Valuation for Environmental Policy: Ecological Benefits

A Workshop sponsored by U.S. Environmental Protection Agency's National Center for Environmental Economics (NCEE) and National Center for Environmental Research (NCER)

Crowne Plaza Washington National Airport Hotel
1480 Crystal Drive
Arlington, VA 22202

April 23-24, 2007

U.S. Environmental Protection Agency (EPA)
National Center for Environmental Economics (NCEE) and
National Center for Environmental Research (NCER)
Valuation for Environmental Policy: Ecological Benefits

Crowne Plaza Washington National Airport Hotel
1480 Crystal Drive
Arlington, VA 22202
(703) 416-1600

April 23-24, 2007

Agenda

April 23, 2007: Valuation for Environmental Policy

8:00 a.m. – 8:30 a.m.	**Registration**	
8:30 a.m. – 8:45 a.m.	**Introductory Remarks** Rick Linthurst, National Program Director for Ecology, EPA, Office of Research and Development	
8:45 a.m. – 11:30 a.m.	**Session I: Benefits Transfer** Session Moderator: Steve Newbold, EPA, NCEE	
	8:45 a.m. – 9:15 a.m.	Benefits Transfer of a Third Kind: An Examination of Structural Benefits Transfer George Van Houtven, Subhrendu Pattanayak, Sumeet Patil, and Brooks Depro, Research Triangle Institute
	9:15 a.m. – 9:45 a.m.	The Stability of Values for Ecosystem Services: Tools for Evaluating the Potential for Benefits Transfers John Hoehn, Michael Kaplowitz, and Frank Lupi, Michigan State University
9:45 a.m. – 10:00 a.m.	**Break**	
	10:00 a.m. – 10:30 a.m.	Meta-Regression and Benefit Transfer: Data Space, Model Space and the Quest for 'Optimal Scope' Klaus Moeltner, University of Nevada, Reno, and Randall Rosenberger, Oregon State University
	10:30 a.m. – 10:45 a.m.	Discussant: Matt Massey, EPA, NCEE
	10:45 a.m. – 11:00 a.m.	Discussant: Kevin Boyle, Virginia Tech University
	11:00 a.m. – 11:30 a.m.	Questions and Discussion
11:30 a.m. – 12:45 p.m.	**Lunch**	

12:45 p.m. – 3:30 p.m. **Session II: Wetlands and Coastal Resources**
Session Moderator: Cynthia Morgan, EPA, NCEE

12:45 p.m. – 1:15 p.m. A Combined Conjoint-Travel Cost Demand Model for Measuring the Impact of Erosion and Erosion Control Programs on Beach Recreation
Ju-Chin Huang, University of New Hampshire; George Parsons, University of Delaware; Min Qiang Zhao, The Ohio State University; and P. Joan Poor, St. Mary's College of Maryland

1:15 p.m. – 1:45 p.m. A Consistent Framework for Valuation of Wetland Ecosystem Services Using Discrete Choice Methods
David Scrogin, Walter Milon, and John Weishampel, University of Central Florida

1:45 p.m. – 2:00 p.m. **Break**

2:00 p.m. – 2:30 p.m. Linking Recreation Demand and Willingness To Pay With the Inclusive Value: Valuation of Saginaw Bay Coastal Marsh
John Whitehead and Pete Groothuis, Appalachian State University

2:30 p.m. – 2:45 p.m. Discussant: Jamal Kadri, EPA, Office of Wetlands, Oceans, and Watersheds

2:45 p.m. – 3:00 p.m. Discussant: John Horowitz, University of Maryland

3:00 p.m. – 3:30 p.m. Questions and Discussion

3:30 p.m. – 3:45 p.m. **Break**

3:45 p.m. – 5:45 p.m. **Session III: Invasive Species**
Session Moderator: Maggie Miller, EPA, NCEE

3:45 p.m. – 4:15 p.m. Models of Spatial and Intertemporal Invasive Species Management
Brooks Kaiser, Gettysburg College, and Kimberly Burnett, University of Hawaii at Manoa

4:15 p.m. – 4:45 p.m. Policies for the Game of Global Marine Invasive Species Pollution
Linda Fernandez, University of California at Riverside

4:45 p.m. – 5:00 p.m. Discussant: Marilyn Katz, EPA, Office of Wetlands, Oceans, and Watersheds

