E. The Rank-Ordered Probit Model

Following the random utility model framework, individuals are assumed to derive utility from (1) each of the coral reef protection programs presented in the survey and (2) all else. Both of these aspects of utility are assumed to have observable components and unobservable, random components. Utility can therefore be expressed as:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \tag{E.1}$$

or, more specifically:

$$U_{ij} = \beta_y y_i + \left(\beta_F F_j + \beta_S S_j + \beta_{FS} F_j S_j\right) X_i + e_{ij} \tag{E.2}$$

where i represents the individual survey respondents ($i = 1 \ldots n$); j represents the four program options in the survey (1 = status quo, 2 = the No-Fishing Zones Program, 3 = the Reef Repair Program, and 4 = the combination of programs 2 and 3); X_i is a $k \times 1$ vector of individual-specific variables, including a 1 to allow for alternative-specific constant terms; y_i is individual i's income; F_j and S_j are scalar indicator variables for whether or not the No-Fishing Zones and Reef Repair programs appear in alternative j; β_y is the marginal utility of money income; and β_F, β_S, and β_{FS} are each $1 \times k$ vectors of the marginal contributions toward utility that individuals with the associated covariates derive from the specific programs.

Letting C_{ij} be the additional cost to individual i's household of program alternative j ($C = 0$ for Program 1, the status quo),[1] and using Equation E.2, individual i's utility for the four programs are:

$$\begin{aligned}
U_{i1} &= \beta_y y_i + e_{i1} \\
U_{i2} &= \beta_y (y_i - C_{i2}) + \beta_F X_i + e_{i2} \\
U_{i3} &= \beta_y (y_i - C_{i3}) + \beta_S X_i + e_{i3} \\
U_{i4} &= \beta_y (y_i - C_{i4}) + (\beta_F + \beta_S + \beta_{FS}) X_i + e_{i4}
\end{aligned} \tag{E.3}$$

1. The experimental design defined 16 cost scenarios that were offered randomly to different survey respondents. Importantly, a zero-cost program – the status quo – is included in all choice sets. This allows for the estimation of total WTP for the programs. See Appendix B for the full experimental design.

The No-Fishing Zones Program at 25% (as opposed to the status quo of 1%), the Reef Repair Program at 10 years of recovery time (as opposed to the status quo of 50 years), and the combination of the two programs are the only programs included in the choice sets. The policy variables, F_j and S_j, are therefore assumed to be binary, "dummy" variables – that is, they can take only the value of zero (for the status quo) or 1 (for the proposed program). This modeling constraint is imposed to conform to the science, which found alternative levels of the policy variables to be unrealistic.

Because of the limited array of policy programs offered, several constraints or limitations had to be imposed on our analysis. First, it had to be acknowledged that the estimated coefficients – the estimated β's – only measure the total contributions that these specific policy changes would have on utility and therefore WTP. In other words, the model cannot be used to estimate the marginal contributions of alternative levels of the policy variables.

Second, inclusion of the interaction term β_{FS} in the utility function for the combination ("both") program gives us a model that is observationally equivalent to one where the utility for each of the program alternatives is represented by a separate program "dummy." In other words, inclusion of the interaction term gives the "both" alternative complete flexibility to be fitted to the data irrespective of the contributions that the separate programs, F_j and S_j, might make to that choice. This essentially boils the "both" alternative down to one that does not explicitly acknowledge the programs it is composed of. Basically, we lose potential information on how respondents valued the individual programs within the "both" alternative. This turned out to be an important issue when we conducted a preliminary analysis and found that the bid values on the individual programs did not do an adequate job of capturing the program values. The "both" program offered a wider range of bids for the two programs, and we found it to be important to the analysis to use this information explicitly. We therefore omitted the interaction term in our final analysis.[2]

Returning to Equation E.3, the ϵ's are the random components of utility and are assumed to be correlated and heteroskedastic across program alternatives. In other words, it is assumed that respondents will have unexplained aspects of their preferences for the programs that are likely to move in the same direction – either positive for all programs or negative.[3] Heteroskedasticity

2. This omission means that we use information gleaned from both the separate program alternatives and the "both" alternative to estimate total WTP for the two, separate programs. Analytically, we are combining the information we have about how people value the programs separately and how they value them jointly. Our estimates of WTP for each of the separate programs will therefore be underestimates of true WTP for those individual programs.

