhn-Tucker conditions. Bockstael et. al. conclude that “without attempting to estimate the corner solutions, there appears to be no consistent way to link independent discrete choice decisions and a macro decision for total trips with a common underlying utility maximization framework” (1986, p. 186).)

They propose an alternative method in which the expected number of trips to all sites over the season T may be interpreted as:

$$E[T] = E[T|T > 0]Pr[T > 0]$$

where the second term on the right-hand side is the probability that the individual engages in any recreation during the season. The equation can be estimated with Tobit, Cragg or Heckman selection procedures. In this method the decision to ever-participate is estimated directly, allowing the researcher to characterize the role of factors such as poor health, adverse financial conditions, or unusually heavy working loads.

The Stochastic Renewal Approach

A major data problem we confront in modeling total trip participation is that we do not know the total number of recreational fishing trips. Therefore, we cannot employ the conventional estimating approaches discussed above! in which the dependent variable is the total number of trips. To accommodate our special data needs, we have developed an alternative framework for modeling the decision about total participation.

As noted above, our information about trips is limited to the duration between trips, and this variable is censored: we only observe the duration from last trip to the survey return date, not to the subsequent trip. Consequently, we estimate a duration mode! of the period between trips, from which we then calculate the expected number of trips. We draw upon a key result in stochastic renewal theory to adapt the duration model to handle the right-censored data.

We incorporate time-varying covariates in the duration model. In addition, we know the length of the most recent trip taken by an angler, which allows us to estimate anglers' demand for trips of different durations. To include this information: we develop a competing risks framework in which individuals may end their spell of no-trips by choosing any one of three trip-lengths (day; weekend, 2-4 days; or vacation, 5+ days). Finally we have some individuals in the sample who took no trips during the period about which they were questioned. We develop procedures to model this right- and left-censored duration data.

The development of the full model requires an extended discussion below due to the many features that have been incorporated. To start, we outline the basic stochastic renewal model, in which the number of trips taken during a period of time is a renewal process. We develop the participation model first for the special case of an exponentially-distributed duration variable and Poisson-distributed trip counts, because the intuition of the model is more accessible with the simpler formulas of the special case. In the next sections of the chapter, we extend the exponential-Poisson model to accommodate: right-censored inter-trip duration data; time-varying covariates: competing risks; and right- and left-censored trip durations.

We then develop the model using the Weibull distribution for inter-trip durations, in order to relax the special assumptions of the exponential-Poisson case. The subsequent four sections follow a similar pattern to the discussion of the exponential model.

The Stochastic Renewal Process

We assume that the number of trips taken during a period of time is a *renewal process:* in which the between-trip durations are independently and identically distributed. Let T be the random variable of independent time spells between successive trips[4] taken by individual *i.* Denote the probability density function (PDF) of T by

$$f(t) \equiv \lim_{\delta \to 0-} \frac{\text{Prob}\{t \leq T < t + \delta\}}{\delta}$$

and the cumulative density function (CDF) by

$$F(t) \equiv \text{Prob}\{T < t\} = \int_0^t f(s)\, ds.$$

---

[4] We ignore the spell of a trip. There are two possible interpretations. First. trips are assumed to be instantaneous events for modeling convenience. Second, when an angler decides to begin a trip on a certain day, he/she decides simultaneous!y not to have another trip during the duration, of the trip.

The *survival* function $S(t)$: which yields the probability that the duration T will be longer than t, is defined as

$$S(t) \equiv \text{Prob}\{T \geq t\} = 1 - F(t) = \int_t^{\infty} f(s)\, ds$$

Hence $S(0) = 1$ and $S(\infty) = 0$, while $F(0) = 0$ and $F(\infty) = 1$. Another conceptually useful function, the *hazard rate* function, is defined as

$$h(t) \equiv \lim_{\delta \to 0+} \frac{\text{Prob}\{t \leq T < t + \delta \mid t \leq T\}}{\delta} = \frac{f(t)}{S(t)} = -\frac{d \log S(t)}{dt},$$

which measures the conditional probability of taking the next trip at time t, given that no trip has been taken before t. A model with a constant hazard rate is said to be *duration independent.*

These functions will be used below to derive the maximum likelihood estimator of T. Note that the PDF and hazard rate are just two different ways of describing the same probability distribution. Given the PDF, the hazard rate function can be uniquely determined, and vice versa.

Let us first look at the special case of the Poisson- Exponential distribution. As Kiefer (1988, p. 652) points out, the exponential distribution is simple to work with and to interpret. However. it may be too restrictive in that no duration dependency is allowed. More flexible distributions, such as Weibull, [5] will be considered next.

The Exponential Distribution

Suppose the time spell T_i between successively taken trips k and $(k + 1)$ by individual i follows the exponential distribution with parameter $\lambda_i > 0$. All durations are independently distributed. The PDF for $T_i (\geq 0)$ is then

$$f_i(t) = \lambda_i e^{-\lambda_i t}$$

[5] The exponential distribution. is a special case of the Weibull distribution. Thus we can conduct a nested model test to check the appropriateness of using the exponential distribution.

and the corresponding CDF is

$$F_i(t) = 1 - e^{-\lambda_i t}.$$

Thus, the hazard rate is

$$h(t) = \frac{f_i(t)}{S_i(t)} = \lambda_i.$$

Since it is constant for an individual i at any time $t > 0$; it is called the *memoryless* property which is unique to the exponential distribution. We assume $\lambda_i = e^{\beta X_i} > 0$, that is, the parameter λ_i is a log-linear function of X_i, which consists of both personal and site variables. β is assumed to be identical across al! individuals.

Given observations of the *completed* durations t for each individual i in the data, the log likelihood function LL can be formed as follows

$$\begin{aligned} LL &= \sum_i \log f_i(t_i) \\ &= \sum_i \log(\lambda_i e^{-\lambda_i t_i}) \\ &= \sum_i [\log \lambda_i - \lambda_i t_i] \\ &= \sum_i [\beta X_i - t_i e^{\beta X_i}]. \end{aligned}$$

The maximum likelihood estimates $\hat{\beta}$ can then be obtained by maximizing the log likelihood function LL with respect to β and setting the first derivatives to zero. This gives us

$$\left.\frac{\partial LL}{\partial \beta}\right|_{\beta=\hat{\beta}} = \sum_i [X_i - t_i X_i e^{\hat{\beta} X_i}] = \sum_i [X_i(1 - t_i e^{\hat{\beta} X_i})] \equiv 0.$$

Note that the expected duration $E[T_i]$ for the exponential distribution $f_i(t)$ given above is just

$$E[T_i] \equiv \int_0^\infty f_i(t)\, t\, dt = \frac{1}{\lambda_i} = \frac{1}{e^{\beta X_i}}.$$

Our goal, however, is the counting process $N_i(S)$, which records the number of trip occurrences in a time period S. In this case, the counting process $N_i(S)$ corresponding to the exponentially distributed between-trip durations is Poisson distributed [6] with

[6] See Ross (1963, pp. 35-36) or Taylor and Karlin (1984, pp. 188-89) for proof.

the discrete PDF:

$$\text{Prob}\{N_i(S) = n\} = \frac{(\lambda_i S)^n e^{-\lambda_i S}}{n!}$$

and expected value

$$E[N_i(S)] = \lambda_i S.$$

The parameter $\lambda_i = e^{\beta X_i}$ (calculated for each individual *i)* has an intuitive interpretation of being the expected number of trips individual *i* will take in one unit of time. Taking days to be the unit of time, the expected number of trips in our Poisson process thus can be readily calculated as the number of days *S* in a fishing season multiplied by λ_i for each individual *i*.

To justify the use of the Poisson-Exponential distribution, we have to refer back to the basic postulates of a Poisson process. It has been proved that a counting process $\{N(S), S \geq 0\}$ is Poisson distributed with parameter $\lambda(>0)$ if the following postulates are satisfied: [7]

1. $N(0)=0$. That is, no trip has occured prier to the start of the time interval $S = 0$.

2. The time intervals between trips are stationary and independently distributed.

   A counting process is *stationary* if the distribution of the number of occurrences (in this case, trips) in any interval of time depends only on the length of the time interval. It is *independent* if the number of occurrences taken in disjoint time intervals are independent.

3. $P\{N(s)=1\} = \lambda s + o(s)$ as $s \to 0$.

   This posits that the probability of having exactly one trip in a very short time interval *s* is proportional to the length of the interval. The function *o(s)* is defined to have the property that

   $$\lim_{s \to 0} \frac{o(s)}{s} = 0.$$

4. $P\{N(s) \geq 2\} = o(s)$ as $s \to 0$.

   This posits that the probability of having at least two trips in a very short time period *s* is very small and can be ignored.

---

[7] See Ross (1983, pp. 32-34) or Taylor and Karlin (1984, pp. 181-184) for proof.

These are reasonable assumptions to make regarding fishing-trip behavior. The advantage of using the Poisson-Exponential pair is that no extra work is necessary to calculate expected total trips $E[N(S)]$.

Data Limitation

A severe limitation with our data is that we do not observe any completed spell T_i. What we have are only the date the last trip began (L_i) and the date the questionnaire was returned (R_i). Consider R_i to be a random censoring point which truncated the spell in question before it was completed.[8] Let the unobserved date of the next trip taken by i (after the questionnaire return date R_i) be V_i, which would have been the endpoint of the sampled duration if it had not been terminated prematurely.

We can then define the following three random variables

$$\begin{aligned} \textit{Age:} \quad & A_i \equiv R_i - L_i \\ \text{Residual life:} \quad & Y_i \equiv V_i - R_i \\ \text{Life of sampled observation:} \quad & B_i \equiv V_i - L_i. \end{aligned}$$

Of these three variables, only age is observed. See figure (IV.2) to illustrate the relationship among the three variables. This is illustrated in figure (IV.2).

It is well known in the stochastic processes literature that the expected length of an inspected duration B_i is greater than that of *a population* duration T_i, due to the greater likelihood of sampling longer intervals. This is called *length-biased sampling*. To distinguish between the sampled interval B_i and the population duration T_i, the latter will be called *normal life* in the discussion below, following the convention in the stochastic processes literature.

For the exponential duration case, it can further be shown that (1) both A_i and

---

[8] The censoring mechanism should be independent of the last trip date L_i.

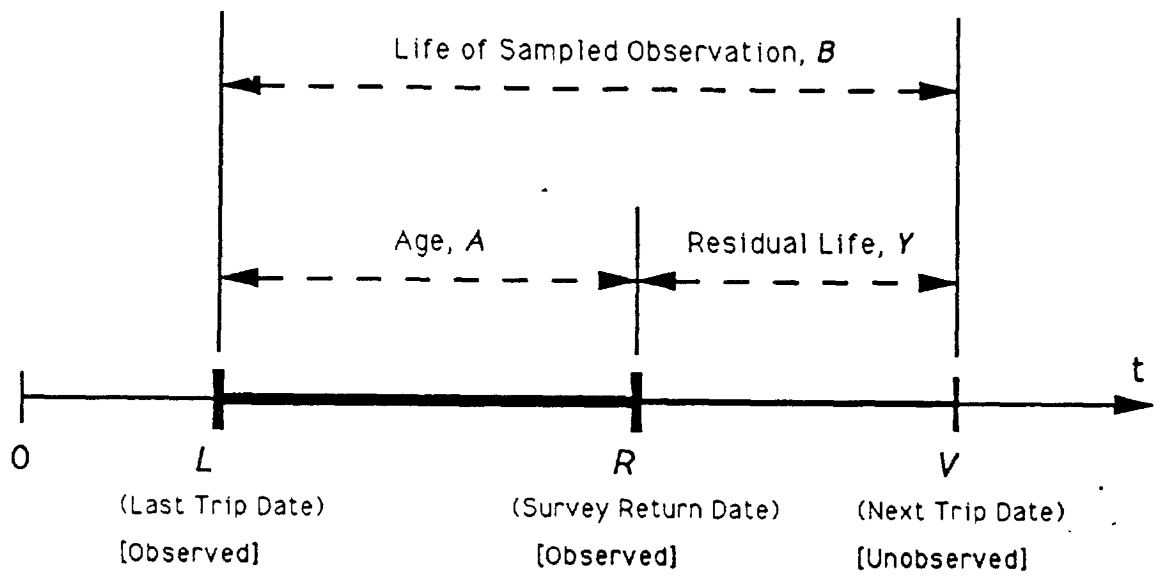


Figure IV.2: The truncated between-trip duration

Y_i have the same distribution as T_i if sampling occurs after the renewal process has been ongoing for a long time. and (2) the length of the sampled interval containing the sampling point R_i is expected to be twice that of a normal life interval T_i, known as the famous *inspection paradox*.[9] Therefore, in the limit,

$$E[A_i] = E[Y_i] = E[T_i] = \frac{1}{\lambda_i}.$$

The solution to our limited data problem will then involve the following steps:

1. We can first estimate the parameters of the age (A_i) distribution using the available age data.

2. Since age A_i and normal life T_i have the same distribution, the parameters obtained for A_i are exactly those for T_i.

[9] See Taylor and Karlin (1984), pp. 282-84.

3. The seasonal total trips can then be calculated using the estimate $\hat{\lambda}_i$.

The procedure outlined above is not limited to the exponential case. We can always derive the distribution of age *A* from any given distribution function for the regular life *T*. Therefore it is genrally the case that parameters in the normal life distribution can be estimated with age data/distribution, if normal life data/distributions are not available.

Time-Varying Covariates

So far we have assumed that each individual *i* has a constant exponential parameter λ_i across time. Since conditions at recreation sites (part of the X_i vector) often vary during a season. both X_i and λ_i should be generalized to be time indexed. The time-varying elements in X_i are called *time-varying covariates*. The probability of an individual taking trips at different times will hence depend on the time-dependent explanatory variables $X_i(t)$.

In the following discussion. a mere statement of interval *t* presupposes implicitly a starting point of time 0. The notation t_s^r will be used when necessary to indicate that the duration *t* runs from time *s* to time r, instead of from 0 to *t*. The endpoints are important now since the parameters λ_{it} are time dependent.

In the case of time-varying parameters, the CDF of $T_i (\geq 0)$ becomes

$$F_i(t) = 1 - \exp\left(-\int_0^t \lambda_i(s)\, ds\right)$$

and the PDF becomes

$$f_i(t) = \lambda_i(t) \exp\left(-\int_0^t \lambda_i(s)\, ds\right)$$

with $\lambda_i(t) > 0$ at any moment of time *t*. The distribution functions still necessarily have the properties that $F_i(0) = 0$ and $F_i(\infty) = 1$. The instantaneous hazard rate $h_i(t) = \lambda_i(t)$ depends solely on the value of parameter λ_i at time t.

The probability that an individual i has a completed duration T_i greater than t and less than $r\,(t \leq r)$ is then

$$\text{Prob}\{t \leq T_i \leq r\} = \exp\left(-\int_0^t \lambda_i(s)\,ds\right) - \exp\left(-\int_0^r \lambda_i(s)\,ds\right) \geq 0.$$

The corresponding Poisson counting process can be shown to have the distribution

$$\text{Prob}\{N_i(S) = n\} = \frac{\left[\int_0^S \lambda_i(u)\,du\right]^n \exp\left[-\int_0^S \lambda_i(u)\,du\right]}{n!} \equiv \frac{(\lambda_i^*)^n e^{-\lambda_i^*}}{n!}$$

where $\lambda_i^* \equiv \int_0^S \lambda_i(u)\,du$. The proof is only a generalization of that for the time-independent version. The expected number of trips taken by i during a season from day 0 to day S is then

$$E[N_i(S)] = \int_0^S \lambda_i(u)\,du = \lambda_i^*.$$

Given observations of the *completed* durations t_i (running from L_i to V_i) and assuming $\lambda_i(t) = e^{\beta X_i(t)}$ for time t, we can construct the log likelihood function LL as follows

$$\begin{aligned} LL &= \sum_i \log f_i(t_{i,L_i}^{V_i}) \\ &= \sum_i \left[\log \lambda_i(V_i) - \int_{L_i}^{V_i} \lambda_i(s)\,ds\right] \\ &= \sum_i \left[\beta X_i(V_i) - \int_{L_i}^{V_i} e^{\beta X_i(s)}\,ds\right]. \end{aligned}$$

The discrete time version of the log likelihood function LL is

$$LL = \sum_i \left[\beta X_i(V_i) - \sum_{s=L_i}^{V_i} e^{\beta X_i(s)}\right].$$

The maximum likelihood estimate $\hat{\beta}$ is then the solution to the equation

$$\frac{\partial LL}{\partial \beta} = \sum_i \left[X_i(V_i) - \sum_{s=L_i}^{V_i} X_i(s)\,e^{\beta X_i(s)}\right] \equiv 0.$$

The expected number of trips of individual i during a season from day 0 to day S is then simply

$$E[N_i(S)] = \sum_{s=0}^{S} \lambda_i(s) = \sum_{s=0}^{S} e^{\beta X_i(s)}$$

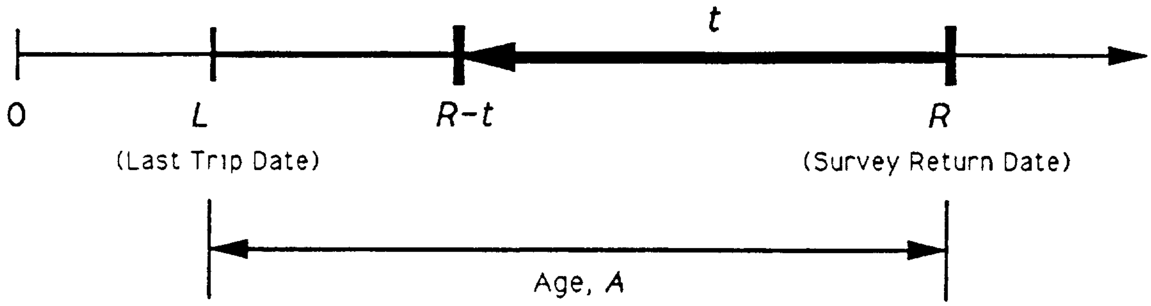


Figure IV.3: Derivation of the age distribution

One more issue wemust address is the estimation of parameters β when we only have age data instead of completed durations. This can be done by first deriving the distribution of age A_i as follows.

$$\begin{aligned}\text{Prob}\{A_i \geq t\} &= \text{Prob}\{N(R_i) - N(R_i - t) = 0\} \\ &= \text{Prob}\{N(t|_{R_i-t}^{R_i}) = 0\} \\ &= \exp\left(-\int_{R_i-t}^{R_i} \lambda_i(u)\, du\right)\end{aligned}$$

for $t \leq A_i$. Note that $\text{Prob}\{A_i \geq t\} = 0$ for $t > A_i$. This is shown in figure (IV.3). Basically, we are looking backwards from the given censoring point R_i to find the time the last trip occurred, not looking forwards in search of the next trip date.

Therefore, the CDF of age A_i is

$$F_A(t|_{R_i-t}^{R_i}) = \text{Prob}\{A_i \leq t|_{R_i-t}^{R_i}\} = 1 - \exp\left(-\int_{R_i-t}^{R_i} \lambda_i(u)\, du\right)$$

and the probability density of having an age $A_i = t$ from the date of last trip L_i (= $R_i - t$) to the survey return date R_i is

$$f_A(t \mid {}^{R_i}_{L_i}) = \frac{dF_A(t \mid {}^{R_i}_{L_i})}{dt} = \lambda_i(L_i) \exp\left(-\int_{L_i}^{R_i} \lambda_i(u)\, du\right)$$

The parameters β can hence be estimated using age data and the age distribution $f_A(t)$. They are exactly those that appear in the normal life distribution. The total trips can then be calculated as

$$E[N_i(S)] = \sum_{s=0}^{S} e^{\beta X_i(s)}$$

Competing Risk Participation

To further enrich our participation model, consider the more complicated situation where an individual can take either a day trip, a weekend trip. or a vacation trip. Trips of unequal lengths are considered to represent different. substitute commodities because their utility trade-offs may be different. In other words, long trips are taken for purposes somewhat different from those of short trips. Here we'll think of them as different *types* of events (or risks) that would terminate the durations and index them as d = 1, 2, 3. In the following discussion, individual index i is omitted for notational simplicity.

The single-type exponential specification can now be extended by defining the *type-specific hazard rate* as

$$h_d(t) \equiv \lim_{\delta \to 0+} \frac{\text{Prob}\{t \le T < t + \delta,\ D = d \mid t \le T\}}{\delta} = \lambda_d(t) = e^{\beta_d X_i(t)}.$$

This is the probability density of an individual i taking a type d trip immediately after time t conditional on no trip occurrence of any type up to time t. The *non-type-specific hazard rate* of individual i taking any type of trip at t is then simply

$$h(t) \equiv \lim_{\delta \to 0-} \frac{\text{Prob}\{t \le T < t + \delta \mid t \le T\}}{\delta} = \sum_d \lambda_d(t) = \sum_d e^{\beta_d X_i(t)}$$

since trips of different types cannot be taken simultaneously on a choice occasion, i.e., they are mutually exclusive. It must necessarily follow that $0 \leq h_d(t) \leq 1$ for all types d and $0 \leq h(t) \leq 1$.

The *non-type-specific CDF* of at least one trip of any type up to time t is

$$F(t) = 1 - \exp\left(-\int_0^t h(s)\, ds\right) = 1 - \exp\left(-\int_0^t \sum_d \lambda_d(s)\, ds\right)$$

and the *non-type-specific survival* function of no trip at all from time 0 to time t is

$$S(t) = 1 - F(t) = \exp\left(-\int_0^t h(s)\, ds\right) = \exp\left(-\int_0^t \sum_d \lambda_d(s)\, ds\right).$$

The *type-specific CDF* of having at least one *type d* trip up to time t is

$$F_d(t) = 1 - \exp\left(-\int_0^t h_d(s)\, ds\right) = 1 - \exp\left(-\int_0^t \lambda_d(s)\, ds\right)$$

and the *type-specific survival function* of no *type d* trip from time 0 to time t is

$$S_d(t) = 1 - F_d(t) = \exp\left(-\int_0^t h_d(s)\, ds\right) = \exp\left(-\int_0^t \lambda_d(s)\, ds\right).$$

do not untrue and

The *non-type-specific PDF*

$$f(t) = \frac{dF(t)}{dt} = \left[\sum_d \lambda_d(t)\right] \exp\left(-\int_0^t \sum_d \lambda_d(s)\, ds\right) = h(t)\, S(t)$$

measures the probability of having no trip up to time t and then a trip of any type at time t. The *type-specific PDF* below

$$f_d(t) = \lambda_d(t) \exp\left(-\int_0^t \sum_j \lambda_j(s)\, ds\right) = h_d(t)\, S(t)$$

gives us the probability of having no trip before t and then a *type d* trip at t. Note that by definition

$$f_d(t) \neq \frac{dF_d(t)}{dt}.$$

As in the previous sections, let R_i and L_i be the observed censoring date and last trip date respectively. Also let D_i be the type of the last trip taken by individual i in our sample data. The likelihood function, incorporating the time-varying covariate results, is

$$L = \prod_i f_{D_i}(t_i|_{L_i}^{R_i}) = \prod_i \left[\lambda_{D_i}(L_i) \exp\left(-\int_{L_i}^{R_i} \sum_d \lambda_d(s)\, ds\right)\right] \qquad \text{(IV.21)}$$

And the log likelihood function is

$$LL = \sum_i \left[\log \lambda_{D_i}(L_i) - \int_{L_i}^{R_i} \sum_d \lambda_d(s)\, ds\right]. \qquad \text{(IV.22)}$$

Alternatively, we can write the likelihood function L as

$$L = \prod_i \left[\lambda_{D_i}(L_i) \exp\left(-\int_{L_i}^{R_i} \lambda_{D_i}(s)\, ds\right) \prod_{d \neq D_i} \exp\left(-\int_{L_i}^{R_i} \lambda_d(s)\, ds\right)\right].$$

Durations corresponding to all types except the chosen type D_i are regarded as censored at individual *i's* survey return date R_i. The parameter β_d for $\lambda_{id}(t) = e^{\beta_d X_i(t)}$ can be estimated by maximizing the above likelihood function.

The expected number of type j trips taken by i during a season from day 0 to day S is readily calculated as

$$E[N_{id}(S)] = \sum_{s=0}^{S} e^{\beta_d X_i(s)}$$

The expected number of total trips taken by i is then $\sum_d E[N_{id}(S)]$.

Note that if individuals have homogeneous (i.e.. not time- varying) hazard λ_d. the log likelihood function in the discrete time context reduces to

$$LL = \sum_i \left[\log \lambda_{D_i} - \sum_{s=L_i}^{R_i} \sum_d \lambda_d\right].$$

Let $t_i(= R_i - L_i)$ denote the observed age of i and N_d be the number of individuals in the sample whose last trips are type *d.* The MLE of λ_d is then

$$\hat{\lambda}_d = \frac{N_d}{\sum_i t_i}.$$

Censored Age Durations

One further complication we address with this model is to accommodate the *censored age duration varible.*

The variable L_i denotes the date individual i took the last trip, as reported in the questionnaire returned on date R_i. For some individuals k, however, L_k is not available, and all we know is that no trip was ever taken from the beginning of sample period C (= April 1, 1983) up to the questionnaire return date R_k. This gives us the *left censored* age data. Recognizing that age duration is essentially right-censored trip duration! the data for these individuals can alternatively be itnerpreted as fight- and left-censored trip durations.

The fact that the age duration has 'survived' the period from C to R_k suggests that we augment the likelihood function (IV.21) with

$$L_0 = \prod_k S_k(t_{k\,C}^{\,R_k}) = \prod_k \exp\left(- \int_C^{R_k} h(s)\, ds \right)$$

to include the non-participants for whom we only have censored age. Therefore the complete log likelihood function is

$$LL = \sum_{i \in P_1} \left[\log \lambda_{D_i}(L_i) - \int_{L_i}^{R_i} \sum_d \lambda_d(s)\, ds \right] - \sum_{k \in P_0} \left[\int_C^{R_k} \sum_d \lambda_d(s)\, ds \right]$$

where P_1 is the sample of participating people, and P_0 is the set of non-participants.

Using Less Restrictive Distributions

Estimation can also be performed using other more flexible functional forms (e.g., Weibull, Log-Logistic. or Box-Cox hazards) for the distribution of between-trip durations if the Poisson-Exponential dual appears too restrictive. Note that the exponential distribution has only one parameter λ, and its mean is equal to its standard deviation $\mathrm{E}(T) = \sqrt{\mathrm{Var}(T)} = 1/\lambda$. Therefore the mean and variance cannot be adjusted separately. As pointed out by Kiefer (1988), the exponential is unlikely to be an adequate description of the data if the sample contains both very long and short durations.

