# APPENDIX B
# ESTIMATION

## B.1 INTRODUCTION

This appendix contains the details of estimation for the full model. The full model uses the binary choice stated preference data, the count data on days to the chosen alternative, and the revealed preference data on the number of Green Bay days. Section B.2 discusses the specification of the conditional indirect utility functions for the Green Bay choices. In Section B.3 we add the elements of the model necessary to incorporate the revealed preference data and the quantitative stated preference data comparing the Green Bay choice to the individual's alternative sites. Section B.4 includes all the pertinent information concerning the maximization of the likelihood, as well as the results from that exercise (parameter estimates, asymptotic t-statistics, etc.). Section B.5 provides a discussion of hypothesis tests and measures of goodness-of-fit. This appendix concludes with a discussion of the exercises the model can be put through, and the results of some of those exercises.

## B.2 THE MODEL AND BINARY CHOICE STATED PREFERENCE DATA

To use the binary choice SP data, a particular form for utility of a fishing day at Green Bay (Equation 1 in Appendix A) must be specified, which means in particular the choice of variables in the linear-in-parameters deterministic portion of Equation 1 in Appendix A. To this end it is assumed that utility is linear in the variables representing catch times for the four species, a set of dummy variables that represent the level of FCAs, and the launch fee, if any. The model is:

$$U_{ij}^{k_{ij}} = \sum_{l=p,s,w,b} b_l c_l^{k_{ij}} + \sum_{q=2}^{9} b_{FCAq} FCA_q^{k_{ij}} + b_y (y_i - TC_i - fee^{k_{ij}}) + e_{ij}^{k_{ij}}, \qquad (1)$$

for $i = 1,..., m$; $j = 1,..., 8$, $k_{ij} = 1$ or 2, and where $y_i$ and $TC_i$ are choice occasion income and travel cost for individual $i$, and average catch times ($c_l$, $l = p,...,b$) are measured as the time (in hours) it takes on average to catch one fish of a particular species (perch, salmon/trout, walleye, bass). For example, the perch catch time is thought to be approximately 0.75 hours. Therefore, it is expected that the coefficients of the catch times will be negative. The nine FCA levels are represented by a set of eight dummy variables, each representing a certain configuration of fish consumption advisories for the four target species. The FCA levels corresponding to the dummy variables generally increase in severity, so that $FCA_3 = 1$ means more (and/or more severe) restrictions than $FCA_2 = 1$, for example. A value of zero for all of the dummy variables ($FCA_2$ through $FCA_9$)

means essentially no restrictions (eat as many of all species as desired), and $FCA_9 = 1$ is a warning not to eat more than one perch meal per month or *any* of the remaining three species at all. The exception is in moving from $FCA_4$ to $FCA_5$ and from $FCA_5$ to $FCA_6$ with the consumption of some species becoming more restricted and others less restricted (see Section 5.4). Since $y_i$ and $TC_i$ do not vary by $k_{ij}$, these variables disappear from the relevant utility difference (see, for example, Equation 3 in Appendix A). In the pairs, the launch fees range from $0-$15. Table B-1 summarizes the data for the 10 versions of the survey, each with eight, paired Green Bay scenarios (160 total scenarios).

**Table B-1**
**Data Summary**

| Variable | Perch | Salmon/ Trout | Walleye | Bass | FCA Level | Launch Fee |
|---|---|---|---|---|---|---|
| Mean | .52 | 5.55 | 5.51 | 5.48 | 4.5 | $7.50 |
| Range | .11 - 1.00 | 1.0 - 12.0 | 1.0 - 12.0 | 1.0 - 12.0 | 1 - 9 | $0 - $15 |