5:00 p.m. – 5:15 p.m. Discussant: Lars Olsen, University of Maryland

5:15 p.m. – 5:45 p.m. Questions and Discussion

5:45 p.m. **Adjournment**

April 24, 2007: Valuation for Environmental Policy

8:30 a.m. – 9:00 a.m.	**Registration**	
9:00 a.m. – 11:45 a.m.	**Session IV: Valuation of Ecological Effects** Session Moderator: William Wheeler, EPA, NCER	
	9:00 a.m. – 9:30 a.m.	Integrated Modeling and Ecological Valuation: Applications in the Semi Arid Southwest David Brookshire, University of New Mexico, Arriana Brand, Jennifer Thacher, Mark Dixon,Julie Stromberg, Kevin Lansey, David Goodrich, Molly McIntosh, Jake Gradny, Steve Stewart, Craig Broadbent and German Izon
	9:30 a.m. – 10:00 a.m.	Contingent Valuation Surveys to Monetize the Benefits of Risk Reductions Across Ecological and Developmental Endpoints Katherine von Stackelberg and James Hammitt, Harvard School of Public Health
10:00 a.m. – 10:15 a.m.	**Break**	
	10:15 a.m. – 10:45 a.m.	Valuing the Ecological Effects of Acidification: Mapping the Extent of Market and Extent of Resource in the Southern Appalachians Shalini Vajjhala, Anne Mische John, and David Evans, Resources for the Future
	10:45 a.m. – 11:00 a.m.	Discussant: Joel Corona, EPA, Office of Water
	11:00 a.m. – 11:15 a.m.	Discussant: David Simpson, Johns Hopkins University
	11:15 a.m. – 11:45 a.m.	Questions and Discussion
11:45 a.m. – 1:00 p.m.	**Lunch**	
1:00 p.m. – 4:15 p.m.	**Session V: Water Resources** Session Moderator: Adam Daigneault, EPA, NCEE	
	1:00 p.m. – 1:30 p.m.	Valuing Water Quality as a Function of Physical Measures Kevin Egan, Joe Herriges, John Downing, and Katherine Cling, Iowa State University
	1:30 p.m. – 2:00 p.m.	Cost-Effective Provision of Ecosystem Services from Riparian Buffer Zones Jo Albers, Oregon State University; David Simpson, Johns Hopkins University; and Steve Newbold, NCEE
2:00 p.m. – 2:15 p.m.	**Break**	
	2:15 p.m. – 2:45 p.m.	Development of Bioindicator-Based Stated Preference Valuation for Aquatic Resources Robert Johnston, Eric Shultz, Kathleen Segerson, Jessica Kukielka, Deepak Joglekar, University of Connecticut; and Elena Y. Besedin, Abt Associates

April 24, 2007 (continued)

	2:45 p.m. – 3:05 p.m.	Comparing Management Options and Valuing Environmental Improvements in a Recreational Fishery Steve Newbold and Matt Massey, NCEE
	3:05 p.m. – 3:20 p.m.	Discussant: Julie Hewitt, EPA, Office of Water
	3:20 p.m. – 3:35 p.m.	Discussant: George Parsons, University of Delaware
	3:35 p.m. – 4:05 p.m.	Questions and Discussions
4:05 p.m. – 4:15 p.m.	**Final Remarks**	
4:15 p.m.	**Adjournment**	

Benefits Transfer of a Third Kind: An Examination of Structural Benefits Transfer

George Van Houtven
RTI International

Subhrendu Pattanayak
RTI International
North Carolina State University

Sumeet Patil
RTI International

Brooks Depro
RTI International

April 2007

Abstract: Most benefit transfer applications use unit value or benefit function transfer approaches, which do not directly impose consistency with an underlying economic structure. This paper examines a third kind of approach – structural benefits transfer (SBT) -- in which the transfer is directly tied to utility theory via the preference structure. It extends previous SBT applications by (1) exploring five different utility function specifications and comparing their implications for predicting benefits, and (2) including nonuse values in each specification. The approach is demonstrated by combining results from a travel cost study and a contingent valuation study for improvements in river water quality. The results show that SBT estimates for selected water quality improvements and conditions are sensitive to preference specification; however, they also highlight the strengths and limitations of different specifications, by providing plausibility checks on the range of predicted outcomes. For example, in this application, SBT functions based on a linear trip demand specification produce more plausible benefit predictions than a log-linear demand or a Stone-Geary framework.

Key Words: Benefit transfer, Structural benefit transfer, Preference calibration

Send all correspondence to:
George Van Houtven, Ph.D.
RTI International
3040 Cornwallis Road
P.O. Box 12194
Research Triangle Park, NC 27709-2194
Voice: (919) 541-7150
E-mail: gvh@rti.org

1. Introduction

The policy community frequently uses benefit transfer methods because they offer a practical and low cost way to provide benefit estimates for benefit-cost analyses, natural resource damage assessments, and other natural resource policy and management analyses. These methods take and adapt results from existing primary valuation studies and apply them to assess the benefits of selected policy changes.

For the most par, benefit transfer approaches fall into two categories-- "unit value" transfers or "value function" transfers—where the key distinction between the two approaches is the degree to which differences between the study and policy contexts are formally accounted for in the transfer. In unit value transfers, a single value or range of values, such as the value per recreation day or per unit change in water quality, is usually transferred with little or no adjustment for differences between the two settings. With benefit, or value, function transfers, information from existing studies is used to identify a functional relationship between the value of interest and the factors that may influence the magnitude of the value (e.g., using meta-regression analysis). This functional relationship allows the analyst to account for differences between the two settings and adapt the transfer estimates accordingly.

Although commonly used for policy analysis, these traditional approaches to benefit transfer do not explicitly impose consistency with the economic theory that is assumed to underlie the value estimates. Moreover, to the extent that they use existing value estimates based on different nonmarket valuation methods, they typically combine them in an ad-hoc manner.