Let $f(t)$ and $F(t)$ be the common PDF and CDF of the independently distributed between-trip intervals T for all individuals. The mean interval between successive trips is then

$$\mu = E[T] = \int_0^\infty f(t)\, t\, dt.$$

The estimation procedure for any $f(t)$ is basically the same as that described in the previous section. To my knowledge, however, no computer package can yet handle the full time-varying competing risk age duration model, though partial estimation can indeed be carried out by some existing commercial programs.[10] In the following sections: the more general Weibull distribution will be employed to illustrate the use of other distribution functions and to test the exponential duration assumption.

The Weibull Distribution

Now assume that the between-trip time intervals are all independent and Weibull-distributed with two parameters: a shape parameter $\gamma(>0)$ and a scale parameter $\lambda(>0)$. The distribution functions are[11]

$$\begin{aligned} PDF: \quad f(t) &= \lambda\gamma(\lambda t)^{\gamma-1}\exp(-(\lambda t)^\gamma) \\ CDF: \quad F(t) &= 1 - \exp(-(\lambda t)^\gamma) \\ Survival: \quad S(t) &= \exp(-(\lambda t)^\gamma) \\ Hazard: \quad h(t) &= \lambda\gamma(\lambda t)^{\gamma-1} \end{aligned}$$

The shape parameter determines the shape of the hazard function $h(t)$. When

[10] For instance, Limdep (1989, chapters 27 and 28) can only handle Cox's proportional hazards model without competing risks, or basic Weibull with neither time-varying covariates nor competing risks.

[11] For a brief discussion, on the Weibull distribution., see Lee (1980), pp. 162-67.

$\gamma > 1$, the hazard rate $h(t)$ increases with t: the case of *positive duration dependence.* When $\gamma < 1$, the hazard rate $h(t)$ declines with t, the case of *negative duration dependence*. When $\gamma = 1$, the model reduces to the exponential case and we have a constant hazard regardless of the value of t. Therefore, the appropriateness of employing the exponential distribution can be empirically tested by formulating a test of the hypothesis $H_0 : \gamma = 1$ [12]

The expected length of between-trip intervals is

$$\mu = \frac{\Gamma(1 + \frac{1}{\gamma})}{\lambda}$$

and the variance is

$$\sigma^2 = \frac{1}{\lambda^2}\left[\Gamma(1 + \frac{2}{\gamma}) - \Gamma^2(1 + \frac{1}{\gamma})\right]$$

where Γ is the gamma function defined as[13]

$$\Gamma(x) = \int_0^\infty u^{x-1} e^{-u}\, du.$$

Note that when $\gamma = 1$, we have $\mu = \Gamma(2)/\lambda = 1/\lambda$ and $\sigma^2 = 1/\lambda^2$, exactly the exponential case.

The three tasks we need to perform to modify the basic Weibull distribution for our estimation problem are

- the derivation of the age distribution,
- the inclusion of the time-varying covariates, and
- the development of the competing risk model.

---

[12] The Weibull hazard is monotonic. Other generalizations that embed Weibull as a special case are Log-Logistic and Box-Cox hazards, for example. Both hazards allow non-monotonic behavior. See Lancaster (1990), chapter 3.

[13] $\Gamma(x)$ is simply $(x-1)!$ when x is a non- negative integer.

The Time-Varying Weibull Age Distribution

To derive the time-varying version of the Weibull distribution, we assume that the scale of the hazard rate is time dependent, i.e., $\lambda(t) > 0$ for all *t*. However, the shape of the hazard function, determined by the value of γ, is preset and not time-varying. This maintains γ as a constant. For the estimation of the Weibull model, we further posit that

$$\begin{aligned} \gamma &= e^{\alpha X_1} > 0 \\ \lambda(t) &= e^{\beta X_2(t)} > 0 \end{aligned}$$

where the explanatory variables X_1 are constant through time while $X_2(t)$ vary with time. Conceptually X_1 consists of variables that determine the shape of the hazard function, and $X_2(t)$ contains the variables that affect the trip-taking probabilities at *t*. There may possibly be overlapping between X_1 and X_2 since X_2 can also have components that do not vary with time. If $\alpha X_1 > 0$ (or equivalently, $\gamma > 1$), an individual is said to have *positive* duration dependence. If $\alpha X_1 = 0$ (or $\gamma = 1$), there is *no* duration dependence. Otherwise, *negative* duration dependence exists.

By modifying the basic Weibull distribution functions, we can derive the time-varying Weibull probability system as follows:

$$\begin{aligned} \textit{Hazard}: \quad h(t) &= \lambda(t)\gamma\left[\int_0^t \lambda(s)\,ds\right]^{\gamma-1} \\ \textit{Survival}: \quad S(t) &= \exp\left(-\left[\int_0^t \lambda(s)\,ds\right]^{\gamma}\right) \\ \textit{CDF}: \quad F(t) &= 1-\exp\left(-\left[\int_0^t \lambda(s)\,ds\right]^{\gamma}\right) = 1 - S(t) \\ \textit{PDF}: \quad f(t) &= \lambda(t)\gamma\left[\int_0^t \lambda(s)\,ds\right]^{\gamma-1}\exp\left(-\left[\int_0^t \lambda(s)\,ds\right]^{\gamma}\right) \end{aligned}$$

It is straightforward to verify that $F(0) = 0$ and $F(\infty) = 1$ and

$$f(t) = \frac{dF(t)}{dt} = h(t)\,S(t),$$

and hence the above equations constitute a consistent distribution definition.

Following the notation of $(t\,|^{R}_{L=R-t})$ used for age duration in previous sections, the age distribution can be derived:

$$\begin{aligned} S_A(t\,|^{R}_{L}) &= \exp\left[-\left(\int_L^R \lambda(s)\,ds\right)^{\gamma}\right] \\ F_A(t\,|^{R}_{L}) &= 1-\exp\left[-\left(\int_L^R \lambda(s)\,ds\right)^{\gamma}\right] = 1 - S_A(t\,|^{R}_{L}) \\ f_A(t\,|^{R}_{L}) &= \frac{dF_A(t\,|^{R}_{L})}{dt} = \lambda(L)\gamma\left[\int_L^R \lambda(s)\,ds\right]^{\gamma-1} S_A(t\,|^{R}_{L}) \end{aligned}$$

The Competing Risk Weibull Model

For the three types of trips (day, weekend and vacation, indexed by j = 1, 2, 3, respectively) that an individual may take, we assume that

$$\lambda_j(t) = e^{\beta_j X_2(t)} > 0$$

We further assume that the shape parameter γ is constant and identical for all types of trips.[14]

The type-specific hazard rate, under the assumption of inter-type dependence is

$$h_j(t) = \lambda_j(t)\gamma\left[\int_0^t \sum_l \lambda_l(s)\,ds\right]^{\gamma-1} \tag{IV.23}$$

The non-type-specific hazard rate, the sum of the type- specific hazards by definition, is then simply

$$h(t) = \sum_j h_j(t) = \left[\sum_j \lambda_j(t)\right]\gamma\left[\int_0^t \sum_j \lambda_j(s)\,ds\right]^{\gamma-1} \tag{IV.24}$$

These hazard functions imply that

[14] There are two reasons for the different treatments. Firstly, we see no reason why different types of trips should have different duration dependencies. Secondly, and more importantly, we need to keep the model under a controllable degree of complexity.

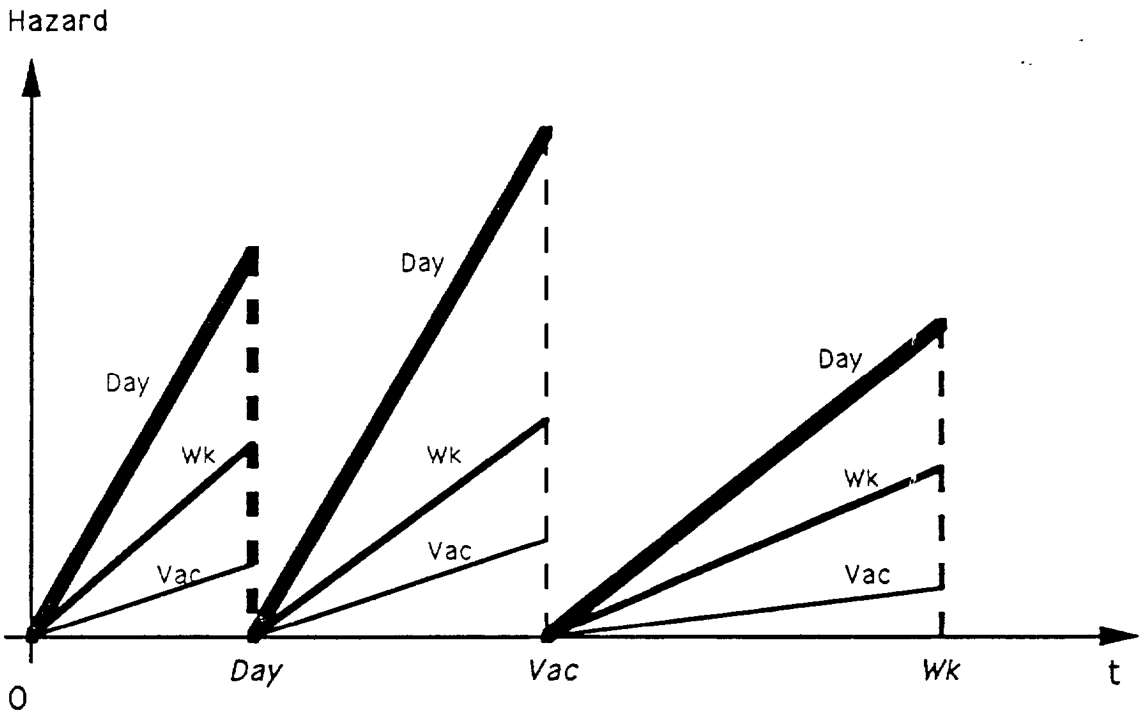


Figure IV.4: Hazard rate with inter-type dependence

- The probability density of taking a type J trip at t, conditional on no trip up to t, depends not only on the history of parameter $\lambda_J(t)$ before time t, but also on the history of parameters of all other types.

- The hazard ratio of having different types of trips at time t is not affected by the parameter values in the past. This can be seen by noting that proportinoaltiy holds as follows:

$$h_1(t) : h_2(t) : h_3(t) = \lambda_1(t) : \lambda_2(t) : \lambda_3(t).$$

Figure (IV.4) shows the hazard rate behaviors of different trip types for the case of positive duration dependence. Note that the hazards of all types become zero with each trip occurrence (i.e., $t = 0$) whatever its type.[15]

[15] In the case of negative duration dependence, all the hazards become infinity the moment after a trip is taken.

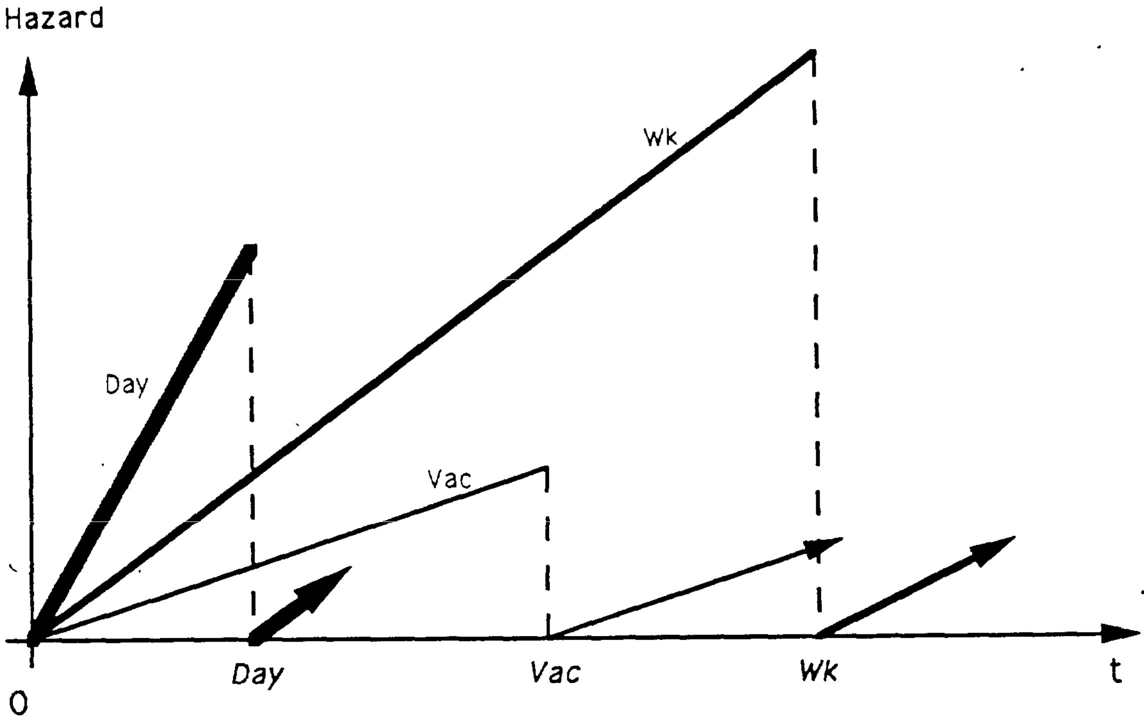


Figure IV.5: Hazard rate without inter-type dependence

The hazard rate based on inter-type dependence discussed above can be contrasted with a hazard rate without inter- type dependence illustrated in figure (IV.5)

$$h_j(t_j) = \lambda_j(t_j)\gamma \left[\int_0^{t_j} \lambda_j(s)\,ds\right]^{\gamma-1}$$

and

$$h(t) = \sum_j h_j(t_j) = \sum_j \left(\lambda_j(t_j)\gamma \left[\int_0^{t_j} \lambda_j(s)\,ds\right]^{\gamma-1}\right).$$

Note that the between-trip duration t is indexed by the trip type j since it is now type-specific. When there is no inter-type dependence among different hazards, the hazard rate of one trip type is not affected by the occurrences of trips of other types. Therefore, the hazard rate of one trip type continues to increase until a trip of its own type is taken, at that time it drops to zero while hazards of other trip types keep increasing.

In our analysis we assume that there is inter-type dependence and employ the hazard rates defined in (IV.23). It is not difficult to verify that the distribution functions corresponding to the hazards (IV.23) and (IV.24) are

$$
\begin{aligned}
S(t) &= \exp\left(-\left[\int_0^t \sum_j \lambda_j(s)\,ds\right]^{\gamma}\right) \\
F(t) &= 1-\exp\left(-\left[\int_0^t \sum_j \lambda_j(s)\,ds\right]^{\gamma}\right) = 1 - S(t) \\
f(t) &= \frac{dF(t)}{dt} = h(t)\,S(t) \\
f_j(t) &= h_j(t)\,S(t)
\end{aligned}
$$

Estimating the Weibull Model

Let $f^A_{J_i}(t_i|^{R_i}_{L_i})$ denote the probability density of individual i taking the most recent trip of type J_i and having an observed age from L_i to R_i. The likelihood function for the sample P_1 for whom we have the last trip data is then

$$
\begin{aligned}
L &= \prod_{i\in P_1} f^A_{J_i}(t_i|^{R_i}_{L_i}) \\
&= \prod_{i\in P_1}\left\{\lambda_{J_i}(L_i)\,\gamma\left[\int_{L_i}^{R_i}\sum_j \lambda_j(s)\,ds\right]^{\gamma-1} \exp\left(-\left[\int_{L_i}^{R_i}\sum_j \lambda_j(s)\,ds\right]^{\gamma}\right)\right\}
\end{aligned}
$$

For the non-participant sample P_0, we know only that no trip was taken from C, the beginning of sample period. up to the questionnaire return date R_k. The likelihood for this sample is

$$
\begin{aligned}
L &= \prod_{k\in P_0} S(t_k|^{R_k}_{C}) \\
&= \prod_{k\in P_0} \exp\left(-\left[\int_C^{R_k}\sum_k \lambda_k(s)\,ds\right]^{\gamma}\right)
\end{aligned}
$$

Combining the participants and the non-participants, the complete log likelihood function is

$$
LL = \sum_{i\in P_1}\left\{\log\lambda_{J_i}(L_i) - \log\gamma + (\gamma-1)\,\log\left[\int_{L_i}^{R_i}\sum_j \lambda_j(s)\,ds\right]\right.
$$

$$
\begin{aligned}
& - \left[\int_{L_i}^{R_i} \sum_j \lambda_j(s)\, ds \right]^{\gamma} \Bigg\} \\
& - \sum_{k \in P_0} \left[\int_C^{R_k} \sum_k \lambda_k(s)\, ds \right]^{\gamma} \\
= & \sum_{i \in P_1} \Bigg\{ \beta_{J_i} X_2(L_i) + \alpha X_1 + (e^{\alpha X_1} - 1) \log \left[\int_{L_i}^{R_i} \sum_j e^{\beta_j X_2(s)}\, ds \right] \\
& - \left[\int_{L_i}^{R_i} \sum_j e^{\beta_j X_2(s)}\, ds \right]^{e^{\alpha X_1}} \Bigg\} \\
& - \sum_{k \in P_0} \left[\int_C^{R_k} \sum_k e^{\beta_j X_2(s)}\, ds \right]^{e^{\alpha X_1}} \\
= & \sum_{i \in P_1} \left\{ \beta_{J_i} X_2(L_i) + \alpha X_1 + (e^{\alpha X_1} - 1) \log S(\beta) - S(\beta)^{e^{\alpha X_1}} \right\} \\
& - \sum_{k \in P_0} N(\beta)^{e^{\alpha X_1}}
\end{aligned}
$$

where

$$
\begin{aligned}
S(\beta) & \equiv \int_{L_i}^{R_i} \sum_j e^{\beta_j X_2(s)}\, ds \quad \text{for a participant } i \text{ in } P_1) \\
N(\beta) & \equiv \int_C^{R_k} \sum_j e^{\beta_j X_2(s)}\, ds \quad \text{(for a non-participant } k \text{ in } P_0).
\end{aligned}
$$

The first derivatives with respect to the parameters α and β are consequently

$$
\begin{aligned}
\frac{\partial LL}{\partial \alpha} = & \sum_{i \in P_1} \left[X_1 + (X_1 e^{\alpha X_1}) \log S(\beta) - S(\beta)^{e^{\alpha X_1}} \log S(\beta)\, e^{\alpha X_1} X_1 \right] \\
& - \sum_{k \in P_0} \left\{ N(\beta)^{e^{\alpha X_1}} \log N(\beta)\, e^{\alpha X_1} X_1 \right\} \\
= & \sum_{i \in P_1} \left\{ X_1 \left[1 + e^{\alpha X_1} \log S(\beta) (1 - S(\beta)^{e^{\alpha X_1}}) \right] \right\} \\
& - \sum_{k \in P_0} \left\{ N(\beta)^{e^{\alpha X_1}} \log N(\beta)\, e^{\alpha X_1} X_1 \right\}
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\partial LL}{\partial \beta_j} = & \sum_{i \in P_1} \left[Q(j, J_i) + (e^{\alpha X_1} - 1) \frac{S_j(\beta_j)}{S(\beta)} - e^{\alpha X_1} S(\beta)^{e^{\alpha X_1} - 1} S_j(\beta_j) \right] \\
& - \sum_{k \in P_0} \left\{ e^{\alpha X_1} N(\beta)^{e^{\alpha X_1} - 1} N_j(\beta_j) \right\} \\
= & \sum_{i \in P_1} \left\{ Q(j, J_i) + \frac{S_j(\beta_j)}{S(\beta)} \left[(e^{\alpha X_1} - 1) - e^{\alpha X_1} S(\beta)^{e^{\alpha X_1}} \right] \right\}
\end{aligned}
$$

$$- \sum_{k \in P_0} \left\{ e^{\alpha X_1} N(\beta)^{e^{\alpha X_1} - 1} N_j(\beta_j) \right\}$$

where

$$\begin{aligned} S_j(\beta_j) &\equiv \frac{\partial S(\beta)}{\partial \beta_j} = \int_{L_i}^{R_i} X_2(s)\, e^{\beta_j X_2(s)}\, ds \quad \text{for a participant } i \text{ in } P_1) \\ N_j(\beta_j) &\equiv \frac{\partial N(\beta)}{\partial \beta_j} = \int_{C}^{R_k} X_2(s)\, e^{\beta_j X_2(s)}\, ds \quad \text{(for a non-participant } k \text{ in } P_0) \end{aligned}$$

and

$$Q(j, J_i) = \begin{cases} X_2(L_i) & \text{if } j = J_i \\ 0 & \text{otherwise.} \end{cases}$$

And the second derivatives (the Hessian matrix) are

$$\begin{aligned} \frac{\partial^2 LL}{\partial \alpha^2} &= \sum_{i \in P_1} \Big\{ X_1^2 e^{\alpha X_1} \log S(\beta) \\ &\quad - S(\beta)^{e^{\alpha X_1}} \left[\log S(\beta)\, e^{\alpha X_1} X_1\right]^2 \\ &\quad - S(\beta)^{e^{\alpha X_1}} \log S(\beta)\, e^{\alpha X_1} X_1^2 \Big\} \\ &\quad - \sum_{k \in P_0} \Big\{ N(\beta)^{e^{\alpha X_1}} \left[\log N(\beta)\, e^{\alpha X_1} X_1\right]^2 \\ &\quad + N(\beta)^{e^{\alpha X_1}} \log N(\beta)\, e^{\alpha X_1} X_1^2 \Big\} \\ &= \sum_{i \in P_1} \left\{ X_1^2 e^{\alpha X_1} \log S(\beta) \left[1 - S(\beta)^{e^{\alpha X_1}} \left(\log S(\beta)\, e^{\alpha X_1} + 1\right)\right] \right\} \\ &\quad - \sum_{k \in P_0} \left\{ X_1^2 e^{\alpha X_1} \log N(\beta)\, N(\beta)^{e^{\alpha X_1}} \left[\log N(\beta)\, e^{\alpha X_1} - 1\right] \right\} \\ \frac{\partial^2 LL}{\partial \alpha\, \partial \beta_j} &= \sum_{i \in P_1} \Bigg[(X_1 e^{\alpha X_1}) \frac{S_j(\beta_j)}{S(\beta)} \\ &\quad - e^{\alpha X_1} S(\beta)^{e^{\alpha X_1} - 1} S_j(\beta_j) \log S(\beta)\, e^{\alpha X_1} X_1 \\ &\quad - S(\beta)^{e^{\alpha X_1}} \frac{S_j(\beta_j)}{S(\beta)} e^{\alpha X_1} X_1 \Bigg] \\ &\quad - \sum_{k \in P_0} \Bigg[e^{\alpha X_1} N(\beta)^{e^{\alpha X_1} - 1} N_j(\beta_j) \log N(\beta)\, e^{\alpha X_1} X_1 \\ &\quad - N(\beta)^{e^{\alpha X_1}} \frac{N_j(\beta_j)}{N(\beta)} e^{\alpha X_1} X_1 \Bigg] \\ &= \sum_{i \in P_1} \left\{ X_1 e^{\alpha X_1} \frac{S_j(\beta_j)}{S(\beta)} \left[1 - S(\beta)^{e^{\alpha X_1}} \left(\log S(\beta)\, e^{\alpha X_1} + 1\right)\right] \right\} \end{aligned}$$

$$- \sum_{k\in P_0} \left\{ X_1 \, e^{\alpha X_1} \, \frac{N_j(\beta_j)}{N(\beta)} \, N(\beta)^{e^{\alpha X_1}} \left(\log N(\beta) \, e^{\alpha X_1} + 1 \right) \right\}$$

$$\begin{aligned}
\frac{\partial^2 LL}{\partial \beta_j^2} &= \sum_{i\in P_1} \Bigg\{ \frac{S_{jj}(\beta_j)\, S(\beta) - S_j(\beta_j)^2}{S(\beta)^2} \left[\left(e^{\alpha X_1} - 1\right) - e^{\alpha X_1} \, S(\beta)^{e^{\alpha X_1}} \right] \\
&\quad - \left(e^{\alpha X_1}\right)^2 \frac{S_j(\beta_j)}{S(\beta)} \, S(\beta)^{e^{\alpha X_1} - 1} \, S_j(\beta_j) \Bigg\} \\
&\quad - \sum_{k\in P_0} \left\{ e^{\alpha X_1} \, N(\beta)^{e^{\alpha X_1}} \left[\left(e^{\alpha X_1} - 1\right) \frac{N_j(\beta_j)^2}{N(\beta)^2} + \frac{N_{jj}(\beta_j)}{N(\beta)} \right] \right\} \\
&= \sum_{i\in P_1} \Bigg\{ \left(e^{\alpha X_1} - 1\right) \frac{S_{jj}(\beta_j)\, S(\beta) - S_j(\beta_j)^2}{S(\beta)^2} \\
&\quad - e^{\alpha X_1} \, S(\beta)^{e^{\alpha X_1}} \left[\left(e^{\alpha X_1} - 1\right) \frac{S_j(\beta_j)^2}{S(\beta)^2} - \frac{S_{jj}(\beta_j)}{S(\beta)} \right] \Bigg\} \\
&\quad - \sum_{k\in P_0} \left\{ e^{\alpha X_1} \, N(\beta)^{e^{\alpha X_1}} \left[\left(e^{\alpha X_1} - 1\right) \frac{N_j(\beta_j)^2}{N(\beta)^2} + \frac{N_{jj}(\beta_j)}{N(\beta)} \right] \right\}
\end{aligned}$$

and, for $j \neq m$,

$$\begin{aligned}
\frac{\partial^2 LL}{\partial \beta_j \partial \beta_m} &= \sum_{i\in P_1} \Bigg\{ (e^{\alpha X_1} - 1) \frac{-S_j(\beta_j) S_m(\beta_m)}{S(\beta)^2} \\
&\quad - e^{\alpha X_1} (e^{\alpha X_1} - 1) \, S(\beta)^{e^{\alpha X_1} - 2} \, S_j(\beta_j) \, S_m(\beta_m) \Bigg\} \\
&\quad - \sum_{k\in P_0} \left[e^{\alpha X_1} \, (e^{\alpha X_1} - 1) \, N(\beta)^{e^{\alpha X_1} - 2} \, N_j(\beta_j) \, N_m(\beta_m) \right] \\
&= \sum_{i\in P_1} \left\{ -(e^{\alpha X_1} - 1) \frac{S_j(\beta_j) S_m(\beta_m)}{S(\beta)^2} \left[1 + e^{\alpha X_1} \, S(\beta)^{e^{\alpha X_1}} \right] \right\} \\
&\quad - \sum_{k\in P_0} \left[e^{\alpha X_1} \, (e^{\alpha X_1} - 1) \, N(\beta)^{e^{\alpha X_1} - 2} \, N_j(\beta_j) \, N_m(\beta_m) \right]
\end{aligned}$$

where

$$\begin{aligned}
S_{jj}(\beta_j) &\equiv \frac{\partial^2 S(\beta)}{\partial \beta_j^2} = \int_{L_i}^{R_i} X_2(s)^2 e^{\beta_j X_2(s)} \, ds \quad \text{(for a participant } i \text{ in } P_1) \\
N_{jj}(\beta_j) &\equiv \frac{\partial^2 N(\beta)}{\partial \beta_j^2} = \int_{C}^{R_k} X_2(s)^2 e^{\beta_j X_2(s)} \, ds \quad \text{(for a non-participant } k \text{ in } P_0).
\end{aligned}$$

The maximum likelihood estimates $\hat{\alpha}$ and $\hat{\beta}$ can be obtained by the Newton-Raphson, algorithm. They are consistent: asymptotically efficient and asymptotically normal. The variance- covariance matrix $(-E[\nabla^2 LL]^{-1})$ can be approximated by

$$-[\nabla^2 LL]^{-1}_{\alpha=\hat{\alpha},\, \beta=\hat{\beta}}$$

If the null hypothesis $H_0 : \alpha = 0$ is rejected by the likelihood ratio test, the use of the exponential distribution cannot be justified, and we have to calculate the expected total seasonal trips for the Weibull model.[16]

Renewal Counting Process for Weibull

If the use of the exponential is rejected in favor of the Weibull, the distribution of the corresponding renewal counting process $N(s), s \geq 0$, has to be determined to calculate the expected number of trips in season.