## B.3 ADDING QUANTITATIVE STATED PREFERENCE AND REVEALED PREFERENCE DATA

Since there are no data on the characteristics of the alternative fishing sites for the respondents, utility for the non-Green Bay alternative site (Equation 4 in Appendix A) is assumed to be constant across individuals and choice occasions, with an additive random disturbance:

$$U_{ij}^{0} = g_{O} + e_{ij}^{0}. \tag{2}$$

This means that variables such as catch time, travel cost, and any fish advisories at other sites (but not income, as it will still drop out) are grouped into the error term. Although a component of travel cost such as distance to the site cannot contribute to a utility difference when the site is Green Bay for both choices, as it is in the binary choice SP data, it could affect the utility differences between other sites and Green Bay. We assume here that the variation in distance to anglers' other sites is not great across anglers. We further assume that any variation is likely to be randomly distributed across anglers (anglers living close to and far from Green Bay have alternative sites both near and far), so that the lack of these data adds noise (in the form of increased variance of the disturbance term in Equation 2), but does not bias parameter estimates.

Other forms of Equation 2 that included person-specific explanatory variables were tested, and these specifications did not materially change results — see Appendix D.

Utility for going to Green Bay under current conditions is given by Equation 3:

$$U_{id}^{G} = \sum_{l=p,s,w,b} b_l c_l + \sum_{q=2}^{9} b_{FCAq} FCA_q + b_y (y_i - TC_i - fee) + e_{id}^{G}, i = 1,\ldots,m, \tag{3}$$

where the values for explanatory variables are perch catch = 0.75, salmon catch = 19.4, walleye catch = 7.4, bass catch = 15.0, FCA Level = 4 ($FCA_4$ = 1, $FCA_m$ = 0 $\forall$ $m \neq 4$), and fee = \$3.

## B.4 OPTIMIZATION AND PARAMETER ESTIMATES

Gauss application model "Maxlik" was used to maximize the likelihood (Equation 11 in Appendix A). To obtain good starting values, we followed a routine: First, the binary probit model treating the 8*m* observations (where *m* is the number of respondents) as independent was fit, with the same catch time and FCA specification discussed above (the stated preference model), and with $s_e^2$ normalized to $\frac{1}{2}$. These estimates were used as starting values, along with zero for the constant and one for the variance ($s_0^2$) when the information on number of days to other sites was added, and the probit/binomial model was fit (a model with the other site included). These estimates in turn became starting values, along with various positive constants for the variance of error differences. Convergence was achieved for a variety of starting values, and always at the same point. Estimation was done on a personal computer with a Pentium III chip operating at 450 megahertz, and took approximately four minutes to converge.

Table B-2 provides the estimated values of the parameters and their estimated asymptotic t-statistics.

All parameters have the expected signs and all are statistically significant by conventional standards. The perch catch time has by far the largest parameter of the four species catch times. The pattern of estimated coefficients on the FCA variables is somewhat striking: as the FCA level increases they increase (in absolute value) nearly uniformly, and where they do not, it is as expected (see Section B.2). The same is true for their precision, as measured by their asymptotic t-statistics.

**Table B-2**
**Parameter Estimates from Main Model**

| Parameter | Estimate | Asy. t-ratio |
|---|---|---|
| $b_y$ | 0.0535 | 20.57 |
| $b_p$ | - 0.5307 | - 14.99 |
| $b_t$ | - 0.0212 | - 7.58 |
| $b_w$ | - 0.0287 | - 11.95 |
| $b_b$ | - 0.0231 | - 11.44 |
| $b_{FCA2}$ | - 0.0972 | - 3.07 |
| $b_{FCA3}$ | - 0.2599 | - 7.653 |
| $b_{FCA4}$ | - 0.5215 | - 12.92 |
| $b_{FCA5}$ | - 0.6017 | - 15.80 |
| $b_{FCA6}$ | - 0.5303 | - 13.08 |
| $b_{FCA7}$ | - 0.7660 | - 18.91 |
| $b_{FCA8}$ | - 1.0581 | - 23.40 |
| $b_{FCA9}$ | - 1.1616 | - 24.79 |
| $g_0$ | - 1.1420 | - 34.40 |
| $s_{0-e}$ | 5.5540 | 33.15[a] |
| $s_{0-G}$ | 3.5257 | 17.32[a] |
| $n$ | 647 | |
| $Log - L$ | - 19.25833 | |

a. t-statistics apply to the logged parameter estimates.