To address these limitations, a third kind of benefits transfer – ”structural benefits transfer” (or“preference calibration”)– has been proposed in which the transfer methodology is directly tied to utility theory via the preference structure (Smith et al. 2002; Bergstrom and Taylor, 2006). Structural benefits transfer is in essence a form of benefit function transfer; where the functional form is specifically derived from an assumed utility function. Although this third approach has the potential to improve and strengthen benefit transfers, it has thus far only been applied and evaluated in a limited number of examples.

This paper further examines and evaluates structural benefits transfer as an alternative transfer method by extending existing applications in two main directions. First, we apply preference calibration using several different utility function specifications and compare their implications for predicting benefits. Second, whereas existing applications have focused on use-related values for environmental improvements, we explicitly include nonuse values in the preference specifications. Through these applications, we examine the generalizability and the robustness of the basic logic of structural benefits transfer.

The paper begins in the next section by providing a background discussion of the structural benefits transfer approach. Section 3 then introduces and describes the preference specifications that will be applied, and Section 4 discusses how estimates from different nonmarket valuation methods can be directly linked to these preference specifications. Section 5 presents a case study application focusing on water quality changes using the five preference specifications. The results and implications of these applications are then discussed in Section 6, along with suggested directions for future research.

2. Background

The main concept underlying preference calibration is that, if one is willing to make explicit assumptions about the functional form of utility with respect to a nonmarket commodity (e.g., environmental quality or health), then information from existing empirical valuation studies can in principle be used to estimate the parameters of the function. When both the utility function parameters and available benefit estimates are few in number, it is possible to calibrate the parameter values such that they produce benefit measures that match the observed empirical estimates. This is the approach used in this paper. As the number of available benefit estimates increases, structural meta-analysis techniques can instead be used to statistically estimate the parameter values (Smith and Pattanayak, 2002; Bergstrom and Boyle, 2006).

Structural benefit transfer recognizes that the selected preference specification has direct implications for both the functional form and the parameters of the corresponding welfare functions (i.e., willingness to pay (WTP), quasi-expenditure, or variation functions, as described for example by McConnell [1990]) and Whitehead [1995]). Therefore, it defines a benefit transfer function with (1) a functional form that is directly derived from the preference specification and (2) parameters that are calibrated from existing empirical estimates.

The preference calibration logic was initially presented and illustrated in studies focusing on water quality changes using a utility specification with a modified constant elasticity of substitution (CES) form (described in more detail below). Using this simple form, which did not, not specifically include nonuse values, Smith et al. (2000) combined travel cost estimates from Englin et al. (1997) and contingent valuation (CV) estimates from Carson and Mitchell (1993) to calibrate preferences. Smith et al. (2002) expanded this approach by including hedonic property value estimates from Boyle et al. (1999), recalibrating the preference parameters, and generating illustrative benefit estimates with the calibrated function. More recently, the general approach has been extended to the area of health valuation related to both morbidity and mortality (Van Houtven et al., 2004; Smith et al., 2006) and visibility benefits (Smith and Pattanayak, 2002).

In all of these cases, the process for developing a structural benefit transfer function generally involves the following steps:

1) Specify a "representative" individual's preference function.
2) Define explicitly the relationships between the available benefits measures and the specified preference function.
3) Derive the structural benefit function that is implied by the assumed preference structure.
4) Adapt the available information from existing benefit studies to assure cross-study compatibility.
5) Calibrate or estimate preference function parameters that are as consistent as possible with the observed benefit measures.
6) Insert the calibrated or estimated parameters into the structural benefit function.

Based on this same general process, in the following sections we apply the preference calibration logic using five different preference specifications to a case study application of water quality changes.

3. Specifying Preferences

To characterize the preferences of a representative individual with respect to changes in water quality, we specify five alternative indirect utility (V) functions, which we refer to as (1) modified constant elasticity of substitution (CES); (2) linear trip demand; (3) semi-log demand ; (4) log-linear demand; and (5) Stone-Geary specifications. Using these alternative specifications allows us to explore the sensitivity of benefit transfer predictions (for changes in water quality) with respect to the assumed functional form of utility.

As shown below, each indirect utility function is specified in terms of income (Y), round-trip travel cost (P), and water quality level (Q).

Modified CES:

$$V_A = \left(\varphi_A Q^{\psi_A}\right) + \left((P - Q^{\gamma_A})^{-\alpha_A} Y^{\delta_A}\right)^{\beta_A} \tag{1}$$

Linear Demand:

$$V_B = \varphi_B Q^{\psi_B} + \left[Y + \frac{1}{\delta_B}\left(\alpha_B - \beta_B P + \gamma_B Q - \frac{\beta_B}{\delta_B}\right)\right] \cdot \exp\left[\frac{\delta_B}{\beta_B}\left(\gamma_B Q - \beta_B P\right)\right] \tag{2}$$

Semi-log Demand:

$$V_C = \varphi_C Q^{\psi_C} + \frac{Y^{(1-\delta_C)}}{(1-\delta_C)} + \left(\frac{1}{\beta_C}\right) \cdot \exp\left(-\beta_C P + \alpha_C + \gamma_C Q\right) \tag{3}$$