Define the random variable W_n as the sum of n consecutive between-trip durations. i.e.,

$$W_n = T_1 + T_2 + \cdots + T_n, \; n \geq 1.$$

The distribution of W_n is then the *convolution* of the CDFs of the n durations $T_1, T_2, \ldots, T_n$.

Definition IV.1 If two independent random variables X and Y have the distribution functions F_X and F_Y, respectively. then the distribution of their sum $Z = X + Y$ is the convolution of F_X and F_Y, defined as

$$F_Z(z) = \int_{-\infty}^{\infty} F_X(z-u)\, dF_Y(u) = \int_{-\infty}^{\infty} F_Y(z-u)\, dF_X(u). \quad \blacksquare$$

Since all between-trip durations T_k have the. common CDF $F(s) = \text{Prob}\{T \leq s\}$, the CDF of W_n is the n-fold convolution of $F(s)$ with itself, denoted by F_n. Hence.

consecutive

$$F_n(s) = \text{Prob}\{W_n \leq s\} = \text{Prob}\{N(s) \geq n\}, \; n \geq 1$$

Given $F_1(s) \equiv F(s)$, any $F_n(s)$ for $n \geq 2$ can be calculated using the recursive formula

$$F_n(s) \;\; = \;\; \int_0^{\infty} F_{n-1}(s-u)\, dF(u)$$

[16] An interesting intermediate case is where the intercept is nonzero, but the slope coefficients are zero.

$$
\begin{aligned}
&= \int_0^\infty F_{n-1}(s-u)\, f(u)\, du \\
&= \int_0^s F_{n-1}(s-u)\, f(u)\, du
\end{aligned}
$$

The probability of having exactly n trips in a time period S is

$$
\begin{aligned}
\text{Prob}\{N(S)=n\} &= \text{Prob}\{W_n \le S\} - \text{Prob}\{W_{n+1} \le S\} \\
&= F_n(S) - F_{n+1}(S).
\end{aligned}
$$

The expected number of trips in the time interval S is then

$$
\begin{aligned}
m(S) &= E[N(S)] \\
&= \sum_{n=1}^{\infty} [n \cdot \text{Prob}\{N(S)=n\}] \\
&= \sum_{n=1}^{\infty} \text{Prob}\{N(S) \ge n\} \\
&= \sum_{n=1}^{\infty} F_n(S).
\end{aligned}
$$

While a closed-form formula for F_n is difficult to obtain. the values of $F_n(S)$ can be approximated by numerical solutions and the calculation of $m(S)$ can be carried out as follows for the discrete time case. First the values of $F_1(s) = F(s)$ for $s = 1,2, \ldots, S$ are calculated and stored. The values of $F_2(s)$ for $s = 1,2, \ldots, S$ can then be computed using the values of F_1 now available.

$$
F_2(s) = \sum_{u=0}^{s} [F_1(s-u)\, f(u)]
$$

This goes on until, for some N, $F_N(S)$ is small enough to be safely ignored. Note that $f(0) = 0$ and $F_n(0) = 0$. $\forall n$. The expected number of seasonal trips is thus approximately

$$
m(S) = \sum_{n=1}^{N} F_n(S) = \sum_{n=1}^{N} \sum_{u=0}^{S} [F_{n-1}(S-u)\, f(u)].
$$

The function $m(S)$, called the *renewal function,* will always converge since it is finite for a finite S, as proved in Ross (1983, p. 57). Note that the simple formula

$m(S) = S/\mu$ does not generally hold for other distributions besides the Poisson-Exponential case, where it occurs due to the special memoryless property of the exponential distribution. Though it is true that both respectively.

$$\frac{N(S)}{S} \rightarrow \frac{1}{\mu} \text{ as } S \rightarrow \infty$$

and

$$\frac{m(S)}{S} \rightarrow \frac{1}{\mu} \text{ as } S \rightarrow \infty,$$

it is a mistake to use (S/μ) as the expected value of $N(S)$ when S is not large enough. In our study, S is the length of a fishing season, the time period during which we count the trips. and is (sibstantially) less than infinity. Hence the extra work of calculating $m(S)$ has to be done if a less restrictive distribution like the Weibull is preferred.

CHAPTER V

CONSUMER WELFARE MEASURE

The calculation of consumer surplus is different with discrete choice travel cost models than with conventional travel cost analysis. With discrete choice models, we estimate the conditional utility functions and then compute directly the Hicksian compensating variation (CV) or equivalent variation (EV). In the conventional travel cost analysis, we generally calculate Marshallian measures of consumer surplus from the estimated demand functions. [1]

The standard consumer surplus measure employed for discrete choice models is based on the assumption that total trips do not change with policy changes. This measure will result in an under- or over-estimate of "true" consumer surplus, depending upon whether total trips increase or decrease. We develop a consumer surplus measure that incorporates the change in trips predicted by the participation model. Additional complexity is added to the measure with a NMNL model when the choice occasion income (budget) is not observed and the marginal utility of income is not

---

[1] Feenberg and Mills (1980, pp. 114-115) calculate the welfare measure C^*, defined as

$$V(p + C^*, q, y, s) = V(p, q', y, s)$$

where V is the indirect utility function. That is, C^* is the amount by which the price would have to be raised in order to offset the effect of the change in the quality. Hanemann (1983, pp. 134-35) argues that CV is more appropriate than C^* since it is a natural generalization to the discrete choice context of the conventional Hicksian compensating variation.

constrained to be constant across alternatives due to the computational complexity of such a procedure. We propose a simplifying procedure that makes the calculation tractable under those circumstances.

Welfare Measure for Individual Choice Occasions

Procedures to calculate the choice occasion welfare changes in the NMNL context have been developed by many researchers.[2] As defined previously by (III.4), let

$$\tilde{V}(\mathbf{P}, \mathbf{Q}, y) = \max_{m}\{\tilde{u}_m\}$$

be the maximum random utility an individual can receive on a choice occasion when facing trip cost P, site quality Q, and choice occasion budget *y*. The expected compensating variation *C* and equivalent variation $\mathcal{E}$ corresponding to a site quality change from $\mathbf{Q}^0$ to $\mathbf{Q}^1$ in the random utility model are defined as

$$E[\tilde{V}(\mathbf{P}, \mathbf{Q}^1, y^1 - \mathcal{C})] = E[\tilde{V}(\mathbf{P}, \mathbf{Q}^0, y^0)] \quad \text{(V.25)}$$

$$E[\tilde{V}(\mathbf{P}, \mathbf{Q}^1, y^1)] = E[\tilde{V}(\mathbf{P}, \mathbf{Q}^0, y^0 + \mathcal{E})]. \quad \text{(V.26)}$$

C is the expected maximum amount of money individuals require to compensate them for the change in site conditions, and $\mathcal{E}$ is the expected minimum amount of money people require to compensate them for foregoing the quality change. Note that both will be positive for quality improvement $(\mathbf{Q}^1 \succ \mathbf{Q}^0)$, and negative for quality deterioration $(\mathbf{Q}^1 \prec \mathbf{Q}^0)$. $\mathcal{C} \neq \mathcal{E}$ for utility functions that yield different values of the marginal utility of income at y^0 and y^1.

It has been shown that for the MNL model

$$E[\tilde{V}(P, Q, y)] = E[\max_{m}\{\tilde{u}_m\}]$$

---

[2] See Small and Rosen (1981) and Hanemann (1982, 1985). Hanemann (1983) also has the calculation for marginal exogenous variable changes. For applications see Carson and Hanemann (1987) and Jones 1988, 1990).

$$= \ln \sum_m e^{u_m(P,Q,y)} + \text{constant}$$
$$= I(P, Q, y) + \text{constant},$$

where $I(P, Q, y)$ is the inclusive value of choices with parameters P, Q and individual choice occasion income y. Therefore, we can rewrite equations (V.25) and (V.26), defining the compensating variation, $\mathcal{C}$, and equivalent variation, $\mathcal{E}$:

$$\begin{aligned} \ln \sum_m e^{u_m(P,Q^1,y^1-\mathcal{C})} &= \ln \sum_m e^{u_m(P,Q^0,y^0)} \\ \ln \sum_m e^{u_m(P,Q^1,y^1)} &= \ln \sum_m e^{u_m(P,Q^0,y^0+\mathcal{E})}. \end{aligned}$$

In general. closed-form solutions for the consumer surplus measures $\mathcal{C}$ and $\mathcal{E}$ are not available, and numerical techniques have to be employed.

However, with a linear-in-income conditional utility u_m that has a constant marginal utility of income η (the coefficient on y): the above equations become

$$\begin{aligned} \ln \sum_m e^{u_m(P,Q^1,y^0)+\eta(y^1-y^0-\mathcal{C})} &= \ln \sum_m e^{u_m(P,Q^0,y^0)} \\ \ln \sum_m e^{u_m(P,Q^1,y^0)+\eta(y^1-y^0)} &= \ln \sum_m e^{u_m(P,Q^0,y^0)-\eta\mathcal{E}}. \end{aligned}$$

which simplify: respectively, to:

$$\begin{aligned} \ln \sum_m e^{u_m(P,Q^1,y^0)} + \eta\,(y^1-y^0) - \eta\,\mathcal{C} &= \ln \sum_m e^{u_m(P,Q^0,y^0)} \\ \ln \sum_m e^{u_m(P,Q^1,y^0)} - \eta\,(y^1-y^0) &= \ln \sum_m e^{u_m(P,Q^0,y^0)} - \eta\,\mathcal{E}. \end{aligned}$$

Therefore,

$$\mathcal{C} = \mathcal{E} = \frac{I(P,Q^1,y^0) - I(P,Q^0,y^0)}{\eta} + (y^1 - y^0). \tag{V.27}$$

This formula presents the consumer surplus per choice occasion. The equality of $\mathcal{C}$ and $\mathcal{E}$ is the result of the linear-in-income indirect utility assumption. When the choice occasion income y is assumed fixed, i.e., $y^1 = y^0$, the second term on the right hand side of equation (V.27) drops out.

The use of this formula is not limited to site quality changes only. As mentioned in Bockstael et al. (1991), the value of adding or deleting sites can also be computed as

$$\mathcal{C} = \frac{I^1(P^1, Q^1, y^0) - I^0(P^0, Q^0, y^0)}{\eta}$$

where $I^0(P^0, Q^0, y^0)$ is the inclusive value before the change, and $I^1(P^1, Q^1, y^0)$ after. However, the site change has to be small so that the choice occasion income y will stay the same.

The above derivation assumes that the marginal utility of income is constant across alternatives (as well as across quality changes.) When that assumption does not apply. the derivation is more complicated. As Hanemann (1982) has shown:

$$\mathcal{C} = \mathcal{E} \approx \frac{\sum_m e^{u_m(P,Q^0,y)} - \sum_m e^{u_m(P,Q^1,y)}}{\sum_m \eta_m e^{u_m(P,Q^1,y)}}$$

Using the approximation that $z \approx \ln(1+z)$, this formula can be re-written to show more clearly its similarity to the constant MCI version:

$$\mathcal{C} = \mathcal{E} \approx \frac{\ln(\sum_m e^{u_m(P,Q^0,y)}) - \ln(\sum_m e^{u_m(P,Q^1,y)})}{\sum_m \eta_m \pi^1_m} = \frac{I^1 - I^0}{\tilde{\eta}}$$

where

$$\pi^1_m = \frac{e^{u_m(P,Q^1,y)}}{\sum_l e^{u_l(P,Q^1,y)}}$$

is the probability of choosing product line m after the quality change, and we define $\tilde{\eta}$ to be the weighted MUI

$$\tilde{\eta} = \sum_m \pi_m \eta_m.$$

Remember that in our framework, we model the trip-duration choice within the macro-level participation model, external to the NMNL analysis. Consequently, we calculate separate compensating or equivalent variations for each of the three trip-duration groups. For simplicity, we have suppressed the subscript d for the trip-duration groups in the formula above - but we will incorporate it explicitly in the calculations below.

Welfare Measure for Multiple Choice Occasions

Since there are multiple choice occasions in a season, most researchers derive the total consumer surplus by first calculating the choice occasion compensating variation C, then multiplying C by the total number of trips N over a season. This yields the seasonal consumer surplus

$$\mathcal{W} = \mathcal{C} \cdot N.$$

Whether N is taken as the number of trips N^0 before a site quality improvement or the predicted number of trips N^1 after an improvement. the calculation is not accurate. In tile former case, the welfare gain associated with the new trips is not included, whereas in the latter case, the formula gives an over-estimate of the true loss.[3] For a quality improvement, we do know that the annual utility gain W is bounded by

$$\mathcal{C} \cdot N^0 \leq \mathcal{W} \leq \mathcal{C} \cdot N^1.$$

The surplus $C \cdot N^0$ is the lower bound on the actual benefits because the increase in total value associated with the increase in trips is not included. The surplus measure $C \,.\, N^1$ is the upper bound because the increase in value is calculated based on the (greater) number of trips that would only be taken under improved site conditions.[4] When there are only marginal changes in site quality and hence the change in total trips N is small, these bounds are tight.

In this section, we propose a procedure to compute the seasonal consumer surplus more precisely for a proposed improvement in site conditions. First, for each trip type d (= day, weekend, or vacation) and each month n (= April - October) during a season, we denote the "true" pre- and post-policy inclusive values by I^0_{nd} and I^1_{nd} (corresponding to the site qualities Q^0_n and Q^1_n), respectively. As discussed above,

[3] See Parsons (1990, p. 14) or Bockstael et al. (1988, p. 18) for a discussion.

[4] See Parsons and Kealy (1990)

we cannot calculate I^0_{nd} or I^1_{nd} because we do not know choice occasion income The pseudo-inclusive values that we can calculate from the MNL parameter estimates are denoted by $\bar{I}^0_{nd}$ and $\bar{I}^1_{nd}$, respectively; as defined in (III.15). The expected number of trips N^0_{nd} and N^1_{nd} can then be estimated with the competing risks duration model proposed in chapter IV. Let y_d be the choice occasion income for a type-d trip. The expected seasonal compensating variation for an individual in the sample will consist of two components: one associated with the trips already taken before the policy, and the other associated with new trips that would only be taken after the policy.

Eased on the derivations above, the expected welfare gain for the N^0_{nd} trips of duration d in month n that occurred before the improvement is

$$\mathcal{W}^0_d = \frac{I^1_{nd} - I^0_{nd}}{\tilde{\eta}_d} \cdot N^0_{nd}$$

If we replace the MUI estimates that vary across product lines η_{md} with the weighted MUI for trip duration d, $\tilde{\eta}_d$, in the formulas for the inclusive value indices I^1_{nd} and I^0_{nd}, then this simplifies to:

$$\mathcal{W}^0_d = \frac{\bar{I}^1_{nd} - \bar{I}^0_{nd}}{\tilde{\eta}_d} \cdot N^0_{nd}$$

For the $(N^1_{nd} - N^0_{nd})$ new trips that would only occur after the site improvement, we assume that the expected no-trip utility u_0 is simply $[\alpha + \tilde{\eta}_d y_d]$ for the linear conditional utility function (III.14) since no travel cost is incurred and no site attributes are enjoyed. The associated compensating variation is thus

$$\begin{aligned} \mathcal{W}^1 &= \frac{I^1_{nd} - E[u_0]}{\tilde{\eta}_d} \cdot (N^1_{nd} - N^0_{nd}) \\ &= \frac{I^1_{nd} - [\alpha + \tilde{\eta}_d y_d]}{\tilde{\eta}_d} \cdot (N^1_{nd} - N^0_{nd}) \end{aligned}$$

Again if we substitute $\tilde{\eta}_d$ for η_{md} in I^1_{nd}, the choice occasion income cancels out in the before-policy and after-policy terms and this simplifies to:

$$\mathcal{W}^1 = \frac{\bar{I}^1_{nd}}{\tilde{\eta}_d} \cdot (N^1_{nd} - N^0_{nd}).$$

Therefore, the total seasonal CV for an individual i is the sum

$$\mathcal{W}^{*} = \mathcal{W}^{0} + \mathcal{W}^{1} = \sum_{n} \sum_{d} \left[\frac{\bar{I}_{nd}^{1} \cdot N_{nd}^{1} - \bar{I}_{nd}^{0} \cdot N_{nd}^{0}}{\bar{\eta}_{d}} \right] \tag{V.28}$$

CHAPTER VI

DATA SOURCES AND DESCRIPTIVE STATISTICS

This chapter describes the data used in this study to estimate the models discussed in previous chapters. Three categories of information have been collected from federal and state sources: angler data: species- and month-specific catch rate data, and other site quality data. Since the units of the site analysis are the 83 Michigan counties: all site quality data and distance measures are defined on a county basis.[1] We describe each category of data in turn.

Angler Survey Data

The primary dataset for estimating the model is a detailed mail surrey of 1% of the anglers Licensed to fish in Michigan during the 1983 and 1984 license-years. This survey was sponsored by the Michigan Department of Natural Resources (MDNR) and had a response rate of 59%. The full sample size is 10,948 licensees, of whom 9,628 fished during 1983 or 1984 prior to their return of their survey.[2] The survey provides detailed information on the angler's most recent fishing trip, including species sought:

---

[1] See map VI.1 for the geographic locations of the 83 Michigan counties. They are numbered alphabetically from 1 to 83.

[2] The earliest survey returns would be from the surveys sent out in November or December 1983 or January 1984, which represent more than 60% of the total. The remaining surveys were sent out in May 1984 and September 1984.

location, trip length, trip expenditures, etc., as well as demographic background and extensive fishing experience and preference information

Sample Definition

The model embodies three nested levels of choice: trip duration; fishing product line; and fishing site. Below, we first explain how we define the anglers' fishing product lines and trip durations. We then explain our sample selection procedures and present descriptive statistics for key analysis variables.

Definition of Product Lines

Kikuchi (1986) performed a factor analysis of the MDNR 1983-84 angler survey which identified eleven distinct market segments of fishing experiences. We refer to the segments as *product lines* (PLs). Key distinctions among the product lines include targeted species (coldwater or warmwater) and destination type (Great Lakes: inland lakes, or inland streams). [3] Other distinctions include a category for ice fishing anglers, a category for anglers targeting "anything biting." and a minor category of smelt anglers. The analysis also examined the role of fishing mode (boat, shore) and method (casting, snagging, fly. ice, etc.), but did not find significant differences along these dimensions. [4]

Because ice fishing is quite limited, we restrict our study to open water angling that occurs between April and October. The anadromous run product lines are further restricted to April, May, September, and October. We combined the anadromous-inland-lake and anadromous-inland-stream product lines due to small sample sizes,

---

[3] Coldwater species consist of mainly salmon and trout, and hence are also called "salmonid."

[4] When estimating the MNL model for product lines in which both modes are well represented, we also examined predictions separately by mode choice to reevaluate the modeling decision. We observed no mode-related pattern to the prediction errors.

particularly for the lake category. Inland lake coldwater and in!and stream coldwater are also combined for the same reason. The six product lines employed in our MNL analysis are, therefore, Great Lakes coldwater *(GLcd)*. Great Lakes warmwater *(GLww)*, anadromous run *(Anad)*. inland lake/stream coldwater *(LScd)*, inland lake warmwater *(ILww)*, and inland stream warmwater *(ISww)*.

Definition of Trip-Duration Groups

The trip-duration categories were chosen on the basis of whether trip destination types were different across the trip duration categories. Based on a χ^2 test, one-, two-, three-, and four-day[5] trips have significantly different destinations from one another. Four- and five-day trips are only different at the 10% level. Destinations of trips of five days are not significantly different from those of trips of greater length. The results suggest that the effect of residential location, which dominates the site choice for day trips, is not completely attenuated until five-day trips. Due to sample size considerations, trips of two to four days are grouped together to form one category; which we label "weekend" for convenience. [6] The three resulting duration groups are hence one-day trips, weekend trips (2-4 days), and vacation trips (at least 5 days and up to the maximum of 16 days).[7] labeled as *Day, Wkn,* and *Vac*, respectively

---

[5] The number of days in a trip is calculated from the date/time people left their homes and the date/time people returned to their homes. If the combined hours in the first trip day and the last trip day are greater than 12, both days count. Otherwise, the first and last days together count as only one trip-day. For example: a trip from 10pm the first day to 7am the second day is considered a l-day trip, even though it invo!ves two calendar days.

[6] Note that the categorization is strictly based on trip length, not on which days of the week are involved, so that "weekend" trips do not necessarily occur over the weekend.

[7] We truncate vacation trips at 16 days because it is two weeks plus the extra weekend days. We delete people from our MNL analysis for whom the most recent trip was of more than 16 days, on the grounds that the longer trips have many other purposes than fishing.

Travel Distance and Cost

For the estimation of the model. we need to calculate the distance between an individual's residence and every county in his cr her choice set. To characterize travel distance between origin (home) and destination, we used the county-to-county distance matrix developed by the Michigan State Department of Transportation, based on highway distance measures The distances are measured between the geographical centers of the 83 Michigan counties.

The travel distance is calculated slightly differently for in-state and out-of-state anglers. For an in-state angler, the distance between the home county and any other county can be obtained from the distance matrix. [8] For an angler from other states, the point where he or she entered the state of Michigan (the entry point) is first assigned according to his or her origin and destination. For the chosen site. the travel distance outside Michigan is then calculated as the difference between the self-reported total travel distance and the entry-destination distance.[9] The distance between his or her home and any other potential fishing site is then computed as the sum of the out-of-Michigan distance and the distance between the entry point and the county in question.

To calculate the travel time from the travel distance, we use the sample average speed of 40.5 miles per hour. To calculate the distance-cost variable for a site, the two-way distance is multiplied by the vehicle operating cost per mile, $.23,[10] and then multiplied by the share of the total fishing party size represented by the respondent's family

---

[8] For people who fished in their home counties, 10 is used as the one-way driving distance.

[9] If the self-reported total travel distance is less than, the calculated entry-destination distance, we use the self-reported driving time (in hours) instead for the calculation.

[10] This is the 1983/1984 estimate provided by the American Automobile Association (AAA).

Sample Selection

To select individuals for the MNL analysis, we defined samples (from the 1% MDNR sample of licensed anglers) that include all individuals whose choices met the definitions of the six product lines and three duration groups, and who indicated fishing as a purpose of the trip.

Of necessity: we deleted individuals if (1) there was an inconsistency between the self-reported travel time/distance data and the values in the Michigan State Department of Transportation distance matrix: because we could not be sure these individuals were properly assigned to the product line or site,[11] or (2) there was incomplete or inconsistent information on trip duration, trip location, or species sought. The resulting sample, called N^* and consisting of 18 PL-duration subsamples (6 PLs by 3 durations). provides the basis for extrapolation of the analysis to the population of licensed anglers.

However, estimation of the MNL model was restricted to a subset of the anglers in each PL-duration group because all individuals with any missing data on the MNL esplanatory variables had to be deleted. The sample used for MNL estimation is called N_{MNL}. To create the analysis sample N_{PART} for estimating the parameters of

[11] We performed three types of data checking to make sure that information in the returned survey questionnaires is internally consistent.

First, we compared the self-reported travel distance/time against the values in the State of Michigan Department of Transportation origin-destination distance matrix. If inconsistency existed, we checked questionnaires for coding errors. In 125 cases, coding or data problems were corrected in home county, destination county, and self-reported distance variables. In some cases, it is obvious that people reported round-trip distances where one-way distance was actually asked. In 41 cases. we could not resolve the inconsistencies. and the data were discarded.

Second. we checked home county against the zip code variable. 16 people had improper values assigned to their home county variable according to their zip codes. Other information, such as travel distance and time, was also used to confirm the corrections.

Third: the destination counties of some people do not provide angling opportunities for the species for which they fished. Me checked this inconsistency between fish species and destination county. and made 30 corrections of coding errors.

the competing risks participation model, (1) we further delete, from the MNL, sample: observations missing either the age duration data[12] or the explanatory variables: and (2) me include the non-participant. To predict total trips in the season: however, only the explanatory variables are needed, so trip predictions are available for a slightly larger sample, N_{PRED}, than N_{PART}.

Table VI.1 presents the classification of the individuals in the MDNR sample and their use in this study. Groups 0 and 1 are the non-participants, people who took no trip from April 1, 1983 up to the time they returned the survey. [13] Groups 2 to 4 are the "pseudo participants" because their trips were longer than 16 days and/or were not for the purpose of fishing, and consequently were considered inappropriate for inclusion in a welfare analysis of recreational fishing Groups 5 to 9 are the "true" participants in our analysis. The MNL sample N_{MNL} consists of groups 6 to 9. The participation analysis sample N_{PART} consists of groups 1, 3, 4, and 9. Since we do net need age data for total trips prediction. sample N_{PRED} contains people in group 8 in addition to those in sample N_{PART}

Tables VI.2 to VI.4 present the means and standard deviations of angler characteristics for the Day, Wkn, and Vac duration samples that we use for estimation. The means for N^{*} are similar to those for the analysis samples. The variables are as follows.

- *HHY* is the annual household income in dollars
- *WkHrs* is the individual weekly work hours. It has a value of 40 if an individual had a full-time job, and 20 for a part-time job. If a second job was also held by the individual, 40 or 20 is further added for full-time and part-time, respectively.
- *Wage* is the individual pre-tax wage rate per hour. It is calculated as the individual's annual income divided by the annual work hours (= WkHrs x 52). If less than \$3.25, it is set to the minimum wage rate of \$3.25 per hour. The

---

[12] Calculated from last trip date and survey return date.

[13] A respondent is classified as a non-participant only if all relevant trip information is missing, including destination site, trip length, fish species sough, and trip date.

post-tax wage rate used in the estimation is obtained by multiplying the pre-tax wage by the individual's tax rate, calculated according to individual income bracket.

- *HmDest* is the one-way distance in miles between an angler's home and his/her chosen fishing destination.
- *Instate* is a (0,1) dummy variable. It is assigned 1 if the angler resided in the state of Michigan: and 0 otherwise.