The parameters $s_{0-e}$ and $s_{0-G}$ are the standard deviations of the differences $e_{ij}^{0} - e_{ij}^{k_{ij}}$ and $e_{ij}^{0} - e_{ij}^{G}$, respectively. They allow for nonzero covariances ($s_{e0}$ and $s_{G0}$) between the random components (see Section A.3).[1]

## B.5 MEASURES OF GOODNESS-OF-FIT

To assess goodness-of-fit we examine three items based on SP data and two items based on a combination of SP and RP data: 1) the conventional 2 x 2 table of predicted versus actual choices, 2) a pseudo-$R^2$ calculation, 3) the distribution of the predicted probability of going to a chosen site, 4) a comparison of the stated number of expected days to a chosen Green Bay alternative with the estimated number of Green Bay days to that alternative, and 5) a comparison of the mean number of days to chosen alternatives to the actual mean Green Bay days under current conditions from the RP data.

1. We adopt the convention that a probability greater than 0.5 is a correct prediction. The prediction rate is high: the model predicts choices correctly 73% of the 5,038 occasions (the sum of the diagonal elements of Table B-3). Table B-3 contains the comparison of actual versus predicted choices (in percentage terms).

**Table B-3**
**Predicted vs. Actual Choices — Number (%)**

| | Predicted A | Predicted B |
|---|---|---|
| Chose A | 2,270 (45%) | 809 (16%) |
| Chose B | 544 (11%) | 1,415 (28%) |

2. A pseudo-$R^2$ measure of goodness-of-fit can be calculated for this discrete data (see Maddala, 1983, pp. 76-77) using the counts from Table B-3 and their row and column sums. Normalized on the unit interval, it is equal to 0.435, which is very reasonable for cross-sectional data.

---

1. Alternatively, one could estimate a model with covariances assumed to be zero. With this restriction, $s_0$ and $s_G$ (the standard deviations of the respective random components) are separately identified (otherwise they are not), and the estimated value of $s_G$ could be compared to the fixed value of $s_e$ to get a notion of the relative amounts of noise in the RP and SP data. However, this is only possible under the assumption that the random components are uncorrelated. In fact, preliminary analysis implied that the covariances are not zero, so the more general model that allows for this correlation statistically dominates.

3. The mean predicted probability of the preferred alternative from the stated preference experiment is 0.63, with a standard deviation of 0.22. This is calculated over 647 individuals x 8 experiments = 5,176 - 138 missing = 5,038 observations.

In Table B-4, the percentage of pairs that alternative A is actually selected is presented for different ranges of the predicted probability of selecting alternative A. These values show that an alternative is infrequently chosen when its probability of being chosen is small, and frequently chosen when its probability is high. For example, when the predicted probability of selecting alternative A is less than 0.1, A is chosen in only 5% of the pairs; but when the predicted probability is greater than 0.9, A is chosen in almost all of the pairs, 96%.

**Table B-4**
**Predicted Probabilities for Alternative A and the Percent of Pairs A Is Selected**

| Predicted Probability of Alternative A | Percent of Pairs in Each Range from which Alternative A Selected |
|---|---|
| 0-0.1 | 5.49% |
| 0.1-0.2 | 18.29% |
| 0.2-0.3 | 24.74% |
| 0.3-0.4 | 34.99% |
| 0.4-0.5 | 45.83% |
| 0.5-0.6 | 55.47% |
| 0.6-0.7 | 65.33% |
| 0.7-0.8 | 74.57% |
| 0.8-0.9 | 84.16% |
| 0.9-1.0 | 96.15% |

4. Respondents indicate the number of days they would spend at Green Bay if conditions at Green Bay were the same as their chosen alternative. The parameter estimates from the model can be used to predict the conditional probability of choosing Green Bay under the hypothetical conditions over the individual's other (real) choices. This is Equation 6 in Appendix A. Multiplying this probability by the actual number of open-water days for the respondent produces an estimate of the number of Green Bay days under hypothetical conditions. The means (standard deviations) of the indicated number of Green Bay days (truncated to be no larger than the total days at all sites for each angler) and estimated number of days are 12.0927 (14.8531) and 12.0917 (12.2885), respectively. The closeness

of these numbers means that the model estimates are in accord with the SP frequency data. While the means are very close, there is substantial variation. Table B-5 is a histogram of the differences across the chosen alternatives. For example, the difference between reported and estimated Green Bay days is between ± 4.5 for almost half of the alternatives (48.7%).