Log-linear Demand:

$$V_D = \varphi_D Q^{\psi_D} + \frac{Y^{(1-\delta_D)}}{(1-\delta_D)} - (1-\beta_D)^{-1} e^{\alpha_D} Q^{\gamma_D} P^{(1-\beta_D)} \tag{4}$$

Stone-Geary:

$$V_E = \varphi_E Q^{\psi_E} + Q^{\gamma_E} \cdot \ln\left(\frac{Q^{\gamma_E}}{-\beta_E P}\right) + \left(1 + Q^{\gamma_E}\right) \cdot \ln\left(\frac{Y^{\delta_E} - \alpha_E - \beta_E P}{1 + Q^{\gamma_E}}\right) \tag{5}$$

Equation 1 is similar to the modified CES indirect utility function used in previous preference calibration analyses (Smith et al ,2002; Smith et al., 2000). Equations 2, 3, and 4 are derived respectively from linear, semi-log, and log-linear trip demand specifications, and Equation 5 is based on a Stone-Geary utility function (see for example, Larson [1991] and Herriges et al. [2004]).[1] To capture nonuse values, each specification includes an additively

[1] To match the number of parameters used in the other functional forms, an additional parameter, δ, is added to the income term in the Stone-Geary model.

separable subcomponent (of the form φQ^{ψ}), which is independent of P and Y. These nonuse values will not be manifested in value estimates based on revealed preference methods, but they are likely to be included in estimates from stated preference studies.

All five preference specifications include six parameters. For each specification, we represent the vector of parameters as $\boldsymbol{\theta}_{\mathbf{j}}$, such that, for example, $\boldsymbol{\theta}_{\mathbf{A}} = (\alpha_A, \beta_A, \gamma_A, \delta_A, \varphi_A, \psi_A)$ is the parameter vector for the modified CES preferences. These are the parameters to be calibrated.

4. Linking Benefit Measures to the Preference Function

In this paper, we calibrate these preference parameters for each specification by combining results from a travel cost and a contingent valuation analysis. The travel cost analysis provide estimates of recreation demand (i.e., number of water based recreation trips per year [X]) and changes in Marshallian consumer surplus (ΔMCS) resulting from changes in water quality. The CV method provides estimates of Hicksian compensating surplus (WTP) for changes in water quality.

Tables 1 and 2 report algebraic expressions for X, ΔMCS , and WTP, which are directly derived from the preference functions listed in Equations 1 to 5. The demand functions are derived by applying Roy's Identity to the indirect utility functions, and the ΔMCS functions are derived by changing the level of water quality (from Q_0 to Q_1) in these demand equations. The WTP functions are derived by solving for the compensating surplus that equalizes indirect utility for different levels of Q. Each expression is a function of the exogenous variables Y, P, and Q, and each one also includes parameters from its corresponding preference specification.

5. Preference Calibration Application

The first objective of this preference calibration application is therefore to identify values for the preference parameters that replicate as closely as possible the observed empirical estimates of X, WTP, ΔMCS (based on conditions defined by Y, P, and Q). We can then insert these calibrated parameter values in the WTP equations shown in Table 2 and use these equations as structural benefit transfer functions.

The two empirical studies used in this application were conducted in the early 1980's as part of a larger research project for EPA.[2] Both studies focused on measuring water quality benefits for households living in the vicinity of the Monongahela River in Southwestern Pennsylvania. The two studies also used data from the same 1981 survey of residents living within the Monogahela River Valley. This survey was based on a stratified sample of 393 households from the five-county area surrounding the Pennsylvania portion of the Monongahela River, including the Pittsburgh metropolitan area. Administration of the survey resulted in 301 completed interviews.

The first study used data from the survey to estimate a recreation demand travel cost model (Smith et al., 1983). This study identified 13 recreation sites along the Pennsylvania portion of the river and 69 respondents who had visited at least one of these sites. The total

[2] For both an overview and detailed summary of this larger research project, see Smith and Desvousges (1986).

number of user-site combinations, each of which represented a single observation, was 94. Smith et al. applied a generalized travel cost model to estimate trip demand functions for each site. They then used these demand functions to estimate the increase in consumer surplus per household per season that would result from increasing water quality levels from boatable conditions to fishable conditions and from boatable to swimmable conditions.

The second study was based on responses to a contingent valuation scenario that was presented as part of the survey (Desvousges et al., 1987). At the time of the survey, the overall water quality levels in the Pennsylvania section of the Monongahela were assumed to be characterized by boatable conditions. Respondents were asked to value three water quality changes:. (1) raising levels from boatable to fishable conditions, (2) raising levels from fishable to swimmable condition, and (3) avoiding a decrease from boatable to nonboatable conditions. The survey used different elicitation methods (iterative bidding, open-ended, and payment card) for different subsamples. For this analysis, we use the open-ended responses, which were collected from 51 respondents, including both users and nonusers of the Monongahela river sites.