Fish Catch Rate Data from the MDNR Creel Survey

Michigan's Great Lakes sport fishery has been monitored by the MDNR Fisheries Division with a statewide contact creel census program since 1983.[14] The objective of the program is to obtain a continuous record of sport catch, catch rate, and catch composition in the Great Lakes and important anadromous river fisheries. Though sampling efforts and study areas are different each year, the creel census methodology remains the same.

The Michigan creel census is based on a stratified design, using simple random sampling within strata. Strata include port fished, month, weekday-weekend (holiday), and mode of fishing. Catch and effort estimates are made for each cell in the stratified design and then combined to give monthly and seasonal figures.

The catch rates used in the analysis are calculated from the angler-party interview data collected for each area sampled. In the creel survey, an angler party is defined as one or more anglers who fished together. Angler parties are interviewed at the end of their fishing trips at various boat launching ramps, marinas, piers, and along the shoreline. Anglers are queried as to the mode of fishing (i.e., boat, shore, pier, open ice, or shanty ice) they just used, where they fished, how long they fished, what they

[14] Information about the census operations can be obtained from the MDNR technical reports written by Rakoczy and Rogers (1987) for the 1986-87 census operations, and Rakoczy and Lockwood (1988) for the 1985-86 year.

fished for, the numbers (by species) of fish they caught, and the number of fishing trips they made or intended to make that day. Additional data are also collected on each angler in the party, such as age and sex of the angler, zip code or county of residence, and the types of angling methods used (casting, still fishing, trolling, etc.). These data are recorded on an angler interview form by census personnel.

The catch rates used in our analysis are broken down by species and by month for each county (the site unit in the analysis.) We combined the data on total catch by species and total angler- hours data for ports in a county to calculate the average hourly catch rates.

When estimating the site choice model for inland lake and inland stream product lines, we did not use catch rate data because of endogeneity between catch rates and participation. (As we explain below, me substituted measures of the quantity of water resources, differentiated by quality.) For the inland PLs, participation appears to adjust slowly to changes in catch rates: previous catch rates appear to drive current participation. Whereas last month's catch rate may have been high, inducing high participation rates, the current catch rate may be low due to the high participation rate last month. However, due to the slow adjustment process, current participation may still be high. (Thus perverse results would occur with catch rates as explanatory variables in the equations.) This is a special phenomenon for areas with limited resources: which can be depleted by too many anglers. It is unlikely that angler participation could have such an effect on large resources like the Great Lakes and their major tributaries.

Great Lakes Coldwater Species Catch Rates

Five salmonid species are considered important for the Great Lakes coldwater product line: chinook salmon, coho salmon, lake trout! rainbow trout, and brown trout. The feasible open-water fishing months are April to October. All 41 Great

Lakes counties provide angling opportunities for this product. line.[15] Map VI.2 shows the location of these Great Lakes counties.

Though not the most abundant Great Lakes species, various species of salmonids are the target of most Great Lakes sport fishing anglers. During the 1985 open-water fishing season, the Lake Michigan salmonid catch was composed of 59% chinook salmon, 13% coho salmon, 16% lake trout, 5% rainbow trout, 6% brown trout, and less than 1% of other salmonids.[16] For the 1986 fishing season: the percentages were 57%, 15%, 15%, 4%, and 8%, respectively. The other Great Lakes have similar catch compositions.[17] Therefore, chinook salmon is the most important salmonid in terms of the numbers of fish harvested. Lake trout and coho salmon are the second and third most numerous salmonid in the Great Lakes sport catch. Table VI.5 reports the means and standard deviations of the catch rates for these Great Lakes salmonid species.

Anadromous Run Species Catch Rates

The same salmon species (chinook, coho, and rainbow) as those of Great Lakes coldwater are adopted here for the anadromous run product line in the Great Lakes river systems. Feasible fishing months are April, May, September, and October only, during which periods salmon migrate down to and back from the Great Lakes. Salmon anadromous run angling is possible in 44 Michigan counties, including most Great Lakes counties and a few inland counties.[18] Map VI.3 indicates the location of these

---

[15] Fishing in the five northernmost Lake Superior counties is, however, still restricted by ice in April, and therefore is only available for six months from May to October. They are Baraga (5), Gogebic (27), Houghton (31), Keweenaw (42), and Ontonagon (66).

[16] Such as pink salmon, Atlantic salmon, brook trout, and spake.

[17] For example, the percentages were 53%, 4%, 33%, 3%, and 7% for Lake Huron in the 1986 season.

[18] The inland counties included are Eaton (23), Ingham (33), Ionia (34), Kent (41), Lake (43), and Newaygo (62). Great Lakes counties excluded are Keweenaw (42), Monroe (58) and Tuscola (79).

anadromous run counties.

Chinook salmon is still the most abundant salmonid in the anadromous run catch, followed by rainbow trout and coho salmon. The catch rates are higher during the fall runs (September and October) than the spring runs (April and May). Table VI.6 reports the means and standard deviations of the catch rates for the anadromous run salmon species.

Great Lakes Warmwater Species Catch Rates

The fish species included in the Great Lakes warmwater product line are: yellow perch, walleye, northern pike, smallmouth bass, and carp.[19] Feasible fishing months are April through October. 40 out of the 41 Great Lakes counties are available for the Great Lakes warmwater fishing.[20]

Yellow perch is the most numerous in the Great Lakes catch of all species, coldwater or warmwater. For instance: it made up 68% (69%) of ail the fish caught in Lake Michigan during the 1985 (1986) fishing season.[21] Table VI.7 shows that the yellow perch catch rate is more than ten times that of any other warmwater species

Data on Other Characteristics of Site Quality

Data from state and federal sources are used to derive site quality indicators. Site properties that are generic to all product lines include

- *AOC* is a dummy variable that indicates whether a county has been designated as an 'Area of Concern' for toxic contamination by the International Joint Commission 21 out of the 83 Michigan counties are designated Areas of Concern, as shown in map VI.5.

---

[19] 'Carp' includes freshwater drum, catfish, and sucker. The 'smallmouth bass' category also includes largemouth bass, bluegill, and pumpkin.

[20] Luce county (48) is excluded due to its extremely low catch throughout all months of the season.

[21] The percentage was as high as 79% for Lake Huron and 88% (including walleye) for Lake Erie during the year 1986.

- % *Forest* measures the percentage of county land that is forested. This variable is used as a proxy for natural beauty.

- A continuous integer-valued variable *Feature* contains the number of unique natural features, such as Pictured Rocks National Lakeshore and Sleeping Bear Dunes National Lakeshore. Only 14 Michigan counties have special landscape features.

Site Data for Great Lakes Counties

Site data specific to the Great Lakes product lines (both coldwater and warmwater) include the following for the 41 Great Lakes counties.

- Number of parking spaces *(GLprkg)* in GL counties. Only 2 GL counties do not have parking facilities.

- Number of harbors *(GLhrbr)* in GL counties. Only 4 GL counties do not have any harbor for boat mooring.

- Number of slips for boat mooring *(GLslip)* in GL counties.

- Number of ramps for boat launching *(GLramp)* in GL counties. 38 out of the 41 GL counties have ramps.

Table VI.8 reports the means and standard deviations of the Great Lakes product line site characteristics.

Site Data for Anadromous Run Counties

Site data specific to the 44 anadromous run product line counties include

- For anadromous run angling, the presence of a lake in an anadromous stream *(ANlake)* provides opportunities for the use of a boat, in addition to the shore angling available in all 44 counties. Only 10 anadromous run counties have lakes.[22]

Table VI.9 reports the means and standard deviations of the anadromous run site data.

---

[22] They are Benzie (10), Berrien (11), Charlevoix (15), Houghton (31), Leelenau (45), Manistee (51), Mason (53), Muskegon (61), Oceana (64), and Ottawa (70).

Site Data for Inland Coldwater Counties

Since this product line is the combination of inland lake coldwater and inland stream coldwater, variables pertaining to both are included. Those 73 counties that offer inland coldwater angling opportunities are shown in map VI.4.

Inland Lakes

Three types of inland lakes are available for recreational angling: coldwater-only lakes, warmwater-only lakes, and two-story lakes. A two-story lake has an upper layer of water that is warm enough to support warmwater fish species, while the water below is cold enough for coldwater angling to be possible. Data from MDNR allow us to measure the acres of lakes in the separate categories. For the inland lake coldwater product line, only the coldwater lake measures are used. A total of 67 Michigan counties have coldwater lakes for trout fishing.

- Acres of the coldwater-only lakes *(ILcdacre)* in the county.
- Acres of the two-story lakes *(IL2story)* in the county.
- Total acres of the coldwater lakes *(ILtotcd)* in the county. This variable is the sum of the above two variables.
- Acres of the coldwater lakes with fishing consumption advisories *(CntmLC)* in the county. Only two counties have contaminated coldwater lakes.[23]

Inland Streams

Inland cold streams can be classified by their fish quality and tributary status. A cold stream may be classified as top quality main stream, top quality tributary stream, second quality main stream, or second quality tributary stream. 69 Michigan counties are available for inland stream trout angling.

[23] They are Houghton (31) and Marquette (52).

Variables pertaining to this specific product line are

- Six counties are listed in the *Michigan Fishing Guide 1983*, section 'Quality Fishing,' as having streams on which fly fishing is allowed to improve the quality of fishing.[24] A dummy variable *IScdFly* is used to indicate this possibility.
- Miles of top quality main streams *(IScd1main)* in county.
- Miles of top quality tributary streams *(IScd1trib)* in county.
- Miles of second quality main streams *(IScd2main)* in county.
- Miles of second quality tributary streams *(IScd2trib)* in county
- Miles not elsewhere classified *(IScdNEC)* in county.
- Miles of coldwater streams contaminated *(CntmSC)* in county. Only two counties have contaminated streams. [25]

Table VI.10 reports the means and standard deviations of these lake and stream variables.

Sire Data for Inland Lake Warmwater Counties

All 83 Michigan counties have warmwater lakes. The fishing resource variables used are

- Acres of the warmwater-only lakes *(ILwwacre)* in the county.
- Acres of the two-story lakes *(IL2story)* in the county.
- Acres of the warmwater lakes with fishing consumption advisories *(CntmLW)*. Only three counties have non-zero values.[26]

Table VI.11 reports the means and standard deviations of these lake acres variables.

---

[24] These counties are Crawford (20), Lake (43), Missaukee (57), Oakland (63), Oscoda (68), and Wexford (83).

[25] They are Marquette (52) and Osceola (67).

[26] They are Allegan (3), Houghton (31), and Ottawa (70).

Site Data for In1and Stream Warmwater Counties

Like inland coldwater streams, the inland warmwater streams can also be classified by their quality and tributary status. A warmwater stream could be top quality main stream. top quality tributary stream, second quality main stream, or second quality tributary stream. All 83 Michigan counties have warmwater streams. Variables specific to this product Line include

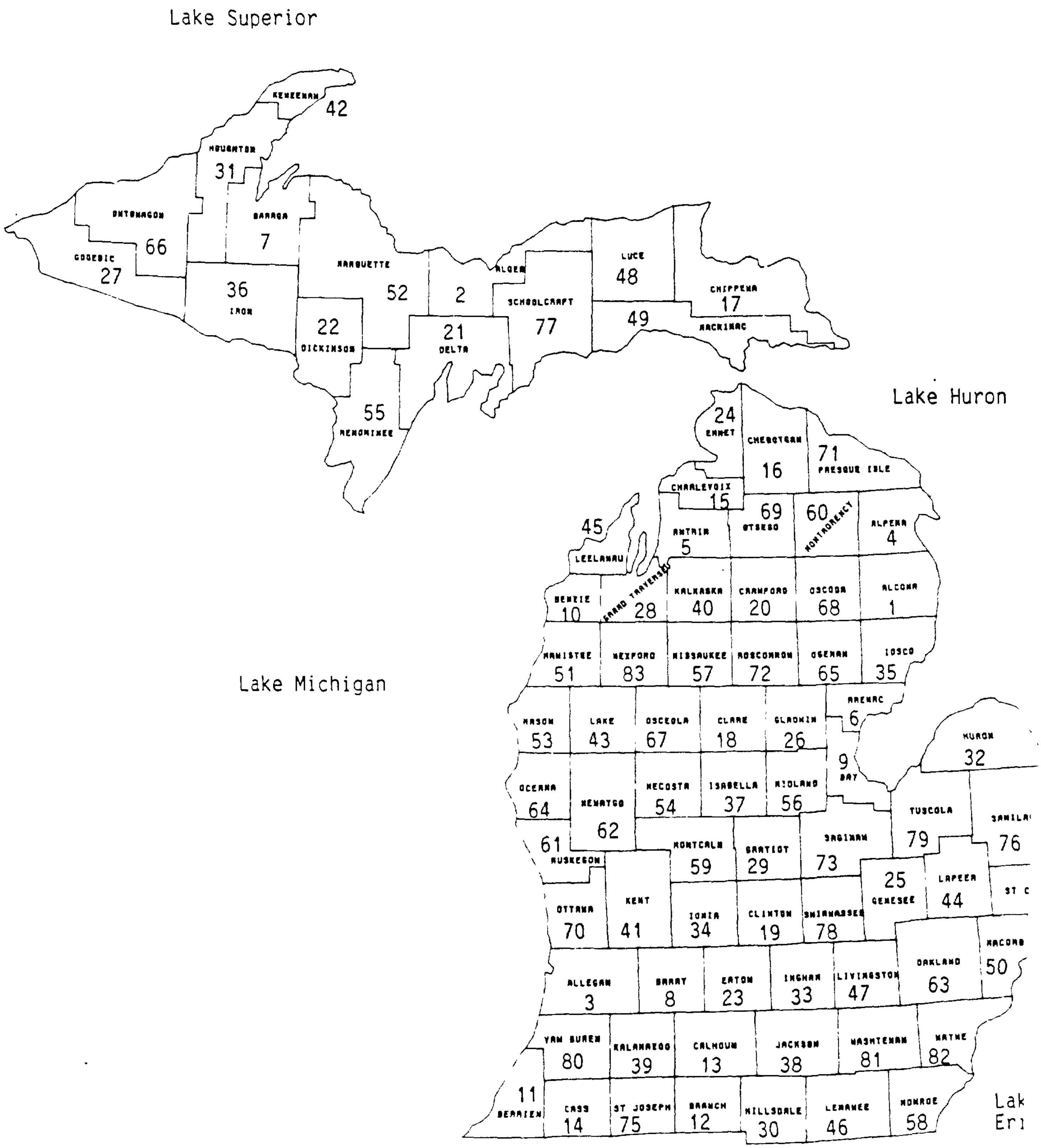
- Miles of top quality main streams *(ISww1main)* in county.
- Miles of top quality tributary streams *(ISww1trib)* in county.
- Miles of second quality main streams *(ISww2main)* in county.
- Miles of second quality tributary streams *(ISww2trib)* in county.
- Miles not elsewhere classified *(ISwwNEC)* in county.
- Miles of warmwater streams contaminated *(CntmSW)* in county. A total of 12 counties have contaminated warmwater streams. [27]

Table VI.12 reports the means and standard deviations of these stream miles variables.

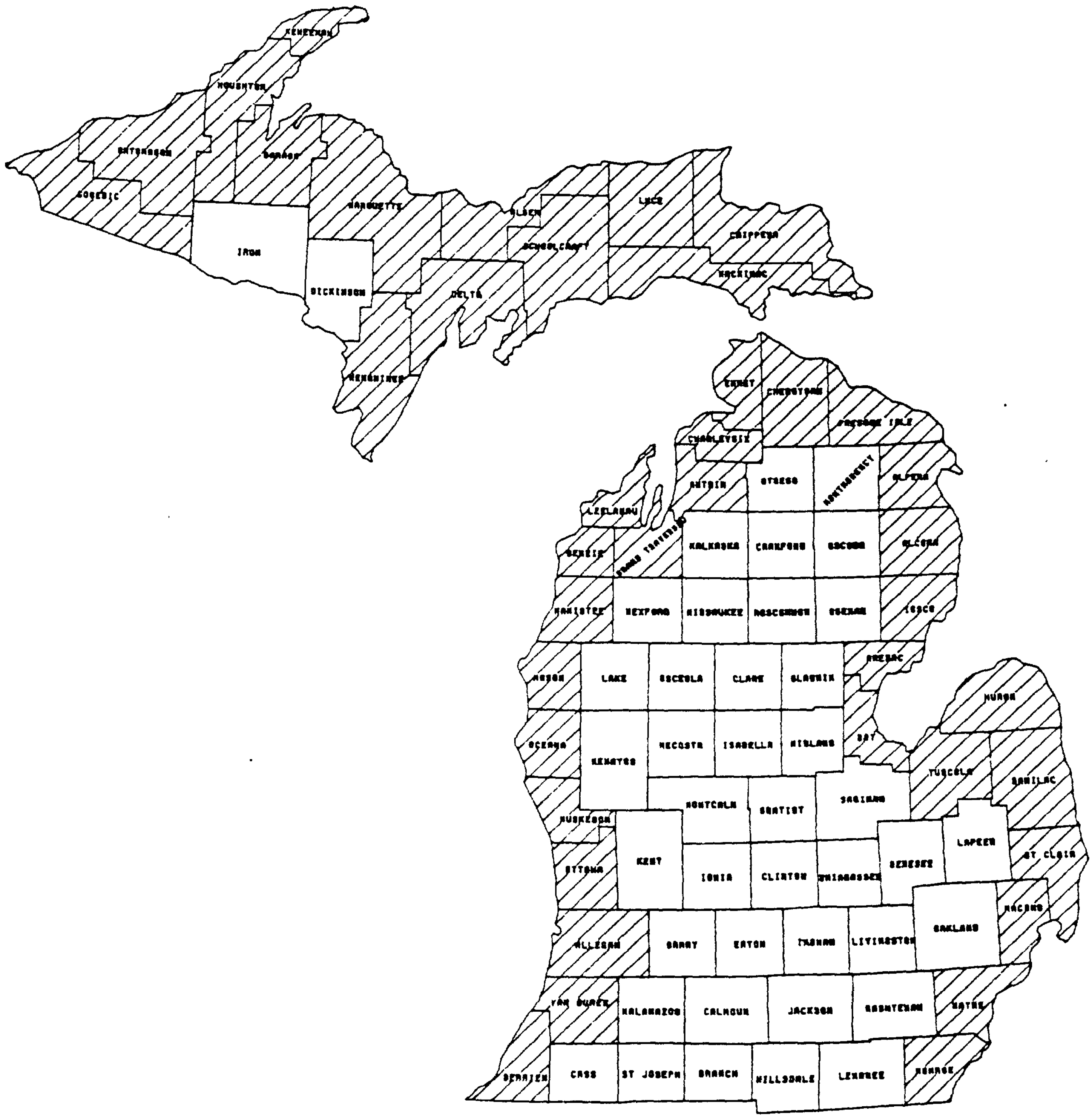
---

[27] They are Allegan (3), Bay (9), Berrien (11), Clinton (19), Gratiot (29), Isabella (37), Kalamazoo (39), Livingston (47), Midland (56), Monroe (58), Saginaw (73), and Shiawassee (78). Many of these counties are in the Saginaw Bay area.

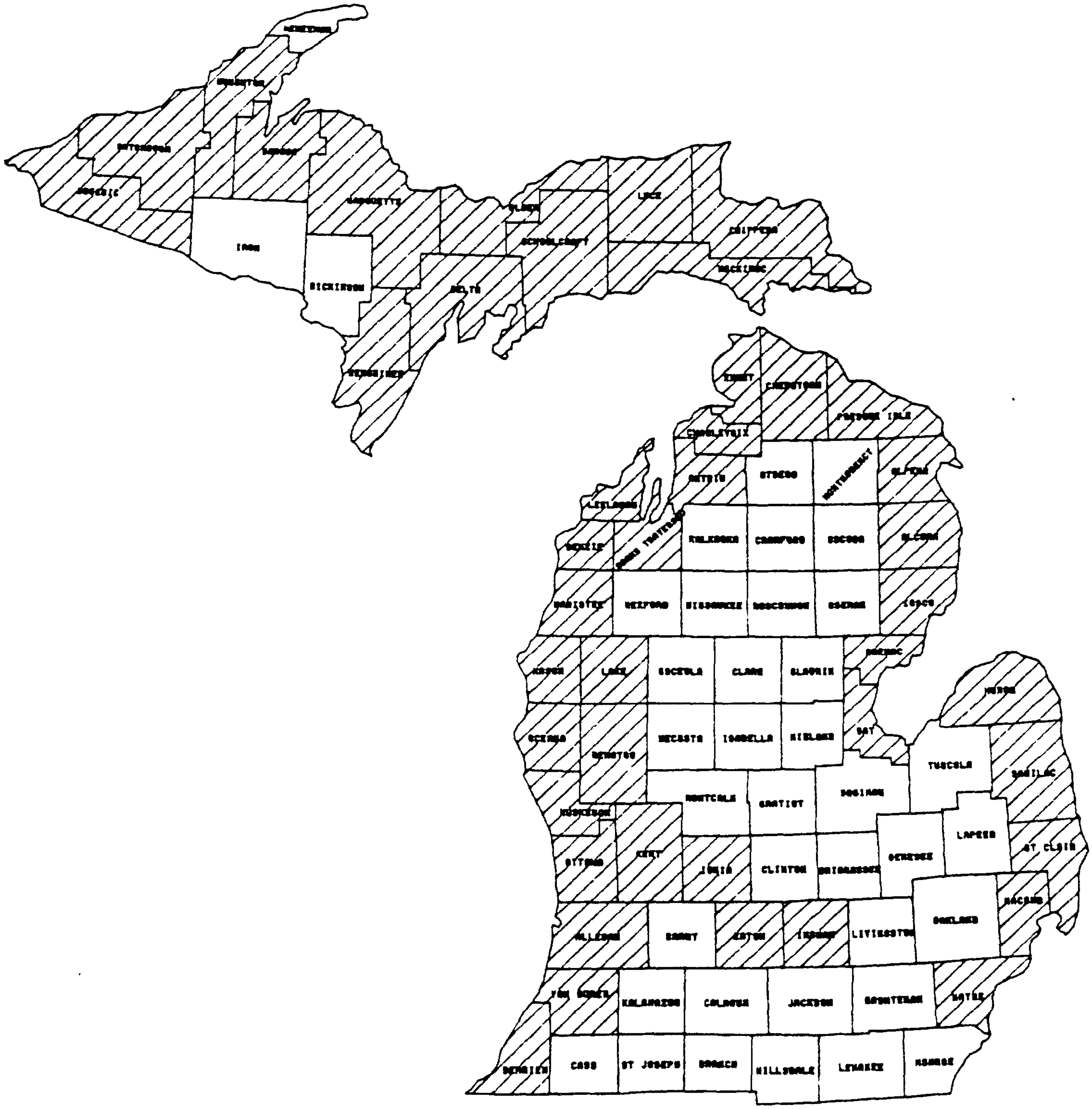
Map VI.1: State of Michigan counties



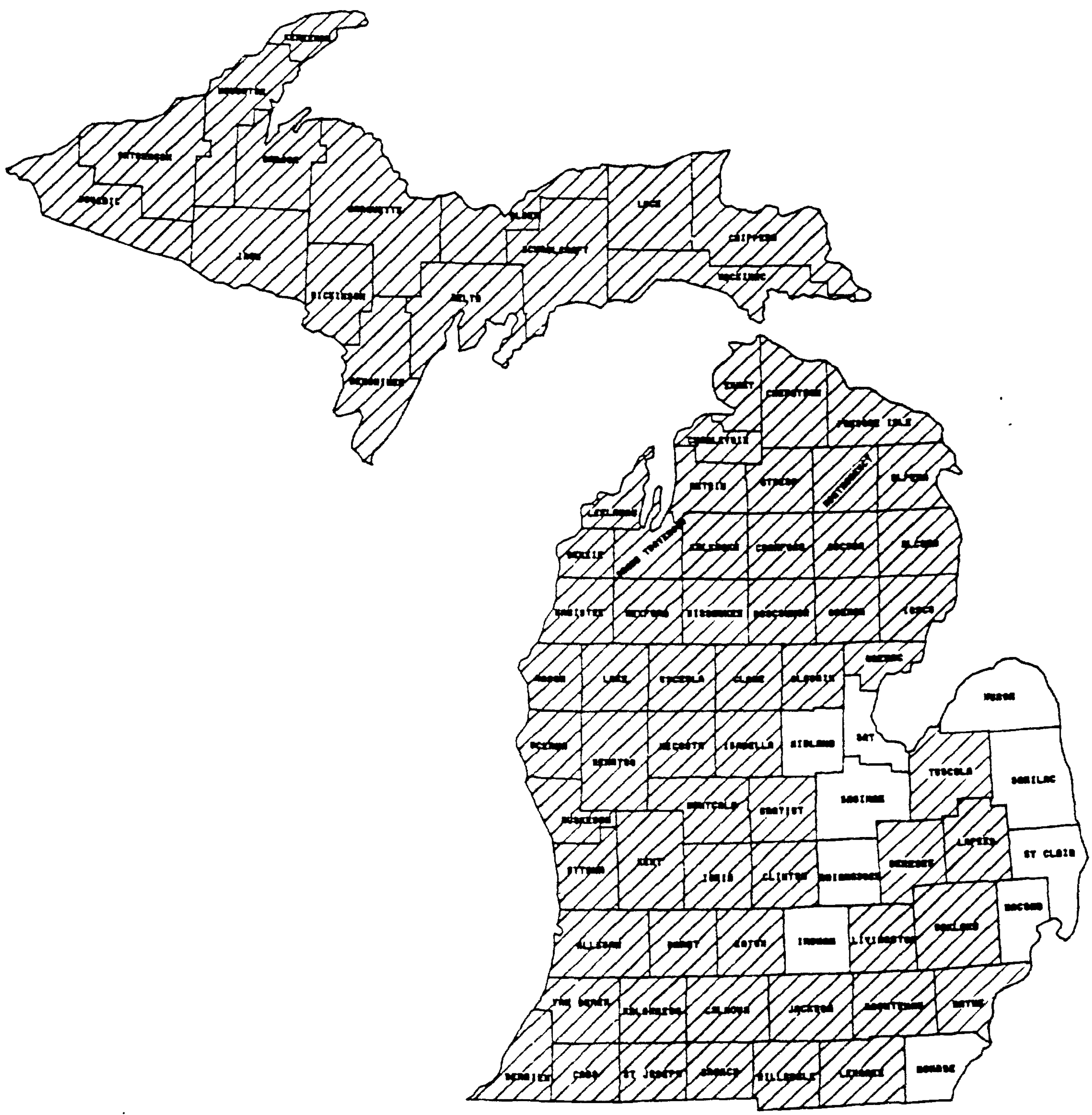
Map VI.2: Great Lakes product line counties