3. One interpretation of this assumed correlation is that different individuals' error terms might contain fixed error components that represent their support or non-support for "any program." This fixed component would feed into the correlation across alternatives. The remaining component of individuals' errors would be specific to the programs offered in each alternative and independent both across and within individuals.

allows for the error variances to vary across alternatives. This allows for the possibility, for example, that the variance of alternative four might be different – probably larger – than the other alternatives.[4]

To estimate model parameters via the maximum likelihood method, the log-likelihood function is specified as the log of the probability that the specific rankings observed in the data would, in fact, occur:

$$loglik = \sum_{i=1}^{n} \sum_{j=1}^{4} log P_{ij} I_{ij} \tag{E.4}$$

where P_{ij} is the probability that individual i will select program j and I_{ij} is an indicator function for whether or not individual i actually chose program j. The probability, for example, that individual i will select program k in the first round of the ranking exercise can be expressed as:

$$Prob(\text{individual } i \text{ chooses program } k) = P_{ik} = P(U_{ik} > U_{ij}, \text{ for all } j \neq k)$$

$$= P(V_{ik} + \varepsilon_{ik} > V_{ij} + \varepsilon_{ij}, \forall\, j \neq k)$$

$$= P(\varepsilon_{ij} - \varepsilon_{ik} < V_{ik} - V_{ij}, \forall\, j \neq k) \tag{E.5}$$

Equation E.5 is in the form of a cumulative density function (CDF) with random terms – the error differences – on the left-hand side of the inequality and a parametric function – differences in observable utilities – on the right-hand side.

Moving to the full sequence of choices, the probability of observing a full ranking of alternatives k, l, m, and n for individual i is:

$$P_i = P_i(U_{ik} > U_{il} > U_{im} > U_{in})$$

$$= P(\varepsilon_{il} - \varepsilon_{ik} < V_{ik} - V_{il}, \varepsilon_{im} - \varepsilon_{il} < V_{il} - V_{im}, \varepsilon_{in} - \varepsilon_{im} < V_{im} - V_{in}) \tag{E.6}$$

4. One way of interpreting heteroskedasticity is by discussion of the scale term. "Scale" is the inverse of the standard deviation, so higher variance for an alternative implies a narrower scale. Swait (2007) discusses the difficulty of interpreting the thinking behind individuals having varying scales across alternatives, but he notes that, analytically, the approach has merits. In our case, it might be that a potentially larger variance could be caused by omission of the interaction term discussed above. Essentially, we are reducing the ability of our model to explain the "both" choice by excluding this term. The larger, unexplained variation goes into the error term, increasing estimated variance. We thank a peer reviewer for making this observation.

which is in the form of a joint CDF in terms of error differences. To estimate model parameters then, an assumption must be made about the joint distribution of these error differences.

Under the rank-ordered probit specification, error terms, the e's, are assumed to be jointly distributed normal with a mean of 0 and a variance-covariance matrix Σ. The matrix Σ is assumed to have non-identical diagonal terms (heteroskedasticity among alternatives) and non-zero, symmetric, off-diagonal terms (non-independence of preferences across alternatives). Error differences are therefore jointly distributed normal with a mean of 0 and a variance-covariance matrix that can be expressed as a quadratic function of Σ (see Train, 2003, pp. 162–163 for details).

Given these assumptions, Equation E.6 can be expressed as a multivariate normal CDF for each individual. These individual CDFs fit directly into the log-likelihood function (Equation E.4), which is maximized to estimate the parameters. In practice, the rank-ordered probit model can be estimated with the "asroprobit" command in Stata 10.[5]

Because Equation E.6 is based only on differences in parametric utilities rather than absolute measurements of the utilities, the model is invariant to "location." That is, we could add or subtract the identical, fixed quantities from the utilities for each program and the same relative rankings would result. This means that we cannot estimate absolute utility levels for every program in the choice set. Rather, we can only estimate how the utilities vary *compared* to one particular program. Estimation therefore requires the selection of a base alternative to compare the other programs to. Selecting such a base alternative essentially reduces the model from a four-way to a three-way structure, with the variance-covariance being reduced from a 4×4 to a 3×3 matrix. Because this 3×3 matrix is symmetric, it has six unique elements. In this study, the base alternative is selected as the status quo.

In addition to this "location" restriction, there is another restriction that must be made before the model is estimable; this one is based on the fact that we cannot independently estimate all of the standard deviations of the error terms. One must be fixed and all else is scaled to this fixed term. This is the same problem that we have with standard logit and probit models and is commonly discussed as the problem of identifying "scale." Standard practice is to assign the value 1 to the

5. The algorithm that Stata 10 uses to approximate the multivariate normal function is called the GHK algorithm (Hajivassiliou and Ruud, 1994). Stata 10 uses a default setting of 200, which draws on the Hammersley sequence to approximate the distribution. The GHK algorithm does not estimate the variance-covariance matrix directly. Rather, to ensure that the matrix remains positive definite and that diagonal elements remain positive over the course of the maximization routine, a square-root transformation is taken on the Cholesky factorization of the variance-covariance matrix. A log transformation is taken on the diagonal elements of this matrix.

unidentified parameter. In this study, the standard deviation of the No-Fishing Zones Program is assigned this value.