5.1.2 Defining Consistent Measures Across the Studies

To define a continuous unit of measure for Q that is consistent across the two studies, we use the same Resources for the Future (RFF) water quality ladder/scale (Vaughan, 1986) that was presented to respondents in the contingent valuation survey to describe water quality changes. According to this 1-to-10 point scale, nonboatable, boatable, fishable, and swimmable water quality levels are assigned values of 0.5, 2.5, 5.1, and 7, respectively.

The summary statistics and benefit estimates used in the calibration applications are summarized in Table 3. The travel cost study provides estimates of the average baseline number of trips ($X = 7.22$), average income, and average travel cost for the sample of 94 recreators. All dollar values from these studies have been converted to 2005 dollars using the consumer price index (CPI). The baseline demand for trips is assumed to be evaluated at a water quality level that is "consistent with supporting boating (the current [1977] recreational use of the river)" (Smith et al., 1983) ($Q_0 = 2.5$). The travel cost study also provides ΔMCS estimates for two water quality improvements -- one to fishable quality ($Q_1 = 5.1$) and the other to swimmable quality ($Q_1 = 7$).

The contingent valuation study also provides estimates of average income and baseline trips for its sample of respondents. Average income is 10 percent lower than for travel cost study sample, and average baseline trips is 67 percent lower, primarily because two thirds of the CV sample are nonusers. Average travel costs (P) are not reported for the CV sample; however, they can be derived by inverting the trip demand functions in Table 3 and expressing P as a function of Y, X, Q and the preference parameters. The WTP estimates for the three water quality changes range from \$26.64 (improving water quality from Q_0=5.1 to Q_1=7) to \$52.64 (avoiding a decrease from Q_1=2.5 to Q_0=0.5).

5.1.3 Calibrating Parameters.

To calibrate parameters for each specification, we define six conditions representing the difference between observed values for X, ΔMCS, and WTP (numbers in bold italics in Table 3) and their predicted values using the equations in Table 1 and 2. From the first column of travel cost results in Table 3 we define:

$$7.22 - X(18, 2.5, 46{,}398\ ; \boldsymbol{\theta}) = \varepsilon_1 * 7.22 \qquad (6.1)$$

$$14.56 - M(18, 2.5, 5.1, 46{,}398 ; \boldsymbol{\theta}) = \varepsilon_2 * 14.56 \quad (6.2)$$

From the second column of travel cost results we define:

$$30.58 - M(18, 2.5, 7.0, 46{,}398 ; \boldsymbol{\theta}) = \varepsilon_3 * 30.58 \quad (6.3)$$

From the three columns of CV results we define:

$$37.81 - M(P, 2.5, 5.1, 46{,}398 ; \boldsymbol{\theta}) = \varepsilon_4 * 37.81 \quad (6.4)$$

$$26.64 - M(P, 5.1, 7.0, 46{,}398 ; \boldsymbol{\theta}) = \varepsilon_5 * 26.64 \quad (6.5)$$

$$52.64 - M(P, 0.5, 2.5, 46{,}398 ; \boldsymbol{\theta}) = \varepsilon_6 * 52.64 \quad (6.6)$$

where P is derived from the inverse demand function $X^{-1}(2.4 \;\; 2.5, 46{,}398 ; \theta)$ (see footnote to Table 1) at baseline conditions for the CV sample..

Ideally, we would identify solutions for the parameter vector $\boldsymbol{\theta}$ that would make each of the six equations exactly equal to zero. However, due to the nonlinearities in this system, no exact solution could be found for any of the five preference specifications. As an alternative, we solved for values of the parameter vector $\boldsymbol{\theta}$ that minimize the sum of squared differences (SSD, with differences expressed in percentage terms) between observed and predicted values in Equations (6.1) to (6.6) -- i.e., minimize $\Sigma_i(\varepsilon_i)^2$.

The calibrated parameter results are reported in Table 4 for each specification. Overall, the linear demand specification provides the closest fit, with an SSD=0.000143, followed by the semi-log demand specification (SSD=0.008). The interpretation of the parameters is often different across preference specifications; however, all indirect utility specifications include an additively separable subcomponent of the form φQ^{ψ} representing nonuse values. As expected, these parameters are always found to have positive values, implying that water quality has a positive effect on nonuse related utility. In all but one specification (log-linear), the calibrated value for ψ is less than one, implying a declining marginal effect of water quality on nonuse values. Also, in all specifications the γ parameter determines the marginal effect of water quality on the use-related component of indirect utility. Its calibrated value is consistently positive across specifications. Similarly, the δ parameter determines the marginal effect of income on utility, and its calibrated value is also consistently positive.

In the linear, semi-log, and log-linear demand models, the β, δ, and γ parameters can also be interpreted as representing the marginal effects of travel cost, income, and water quality on trip demand. When the calibrated value for β has a positive sign, as it does in the three specifications, it implies a negative effect of P on trip demand, which is consistent with expectations. The log-linear demand model implies an almost unit elastic trip demand with respect to P, and the linear demand implies that each dollar decrease in round trip costs increase the annual number of by almost 5. Similarly, the positive calibrated values for the δ parameter imply that trips are a normal good, with an income elasticity between 0.45 and 0.68 in the semi-log and log-linear models.