Map VI.3: Anadromous run product line counties



Map VI.4: Inland coldwater product line counties



Map VI.5: Areas of Concern in Michigan

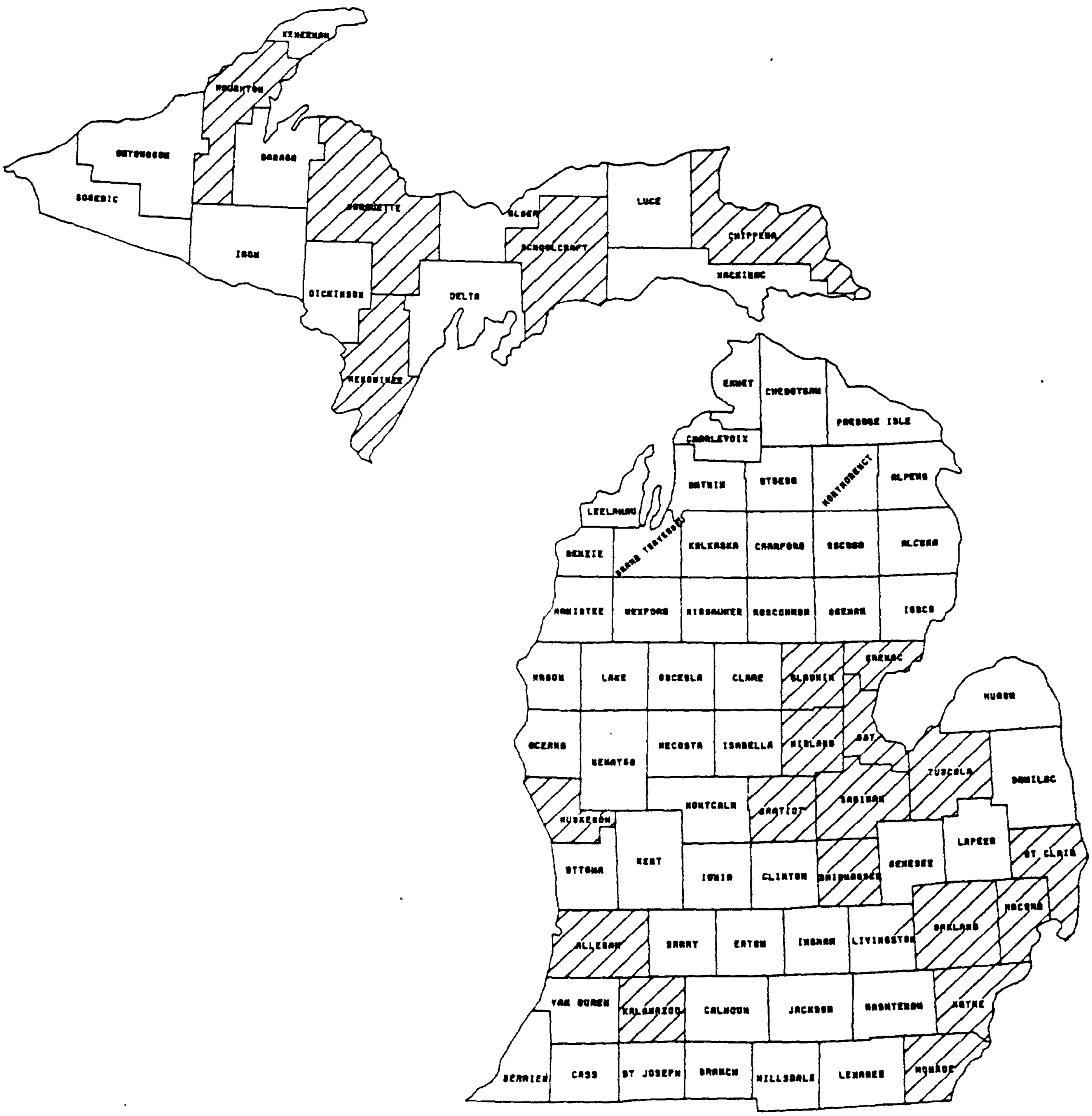


Table VI.1: Classification of sample observations

Group	N	Definition
Non-participants:		
0	738	No trip; no data for participation estimation
1	582	No trip; have data for participation estimation
Pseudo-participants:		
2	1707	Trip invalid, no data for participation estimation
3	137	Trip > 16 days; have data for participation estimation
4	148	Fishing not a trip purpose; have data for participation estimation
Participants:		
5	1817	Trip valid; missing data for MNL analysis
6	556	MNL sample (trip hours missing)
7	224	MNL sample (trip hours available)
8	358	MNL sample; have data for participation prediction only
9	4681	MNL sample; have data for participation estimation
Total	**10948**	**Full MDNR sample**

Table VI.2: Angler characteristics of the Day group

Sample	HHY	Wage	WkHrs	HmDest	Instate
N^*	25937.33	8.45	29.02	29.07	0.96
[N=4067]	(15002.19)	(7.46)	(18.95)	(39.48)	(0.20)
N_{MNL}	26407.20	8.65	28.61	29.23	0.96
[N=2666]	(15216.03)	(7.64)	(19.12)	(40.32)	(0.19)
N_{PART}	26656.12	8.98	29.74	28.01	0.96
[N=2370]	(15255.66)	(7.40)	(18.49)	(38.14)	(0.19)

Table VI.3: Angler characteristics of the Wkn group.

Sample	HHY	Wage	WkHrs	HmDest	Instate
N^*	30092.23	10.65	32.76	131.14	0.84
[N=1653]	(16049.09)	(7.93)	(16.76)	(93.36)	(0.37)
N_{MNL}	30106.46	10.64	32.27	131.05	0.83
[N=1228]	(15922.25)	(7.96)	(17.10)	(94.45)	(0.36)
N_{PART}	30209.84	10.85	32.98	132.17	0.85
[N=1108]	(15658.39)	(7.61)	(16.16)	(94.53)	(0.36)

Table VI.4: Angler characteristics of the Vac group.

Sample	HHY	Wage	WkHrs	HmDest	Instate
N^*	29357.35	9.71	29.40	235.16	0.73
[N=1915]	(15108.86)	(7.94)	(18.33)	(159.97)	(0.45)
N_{MNL}	29289.01	9.77	28.70	237.29	0.74
[N=1369]	(15004.55)	(7.94)	(18.45)	(164.07)	(0.44)
N_{PART}	29750.62	10.29	30.52	241.34	0.73
[N=1203]	(14970.86)	(7.53)	(17.44)	(163.41)	(0.44)

Note: Numbers in parentheses are standard deviations.

Table VI.5: Means and standard deviations of the GLcd catch rates

Month	N	Chinook	Coho	LakeT	Rainbow	BrownT
April	36	0.026	0.019	0.001	0.015	0.028
		(0.045)	(0.044)	(0.005)	(0.023)	(0.044)
May	41	0.030	0.018	0.065	0.010	0.009
		(0.053)	(0.045)	(0.106)	(0.025)	(0.016)
June	41	0.022	0.007	0.073	0.003	0.005
		(0.031)	(0.019)	(0.087)	(0.008)	(0.012)
July	41	0.044	0.003	0.056	0.001	0.003
		(0.041)	(0.006)	(0.085)	(0.004)	(0.013)
August	41	0.056	0.005	0.042	0.001	0.002
		(0.053)	(0.010)	(0.093)	(0.003)	(0.011)
September	41	0.039	0.018	0.014	0.004	0.001
		(0.030)	(0.046)	(0.043)	(0.012)	(0.004)
October	41	0.037	0.022	0.002	0.020	0.017
		(0.042)	(0.059)	(0.008)	(0.033)	(0.078)

Table VI.6: Means and standard deviations of the Anad catch rates

Month	N	Chinook	Coho	LakeT	Rainbow	BrownT
April	44	0.0	0.0	0.0	0.090	0.0
		(0.0)	(0.0)	(0.0)	(0.102)	(0.0)
May	44	0.0	0.0	0.0	0.079	0.0
		(0.0)	(0.0)	(0.0)	(0.088)	(0.0)
September	44	0.112	0.037	0.0	0.011	0.0
		(0.177)	(0.096)	(0.0)	(0.023)	(0.0)
October	44	0.107	0.021	0.0	0.039	0.0
		(0.163)	(0.039)	(0.0)	(0.054)	(0.0)

Table VI.7: Means and standard deviations of the GLww catch rates

Month	N	Y.Perch	Walleye	N.Pike	SM Bass	Carp
April	40	1.073	0.002	0.000	0.000	0.007
		(1.431)	(0.008)	(0.000)	(0.002)	(0.031)
May	40	0.524	0.034	0.005	0.004	0.009
		(0.735)	(0.051)	(0.011)	(0.011)	(0.021)
June	40	0.781	0.028	0.002	0.006	0.015
		(0.816)	(0.057)	(0.006)	(0.015)	(0.042)
July	40	0.570	0.048	0.003	0.004	0.021
		(0.589)	(0.104)	(0.007)	(0.007)	(0.070)
August	40	0.621	0.033	0.011	0.002	0.009
		(0.714)	(0.081)	(0.026)	(0.005)	(0.022)
September	40	0.921	0.023	0.002	0.004	0.010
		(1.073)	(0.066)	(0.006)	(0.013)	(0.030)
October	40	1.254	0.010	0.004	0.001	0.007
		(1.775)	(0.028)	(0.011)	(0.004)	(0.018)

Table VI.8: Descriptive statistics: Great Lakes site attributes

Variable	N	Mean	Std Dev.	Minimum	Maximum
%Forest	41	0.547	0.291	0.7	0.97
Feature	41	0.366	0.733	0.0	3.00
AOC	41	0.341	0.480	0.0	1.00
GLprkg	41	205.512	169.034	0.0	720.00
GLhrbr	41	1.683	1.105	0.0	5.00
GLslip	41	893.927	1633.222	0.0	7951.00
GLramp	41	8.707	10.530	0.0	44.00

Table VI.9: Descriptive statistics: Anad site attributes

Variable	N	Mean	Std Dev.	Minimum	Maximum
AOC	44	0.273	0.451	0.0	1.0
ANlake	44	0.227	0.424	0.0	1.0

Table VI.10: Descriptive statistics: LScd site attributes

Variable	N	Mean	Std Dev.	Minimum	Maximum
AOC	73	0.192	0.396	0.0	1.0
Feature	73	0.260	0.602	0.0	3.0
IScdFly	73	0.082	0.277	0.0	1.0
IScd1main	73	19.671	27.767	0.0	112.0
IScd1trib	73	24.877	29.705	0.0	144.0
IScd2main	73	59.000	79.890	0.0	478.0
IScd2trib	73	79.945	98.199	0.0	456.0
IScdNEC	73	3.630	5.934	0.0	30.0
CntmSC	73	0.315	2.101	0.0	17.0
ILtotcd	73	2817.712	6159.408	0.0	33942.0
CntmLC	73	17.233	138.062	0.0	1178.0

Table VI.11: Descriptive statistics: ILww site attributes

Variable	N	Mean	Std Dev.	Minimum	Maximum
AOC	83	0.253	0.437	0.0	1.0
Feature	83	0.229	0.570	0.0	3.0
ILwwacre	83	7652.084	6755.713	204.0	29219.0
IL2story	83	2414.048	5830.382	0.0	33897.0
CntmLW	83	50.096	265.674	0.0	1780.0

Table VI.12: Descriptive statistics: ISww site attributes

Variable	N	Mean	Std Dev.	Minimum.	Maximum
AOC	83	0.253	0.437	0.0	1.0
Feature	83	0.229	0.570	0.0	3.0
ISww1main	83	28.578	30.022	0.0	109.0
ISww1trib	83	21.711	28.624	0.0	115.0
ISww2main	83	22.639	35.032	0.0	181.0
ISww2trib	83	105.446	89.374	0.0	330.0
ISwwNEC	83	3.976	6.912	0.0	33.0
CntrnSW	83	3.042	10.170	0.0	55.0

CHAPTER VII

EMPIRICAL RESULTS: MULTINOMIAL LOGIT MODEL

We present the multinominal logit estimation results in this chapter. First, we discuss the empirical basis for our definition of the number of active hours in a day. This parameter choice then affects the definition of the choice set for individuals, particularly on the shorter duration trips of one and two days. We then report estimates of the nested multinomial logit model. The major portion of the chapter focuses on estimates of the determinants of site choice for each of the six product lines. The final segment presents estimates of the second level of the nested structure, the choice of fishing product line.

In this chapter the model we estimate is based on the assumption that total trip time is exogenous for individuals when they are making their choice of which site to visit. In other words, if they had chosen any other feasible alternative in their choice set, the total trip time would have been the same as it was for the chosen site. As we argued in Chapter III we have chosen this model as our base case because we believe it incorporates a more consistent treatment of time than the others appearing in the literature. In the Appendix we explore the sensitivity of the model to the alternative assumptions about trip time discussed in Chapter III.

Choice Set Computation: Implementation Details

To define the individual choice set of feasible sites, as discussed in chapter III, we choose a value for the active hours h per day to include people whose trip durations are within two standard deviations above the mean of all people having the same number of trip days.[1] Consider the means and standard deviations of the observed trip time for people whose trips last from 1 to 7 days below,

Trip days, D	1	2	3	4	5	6	7
Trip time, T	7.56	25.77	52.65	77.84	101.21	126.46	152.17
	(3.85)	(11.05)	(6.36)	(6.43)	(6.59)	(5.76)	(6.03)

Let μ_D and σ_D denote the mean and standard deviation of the trip time T in hours for the observed D-day trips. The criterion we adopt is to select an h such that, for any D.

$$(D-1)\cdot 24 + h \approx \mu_D + 2\sigma_D.$$

Consequently h = 17 is used for the calculation of individual choice sets in the MNL site analysis. People whose trip durations are greater than $(D-1)$ 24-17 are deleted as outliers. [2]

The Site Choice MNL Estimation

The MNL model is first applied repeatedly to the 18 PL-duration subsamples (6 PLs by 3 durations) to estimate the parameters of the PL-specific site utility functions. The inclusive value of the sites for each PL-duration group is calculated from this analysis. For each duration group, the six PL subsamples are then pooled together and MNL is again used to model the PL decision.

---

[1] This is a somewhat arbitrary decision. If a small value is used, we not only lose a lot of observations for individuals who chose sites we define as infeasible for them but we may also exclude sites that people actually consider visiting when making site choices. On the other hand, for some people we may be defining the choice set too broadly.

[2] Fewer than 2% of the total observations were excluded due to this time constraint violation.

Great Lakes Coldwater Product Line.

Table VII.1 presents the MNL estimation results for the Great Lakes cold water product line. All parameters are estimated with signs consistent with *a priori* predictions, except for the *Lake Trout* catch rate in the Wkn sample. The *Forest* variable is significant for all three duration groups, as is the *AOC* indicator. The *Feature* variable is significant for the longer trips (i.e., Wkn and Vac groups). Most catch rate variables are significant except *Rainbow Trout*. The catch rate of *Brown Trout* is not used because it does not have much variation across counties. The access variables (parking, harbor, slips, and ramps) are also excluded due to concern they may be endogenously determined.[3] Due to the limited observations we have in some of the 18 PL- duration subsamples, we cannot incorporate site dummies to estimate the site-specific constant terms in the utility function.[4]

The travel cost coefficients provide estimates of the marginal utility of income (MUI) for each group. As expected, the MUI decreases as the constraint on available opportunities relaxes from the Day group to the Wkn and Vac groups.

The model predicts 51% of the actual choices made by the *Day* group, which indicates that destination decisions for short trips are substantially driven by geographical proximity. As the importance of geographical proximity declines with increasing trip duration, the model is less effective in capturing the other factors influencing choices: correct predictions are 15% and 14%, for Wkn and Vac groups respectively.[5]

---

[3]The State has attempted to construct access facilities at popular sites to accommodate demand, which suggests a reverse causation as well as a direct causation from available facilities.

[4]Some subsamples have even fewer observations than the number of sites

[5]These low percentages can also be attributed to the larger choice sets of Wkn and Vac trips. On average, number of feasible sites in a Wkn or Vac choice set is about twice that in a Day choice set.

Great Lakes Warmwater Product Line

Table VII.2 presents the MNL estimation results for the Great Lakes warmwater product line. The *Forest* variable is not significant for the Day and Wkn groups, but is positive and significant for the Vac group. *Feature* has a positive sign and is significant for the longer trips. These parameter estimates are consistent with our expectation that people who stay longer on a site will care more about environmental amenities. *AOC* is not significant for the shorter trips. For the Vac group, *AOC* is significant but the sign is opposite our expectation.[6]

The catch rate estimates are mostly significant for the Vac trips. Only *Northern Pike* is significant for the Wkn group and none is significant for the Day group. Some species have a negative sign for some duration groups, possibly due to high correlation among the catch rates, but they are not significant in those cases.

Anadromous Run Product Line

Table VII.3 presents the MNL estimation results for the Anadromous run product line. The presence of lakes or reservoirs is important, as manifested by the significance of the *Lake* dummy variable: when lakes occur in anadromous streams in a county, anglers can choose either boat- or shore-based angling. The *AOC* variable is significant for all trips, but has a positive effect on day-trip anglers. The positive *AOC* parameter estimate for Day group may be picking up non-linearities in the utility function with respect to travel distance (and cost). Most of the population is in the southeast Michigan, and for those individuals, most nearby counties are Areas of Concern.

[6]According to Douglas Jester of the MDNR Fisheries Division, the reason for this perverse result is that Saginaw Bay, where people go on longer trips, is one of the few areas with high walleye catch rates, and also is one of the Areas of Concern.

Chinook and *Rainbow* salmon catch rates are also important factors affecting people's site choice decision, whereas *Coho* is not.

Inland Coldwater Product Line

Table VII.4 presents the MNL estimation results for the inland coldwater (lakes and streams) product line. Most parameters are imprecisely estimated, perhaps in part because we have pooled two product lines, each with small sample sizes. Unexpectedly, we do not pick up a significant positive effect for the opportunity to do fly fishing at a site, as the lackluster results for the *IScdFly* dummy shows. *AOC* is not significant for any group. The Area of Concern designation primarily applies to Great Lakes contamination, so it is not surprising that it is not as much of a concern for inland anglers as for Great Lakes anglers.

The coldwater stream miles variables and the Forest variable are highly correlated ($p > .6$); we attribute the mixed performance of the stream miles variables to the correlation. The negative parameter estimates for the stream contamination measure *CntmSC* indicate that anglers avoid contaminated streams when making a site decision, except for the Wkn sample (which has an insignificant positive parameter estimate). The fish consumption advisory variable for coldwater lakes *CntmLC* is not significant (except for Wkn where it has a significant and perverse effect). These results probably reflect the lack of variation in the variable: only two counties have fish consumption advisories on coldwater lakes.

Inland Lakes Warmwater Product Line

Table VII.5 presents the MNL estimation results for the inland lake warmwater product line. The parameters of the environmental amenities variables *(AOC, Forest, and Feature)* are precisely estimated with predicted signs. Note this is the only inland product line for which *AOC* parameter estimates are significant. The acres of inland warmwater lakes variables are very significant; the variables measuring acres of two-story lakes are only significant for the Vac group. *GntmLW* is not significant, which again probably reflects the lack of variation in the variable: only three counties have fish consumption advisories on warmwater lakes.

Inland Streams Warmwater Product Line

Table VII.6 presents the MNL estimation results for the inland stream warmwater product line. *Feature* is significant, with the predicted sign, for the longer trips, and *Forest* is significant for all duration groups. The *AOC* dummy is not significant for any duration group. The negative coefficients on the fish consumption advisory - variable border on significance at the 10% level for the Wkn and Vac groups. All the top quality stream parameter estimates have positive signs: the main stream variables are significant; the tributary stream variables are not, though in the Vac group it borders on 10% significance. All the second quality stream variables have negative parameter estimates though none is precisely estimated.

The Product Line Choice MNL Estimation

The following variables are considered relevant when people make product line choices:

- The inclusive value *(SiteIV)* as an index of the potential utility the sites of a PL can offer.
- The favorite catch species *(FavCatch)* dummy. People indicated in the questionnaire which species (out of a total of 16 species) they like to catch most.[7] The product lines that contain the individual's favorite species have a value of 1, otherwise they are set to 0.[8]

---

[7]The species composition of the six product lines are not mutually exclusive. For example, the Great Lakes coldwater and the inland coldwater product lines overlap. The Great Lakes warmwater and the inland warmwater product lines also share most of their fish species.

[8]People also indicated in the survey what their favorite eating species are. This variable is not used because it is highly correlated with the favorite catch species.

- The favorite water type *(FavWater)* dummy. people revealed their water type preference among Great Lakes, inland lakes, and streams/rivers. Product lines whose water type is favored have a value of 1, and 0 otherwise.[9]

- The expected supply costs *(Supplies$)*. To predict the supply costs of the different product lines, the self-reported supply costs are regressed on number of days in trip, angler party size, and the interaction of them separately for each product line. The estimated parameters are then used to calculate the projected product line costs.

- The expected boat costs (*Boat* $). similar regressions are run for the calculation of the boat costs as those for the supply costs. This is done, however, separately for people who own a boat and those who do not. The Great Lakes coldwater product line is the most costly in terms of both supply costs and boat use costs.

Because the favorite species and favorite catch variables are so closely correlated with the product line choices, we report regression results without those variables. When those variables (with an average value of .8) are included, their power overwhelms the effect of some of the other variables. We include five product line dummies, with the Great Lakes coldwater (GLcd) dummy omitted as the base case.

Table VII.7 presents the MNL estimation results for anglers' product line choice. All parameter estimates of *SiteIV* are within the unit interval [0,1], which assures us that the NMNL model is not violating the consumer random utility maximization assumption. Also the coefficients of *SiteIV* are significantly different from 1 at 1% level for all three duration groups, Therefore, the simple logit, in which the *SiteIV* coefficient is assumed to be 1, is rejected.

The parameters of the supply and boat use costs are positive for some duration groups, contrary to expectations. It is possible that by incurring higher costs of some sort, people will also produce a higher quality experience which is not captured through any of the other variables. For example, the use of bigger boats may provide

[9]We also know the fishing mode and method that are favored by anglers. This information, however, is not utilized because most PLs offer opportunities for the use of most modes/methods. Angler experience is not included as a variable either since only inland cold stream angling demands some skill.

access to fishing opportunities that otherwise are not available.

Table VII.1: MNL estimates for the GLcd product line

	Day Anglers	Wkn Anglers	Vac Anglers
Dist$/100	–17.39	–4.20	–2.40
	(-16.49)	(-10.77)	(-8.52)
AOC	–1.54	–1.75	–.98
	(-8.60)	(-8.07)	(-4.26)
%Forest	2.34	1.23	2.26
	(4.13)	(3.14)	(4.74)
Feature	0.08	0.51	0.60
	(0.38)	(3.15)	(3.62)
Chinook Salmon	8.37	8.93	9.73
	(3.88)	(5.66)	(5.62)
Coho Salmon	3.93	5.37	5.06
	(1.90)	(3.34)	(3.20)
Lake Trout	3.31	–1.27	3.93
	(1.66)	(-0.49)	(3.40)
Rainbow Trout	2.21	2.47	3.45
	(0.42)	(0.70)	(0.88)
Log Likelihood	–520.2	–795.7	–625.2
x^2- test	737.5	321.3	205.3
%Choices Right	50.9	15.3	13.8
#People[MI/non-MI]	338[327/11]	195[151/44]	195[151/44]
#Choices	5624	10201	7785

Note: Numbers in parentheses are *t*- statistics.

Table VII2: MNL estimates for the GLww product line

	Day Anglers	Wkn Anglers	Vac Anglers
Dist$/100	-17.34	-3.49	-1.47
	(-22.17)	(-7.86)	(-5.01)
AOC	-0.05	-0.23	0.60
	(-0.30)	(-0.90)	(2.92)
%Forest	-0.93	-0.34	2.21
	(-1.40)	(-0.59)	(3.06)
Feature	0.20	0.84	1.62
	(0.53)	(3.01)	(6.34)
Yellow Perch	0.08	0.13	0.12
	(1.09)	(1.36)	(0.94)
Walleye	-0.20	0.53	3.23
	(-0.34)	(0.46)	(2.32)
Northern Pike	0.73	10.45	26.91
	(0.08)	(2.52)	(7.46)
Smallmouth Bass	9.37	-22.61	21.76
	(1.00)	(-1.28)	(2.36)
Carp	2.97	-5.43	3.92
	(1.95)	(-1.48}	(2.08)
Log Likelihood	-824.2	-424.3	-389.3
x^2- test	1743.3	159.9	195.3
%Choices Right	62.7	12.9	18.2
#People[MI/non-MI]	668[654/14]	140[133/7]	132[105/27]
#Choices	10179	5250	5280

Note: Numbers in parentheses are t- statistics.

Table VII.3: MNL estimates for the Anad product line

	Day Anglers	Wkn Anglers	Vac Anglers
Dist$/100	-15.72	-2.66	-1.45
	(-10.10)	(-5.89)	(-3.09)
AOC	0.66	-0.86	-1.91
	(2.56)	(-2.64)	(-3.18)
Lake	0.63	0.86	0.79.
	(2.44)	(3.89)	(3.01)
Chinook Salmon	2.53	4.04	4.42
	(3.53)	(9.33)	(9.15)
Coho Salmon	-0.85	-8.80	0.88
	(-0.27)	(-1.72)	(0.43)
Rainbow Trout	9.49	6.13	6.15
	(4.63)	(4.67)	(3.93)
Log Likelihood	-173.5	-309.4	-224.2
X^2- test	276.7	177.8	157.0
%Choices Right	67.5	15.9	13.8
#People[MI/non-MI)	123[113/10]	107[77/30]	80[40/40]
#Choices	2133	4475	3520

Note: Numbers in parentheses are *t*- statistics.

Table VII.4: MNL estimates for the LScd product line

	Day Anglers	Wkn Anglers	Vac Anglers
Dist$/100	-24.49	-5.08	-2.13
	(-14.44)	(-9.20)	(-6.11)
AOC	0.13	0.18	-0.12
	(0.39)	(0.48)	(-0.30)
Forest	4.90	7.36	5.76
	(4.40)	(9.09)	(6.81)
Feature	-0.21	-0.22	-0.38
	(-0.64)	(-0.82)	(-1.26)
IScdFly	0.32	-0.19	0.23
	(0.86)	(-0.63)	(0.70)
IScd1main/100	-0.32	1.11	0.27
	(-1.05)	(3.15)	(0.69)
IScd1trib/100	-0.59	0.40	0.15
	(-0.92)	(0.76)	(0.30)
IScd2main/100	0.48	0.03	0.23
	(2.19)	(0.18)	(1.45)
IScd2trib/100	0.44	-0.08	0.06
	(1.66)	(-0.44)	(0.33)
SCnec/100	-0.27	-6.97	1.52
	(-0.11)	(-2.34)	(0.75)
CntmSC	-0.13	0.02	-.14
	(-1.88)	(0.39)	(-1.84)
ILTotCd/100	0.003	0.001	0.002
	(1.49)	(0.74)	(1.08)
CntmLC/100	0.10	0.43	0.11
	(0.37)	(2.10)	(0.55)
Log Likelihood	-243.0	-466.0	-392.3
X^2- test	671.1	383.8	176.6
%Choices Right	69.3	20.6	9.8
#People[MI/non-MI]	192[187/5]	155[143/12]	112[97/15]
#Choices	5343	10977	8176

Note: Numbers in parentheses are *t*- statistics.