**Table B-5**
**Difference between Reported and Estimated Green Bay Days**

| Difference | Frequency (%) |
|---|---|
| < -20 | 267 (5.2%) |
| -20 to -15 | 156 (3.0%) |
| -15 to -10 | 290 (5.8%) |
| -10 to -4.5 | 507 (10.0%) |
| -4.5 to -3.5 | 97 (1.9%) |
| -3.5 to -2.5 | 126 (2.5%) |
| -2.5 to -1.5 | 157 (3.1%) |
| -1.5 to -0.5 | 179 (3.6%) |
| -0.5 to 0.5 | 507 (9.8%) |
| 0.5 to 1.5 | 442 (8.8%) |
| 1.5 to 2.5 | 454 (9.0%) |
| 2.5 to 3.5 | 348 (6.9%) |
| 3.5 to 4.5 | 157 (3.1%) |
| 4.5 to 10 | 851 (16.9%) |
| 10 to 15 | 360 (7.1%) |
| 15 to 20 | 145 (2.9%) |
| > 20 | 133 (2.6%) |

5. The estimated mean number of expected days to the chosen Green Bay alternatives (12.09) is larger than the reported number of days (9.95). Because current conditions are, on average, inferior to the average conditions over the chosen alternatives, this result shows that model estimates are consistent with actual behavior.

The probability of going to Green Bay under current conditions is estimated by the model at 0.40, which when multiplied by the total number of days is exactly equal to the actual number of Green Bay days: 9.95. This is required by the choice of statistical model for number of days (binomial). The standard deviation of the number of days in the data is 14.10, compared to 9.91 for the distribution of the predictions.

## B.6 USING THE MODEL

The model can be used to estimate how the probability of fishing Green Bay will change (and hence how the number of days fishing Green Bay will change, holding constant total fishing days) from either a change in catch times or FCA levels. This is done with Equation 9 in Appendix A, repeated here:

$$P_i^G = F\,[(b\,'(x_i^G - x_{ij}^0)\,/\,s_{\,0-G}\,], \qquad (4)$$

employing estimates for $b$ and $s_{\,0-G}$. For example, suppose there is an hour increase in the time it takes to catch a perch. The argument of the normal CDF in Equation 4 would then decrease by 0.5307/12.43 = 0.04270, causing a 40% chance of going to Green Bay to decrease to about 38%. Reducing the FCA level from four to one causes the 40% probability of going to Green Bay to increase to about 46% (and the mean number of days to increase 14.5%). Table B-6 provides some behavioral responses to site changes.

**Table B-6**
**Behavioral Response to Changes in Site Characteristics**

| Change in Site Characteristic | Change in Probability of Going to Green Bay | Change in Mean Number of Green Bay Days |
|---|---|---|
| $FCA_4 \rightarrow FCA_1$ | 40% → 46% | 14.5% |
| $FCA_3 \rightarrow FCA_1$ | 43%→ 46% | 6.8% |
| $FCA_2 \rightarrow FCA_1$ | 45% → 46% | 2.4% |
| Double perch catch rate | 40% → 42% | 5.5% |
| Quadruple perch catch rate | 40% → 43% | 8.3% |
| Multiply perch catch by 10 | 40% → 44% | 10.0% |
| Double all catch rates | 40% → 47% | 19.1% |

For example, holding constant other site characteristics, the probability of going to Green Bay would increase by 6% if FCAs were eliminated. However, at an existing FCA Level of four, doubling the catch rate for perch would only cause a 2% increase in the probability of going to Green Bay.

The model can also be used to estimate compensating variation resulting from changes in catch rates, fees, or FCA levels (Appendix C).