5.1.4 Predicting Values with the Calibrated Parameters.

To further evaluate the calibrated parameters, we insert them back into the equations in Tables 1 and 2, and we predict X, ΔMCS, and WTP for selected combinations of individual characteristics (Y and P) and changes in water quality (Q_0 and Q_1). The predictions, which are shown in Table 4, provide important additional internal validity checks on the calibrated parameters. For each preference specification, the six numbers in shown in bold italics are the predicted values associated with Equations 6.1 to 6.6. Since these are the equations that were used to calibrate the parameters, the predicted values all match closely with the corresponding values in Table 3. The other values reported in Table 4 include (1) predicted average travel cost for the CV sample, (2) predicted trips for the two samples under different water quality conditions, (3) predicted ΔMCS values for the CV sample, and (4) predicted WTP values for the travel cost sample.

In the linear and log-linear demand models, the predicted average travel cost for the CV sample is respectively 1 percent and 15 percent higher than for the travel cost sample, and in the Stone-Geary model it is 9 percent lower. In contrast, the modified CES and log-linear models predict average travels costs for the CV sample that are more than double. Two opposing effects make it difficult to form strong priors about the expected sign and magnitude of these differences. On the one hand, the predicted average travel cost for the CV sample should be higher than for the travel cost sample because the former includes nonusers who are expected on average to live farther from the water resource. On the other hand, the CV sample's average income is 10 percent lower, which implies a lower opportunity cost for travel. Nevertheless, the Stone-Geary results, with 9 percent lower travel costs for the CV sample, do not seem plausible.

Compared to the linear, semi-log, and Stone-Geary models, the modified CES and log-linear models also predict that trip demand is much less sensitive to water quality changes. The log-linear model shows virtually no changes in trips even for large changes in water quality, whereas the linear and semi-log models predict that trips for the travel cost (user) sample would more than double (from 7.22 to almost 19 trips per year) if water quality increased from boatable to swimmable. On the other end of the spectrum, the Stone-Geary specification predicts an almost tenfold increase in trips. The lack of sensitivity of the log-linear model to water quality changes and oversensitivity of the Stone-Geary model cast doubt on the validity of these calibrated preferences for benefits transfer.

For similar water quality changes, all models predict lower ΔMCS for the CV sample, and higher WTP for the travel cost sample. These differences occur because the average income for the CV sample is lower and because all of the models predict higher travel costs and fewer trips for the CV sample. Again, the smallest differences come from log-linear model.

The values reported in Table 4 are fundamentally "in-sample" predictions, because they are based on observed conditions in the two source studies. In Figures 1 through 3, we use a broader set of conditions to evaluate the calibrated models as transfer functions for predicting WTP. For these figures, we selectively vary water quality changes, income, and travel cost, and we compare WTP predictions across the preference specifications.

Figure 1 shows how predicted WTP for a 1 unit change in water quality (on a 10 point scale) varies with respect to baseline water quality (Q_0). Income is held constant at $45,000 and travel cost at $18 for all predictions. In each case, WTP is highest when starting from the lowest baseline level (Q_0=1), and it decreases as long as Q_0 is less than 5 (below fishable). Above the fishable level, however, predicted WTP for a unit change is U-shaped for all specifications, except the log-linear and Stone-Geary models which predict monotonically declining WTP. None of these WTP predictions are implausible, but the semi-log model is distinctly more convex than the other models.

Figure 2 shows predicted WTP for a change in water quality from fishable (Q_0=5.1) to swimmable (Q_1=7) conditions when average annual household income is varied between

$30,000 and $70,000 and average travel cost is held constant at $18 per round trip. Figure 3 shows predicted WTP when average travel cost is varied between $16 and $20 and average household income is held constant at $45,000. As expected, all specifications predict increasing WTP with respect to income and decreasing WTP with respect to travel costs. The log-linear demand model is least sensitive to both types of variation. In particular, it shows almost no sensitivity to changes in travel cost, which again casts doubt on the validity of this specification for representing preferences for water quality changes. Again, on the other end of the spectrum, the Stone-Geary model exhibits extreme sensitivity to both income and travel cost changes.[2] In contrast, the linear demand model predicts roughly unit elasticity of WTP with respect to income variation and declining WTP (from $56 to $30) when travel cost increases by 25 percent from $16 to $20. The WTP predictions from the semi-log demand are only slightly less sensitive to income and travel cost changes than the linear demand model. The modified CES shows similar sensitivity to income, but is relatively insensitive to travel cost changes.

5. Discussion and Conclusions

This paper demonstrates how the preference calibration method for developing structural benefit transfer functions can be generalized to several alternative preference specifications and can be expanded to include nonuse values. In addition to using a modified CES utility specification, similar to the one used in previous applications, we calibrated preference parameters using four other specifications. In each case, we combined summary data and estimates from a travel cost study with estimates from a contingent valuation study. Both of these source studies estimate nonmarket values for specific improvements in river water quality,

[2] Below $45,000 income and above $18 travel cost, which are close to the values where the model was calibrated, the Stone-Geary model predicts 0 trips. As a result, the Stone-Geary curves "flatten out" in these regions and only reflect nonuse values.

and they also provide information on average use levels (trips), travel costs, and incomes for their respective samples.