Table VII.5: MNL estimates for the ILww product line

	Pay Anglers	Wkn Anglers	Vac Anglers
Dist$/100	-23.66	-5.15	-1.71
	(-35.29)	(-18.60)	(-14.20)
AOC	-0.43	-0.81	-0.65
	(-4.31)	(-5.34)	(-4.88)
Forest	2.83	3.62	2.83
	(6.79)	(12.06)	(11.59)
Feature	0.21	0.33	0.54
	(1.53)	(2.59)	(5.68)
ILwwacre/100	0.007	0.006	0.007
	(12.14)	(9.74)	(14.84)
IL2story/100	0.0004	0.0003	0.0027
	(0.35)	(0.37)	(5.66)
CntmLW/100	-0.0006	0.006	0.016
	(-0.40)	(0.24)	(0.74)
Log Likelihood	-1645	-1613	-2455
x^2- test	3353.2	791.5	870.7
%Choices Right	61.2	10.5	8.6
#People[MI/non-MI]	989[949\40]	459[369/90]	654[470/184]
#Choices	35284	36941	54282

Note: Numbers in parentheses are *t*- statistics.

Table VII.6: MNL estimates for the ISww product line

	Day Anglers	Wkn Anglers	Vac Anglers
Dist$/100	-26.47	-6.08	-3.75
	(-17.46)	(-7.95)	(-7.70)
AOC	-0.06	0.15	-.14
	(-0.26)	(0.44)	(-0.40)
Forest	1.81	3.88	6.37
	(2.09)	(4.79)	(7.46)
Feature	-.29	0.79	0.64
	(-0.83)	(2.54)	(2.48)
ISww1main/100	0.82	1.67	1.10
	(2.55)	(3.63)	(2.48)
ISww1trib/100	0.56	0.64	0.87
	(1.12)	(0.93)	(1.57)
ISww2main/100	-0.16	-1.02	-0.60
	(-0.61)	(-1.88)	(-1.29)
ISww2trib/100	-0.24	-0.37	-0.02
	(-1.73)	(-1.79)	(-0.10)
ISwwNEC/100	-1.39	2.56	4.03
	(-0.93)	(1.23)	(2.01)
CntmSW	-.002	-0.028	-0.098
	(-0.19)	(-1.63)	(-1.59)
Log Likelihood	-349.9	-249.5	-3112.8
x^2- test	887.2	136.7	160.9
%Choices Right	73.2	21.9	14.6
#People[MI/non-MI]	246[240/6]	73[67/6]	89[67/22]
#Choices	8100	5747	7387

Note: Numbers in parentheses are *t*- statistics.

Table VII.7: MNL estimates for the product line choice

	Day Anglers	Wkn Anglers	Vac Anglers
SiteIV	0.925	0.438	0.246
	(26.66)	(5.58)	(2.80)
Supplies$/100	0.034	-0.021	0.003
	(3.37)	(-1.46)	(0.72)
Boat$/100	0.015	0.001	0.002
	(1.97)	(0.80)	(0.48)
GLww(dummy)	1.579	-0.899	-0.657
	(13.31)	(-3.67)	(-2.88)
Anad(dummy)	-0.995	-0.902	0.285
	(-5.594)	(-3.94)	(1.53)
LScd(dummy)	-1.975	-2.707	-1.070
	(-12.90)	(-6.40)	(-3.06)
ILww(dummy)	0.359	-0.868	0.922
	(2.63)	(-2.52)	(3.45)
ISww(dummy)	-.191	–2.461	-1.170
	(-1.451)	(-7.45)	(-3.95)
Log Likelihood	-3422	-1844	-1730
x^2- test	1928.8	429.8	770.2
%Choices Right	50.5	38.9	51.8
#People	2580	1196	1262
#Choices	14213	6715	6769

Note: Numbers in parentheses are *t*- statistics.

CHAPTER VIII

EMPIRICAL RESULTS: THE PARTICIPATION MODEL

We report the estimation results of the competing risks model in this chapter. We first present the exponential model estimates. We then report the Weibull model estimates, which allow us to test the duration independence assumption of the exponential model,

Variable Definitions and Analysis Sample

For the participation model, we estimate the determinants of individuals' choices about how many trips of different durations to take during a fishing season. The explanatory variables include

- *IV:* The inclusive value of the product lines and sites available to an individual. This variable varies with time and is computed for each month in the open-water season (April - October) from the NMNL estimates.

  We showed above in chapter III, equation (III.15), that the inclusive value index could be decomposed into the sum of three terms: the pseudo-IV (which does not include the unmeasured choice occasion income), the choice occasion income, and an individual specific constant term. This variable is the pseudo-IV. As noted below, we substitute an alternative income measure below for the (unmeasured) choice occasion budget. The individual-specific term is captured by the variables measuring individual characteristics. Because we substitute for the components of the IV, we do not constrain the parameter estimates of the various substitute variables to be equal.

- *WageCost:* The measure of the opportunity wage cost of taking a fishing trip, which equals the after-tax wage rate times the number of trip-hours. (In Table VIII.1, we report the values of the after-tax wage variable.)
- *HHY:* Individual annual household income is substituted for the unknown choice-occasion budget for recreational fishing.[1]
- *ExpRate:* (0,1) dummy, =1 for people self-reported as "experienced" or "expert"; =0 for people self-reported as "beginner" or "somewhat experienced".
- *No Work:* (0, 1) dummy, =1 if the respondent is a student, unemployed, or retired.
- *NoSpouse:* (0, 1) dummy, =1 if the respondent does not have a spouse.
- *SpNoFish:* (0, 1) dummy, =1 if the respondent is married and his/her spouse does not fish.
- *NoKids:* (0, 1) dummy, = 1 if the respondent has no child under 16 years old.

Table VIII.1 presents the descriptive statistics of the age duration and the explanatory variables for the analysis sample. The mean and standard deviation of the *IV* variable are not reported because *IV* varies with time. The sample sizes in this table represent all observations included in the participation estimation. The full sample of the 5376 observations comprises four subsamples:

1. The *Day* group — 2258 people whose last trips are day trips.
2. The *Wkn* group — 1070 people whose last trips are weekend trips.
3. The *Vac* group — 1105 people whose last trips are vacation trips
4. The *Non-Partic* group — 867 people for whom no trip was observed taken.

We split the *Non-Partic* group into two types of people: (1) the 582 people who did not report a trip *(No- Trip)* and (2) the 285 people whose reported trip is not suitable for being counted in the welfare analysis ***(Inelig Trip)*** .[2] The "true" non-participants

[1] We can interpret this substitution to be in the spirit of the lifetime income framework, which posits that people can borrow and lend freely across time periods.

[2] A trip is labeled as uncountable either because fishing was not a purpose of the trip or because the trip was longer than 16 days, and so was motivated by many purposes besides recreational fishing.

did not take a trip between the beginning of the 1983 open-water fishing season and their questionnaire return date, which ranges from seven to fourteen months. Therefore, their sampled between-trip duration is left truncated at April 1, 1983 and fight truncated at the questionnaire return date. For the people last trip is not a "countable" fishing trip, we can only conclude that no "countable" fishing trip has occurred between the time the uncounted trip was taken and the time the questionnaire was returned, which ranges from one to nine months (with one outlier at thirteen months). Therefore, the sampled between-trip duration is left-truncated by the trip that is not countable and right-truncated at the return of the questionnaire.

The distribution of the age durations (or the censored ages for the non-participants) in our sample is presented in table VIII.2. The different selection processes for the two sub-groups of non-participants are clearly evident in the lower means and ranges of the age variable for the *Pseudo-Trip* group. Since we only consider the open-water fishing season (from April to October), a "year" consists of only seven months.

The Day, Wkn and Vac samples are smaller than their counterparts used in the MNL analysis because we had to delete the people for whom we did not have the questionnaire return dates or the last trip dates. For the MNL analysis, we only need to know the month during which the trip was taken. For the participation analysis, however, we require the complete month/day/year information in order to calculate the age duration.[3] Also, people whose questionnaires were returned on the same day that their trips ended are excluded because they might have waited until after their next trip to fill out (and mail) the questionnaires. In such cases, these two dates are not independent and our random censoring assumption is violated.

---

[3] Actually we can further exploit interval information on age. For anglers for whom we only have month and year data, for example, we can calculate upper and lower bounds of the age duration and include an integral term (instead of a density term) in the likelihood function.

Participation Model Estimates

Table VIII.3 presents the parameter estimates of the exponential competing risks model. The *IV* parameter indicates that there is a positive relationship between trip value and number of day and weekend trips taken by anglers, as predicted; however; the relationship is negative for vacation trips. The parameter estimates are significant for all trip lengths, Therefore, people are likely to take more day and weekend trips, but fewer vacation trips, when trip value is higher.

The *WageCost* variables are only significant (with the predicted negative effect) for Day trips; the probabilities of taking a weekend or vacation trip do not appear to be sensitive to the wage costs of the trips. People with higher household incomes *(HHY)* tend to take more trips of all durations than lower income people. Greater angler experience *(ExpRate)* is also associated with greater participation intensity.

To interpret the effect of not working *(NoWork* = 1, for students, unemployed and retired people), we need to consider the combined impact of *NoWork* and *WageCost*, where the wage cost is positive only for employed anglers. Though the *NoWork* coefficient estimates are negative for all three duration groups, the net effect of not working is negative only for Wkn and Vac (for which *WageCost* is not significant). Consequently, the results indicate that non-working individuals are more likely than working individuals to take day trips. The negative parameter estimate on *NoWork* indicates there is a non-linearity in the relationship between participation probabilities and *WageCost* when an individual does not earn wages.

To interpret the effect of marital status and whether a spouse fishes, we need to look at both *NoSpouse* and *SpNoFish* variables. An angler is assigned to one of three categories: (1) single, (2) married and spouse does not fish, or (3) married and spouse fishes (the excluded category). For the single anglers, *NoSpouse* = 1 and *SpNoFish* = 0. For the married anglers (with *NoSouse* = 0), *SpNoFish* = 1 if their spouses do not fish, and = 0 otherwise. The parameter estimates indicate that, relative to having a spouse who fishes, having a spouse who does not fish significantly lowers the probability of taking the longer trips. Being single does not have significantly different effects than having a spouse who fishes, though a dampening effect on the vacation probability does approach significance at the 10% level. Referring to table VIII. 1, we know that the four analysis subsamples have the following distribution:

Status	No-Trip	Pseudo-Trip	Day	Wkn	Vac
Single	26%	22%	23%	20%	19%
Married, SpNoFish=1	31%	27%	28%	27%	24%
Married, SpNoFish=0	43%	51%	49%	53%	57%

Note a larger proportion of the non-participants are single or have spouses who do not fish: this contributes to the negative parameter estimates of the *NoSpouse* and *SpNoFish* variables. Having no children under 16 years old also reduces the probability of taking a day trips but has no effect on the probabilities of weekend or vacation trips. [4]These family variables do not have the large and significant effects of *IV*, *HHY*, and *ExpRate*.

The Weibull model reported in table VIII.4 yields scale parameter estimates very similar to, the values imposed in the exponential model. The existence of negative duration dependence is suggested by the negative shape exponent $\alpha = -0.022 < 0$. The shape parameter is calculated as $\gamma = \epsilon^{-0.022} = 0.978 < 1$. However, since the shape exponent α is not significantly different from 0, the exponential model is not rejected.

From the estimated parameters in the exponential and Weibull models, we calculate the predicted numbers of day, weekend, and vacation trips for the anglers in the analysis sample, which are reported in table VIII.5 The negative duration dependence in the Weibull model is sufficiently small that the predicted numbers of trips

[4]The estimated effect is only significant on Day trips.

of each type are essentially the same for both models.

In future work, it would be desirable to evaluate whether we should treat the non-participants differently from the participants in our analysis. In a population-based sample, non-participants will include respondents who clearly did not intend to participate (and who realized their intentions.) In contrast, our sample is restricted to people who purchased Michigan fishing licenses during the survey years: non-participants wanted the option to fish, but chose not to exercise the option. Nonetheless, it is possible that sample non-participants differ substantially from participants in their unobserved characteristics. Unfortunately, we do not have the data to identify special circumstances such as illness or unusually heavy work or family obligations, which could have been unanticipated at the time of the license purchases.[5] The suggestion of negative, duration dependence in our sample could be due to these or other sources of unobserved heterogeneity between the participants and the non-participants.

To test this hypothesis, a *conditional Weibull* model could be estimated with only the participants (conditional on participation during the survey seasons). If the original competing risks model is correct (i.e., no unobserved heterogeneity exists), the conditional Weibull estimates should be very close to those of the unconditional Weibull since the conditional Weibull involves only loss of efficiency.[6] This additional analysis was beyond the scope of this study.

---

[5]Table VII. 1 shows that the sample means for the non-participant groups are different from those of the other three groups for key observed characteristics, though the differences are not significant (1) on average, the non-participants have lower wages and household incomes (2) they have less angling experience than people in the other groups; (3) a larger proportion of them do not work; (4) a larger proportion of them do not have a spouse; and (5) a larger proportion of them do not have children under 16 years old. In Table VIII.1, we also can see that the mean censored age duration of the non-participants is over three times the mean (uncensored) age for the Day and Wkn samples, and twice that of the Vac sample.

[6]Defining the distinction between participants and non-participants is more complicated in our dataset than with a more typical survey, in which total trips are measured for a fixed time period across all individuals. In our dataset, we observe “no trip” outcomes over very different time periods, ranging from one to fourteen months. To model “no-participation”, we must confront the question, “over what time period must a licensed angler not-participate to be considered a different type of person?”

External Validation of the Participation Model

In order to validate the participation estimates from the model, we compare total participation predicted by the model against participation estimates derived from another data source: mail surveys collecting participation diary data, sponsored by the Michigan Department of Natural Resources during years 1980-82. Because the process and criteria for counting trips and days are different in the two datasets, the comparison is not suited to statistical testing. The diary mail survey was a 1% sample of all licensed anglers, with a 62% response rate. The questions about trip participation elicited counts of angler-days over the prior three months, in single-month increments. In the table below, we report the number of angler days calculated directly from the data, by fishing product line. Two adjustments are then made to these estimates to provide a more appropriate estimate of open-water angler days. First, the MDNR estimates that the calculations overstate the number of annual angler days by 35% on average, based on comparison of these trip estimates against direct observation of participation in small area surveys.[7] To correct the overstatement, we divide the numbers by 1.35. Second, we adjust the total annual days by a factor of .9 in order to limit the estimate to open-water angling days only, for comparability with the sample employed in participation modeling. The adjusted totals are reported at the bottom of the table.

---

[7] Personal communication with Douglas B. Jester, Fisheries Division, Michigan Department of Natural Resources, 1992.

Angler-Days Estimated from the MDNR Diary Survey[8]

Product Line	1980	1981	1982
Great Lakes Coldwater	2,150	2,575	2,220
Great Lakes Warmwater	4,620	4,690	4,710
Anadromous Runs	1,430	1,735	1,270
Inland Coldwater	2,000	2,250	1,590
Inland Warmwater	11,200	12,150	11,010
Annual Total	21,400	23,400	20,800
Adjusted Total	14,267	15,600	13,867

To compare model predictions against the diary data estimates above, we translate the predicted number of fishing trips from the model reported in Table VIII.5 into angler-days and extrapolate from the analysis sample to the population of anglers. However, as noted above, the participation concepts are different in the two samples: the analysis of the MDNR diary mail survey is designed to be all-inclusive of fishing days, whereas our demand analysis is restricted to trips suitable for inclusion in a welfare analysis of the benefits of recreational fishing.[9] We incorporate a partial adjustment for this exclusion, as described below.

To calculate estimated angler days based on the recreational fishing model, we translate trips T into days D by multiplying the number of weekend trips by the sample mean weekend trip-length of 3.05 days, and the number of vacation trips by the sample mean vacation trip-length of 9.12 days.[10] The estimated total trips and angler-days for the 4433 participants and the 867 non-participants in our sample are presented in Table VIII.6, separately for each model.

Denote the predicted total number of (eligible) angler-days of the 867 non-participants in the sample by D_0 and that of the 4433 participants by D_1. We then

---

[8]The unit is thousand angler-days.

[9]On the other hand, the variable measured in the angler survey used in the participation modeling is number of trip-days.

[10]This is the average length of all trips between 5 days and 16 days in our sample.

extrapolate as follows:

$$D^* = \frac{P}{N}\left[\frac{N_{0,N}}{S_{0,N}} \times D_{0,N} + \frac{N_{0,I}}{S_{0,I}} \times D_{0,I} + \frac{N_1}{S_1} \times D_1\right]$$
$$= \frac{1414914}{10948}\left[\frac{1320}{582} \times D_{0,N} + \frac{1992}{285} \times D_{0,I} + \frac{7636}{4433} \times D_1\right],$$

where P = 1,414,914 is the total angler population in 1983,
$N_{0,N}$ = 1320 is the number of no-trip people in the MDNR data,
$N_{0,I}$ = 1992 is the number of ineligible-trip people in the MDNR data,
N_1 = 7636 is the number of participants in the MDNR data, and
N = $N_{0,N} + N_{0,I} + N_1$ = 10,948 is the total sample size of the MDNR data.
$S_{0,N}$ = 582 is the number of no-trip people in our analysis sample,
$S_{0,I}$ = 285 is the number of ineligible-trip people in our analysis sample, and
S_1 = 4433 is the number of participants in our analysis sample.[23]

The total predicted (eligible) trip-days for the population, corresponding to the different models, are calculated to be

Model	$D_{0,N}$	$D_{0,I}$	D_1	D^*
Exponential	3,545	2,123	33,274	10,364,296
Weibull	3,518	2,122	33,124	10,322,086

The total angler-days estimated from our competing-risk model appears to be about 30% less than the MDNR diary mail survey estimate, *without any adjustment* for deletion from the sample of trips of longer than 16 days or trips not originally planned for the purpose of fishing. To take account of these 'ineligible' trips, we also report in Table VIII.6 the number of days of ineligible fishing trips *measured in the sample.* This is a very limited measure of omitted fishing days for the season: whereas all 5300 individuals in the sample may take multiple trips per season that would not be counted in the welfare analysis,[11] we only count the days of one trip and we count them only for those 285 people whose most recent trip was ineligible. We calculated the total trip-days of the 285

[11] The exception is the 582 people who took no trips over one 7-month open water season (or longer.)

people in the sample whose last trip was ineligible as $DNE_{0,1} = 3931$. We extrapolate[12] this estimate to the population of anglers as follows:

$$DNE^{*} = \frac{P}{N} \times \frac{N_{0,I}}{S_{0,I}} \times DNE_{0,I} = \frac{1414914}{10948} \times \frac{1992}{285} \times 3931 = 3,550,935.$$

With the partial adjustment, the prediction is about the same magnitude as the MDNR figures. We infer that the differences in the definition of fishing days included in the estimates may account for the differences between estimates, though we have insufficient data to test fully the hypothesis. Though the differences in the concepts being measured limit our ability to compare the estimates, we conclude that the similarity of predicted participation between the model and the annual diary data provides some evidence corroborating the participation model.

[12]It is possible that sampling bias or selection bias exists in our extrapolation procedures. The individuals in both MDNR datasets were randomly sampled from the total licensed angler population; however, according to Douglas Jester of MDNR Fisheries Division, people who are less experienced or who fish less frequently do tend to be slower (and less likely) to return the questionnaires. As a consequence, we may have an upward bias in our participation calculation from both data sources.

Table VIII.1: Attributes of the participation analysis sample

	No-Trip	Inelig Trip	Day Trip	Wkn Trip	Vac Trip
[Censored] Age	246.96	98.01	58.89	64.10	92.11
(in days)	(54.03)	(56.21)	(50.83)	(54.57)	{52.46)
Wage	5.45	5.96	7.23	8.70	8.34
	(6.09)	(5.83)	(5.66)	(5.73)	(5.80)
$HHY/10^4$	2.20	2.45	2.65	3.02	2.98
	(1.54)	(1.47)	(1.52)	(1.57)	(1.50)
ExpRate	0.38	0.40	0.53	0.51	0.50
	(0.43)	(0.48)	(0.50)	(0.50)	(0.50)
NoWork	0.43	0.39	0.25	0.17	0.23
	(0.50)	(0.49)	(0.44)	(0.38)	(0.42)
NoSpouse	0.26	0.22	0.23	0.20	0.19
	(0.44)	(0.42)	(0.42)	(0.40)	(0.39)
SpNoFish	0.31	0.27	0.28	0.27	0.24
	(0.46)	(0.45)	(0.45)	(0.44)	(0.43)
NoKids	0.71	0.72	0.55	0.56	0.60
	(0.46)	(0.45)	(0.50)	(0.50)	(0.49)
N	**582**	**285**	**2258**	**1070**	**1105**

Note: Numbers in parentheses are standard deviations.

Table VIII.2: Distribution of the age (or censored age) duration length

#Months	No-Trip	Inelig Trip	Day Trip	Wkn Trip	Vac Trip	**Sub-Total**
1	0	22	718	274	122	**1136**
2	0	64	590	366	207	**1227**
3	0	56	423	151	224	**854**
4	0	58	253	108	243	**662**
5	0	36	142	64	152	**394**
6	0	24	67	61	96	**248**
7	372	14	31	25	38	**480**
8	0	5	21	13	17	**56**
9	82	5	10	6	5	**108**
10	57	0	2	1	1	**61**
11	11	0	1	1	0	**13**
12	0	0	0	0	0	**0**
13	53	1	0	0	0	**54**
14	7	0	0	0	0	**7**
Total	**582**	**285**	**2258**	**1070**	**1105**	**5300**

Table VIII.3: Competing risks exponential model estimates

	Day Trips	Wkn Trips	Vac Trips
Intercept	-6.459	-6.812	-5.791
	(-61.56)	(-50.22)	(-51.25)
IV	0.780	0.312	-0.437
	(17.53)	(5.71)	(-16.25)
WageCost/10^2	-1.879	-0.128	0.400
	(-2.50)	(-0.13)	(0.40)
HHY/10^4	0.045	0.123	0.096
	(2.30)	(4.41)	(3.51)
ExpRate	0.372	0.295	0.204
	(8.73)	(4.76)	(3.36)
NoWork	-0.326	-0.587	-0.213
	(-4.21)	(-4.94)	(-1.82)
NoSpouse	-0.019	-0.113	-0.130
	(-0.34)	(-1.30)	(-1.52)
SpNoFish	-0.048	-0.143	-9.264
	(-0.96)	(-1.96)	(-3.58)
NoKids	-0.259	-0.021	0.074
	(-5.43)	(-0.31)	(1.11)
Log Likelihood		-29219.5994	
x^2- test (DOF=27)		2791554.8012	
Likelihood Ratio Index		0.9795	

Note: Numbers in parentheses are t- statistics.

Table VIII.4: Competing risks Weibull model estimates

	Day Trips	Wkn Trips	Vac Trips
Shape exponent α		-0.022	
		(-1.78)	
Intercept	-6.455	-6.811	-5.787
	(-60.88)	(-49.93)	(-50.78)
IV	0.780	0.312	-0.440
	(17.36)	(5.69)	(-16.07)
WageCost/10^2	-1.909	-0.141	0.395
	(-2.51)	(-0.14)	(0.39)
HHY/10^4	0.046	0.124	0.096
	(2.31)	(4.41)	(3.51)
ExpRate	0.376	0.299	0.208
	(8.71)	(4.80)	(3.41)
NoWork	-0.331	-0.589	-0.215
	(-4.21)	(-4.94)	(-1.83)
NoSpouse	-0.020	-0.113	-0.129
	(-0.35)	(-1.30)	(-1.49)
SpNoFish	-0.049	-0.144	-0.265
	(-0.97)	(-1.96)	(-3.56)
NoKids	-0.260	-0.022	0.073
	(-5.39)	(-0.33)	(1.08)
Log Likelihood		-29217.9915	
X^2- test (DOF=28)		2791558.0169	
Likelihood Ratio Index		0.9794	

Note: Numbers in parentheses are *t*- statistics,

Table VIII.5: Predicted number of trips per angler in an open-water season

	N	Day Trip	Wkn Trip	Vac Trip
Exponential model:				
Day Sample	2258	1.255	0.549	0.442
Wkn Sample	1070	1.157	0.568	0.516
Vac Sample	1105	0.905	0.495	0.671
No Trip	582	0.916	0.422	0.426
Inelig Trip	285	0.875	0.426	0.580
Weibull model				
Day Sample	2258	1.248	0.546	0.440
Wkn Sample	1070	1.150	0.565	0.514
Vac Sample	1105	0.900	0.492	0.669
No Trip	582	0.910	0.419	0.423
Inelig Trip	285	0.869	0.423	0.578

Table VIII.6: Predicted number of total angler-days in an open-water season

	Day Trip	Wkn Trip	Vac Trip	**Total**
Exponential model:				
Total Trips (Participants), T_1	5072	2394	2292	– -
Total Days (Participants), D_1	5072	7303	20899	**33274**
Total Trips (No Trip), $T_{0,N}$	533	246	248	—
Total Days (No Trip), $D_{0,N}$	533	750	2262	**3545**
Total Trips (Inelig Trip), $T_{0,P}$	249	121	165	– -
Total Days (Inelig Trip), $D_{0,P}$	249	369	1505	**2123**
Weibull model:				
Total Trips (Participants), T_1	5043	2381	2283	– -
Total Days (Participants), D_1	5043	7262	20819	**33124**
Total Trips (No Trip), $T_{0,N}$	530	244	246	—
Total Days (No Trip), $D_{0,N}$	530	744	2244	**3518**
Total Trips (Inelig Trip), $T_{0,P}$	248	121	165	– -
Total Days (Inelig Trip), $D_{0,P}$	249	369	1505	**2122**

CHAPTER IX

POLICY APPLICATION: LUDINGTON PUMPED-STORAGE PLANT

Biological Scenarios

The economic principles of natural resource damage assessment can be illustrated in the context of an important liability case in which the State of Michigan is suing for damages as a result of fishkills attributable to the operation of the Ludington Pumped-Storage plant on Lake Michigan.[1] A related action has also been brought by the State and the National Wildlife Federation before the Federal Energy Regulatory Commission, which licenses hydro-power plants.