This leaves five elements of the variance-covariance matrix that are identified: two variances and three covariances, or correlations. The scale alternative in this study is specified to be Alternative Two, and the standard deviation of Alternative Two is set to unity.

Once parameter estimates are available, individual *i*'s WTP for program *j* can be estimated as (omitting the interaction term):

$$WTP_{ij} = \frac{-(\beta_F F_j + \beta_S S_j) X_i}{\beta_y} \tag{E.7}$$

Once individual WTP amounts are estimated, mean WTP can be calculated by taking the mean of the individual estimates, weighted by the sample probability weights:

$$mean(WTP_j) = \frac{\sum_{i=1}^{n} w_i WTP_{ij}}{n} \tag{E.8}$$

where w_i is the sample probability weight for individual *i*.

Applying Equation E.7, Equation E.8 can be re-expressed as:

$$mean(WTP_j) = \frac{-(\beta_F F_j + \beta_S S_j)\overline{X}_w}{\beta_y} \tag{E.9}$$

where $\overline{X}_w$ is the vector of the weighted means of the individual vectors, x_i.[6]

The estimated variance of mean WTP for program *j* can be calculated by applying the delta method (see Alberini et al., 2007):

$$var[mean(WTP_j)] = d_j' var(\beta) d_j \tag{E.10}$$

6. The normal distribution is a symmetric distribution that covers the domain from negative to positive infinity. Error differences also span this range, and these affect the estimation of utility differences and therefore model parameters. Underlying the estimation of mean WTP then is the possibility that some individual, predicted WTP amounts could be negative, which is counter-intuitive. As Equation E.9 shows, however, mean WTP can be estimated based on the "average" individual; it is a measure of central tendency. So, while the normal assumption might provide counter-intuitive results for some individuals, it can be used to estimate the overall mean – the central tendency of the distribution – when the bulk of the distribution is in positive territory.

where *var*(β) is the estimated variance-covariance matrix of the estimated parameter vector (β_F, β_S, and β_y) and d_j is the *j*-specific value of the derivative of Equation E.9 with respect to the parameter vector. All elements of Equation E.10 are estimated using the maximum likelihood estimates of the parameters and, from Equation E.9, the weighted means of the covariates.

Ninety-five percent confidence intervals are estimated as:

$$CI_j = mean(WTP_j) \pm 1.96 \times SE[mean(WTP_j)] \tag{E.11}$$

where $SE[mean(WTP_j)]$ is the square root of the variance of $mean(WTP_j)$.

The income elasticity of WTP_j, or the percentage change in WTP_j due to a percentage change in income, can be estimated as the ratio of the estimated program-specific coefficient on income over the coefficient on program cost, scaled by the ratio of mean income over mean estimated WTP_j. Specifically, the income elasticity of WTP_j is estimated as:

$$E = -\left(\frac{\beta_j^{income}}{\beta_y}\right)\left(\frac{mean(income)}{mean(WTP_j)}\right) \tag{E.12}$$

The delta method can be used to estimate standard errors and confidence intervals.

For discrete variables, such as "being a strong environmentalist," the marginal impact of the *k*th program-specific variable WTP_j can be estimated by taking the negative of the ratio of the *k*th program-specific coefficient over the coefficient on program cost:

$$E = \left(\frac{\beta_j^k}{\beta_y}\right) \tag{E.13}$$

Again, the delta method is used to estimate standard errors and confidence intervals.

References

Alberini, A., A. Longo, and M. Veronesi. 2007. Basic statistical models for stated choice studies. In *Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice,* B. Kanninen (ed.). Springer, Dordrecht, the Netherlands.

Hajivassiliou, V.A. and P.A. Ruud. 1994. Classical estimation methods for LDV models using simulation. In *Handbook of Econometrics, Volume IV,* R.F. Engle and D.L. McFadden (eds.). *Elsevier Science B.V.* 2383–2441.

Swait, J. 2007. Advanced choice models. In *Valuing Environmental Amenities Using Stated Choice Studies: A Common Sense Approach to Theory and Practice,* B. Kanninen (ed.). Springer, Dordrecht, the Netherlands.

Train, K.E. 2003. *Discrete Choice Methods with Simulation.* Cambridge University Press, Cambridge, UK.