For each preference specification, we calibrated six preference parameters. These parameters have somewhat different interpretations and roles in the respective specifications; however, their calibrated values all have plausible signs. For example, the parameters φ, ψ, and γ are all directly related to the marginal utility of water quality improvements, and as expected they are all calibrated with positive signs.

To more thoroughly evaluate the parameter estimates and their implications for benefits transfer, we apply them to the Hicksian WTP functions derived from each specification. In effect, this gives us a calibrated benefit transfer function for each specification, which we use to predict average WTP for selected combinations of water quality levels and changes, income, and travel costs. This process mimics how the functions would be used to estimate benefits for selected policy conditions and changes.

The results show that the structural benefit transfer estimates can be very sensitive to the selection of preference specification. However, they also highlight the strengths and limitations of different specifications, by providing plausibility checks on the range of predicted outcomes.

The linear demand model provides the most consistently plausible results, with (1) positive WTP between \$15 and \$25 for each unit increment in water quality, (2) close to unit elasticity of WTP with respect to income variation, and (3) declining WTP with respect to travel cost. The semi-log demand and modified CES specifications also produce sensible estimates of WTP; however, the semi-log demand model produces WTP estimates that are notably more convex with respect to baseline water quality than other specifications, and the modified CES estimates are relatively insensitive to differences in travel cost. In contrast, the Stone-Geary model produces the least reliable results. In particular, the WTP estimates are implausibly sensitive to both income and travel cost differences. The results from the log-linear model are also somewhat suspect for opposite reasons -- they show virtually no sensitivity to travel cost differences and very low sensitivity to income changes.

In addition to providing structural WTP functions, the preference calibration results can also be used to specify functions for predicting trip demand (or travel costs) and Marshallian consumer surplus. These predictions are not only relevant for policy analysis (as measures of behavioral changes and use values), but they also provide secondary checks on the plausibility of the calibrated results. These secondary predictions (reported in Table 4) confirm the findings from the WTP functions – that the linear demand, semi-log demand, and modified CES specifications generate more plausible estimates than the log-linear and , in particular, the Stone-Geary specifications.

A main advantage of structural benefits transfer is that it imposes a degree of internal validity on the benefit transfer process, by requiring consistency with preferences and economic theory. This paper demonstrates how benefit transfer functions that are internally consistent with different preference specifications can be developed. However, more research is required to determine whether these advantages extent to convergent validity. The existing empirical research evaluating the convergent validity of traditional benefit transfer approaches, where "out of sample" benefit transfer estimates are compared to benefit estimates using original valuation results, has yielded at best mixed results (Shrestha and Loomis, 2003; Downing and Ozuna 1996; Kirchhoff et al., 1997). It remains to be seen whether structural benefits transfer can improve on these results.

An inherent feature of the preference calibration approach described in this paper is that it is most applicable when there are a limited number of available benefit estimates from different nonmarket valuation studies. In this case we use two ΔMCS estimates from a travel cost study and three WTP estimates from a CV study. However, when a large number of such estimates are available (e.g., values related to mortality risks) the logic of preference calibration can in principle be extended to statistical estimation and meta-regression analysis. This concept of "structural meta-analysis," as introduced by Smith and Pattanayak (2002) and discussed in more detail in Smith et al. (2006) and Bergstrom and Taylor (2006), presents a number of empirical challenges, but it continues to be a potentially fruitful area for future research.

References

Bergstrom, John C. and Laura O. Taylor. 2006. "Using Meta-Analysis for Benefits Transfer: Theory and Practice", *Ecological Economics* 60:351-360.

Boyle, Kevin J., P. Joan Poor and Laura O. Taylor. 1999. "Estimating the Demand for Protecting freshwater Lakes from Eutrophication", *American Journal of Agricultural Economics*, Vol. 81 (5): 1118-1122.

Cameron, Trudy A. 1992. "Combining Contingent Valuation and Travel Cost Data for the Valuation of Nonmarket Goods." *Land Economics* 68(3): 302-317.

Carson, Richard T., and Robert Cameron Mitchell. 1993. "The Value of Clean Water: The Public's Willingness to Pay for Boatable, Fishable, Swimmable Quality Water." *Water Resources Research* 29(July): 2445-2454.

Downing, M. and T. Ozuna, Jr. 1996. "Testing the Reliability of the Benefit Function Transfer Approach." *Journal of Environmental Economics and Management* 30(May): 316-322.

Englin, Jeffery, David Lambert, and W. Douglass Shaw. 1997. "A Structural Equations Approach to Modeling Consumptive Recreation Demand." *Journal of Environmental Economics and Management* 33(May): 31-43.

Kennard, D.K. "Depth of Burn." Forest Encylopedia. (Available at http://www.forestencyclopedia.net; accessed February 4, 2004), 2004.

Kirchhoff, S., B. Colby, and J. LaFrance. 1997. "Evaluating the Performance of Benefit Transfer: An Empirical Inquiry." *Journal of Environmental Economics and Management* 33(May): 75-93.