The largest hydropower facility of its kind in the country, the pump-storage plant is responsible for the largest continuous fishkill in Michigan waters. Designed to serve the peak load requirements of Michigan electric consumers, it pumps water from Lake Michigan to a storage reservoir 360 feet above lake level during low-demand periods and releases it back to the lake through six power-generating turbines during peak-demand hours. When operating at full capacity, the plant is capable of producing 1.8 million kilowatts of electricity. Millions of fish are killed every year as they are pumped in with water currents traveling at up to 6 ft/sec and later released through

---

[1] Civil Action No, 86-7075-CE, State of Michigan Circuit Court.

the pump-turbines. Death occurs as a result of pressure changes, direct contact with the pump-turbine blades, and associated stress.

A study commissioned by the utilities that own the power plant estimated that in 1980 the plant killed, among other fish, 1.1–3.2% of the entire biomass in Lake Michigan of alewife, a forage species necessary to support the stocked recreational trout and salmon fisheries, and 5.6% of adult-equivalents for combined angler harvests of five trout and Salinon species. The trout and salmon fisheries most heavily affected by the power plant are completely allocated to recreational uses.

Baseline: Current Plant Operation

This baseline situation represents the operation of the Ludington plant without fish protection measures, from the initial plant startup in 1971 through 1988. Catch rates under this situation are adequately represented by the catch rates used in the estimation of the discrete choice NMNL model.

Termination of Plant Operation

This scenario[2] is designed to capture the change in fishing quality that would occur if all fish mortalities associated with plant operations were eliminated, either by fish protection measures or termination of plant operations. Catch rates under this scenario are higher than in the base scenario for two reasons: sport fish killed by the plant would remain in the stock, and, more importantly, forage fish killed by the plant would be available to support additional stocks of sport fish. Forage is a limiting factor in the current State program for stocking trout and salmon in the Great Lakes.

---

[2]This Ludington pumped-storage plant biological scenario is provided by Douglas B. Jester of the Michigan Department of Natural Resources Fisheries Division.

Salmon and trout catch rates throughout Lake Michigan would be affected since migration of the salmonid species and the forage fish would rapidly diffuse the effects of a change in mortalities at the Ludington-Pumped Storage plant.[3] Catch rates in anadromous fisheries for trout and salmon upstream from Lake Michigan would also be affected in the same Lake Michigan counties and in some inland counties.[4]

According to the MDNR scenario: the termination of plant operations would improve both Great Lakes and anadromous fisheries for trout and salmon in and along Lake Michigan at the following rate:

Product Lines	Species	Increase in Catch Rate
GLcd, Anad	Chinook Salmon	10.0%
	Coho Salmon	3.3%
	Lake Trout	13.7%
	Rainbow Trout	8.6%

Catch rates of other species are unlikely to change outside the immediate plant area of Mason (53) and Oceana (64) counties. Among the warmwater species killed by the plant, yellow perch is the only recreational species killed in significant numbers and included in the MNL model specifications. For these two counties, the scenario specifies that yellow perch catch rates would increase by approximately 7% if yellow perch were not killed by operations of the plant. All of the above catch rate changes are predicted to occur across all months of the year. The 22 Michigan counties affected by the operation of the Ludington plant are shown in map IX.1.

[3]Lake Michigan counties affected include: Allegan (3), Antrim (5), Benzie (10), Berrien (11), Charlevoix (15), Emmet (24), Grand Traverse (28), Leelanau (45), Mackinac (49), Manistee (51), Mason (53), Muskegon (61), Oceana (64), Ottawa (70), Schoolcraft (77), and Van Buren (80). Two Green Bay counties, Delta (21) and Menominee (55), are the only exceptions.

[4]These inland counties are Eaton (23), Ingham (33), Ionia (34), Kent (41), Lake (43), and Newaygo (62).

Consumer Surplus Calculation

To estimate people's willingness-to-pay (WTP) for the termination of Ludington plant operations, we first calculate the seasonal compensating variation for the analysis sample according to formula V.28 in chapter V:

$$W = \sum_i \sum_m \sum_d \left[\frac{T^1_{imd} \cdot \bar{I}^1_{imd} - T^0_{imd} \cdot \bar{I}^0_{imd}}{\tilde{\eta}_d / 100} \right]$$

where

i indexes individuals in the sample of our consumer surplus analysis.

m indexes months (April — October) in an open-water season.

d indexes types of trips (= Day, Wkn, Vat).

0 refers to the "with plant operations" case.

1 refers to the "no plant operations" case,

$\tilde{\eta}_d$ is the weighted (across product lines) MUI per $100 for trip duration type *d*.

T is the predicted number of total trips in a season.

$\bar{I}$ is the pseudo-IV variable defined in chapter III.

The multiplication by 100 is to correct for the fact that the unit of the marginal utility of income parameter $(\tilde{\eta}_d)$ is utility-per-$100 because the distance cost variable is divided by 100. Table IX.1 presents, separately for each duration group, the compensating variation per trip $(\bar{I}^1_{id} - \bar{I}^0_{id}) \times 100/\tilde{\eta}_d$ in 1984 dollars (averaged over the seven open-water fishing months) associated with the fishkill caused by the Ludington operations, conditional on the trip type being chosen. The expected increase in value per trip is small since only a few product lines and sites are affected. Table IX.2 reports the predicted number of season trips T^0 with the plant operating. Since the exponential model is not rejected when tested by estimating the Weibull model, we use the exponential estimates in the calculation. Tables IX.3 and IX.4 report the predicted change in total trips $(T^1_{id} - T^0_{id})$ and the total compensating variation (W_d)

in 1984 dollars for one open-water season if the operation of Ludington plant were terminated. We predict fewer vacation trips from the termination of the plant because the *IV* parameter estimate of the participation model is negative for the vacation trip. Therefore, higher *IV* will lead to a reduction in the vacation trips. The total seasonal compensating variation for the sample is thus calculated to be W = $1939. 71 (in 1984$) from the subtotals in table IX.4.

We then extrapolate the sample CV to the population as

$$\begin{aligned} W^* &= \text{CPI}\left(\frac{91}{84}\right) \times \frac{P}{N} \times \frac{N}{S} \times W \\ &= 1.348 \times \frac{1414914}{10948} \times \frac{10948}{4824} \times 1939.71 \\ &= \$766{,}219.03 \end{aligned}$$

where P = 1, 414, 914 is the total population of licensed anglers in 1984, N = 10,948 is the sample size of the MDNR data, and S = 4824 is the number of people in our consumer surplus analysis. N/S is the factor for extrapolating from the consumer surplus sample to the MDNR sample. P/N is the factor for extrapolating from the MDNR sample to the total population of licensed anglers. Because the trip choices and associated expenditures were incurred in 1983 and 1984, the measure is in 1983 or 1984 dollars until corrected with a current price index.[5]

Therefore, the final extrapolation from the sample to the population of licensed anglers yields an annual damage estimates of $0.77 million (in 1991$) from the operation of the Ludington Pumped-Storage plant.

[5]The current consumer price index 1.348 we use here is the 1991 (February) price relative to the base years 1982-84. This information is obtained from the "Consumer Price Index for All Urban Consumer (CPI-U)" table in the *Summary Data from the Consumer Price Index News Release February 1991,* published by the Bureau of labor statistics, US Department of Labor.

Comparison With Other Estimates

Previous recreation demand studies have generated consumer surplus measures corresponding to various site condition changes. For example, Bockstael et al. (1988) estimate that the consumer surplus per choice occasion associated with a 20% increase in the game fish catch rates ranges from $.32 to $1.56, depending on species affected, in their Florida (Atlantic coast) sport fishery study. For a 25% increase in fish catch rates on Aibemarle and Pamlico Estuaries, the nested logit estimation of Smith and Palmquist (1988) yields an angler welfare change of $2.43 per trip when all sites are affected, or $ .60 when only the closest four sites are affected. To calculate. a consumer surplus per choice occasion from the Michigan recreational fishing study that is comparable in definition to other estimates, we proceed as follows.[6]

First, under the competing risks framework, an angler has a certain probability of taking a trip of each type on any choice occasion. The compensating variations reported in table IX.1 are conditional compensating variations per choice occasion ($CCOCV_d$), for specific trip types *d* being chosen, (conditional upon participation). To calculate the average conditional compensating variation per trip (*CCOCV*), we have to weight each $CCOCV_d$ by its corresponding probability. Here we use the angler distribution in the sample of the participants (N=4824) as an approximation of the probabilities.

---

[6]We cannot compare seasonal consumer surplus because people living in different geographical locations are likely to have different participation rates due to different fishing opportunities. Therefore, consumer surplus per choice occasion, conditional upon participation, is the only measure that we can reasonably compare across studies. We expect this measure to vary across contexts, because it is influenced by the distance between the population of anglers and the fishing sites in consideration and by quality levels.

Trip Type	Probability	$CCOCV_d$	Prob x $CCOCV_d$
Day	2463/4824=51%	.04	.02
Wkn	1159/4824=24%	.17	.04
Vac	1202/4824=25%	.24	.06
Sum			**.12**

The probability that an individual will choose GLcd or Anad, the two product lines most affected by the Ludington scenario, is approximated by dividing the number of anglers in GLcd (N=769) and Anad (N=299) samples by the number of total participants (N=4824). The compensating variation per choice occasion of an angler targeting GLcd or Anad is, therefore.

$$\$.12 \div \left(\frac{769 + 299}{4824}\right) = \$.542$$

Since only half of the 41 Great Lakes counties are affected, we have \$.542 x 2 = \$1.084 (in 1984\$) as the projected average CCOCV of a GLcd or Anad angler if all Great Lakes sites were affected. This is the CCOCV corresponding to a roughly 10% increase in all salmonid catch rates (10% for Chinook, 3.3% for Coho, 13.7% for Lake Trout, and 8.6% for Rainbow). The CCOCV for a 20% catch rate increase in 1991\$ will thus be about

$$\$1.084 \times \frac{20\%}{10\%} \times 1.348 = \$2.922$$

This estimate is of the same order of magnitude as those obtained by other researchers.

Map IX.1: Michigan counties affected by the Ludington scenario

Table IX.1: Ludington: Mean compensating variation per trip in 1984 dollars

	N	Day Trip	Wkn Trip	Vac Trip
Day Sample	2463	0.0418	0.1657	0.2417
Wkn Sample	1159	0.0359	0.1666	0.2437
Vac Sample	1202	0.0306	0.1707	0.2427

Table IX.2: Ludington: Total trips per person with plant operation

	N	Day Trip	Wkn Trip	Vac Trip
Day Sample	2463	1.2513	0.5502	0.4423
Wkn Sample	1159	1.1506	0.5675	0.5171
Vac Sample	1202	0.9044	0.4960	0.6831
Total	**4824**	**5502.47**	**2609.01**	**2509.78**

Table IX.3 Ludington: Mean change in season trips

	N	Day Trip	Wkn Trip	Vac Trip
Day Sample	2463	0.0074	0.0013	-0.0014
Wkn Sample	1159	0.0067	0.0013	-0.0017
Vac Sample	1202	0.0053	0.0012	-0.0024
Total	**4824**	**32.34**	**6.10**	**–8.39**

Table IX.4: Ludington: Mean season compensating variation in 1984 dollars

	N	Day Trip	Wkn Trip	Vac Trip
Day Sample	2463	0.1771	0.1407	0.1062
Wkn Sample	1159	0.1328	0.1402	0.1251
Vac Sample	1202	0.0857	0.1150	0.1603
Total	**4824**	**693.179**	**647.25**	**599.28**

CHAPTER X

POLICY APPLICATION: KALAMAZOO RIVER CONTAMINATION

Biological Scenarios

The Kalamazoo river, located in the southwestern portion of the lower Peninsula of Michigan, flows in a westerly direction and discharges into Lake Michigan. High levels of PCBs contaminate approximately 80 miles of the river upstream from Lake Michigan, affecting the biota (particularly fish),water and sediment. The site, listed on the Superfund National Priorities List, is identified as the third worst contamination site in Michigan. Evidence suggests that contaminated sediments in natural depositional areas and behind both drawn-down and operating hydroelectric dams[2] are continuing sources of PCBs to the water column and to fish. A fish consumption advisory is in place for the stretch of the river with upstream mobility. The International Joint Commission has identified the Kalamazoo rive: as one of 14 *Areas of Concern* in Michigan.

The Michigan Department of Natural Resources has proposed a multi-action management plan for the Kalamazoo River. This plan includes passing anadromous fish

[1] The description of AOC below is based on the 1989 Report on Great Lakes Water Quality, Appendix A, by the Great Lakes Quality Board of the International Joint Commission.

[2] An estimated 104,000 kg of PCBs reside in the sediments.

over several dams, rehabilitating the resident fish community in a large reach of the river, and reducing problems of chemical contamination (mostly PCB's) in the River. Because the fishery management actions will take place if and only if the PCB cleanup occurs, the benefits of the plan should be evaluated as a single policy option.

Baseline:

The baseline for the policy scenario is the current situation, defined by the base data with which the discrete choice model is estimated.

Scenario: PCB Cleanup

The scenario is designed to capture the expected results from implementation of the Kalamazoo River Remedial Action plan.

Contamination:

Cleanup of the PCB contaminated sediments in the river will eliminate the designated Areas Of Concern in Allegan (3) and Kalamazoo (39) counties.[3] In addition, fish contamination advisories can be eliminated on warmwater river fisheries in both of these counties. Fish contamination advisories are expected to remain in effect on Great Lakes and anadromous fisheries in these counties since the contaminants in these fish are accumulated during life in Lake Michigan. Containment of contaminants in the Kalamazoo River will reduce discharge of these contaminants into Lake Michigan but the reduction will be only a marginal change in total loading on Lake Michigan.

---

[3] This will potentially affect all product lines.

Product Line	County	Variable	[Baseline]	[Policy]
All	Allegan	AOC	1	0
	Kalamazoo	AOC	1	0
ISww	Allegan	CntmSW	55	0
	Kalamazoo	CntmSW	15	0
ILww	Allegan	CntmLW	1200	0

Anadromous Product Line: Catch Rates

Containment of contaminated sediments will permit removal of three state-owned dams from the Kalamazoo River. Construction of fish ladders on remaining dams would open 44 miles of river to anadromous trout and salmon fishing, with 18 miles in Allegan county and 26 miles in Kalamazoo county. Reservoirs in both counties would support inland lake fishing for anadromous trout and salmon. Catch of anadromous trout and salmon rates in Allegan county should increase modestly, perhaps 20% for each species. Catch rates of anadromous fish in Kalamazoo county (currently non-existent) should compare to these increased catch rates in Allegan county as follows:

Product Line	County	Species	Month	CR = Allegan CR x
Anad	Kalamazoo	Chinook	September	0.25
		Chinook	October	0.90
		Coho	September	1.00
		Coho	October	1.00
		Rainbow	April	2.00
		Rainbow	May	1.50
		Rainbow	September	1.50
		Rainbow	October	2.00

Other Product Lines: Quantity of Fishing Resources

Rehabilitation of the warmwater fish community in the Kalamazoo River, combined with PCB containment and dam removal should convert 34 miles of second quality, mainstream, warmwater river to top quality, mainstream, warmwater river. Of these 34 miles, 18 miles are in Allegan county and 16 miles are in Kalamazoo county. In addition, 10 miles of a second quality, warmwater tributary in Allegan county would be converted to a second quality trout tributary.

The product lines and variable affected are shown below:

Product Line	County	Variable	Change in value
ISww	Allegan	ISww1main	+18
ISww	Allegan	ISww2main	–18
ISww	Kalamazoo	ISww1main	+16
ISww	Kalamazoo	ISww2main	–16
ISww	Allegan	ISww2trib	–10
IScd	Allegan	IScd2trib	+10

The two Michigan counties affected by the Kalamazoo river cleanup plan are shown in map X.1.

Consumer Surplus Calculation

In this section, we carry out similar calculations as we do for the Ludington case to estimate people's willingness-to-pay for the cleanup plan of the Kalamazoo river contamination. The compensating variation for the open-water fishing season according to formula V.28 in chapter V is still computed as

$$W = \sum_i \sum_m \sum_d \left[\frac{T^1_{imd} \cdot \bar{I}^1_{imd} - T^0_{imd} \cdot \bar{I}^0_{imd}}{\bar{\eta}_d / 100} \right]$$

where

i indexes individuals in the sample of our consumer surplus analysis.

m indexes months (April — October) in an open-water season,

d indexes trip durations (= Day, Wkn, Vac).

0 refers to the "before cleanup" case.

1 refers to the "after cleanup" case.

$\tilde{\eta}_d$ is the weighted MUI per \$100, for trip duration type d.

T is the number of total trips in the season.

$\bar{I}$ is the pseudo-IV defined in chapter III.

Table X.1 presents the conditional compensating variation per trip $(\bar{I}^1_{id} - \bar{I}^0_{id}) \times 100/\tilde{\eta}_d$ in 1984 dollars (averaged over the seven open-water fishing months) associated with the Kalamazoo river cleanup. The expected increase in value per trip is larger than that of the Ludington case. Table X.2 reports the predicted number of season trips T^0 without the cleanup using the exponential model estimates. Tables X.3 and X.4 report the predicted change in total trips $(T^1_{id} - T^0_{id})$ and the total compensating variation (W_d) in 1984 dollars for one open-water season if the cleanup plan is implemented. Again, we predict that more day and weekend trips and fewer vacation trips will be taken as a result of the cleanup. The total seasonal compensating variation for the sample is calculated to be W = \$2920.63 (in 1984\$) from the subtotals in table X.4.

We then extrapolate the sample CV to the population similarly as

$$\begin{aligned} W^{**} &= \text{CPI}\left(\frac{91}{84}\right) \times \frac{P}{N} \times \frac{N}{S} \times W \\ &= 1.348 \times \frac{1414914}{10948} \times \frac{10948}{4824} \times 2920.63 \\ &= \$1{,}153{,}699.41 \end{aligned}$$

where P = 1,414,914 is the total population of licensed anglers in 1984. N = 10,948 is the sample size of the MDNR data, and S = 4824 is the number of people in our consumer surplus sample. N/S is the factor for extrapolating from the consumer surplus sample to the MDNR sample. P/N is the factor for extrapolating from the MDNR sample to the total population of licensed anglers.

Therefore, the final extrapolation from the sample to the population of licensed anglers yields an annual consumer surplus of \$1.15 million (in 1991\$) from the implementation of the Kalamazoo river PCB cleanup plan. Because no other studies have been conducted for site quality changes of this nature in the past, we have no outside estimates against which to compare these numbers.

Map X.1: Michigan counties affected by the Kalamazoo scenario

Table X.1: Kalamazoo: Mean compensating variation per trip in 1984 dollars

	N	Day Trip	Wkn Trip	Vac Trip
Day Sample	2463	0.1048	0.3093	0.1999
Wkn Sample	1159	0.1058	0.3245	0.2103
Vac Sample	1202	0.1115	0.3680	0.2270

Table X.2: Kalamazoo: Total trips per person before PCB cleanup

	N	Day Trip	Wkn Trip	Vac Trip
Day Sample	2463	1.2513	0.5502	0.4423
Wkn Sample	1159	1.1506	0.5675	0.5171
Vac Sample	1202	0.9044	0.4960	0.6831
Total	**4824**	**5502.47**	**2609.01**	**2509.78**

Table X.3: Kalamazoo: Mean change in season trips

	N	Day Trip	Wkn Trip	Vac Trip
Day Sample	2463	0.0118	0.0023	–0.0029
Wkn Sample	1159	0.0112	0.0024	–0.0037
Vac Sample	1202	0.0100	0.0024	–0.0054
Total	**4824**	**54.10**	**11.34**	**–17.94**

Table X.4: Kalamazoo: Mean season compensating variation in 1984 dollars

	N	Day Trip	Wkn Trip	Vac Trip
Day Sample	2463	0.2924	0.2436	0.0942
Wkn Sample	1159	0.2310	0.2523	0.1172
Vac Sample	1202	0.1705	0.2272	0.1615
Total	**4824**	**1192.99**	**1165.58**	**562.06**

APPENDIX

Sensitivity Analysis of Trip Time Costs

We perform a sensitivity analysis to the alternative treatments of travel time discussed above in Chapter III for the Great Lakes coldwater product line. To make the estimates comparable, we have to restrict the sample sizes to be the same across runs. The number of anglers in the samples are the same, though the choice sets for each of the anglers are different under each hypothesis. Therefore, the difference in the estimates will come from the different definitions of the choice set and the different definitions of the travel cost to a site.

We estimate the three models derived in Chapter III. The sample without missing data for the exogenous trip days model is larger than for the other two models, because it does not require use of the variable measuring trip hours, which has numerous missing values. To separate out the effect of the different samples, we estimate that model twice: once for the restricted sample used for the other two models and once for its full sample.

1. The exogenous on-site time model *(SiteTime).*

2. The exogenous trip time model *(TrpTime).*

3. The exogenous trip duration in days, using a sample defined by the above models *(TrpDays-Subset).*

4. The exogenous trip duration in days, using the sample defined by its own time constraints *(TrpDays-Full).*

The estimates for the three trip durations are presented in the tables in this Appendix. The travel time cost variable is only included in the site choice portion of the NMNL model for the SiteTime model. For the other treatments of travel time, the travel time cost becomes part of the total cost of choosing a trip duration, and is included (along with on-site time costs) in the WageCost variable in the Participation model. Due to the correlation between the distance cost variable and the travel time cost variable, the estimated marginal utility of income, (the parameter of the distance cost variable), is much smaller for the SiteTime version than for the other three models.

The parameter estimates are, in general, quite different across the four models. To compare across the specifications the contribution of each quality attribute to angler value during the choice occasion, we translate the effects into monetary terms by dividing by the MUI. See the bottom of these tables for the calculations. The contributions of most quality attributes increase in monetary terms as the trip length increases. The exogenous SiteTime model predicts higher (in absolute value terms) contributions from the quality attributes than the other models. partly because its MUI is smaller.

Since the exogenous on-site time hypothesis is theoretically flawed and the trip days model may have substantial measurement error, we use the exogenous Trip Time model for our NMNL analysis.

Table A.1 MNL estimates for the GLcd-Day sample

Variable	SiteTime	TrpTime	TrpDays (Subset)	TrpDays (Full)
Dist$/100	-14.51	-17.27	-18.28	-16.01
	(-12.04)	(-16.42)	(-17.76)	(-19.11)
Time$/100	-2.33	N.A.	N.A.	N.A.
	(-3.77)			
AOC	-1.58	-1.53	-1.55	-1.53
	(-8.82)	(-8.53)	(-8.61)	(-9.35)
%Forest	2.87	2.34	2.22	1.69
	(4.89)	(4.12)	(3.93)	(3.42)
Feature	0.09	0.08	0.08	0.26
	(0.41)	(0.39)	(0.38)	(1.39)
Chinook Salmon	9.10	8.36	8.80	10.59
	(4.17)	(3.88)	(4.20)	(5.74)
Coho Salmon	4.07	3.87	4.08	3.41
	(1.96)	(1.86)	(2.00)	(1.90)
Lake Trout	3.70	3.32	3.48	4.28
	(1.81)	(1.67)	(1.77)	(2.47)
Rainbow Trout	1.75	2.19	1.80	1.83
	(0.35)	(0.42)	(0.36)	(0.41)
Log Likelihood	-509.1	-518.5	-528.9	-657.2
χ^2-test	943.7	727.3	1263.8	1355.9
%Choices Right	50.6	50.6	50.3	51.2
#People	336	336	336	387
#Choices	7012	5565	10743	12326
$(\partial V/\partial AOC)/MUI$	-0.11	-0.09	-0.09	-0.10
$(\partial V/\partial Forest)/MUI$	0.20	0.14	0.12	0.11
$(\partial V/\partial Feature)/MUI$	0.01	0.01	0.00	0.02
$(\partial V/\partial Chinook)/MUI$	0.63	0.48	0.48	0.66
$(\partial V/\partial Coho)/MUI$	0.28	0.22	0.22	0.21
$(\partial V/\partial LakeT)/MUI$	0.26	0.19	0.19	0.27
$(\partial V/\partial RainbowT)/MUI$	0.12	0.13	0.10	0.11

Note: Numbers in parentheses are *t*- statistics.

Table A.2 MNL estimates for the GLcd-Wkn sample

Variable	SiteTime	TrpTime	TrpDays (Subset)	TrpDays (Full)
Dist$/100	-2.64	-4.20	-4.27	-4.51
	(-5.33)	(-10.77)	(-10.95)	(-11.88)
Time$/100	-0.75	N.A.	N.A.	N.A.
	(-2.93)			
AOC	-1.67	-1.75	-1.76	-1.70
	(-7.76)	(-8.07)	(-8.11)	(-8.34)
%Forest	1.80	1.23	1.19	1.24
	(4.34)	(3.14)	(3.04)	(3.30)
Feature	0.51	0.51	0.51	0.57
	(3.10)	(3.15)	(3.18)	(3.71)
Chinook Salmon	9.09	8.93	8.99	10.02
	(5.49)	(5.66)	(5.71)	(6.75)
Coho Salmon	6.33	5.37	5.24	5.99
	(3.60)	(3.34)	(3.27)	(3.98)
Lake Trout	0.33	-1.27	-1.58	-0.86
	(0.12)	(-0.49)	(-0.61)	(-0.35)
Rainbow Trout	2.59	2.47	2.06	4.22
	(0.73)	(0.70)	(0.58)	(1.34)
Log Likelihood	-740.9	-795.7	-800.8	-878.6
χ^2-test	229.1	321.3	341.5	393.4
%Choices Right	16.4	15.3	14.5	15.5
#People	262	262	262	290
#Choices	7638	10201	10690	11828
$(\partial V/\partial AOC)/MUI$	-0.63	-0.42	-0.41	-0.38
$(\partial V/\partial Forest)/MUI$	0.68	0.29	0.28	0.28
$(\partial V/\partial Feature)/MUI$	0.19	0.12	0.12	0.13
$(\partial V/\partial Chinook)/MUI$	3.44	2.13	2.11	2.22
$(\partial V/\partial Coho)/MUI$	2.40	1.28	1.23	1.33
$(\partial V/\partial LakeT)/MUI$	0.13	-0.30	-0.37	-0.19
$(\partial V/\partial RainbowT)/MUI$	0.98	0.59	0.48	0.94

Note: Numbers in parentheses are *t*- statistics.

Table A.3: MNL estimates for the GLcd-Vac sample

Variable	SiteTime	TrpTime	TrpDays (Subset)	TrpDays (Full)
Dist$/100	-1.61	-2.41	-2.41	-2.41
	(-4.17)	(-8.67)	(-8.67)	(-9.30)
Time$/100	0.21	N.A.	N.A.	N.A.
	(0.93)			
AOC	-0.86	-1.03	-1.03	-1.05
	(-3.72)	(-4.48)	(-4.48)	(-4.94)
%Forest	2.33	2.22	2.22	2.05
	(4.79)	(4.73)	(4.73)	(4.78)
Feature	0.63	0.62	0.62	0.73
	(4.79)	(3.79)	(3.79)	(4.89)
Chinook Salmon	8.87	9.73	9.73	9.83
	(5.12)	(5.69)	(5.69)	(6.16)
Coho Salmon	5.39	4.50	4.50	4.52
	(3.19)	(2.77)	(2.77)	(2.98)
Lake Trout	4.84	4.00	4.00	3.68
	(3.84)	(3.50)	(3.50)	(3.37)
Rainbow Trout	2.09	2.77	2.77	2.01
	(0.52)	(0.71)	(0.71)	(0.52)
Log Likelihood	-589.7	-640.8	-640.8	-748.1
χ^2-test	153.7	203.0	203.0	240.8
%Choices Right	14.5	14.0	14.0	14.5
#People	200	200	200	234
#Choices	5935	8185	8185	9574
$(\partial V/\partial AOC)/MUI$	-0.53	-0.43	-0.43	-0.44
$(\partial V/\partial Forest)/MUI$	1.45	0.92	0.92	0.85
$(\partial V/\partial Feature)/MUI$	0.39	0.26	0.267	0.30
$(\partial V/\partial Chinook)/MUI$	5.51	4.04	4.04	4.08
$(\partial V/\partial Coho)/MUI$	3.35	1.87	1.87	1.88
$(\partial V/\partial LakeT)/MUI$	3.01	1.66	1.66	1.53
$(\partial V/\partial RainbowT)/MUI$	1.30	1.15	1.15	0.83

Note: Numbers in parentheses are *t*- statistics.

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] Takeshi Amemiya. Qualitative response models: A survey. *Journal of Economic Literature*, 19:1483-1536, 1981.

[2] Takeshi Amemiya. *Advanced Econometrics*. Cambridge: Harvard University Press, 1985.

[3] Moshe Ben-Akiva and Steven R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge: MIT Press, 1985.

[4] Clark S. Binkley and W. Michael Hanemann. The recreation benefits of water quality improvement: Analysis of day trips in an urban setting. Washington, D. C.: Environmental Protection Agency. EPA-600 /5-78-010, 1978.

[5] Nancy E. Bockstael, W. Michael Hanemann, and Catherine L. Kling. Modeling recreational demand in a multiple site framework. *Water Resources Research*, 23(5):951-60, May 1987.

[6] Nancy E. Bockstael, W. Michael Hanemann, and Ivar E. Strand, Jr. Time and the recreational demand model. *American Journal of Agricultural Economics*, 69:293-302, May 1987.

[7] Nancy E. Bockstael, W. Michael Hanemann, and Ivar E. Strand, Jr. Measuring the benefits of water quality improvements using recreation demand models, Volume II. Prepared for US EPA under Cooperative Agreement No. CR-81143-01-1, 1986.

[8] Nancy E. Bockstael, Kenneth E. McConnell, and Ivar E. Strand, Jr. A random utility model for the Middle and South Atlantic sport fishery: Preliminary results. Presented at the 4th Annual AERE Workshop on Marine Recreational Fishing, Seattle, Washington, June 1988.

[9] Nancy E. Bockstael, Kenneth E. McConnell, and Ivar E. Strand, Jr. Recreation. In John B. Braden and Charles D. Kolstad, editors, *Measuring the Demand for Environmental Quality*, chapter 8. New York: North-Holland, January 1991.

[10] Nicholas S. Cardell and Danny Steinberg. A Gauss-Newton FIML estimator for the nested multinominal logit model. Paper presented at the Annual Meeting of the American Statistical Association, New Orleans, August 1988.

[11] Richard T. Carson, W. Michael Hanemann, Russell Gum, and Robert Mitchell. Random utility model of the Alaska sport fishery. Draft manuscript, 1987.

[12] Peter P. Caulkins. An empirical study of the recreational benefits generated by a water quality improvement. PhD dissertation. University of Wisconsin-Madison, 1982.

[13] Peter P. Caulkins, Richard C. Bishop, and Nicolaas W. Bouwes. The travel cost model for lake recreation: A comparison of two methods for incorporating site quality and substitution effects. *American Journal of Agricultural Economics,* 68(2):291-97, 1986.

[14] Frank J. Cesario. Value of time in recreation benefit studies. *Land Economics,* 52:32-41, 1976.

[15] Frank J. Cesario and Jack L. Knetsch. Time bias in recreation benefit estimates. *Water Resources Research*, 6:700-704, 1970.

[16] Marion Clawson and Jack L. Knetsch. *Economics of Outdoor Recreation*. Washington, DC.: Resources for the Future, 1966.

[17] Angus Deaton and John Muellbauer. *Economics and Consumer Behavior.* New York: Cambridge University Press, 1983.

[18] Daniel Feenberg and Edwin S. Mills. *Measuring the Benefits of Water Pollution Abatement.* New York: Academic Press, 1980.

[19] William H. Greene. *Limdep Version 5.1 Documentaton*. Econometric Software, Inc., March 1989.

[20] W. Michael Hanemann. A methodological and empirical study of the recreation benefits from water quality improvement. PhD dissertation, Department of Economics, Harvard University, 1978.

[21] W. Michael Hanemann. Applied welfare analysis with qualitative response models. Working Paper No. 241, Giannini Foundation of Agricultural Economics, University of California, October 1982.

[22] W. Michael Hanemann. Marginal welfare measures for discrete choice models. *Economic Letters,* 13:129-36, 1983.

[23] W. Michael Hanemann. Discrete/continuous models of consumer demand. *Econometrica,* 52(3):541-61, May 1984.

[24] W. Michael Hanemann. Applied welfare analysis with discrete choice models. Working paper, Department of Agricultural and Resource Economics, University of California, Berkeley, March 1985.

[25] David A. Hensher and Lester W. Johnson. *Applied Discrete Choice Modelling.* London: Croom Helm, 1981.

[26] Carol A. Jones. Valuing nonmarket goods: Contingent behavior and contingent valuation measures of the benefits of recreational fishing. School of Natural Resources and Department of Economics, University of Michigan, October 1988.

[27] Carol A. Jones, Theodore Graham-Tomasi, Yuc-Sheng Sung, and Anne Wittenberg. The economic value of damages to Lake Michigan fish populations due to the Ludington pump-storage plant. Report to the Michigan Department of Natural Resources, June 1988.

[28] Carol A. Jones, Douglas B. Jester, Theodore Graham-Tomasi, and Yuc-Sheng Sung. Valuation of changes in the quality of Lake Michigan recreational fisheries. Working paper, October 1989.

[29] Carol A. Jones and Yuc-Sheng Sung. Use of discrete choice models to value natural resource damages in recreational fisheries. Paper presented at the AAEA/AERE 1990 Annual Meeting, University of British Columbia, Vancouver, August 1990.

[30] Mary Jo Kealy and Richard C. Bishop. Theoretical and empirical specification issues in travel cost demand studies. *American Journal of Agricultural Economics,* 68:660-67, August 1986.

[31] Nicholas M. Kiefer. Economic duration data and hazard functions. *Journal of Economic Literature,* 26:646-79, June 1988.

[32] Hideo Kikuchi. Segmenting Michigan's sport fishing market: Evaluation of two approaches. Michigan State University PhD dissertation, Department of Parks and Recreation Resources, 1986.

[33] Catherine L. Kling. Measuring the recreational benefits of environmental amenities using multiple site models: An evaluation of techniques. University of Maryland PhD dissertation, 1986.

[34] Catherine L. Kling. Comparing welfare estimates of environmental quality changes from recreation demand models. *Journal of Environmental Economics and Management*, 15:331-40, September 1988.

[35] Tony Lancaster. *The Econometric Analysis of Transition Data.* Econometric Society Monographs No. 17. New York: Cambridge University Press, 1990.

[36] Elisa T. Lee. *Statistical Methods for Survival Data Analysis.* Belmont, California: Lifetime Learning Publications, 1980.

[37] R. D. Luce. *Individual Choice Behavior:* A Theoretical Analysis. New York: John Wiley, 1959.

[38] G. S. Maddala. *Limited-Dependent and Qualitative Variables in Econometrics.* Econometric Society Monographs No. 3. New York: Cambridge University Press, 1983.

[39] Karl- Goran Maler. *Environmental Economics: A Theoretical Inquiry.* Johns Hopkins Press for Resources for the Future, 1974.

[40] Kenneth E, McConnell. Revisiting the problem of on-site time in the demand for recreation. Draft paper, Department of Agricultural and Resource Economics, University of Maryland, June 1990.

[41] Kenneth E. McConnell, Lynne Blake-Hedges, and Ivar E. Strand, Jr. Modelling catch rates in the RUM framework. Paper presented at the 1990 Annual Meeting of American Agricultural Economics Association, Vancouver, British Columbia, August 1990.

[42] Daniel McFadden. Conditional logit analysis of qualitative choice behavior. In Paul Zarembka, editor, *Frontiers in Econometrics.* New York: Academic Press, 1974.

[43] Daniel McFadden. Quantal choice analysis: A survey. *Annals of Economic and Social Measurement,* 5:363-90, 1976.

[44] Daniel McFadden. Modelling the choice of residential location. In Anders Karlqvist, Lars Lundqvist, Folke Snickars, and Jorgen W. Weibull, editors, *Spatial Interaction Theory and Planning Models,* chapter 3, pages 75-96. Amsterdam: North-Holland, 1978.

[45] Daniel McFadden. Econometric models of probabilistic choice. In Charles F. Manski and Daniel McFadden, editors, *Structural Analysis of Discrete Data with Econometric Applications.* Cambridge: MIT Press, 1981.

[46] Daniel McFadden. Qualitative response models. In Werner Hildenbrand, editor, *Advances in Econometrics.* New York: Cambridge University Press, 1982.

[47] Daniel McFadden. Econometric analysis of qualitative response models. In Zvi Griliches and Michael D. Intriligator, editors, *Handbook of Econometrics,* volume 2, Amsterdam: North-Holland, 1984,

[48] Edward R. Morey. Confuser surplus. *American Economic Review,* 74(1):163–73, 1984.

[49] Edward R. Morey. Derivation of expected consumer's surplus from the nested-logit model of consumer choice. Department of Economics, University of Colorado, Boulder, February 1989.

[50] W. Douglass Shaw Morey, Edward R. and Robert D. Rowe. A discrete choice model of recreational participation, site choice, and activity valuation when complete trip data are not available. *Journal of Environmental Economics and Management,* 20:181-201, 1991.

[51] George R. Parsons. The participation decision in random utility models of recreation choice. Draft manuscript, 1990.

[52] George R. Parsons and Mary Jo Kealy. Measuring water quality benefits using a random utility model of lake recreation in Wisconsin. Draft manuscript, January 1990.

[53] Gerald P. Rakoczy and Roger N. Lockwood. Sportfishing catch and effort from the michigan waters of lake michigan and their important tributary streams, january 1, 1985 - march 31, 1986. Technical report, Michigan Department of Natural Resources Fisheries Division, December 1988. Fisheries Technical Report No. 88-11a.

[54] Gerald P. Rakoczy and Richard D. Rogers. Sportfishing catch and effort from the michigan waters of lake michigan, huron, and erie, and their important tributary streams, april 1, 1986 – march 31, 1987. Technical report, Michigan Department of Natural Resources Fisheries Division, November 1987. Fisheries Technical Report No. 87-6a.

[55] Sheldon M. Ross. *Stochastic Processes.* Wiley Series in Probability and Mathematical Statistics. New York: John Wiley, 1983.

[56] Robert D. Rowe, Edward R. Morey, Arthur D. Ross, and W. Douglass Shaw. Valuing marine recreation fishing on the Pacific coast. Energy and Resource Consultants, report prepared for National Oceanic and Atmospheric Administration, 1985.

[57] Richard Schmalensee and Paul L. Joskow. Estimated parameters as independent variables: An application to the costs of electric generating units. *Journal of Econometrics,* 31:275-305, 1986.

[58] Kenneth A. Small and Harvey S. Rosen. Applied welfare economics with discrete choice models. *Econometrica,* 49:105–30, January 1981.

[59] V. Kerry Smith, William H. Desvousges, and Matthew P. McGivney. The opportunity cost of travel time in recreation demand models. *Land Economics,* 59(3):259–78, 1983.

[60] V. Kerry Smith and Yoshiaki Kaoru. Black mayonnaise and marine recreation: Methodological issues in valuing a cleanup. Resources for the Future Quality of the Environment Division discussion paper QE91-02, 1990.

[61] V. Kerry Smith and Raymond J. Kopp. The spatial limits of the travel cost recreational demand model. *Land Economics,* 56(1):64–71, February 1980.

[62] V. Kerry Smith and Raymond B. Palmquist, The value of recreational fishing on the Albemarle and Pamlico estuaries. Prepared for the Office of Policy Planning and Evaluation, US Environmental Protection Agency under cooperative agreement CX814569-01, 1988.

[63] Howard M. Taylor and Samuel Karlin. *An Introduction to Stochastic Modeling.* New York: Academic Press, 1984.

[64] Kenneth Train. *Qualitative Choice Analysis: Theory, Econometrics, and an Application to Automobile Demand.* Cambridge: MIT Press, 1986.

[65] Hal R. Varian. *Microeconomic Analysis.* W. W. Norton and Company, Inc., 2nd edition, 1984.

MDNR ANGLER SURVEY QUESTIONNAIRE

1983 and 1984

MICHIGAN SPORT FISHING SURVEY

Dear Angler:

Each year the Department of Natural Resources (DNR) must gather information on recreational fishing in Michigan. One of the best methods is to obtain information directly from the angler. This information will be used to improve fishing opportunities and document the importance of fishing to the state's economy.

Your name has been selected at random from fishing license records. Would you please take a few minutes to answer all the questions. A prompt return of your questionnaire in the postpaid return envelope will be appreciated.

Questionnaires are being sent to a number of anglers but there can be no substitute for the information you, yourself, provide. Your response is needed even if you did not fish or did not catch anything. Be assured that your reply is confidential and will be used only for better management of Michigan's fish resources.

Thank you for your cooperation.

Sincerely,

John A. Scott
Chief, Fisheries Division

---

1 a. Where is your permanent residence? County ________ State ________ Zip Code ________

b. How long have you lived there? ______years. c. How long have you lived in Michigan? ______years.

2. Are you married? ☐ Yes (go to question 2a) ☐ No (go to question 2b)

   2a. Does your spouse fish? ☐ Yes ☐ No

   2b. Do you have any children age 16 or younger? ☐ Yes ☐ No (go to question 3)

   2c. Please indicate their ages and whether or not they fish:

Ages	Male	Female	Do they fish? Yes	No
____	☐	☐	☐	☐
____	☐	☐	☐	☐
____	☐	☐	☐	☐
____	☐	☐	☐	☐

3. Please indicate when you work:

   ☐ Full-Time Days ☐ Full-Time Nights ☐ Part-Time Days ☐ Part-Time Nights ☐ Retired ☐ Unemployed ☐ Student

4. How long have you been fishing? ______ years. How long have you fished in Michigan? ______ years.

5. How do you rate yourself as an angler? ☐ Beginner ☐ Somewhat experienced ☐ Experienced ☐ Expert

6. Did you fish in any other state or foreign country last year? ☐ Yes ☐ No

   If yes, where? ________, ________, ________

   and for what species? (e.g. trout)________, ________, ________.

7. Please check one box indicating with whom you fish most often:

   ☐ Alone ☐ Spouse ☐ Son(s) ☐ Daughter(s) ☐ Other Relatives ☐ Friends

8. Do you own a boat(s) or canoe(s) used for fishing in Michigan? ☐ No ☐ Yes Please complete table below

		Length In Feet	Total Days Used Per Year	Days Per Year For Fishing
Boats	#1			
	#2			
Canoes	#1			
	#2			

9 Over the last two years, we would like to know (1) what species did you fish; (2) where did you fish for these species, (3) modes of fishing, eg shore, pier, and (4) fishing method(s) you used. If you fished in more than one location or used more than one method, check all appropriate boxes (see example) Please answer only for the last two years.

Did you fish for these species over the last two years?			Location of Fishing			Mode of Fishing					Fishing Method							
Fish	Yes	No	Inland Lakes	Great Lakes	Stream/ River	Shore or Wading	Pier or Dock	Rental or Private Boat	Charter Boat	Ice Fishing	Casting	Spin or Spincasting	Bait Fishing	Trolling	Fly Fishing	Spearing	Dipping	Snagging
(EXAMPLE) Bass	☑	☐	☑	☑	☐	☑	☐	☑	☐	☐	☐	☐	☑	☐	☐	☐	☐	☐
Yellow Perch	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Panfish	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Bass	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Walleye or Sauger	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Pike or Musky	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Lake Trout	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Steelhead	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Rainbow Trout	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Brown Trout	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Brook Trout	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Chinook Salmon	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Coho Salmon	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Catfish or Bullhead	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Suckers or Carp	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Smelt	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

10 Which one of the above species do you most prefer to catch? _____ (name one only) Which one do you most prefer to eat? _____ (name one only)

11 Please name the one mode (listed above) of fishing you most prefer _____ (name one only)

12 Please name the one method (listed above) of fishing you most prefer _____ (name one only)

CONTINUE

15. In orde[illegible] improve fishing opportunities, we need to know what factors are important to you in selecting where and when to fish. Please check one box indicating the importance you place on the factors shown in the table below:

	Crucial	Very Important	Important	Somewhat Important	Not Important
Angler crowding	☐	☐	☐	☐	☐
Competition with other recreationists, e.g., canoes, sailboats	☐	☐	☐	☐	☐
Places to fish from shore	☐	☐	☐	☐	☐
Boat launching facilities	☐	☐	☐	☐	☐
Marina facilities and services	☐	☐	☐	☐	☐
Availability of parking facilities	☐	☐	☐	☐	☐
Nearness of restaurants	☐	☐	☐	☐	☐
Nearness of bait and tackle shops	☐	☐	☐	☐	☐
Nearness of overnight accommodations, e.g., motels, campgrounds	☐	☐	☐	☐	☐
Natural beauty of the area	☐	☐	☐	☐	☐
Solitude	☐	☐	☐	☐	☐
Water clarity	☐	☐	☐	☐	☐
Presence of contaminants in fish	☐	☐	☐	☐	☐
Catch rate of keepable fish	☐	☐	☐	☐	☐
Catch rate of all fish	☐	☐	☐	☐	☐
Presence of favorite fish (species)	☐	☐	☐	☐	☐
Size of fish	☐	☐	☐	☐	☐
Diversity of fish species which can be caught	☐	☐	☐	☐	☐
Nearness to home (travel distance)	☐	☐	☐	☐	☐
Information about the area, e.g. catch rates, best fishing methods, hot spots	☐	☐	☐	☐	☐
Nearness to second home/cottage camp	☐	☐	☐	☐	☐

We would also like to know some of the reasons why you fish. Please ind[illegible] the importance of the following reasons. Please check the box indicating the imp[illegible]nce you place on each reason

	Crucial	Very Important	Important	Somewhat Important	Not Important
To catch fish to eat	☐	☐	☐	☐	☐
For relaxation	☐	☐	☐	☐	☐
For companionship	☐	☐	☐	☐	☐
To enjoy nature	☐	☐	☐	☐	☐
For the challenge and excitement	☐	☐	☐	☐	☐
To be alone	☐	☐	☐	☐	☐
To improve my fishing skill	☐	☐	☐	☐	☐
To get away	☐	☐	☐	☐	☐
For exercise	☐	☐	☐	☐	☐
Family togetherness	☐	☐	☐	☐	☐
To catch a trophy fish	☐	☐	☐	☐	☐
For a sense of achievement	☐	☐	☐	☐	☐

17. What sources of information do you use in selecting where and when to fish?

	Often	Occasionally	Never
Comments and opinions of other anglers	☐	☐	☐
Information provided by the DNR	☐	☐	☐
Newspaper articles	☐	☐	☐
Magazine articles	☐	☐	☐
Bait and tackle shops	☐	☐	☐
Radio or TV	☐	☐	☐

CONTINUE

Now we would like to ask you some questions about the LAST TIME you went fishing in Michigan, even if fishing wasn't the primary purpose of the trip. We are interested in your last trip even if you walked to a fishing site located near or adjacent to your home

18. When did you leave home on this trip?

	Month	Day	Year	Time
	______	______	______	______
(Example)	June	5	1983	8 a.m.

19. When did you arrive back home from this trip?

	Month	Day	Year	Time
	______	______	______	______
(Example)	June	6	1983	9:30 p.m.

20. Where did the majority of fishing on this trip take place? It is important that you are as specific as possible.

Name of Lake or Stream ______________ County ______________ Nearest Town or City ______

21. How many total hours did you fish at this location while on this trip? ________ hours.

22. Approximately (your best estimate) how long did it take you, including rest stops to travel (one way) to this location from you permanent home? ________ hours ________ minutes

23. Approximately (again your best estimate) how many miles is the one-way driving distance from your permanent home to thi location? ________ miles one way (enter 0 if you walked to the site from home).

24. Did you fish at any other location(s) while on this trip?

☐ Yes (if yes, please answer 24a) ☐ No

24a.

Name of Lake/Stream	County	Nearest Town/City	Hours Fished There
______	______	______	______
______	______	______	______
______	______	______	______

25. Which of the following best describes the purpose of this trip?

☐ Fishing was the primary and only purpose of the trip.

☐ Fishing was the primary but not only purpose for the trip. What was the secondary purpose? ______
Would you have made the trip to this location if fishing opportunities were not available nearby? ☐ Yes ☐ No

☐ The trip was primarily for another purpose but I planned to fish when I left home. What was the primary purpose ______________ Would you have made the trip to this location if fishing opportunities were not available nearby? ☐ Yes ☐ No

☐ The trip was primarily for another purpose, and even though I fished, I did not plan to do so before I left home. What was the primary purpose? ______________

26. What percent (%) of the reason for making this trip could be attributed to fishing ______%.

27. How many other people accompanied you on this trip whether or not they fished? ______________
(If you went alone, go to question 28.)

Relationship	Are they 16 or younger? Yes	No	Did they fish on the trip? Yes	No	Was fishing the primary activit they engaged in on the trip? Yes	No
(Example) Son	☑	☐	☑	☐	☑	☐
______	☐	☐	☐	☐	☐	☐
______	☐	☐	☐	☐	☐	☐
______	☐	☐	☐	☐	☐	☐
______	☐	☐	☐	☐	☐	☐

CONTINUE ➧

28 If it was an overnight trip, what type of lodging did you use?

Type of Lodging	Number of Nights	Type of Lodging	Number of Nights
☐ Hotel or motel	______	☐ Rental cottage	
☐ A second home/cottage/ camp that you own	______	☐ Lodge	
☐ Relative's or friend's home or second home	______	☐ Campground	
		☐ Other, please specify: ______	

29 What was the primary species you were fishing for while on this trip?

- ☐ Yellow Perch
- ☐ Panfish
- ☐ Bass
- ☐ Walleye or Sauger
- ☐ Pike or Musky
- ☐ Lake Trout
- ☐ Steelhead
- ☐ Rainbow Trout
- ☐ Brown Trout
- ☐ Brook Trout
- ☐ Chinook Salmon
- ☐ Coho Salmon
- ☐ Catfish or Bullhead
- ☐ Suckers or Carp
- ☐ Smelt
- ☐ Anything that was biting

30. During what time was the trip taken?

- ☐ Regularly scheduled time off (e.g., week-ends, after work)
- ☐ Time off without pay
- ☐ Vacation time (off with pay)
- ☐ Other time off with pay (e.g., sick time, personal time)
- ☐ Other, please specify: ______

31. If you hadn't taken this trip to this location, what would you have likely done instead?

- ☐ Worked—regular time at main job
- ☐ Worked—over-time at main job
- ☐ Worked—a second job
- ☐ Fished somewhere else
- ☐ Participated in another recreation activity, please specify: ______
- ☐ Worked around the house
- ☐ Other, please specify: ______

32. Which mode of fishing did you use a majority of the time on this trip?

- ☐ Shore or Wading
- ☐ Pier or Dock
- ☐ Private Boat →
- ☐ Charter Boat
- ☐ Rented Boat
- ☐ Ice Fishing

> How long was the boat used on this trip?
> ______ ft.
>
> Was the boat:
>
> ☐ Transported to the fishing site
>
> ☐ Moored or stored near the fishing site

33. Which fishing method did you use most frequently on this trip?

- ☐ Casting
- ☐ Spin or Spin Casting
- ☐ Bait Fishing
- ☐ Trolling
- ☐ Fly Fishing
- ☐ Spearing
- ☐ Dipping
- ☐ Snagging

CONTINUE ➧

34. Next, we would like to know your out-of-pocket expenses for goods and services, including travel, on this entire trip. This includes purchases at home made especially for this trip. By out-of-pocket we mean all your expenditures whether you spent money for yourself or others in your party.

No matter what your age, we only want your expenditures. Do not ask other people (e.g., father) what they spent for you. For example, if you paid for the gas and someone else in your travel party paid for the motel room, then record the amount you paid for the gas (and anything else you bought) but not the cost of the motel.

Include all of your trip expenditures whether or not they relate to fishing.

Category	1 At Home For This Trip	2 On The Trip To And From The Area	3 Near The Fishing Site
Rods, reels, downriggers, bait, fishing line, lures, hooks, weights and other fishing supplies	$	$	$
Charter fees			
Lodging—motels, hotels, resorts, cottage rentals, or camping fees			
Restaurants			
Groceries, food & snacks, take-out beverages (including alcohol)			
Boat gas and oil			
Auto gas and oil			
Boat rentals, daily transient slip fees, launching fees			
Entertainment and other recreation (including bars, night clubs)			
Other trip expenditures (e.g., parking, shopping)			

The remaining questions on yourself and your family are needed so that we can generalize our findings to all other anglers. Again be assured that the information you provide will remain strictly confidential.

35. What is your race? ☐ White ☐ Black ☐ Native American ☐ Hispanic ☐ Oriental
☐ Other, please specify ______________________

36. What is the highest level you completed in school?

☐ Grade School ☐ High School Diploma ☐ College Degree (B.S. or B.A.) ☐ Advanced Degree (M.S., Ph.D., M.D., D.O., D.D.S., D.V.M., J.D.)

☐ Some High School ☐ Some College ☐ Some Graduate Medical or Law School

37. What is your present primary occupation? If you are unemployed or retired, tell us your last occupation: ______________

38. What is your individual income before taxes?

☐ Under $10,000 ☐ $20,000 to $24,999 ☐ $35,000 to $39,000 ☐ $50,000 or over
☐ $10,000 to $14,999 ☐ $25,000 to $29,999 ☐ $40,000 to $44,999
☐ $15,000 to $19,999 ☐ $30,000 to $34,999 ☐ $45,000 to $49,999

39. If there is more than one wage earner in your household, what is your total family income before taxes?

☐ Under $10,000 ☐ $20,000 to $24,999 ☐ $35,000 to $39,000 ☐ $50,000 or over
☐ $10,000 to $14,999 ☐ $25,000 to $29,999 ☐ $40,000 to $44,999
☐ $15,000 to $19,999 ☐ $30,000 to $34,999 ☐ $45,000 to $49,999