Larson, D.M. 1991. "Recovering Weakly Complementary Preferences." *Journal of Environmental Economics and Managemen.t* 21: 97-108.

Loomis, J and A. Gonzalez-Caban. 1997. "Testing for Differential Effects of Forest Fires on Hiking and Mountain Biking Demand and Benefits." *Journal of Agricultural and Resource Economics.* 26(2):508-522.

Loomis, J and A. Gonzalez-Caban. 1997. "Comparing the Economic Value of Reducing Fire Risk to Spotted Owl Habitat in California and Oregon." *Forest Science.* 43(4):473-482.

Shrestha, Ram K. and John B. Loomis. 2003. "Meta-Analytic Benefit Transfer of Outdoor Recreation Economic Values: Testing Out-of-Sample Convergent Validity". *Environmental and Resource Economics* 25(1): 79-100.

Smith and Desvousges. 1986. *Measuring Water Quality Benefits*. Boston: Kluwer-Nijhoff.

Smith, V.K., W.H. Desvousges, and M.P. McGivney. 1983. "Estimating Water Quality Benefits: An Economic Analysis." *Southern Economic Journal* 50(2):422-437.

Smith, V.K., G. Van Houtven, and S. Pattanayak. 2002. "Benefit Transfer Via Preference Calibration: "Prudential Algebra" for Policy." *Land Economics.* 78(1):132-152.

Smith, V.K. Van Houtven, G.L., S.K. Pattanayak, and T.H. Bingham. 2000. "Improving the Practice of Benefit Transfer: A Preference Calibration Approach." Prepared for the U.S. Environmental Protection Agency, Office of Water, February 2000.

Smith, V.K., and S. Pattanayak. 2002. "Is Meta-Analysis a Noah's Ark for Non-Market Valuation?" Environmental and Resource Economics 22:271–296.

Smith, V.K., S.K. Pattanayak, and G.L. Van Houtven. 2006. "Structural Benefit Transfer: An Example Using VSL Estimates." *Ecological Economics* 60:361-371.

Van Houtven, G.L., S.K. Pattanayak, and V.K. Smith. 2004a. "Benefit Transfer Functions for Avoided Morbidity: A Preference Calibration Approach." National Center for Environmental Economics, Working Paper #04-04.

Vaughan, William J. 1986. *The Water Quality Ladder.* Included as Appendix B in Mitchell, R.C., and R.T. Carson. 1986. *The Use of Contingent Valuation Data for Benefit/Cost Analysis in Water Pollution Control*. CR-810224-02. Prepared for U.S. Environmental Protection Agency, Office of Policy, Planning, and Evaluation. Washington, DC.

Wade, D.D. 1986. Linking fire behavior to its effects on living plant tissue. In: Proceedings, Society of American Foresters Convention, Birmingham, AL. October 5-8. p. 112-116.

Table 1. Algebraic Expressions for Trip Demand and Change in Marshallian Consumer Surplus

Preference Specification	**Trip Demand** $X = X(P, Q, Y; \theta)$ [a]	**Change in Marshallian Consumer Surplus (ΔMCS)** $\Delta MCS = M(P, Q_o, Q_1, Y; \theta)$
Modified CES:	$X_i = \frac{\alpha_A}{\delta_A} \cdot \frac{Y}{\left(P - Q_i^{\gamma_A}\right)}$	$\Delta MCS = \frac{\alpha_A}{\delta_A} \cdot Y \cdot \left[\ln\left(P - Q_0^{\gamma_A}\right) - \ln\left(P - Q_1^{\gamma_A}\right)\right]$
Linear Demand:	$X_i = \alpha_B - \beta_B P + \delta_B Y + \gamma_B Q_i$	$\Delta MCS = \frac{\left(X_1^2 - X_0^2\right)}{2\beta_B}$
Semi-log Demand:	$X_i = \exp\left(\alpha_C - \beta_C P + \delta_C \ln(Y) + \gamma_C Q_i\right)$	$\Delta MCS = \frac{\left(X_1 - X_0\right)}{\beta_C}$
Log-linear Demand:	$X_i = e^{\alpha_D} Q_i^{\gamma_D} P^{-\beta_D} Y^{\delta_D}$	$\Delta MCS = \frac{\left(X_1 - X_0\right) \cdot P}{\left(\beta_D - 1\right)}$
Stone-Geary:	$X_i = \left(\frac{Q_i^{\gamma_E}}{1 + Q_i^{\gamma_E}}\right) \cdot \left(\frac{Y^{\delta_E} - \alpha_E}{P \cdot \left(\delta_E Y^{\delta_E - 1}\right)}\right) + \frac{\beta_E}{\left(1 + Q_i^{\gamma_E}\right) \cdot \left(\delta Y^{\delta_E - 1}\right)}$	$MCS_i = \left(\frac{Q_i^{\gamma_E}}{1 + Q_i^{\gamma_E}}\right) \cdot \left(\frac{Y^{\delta_E} - \alpha_E}{\delta_E Y^{\delta_E - 1}}\right) \bullet \left\{ \ln\left[Q_i^{\gamma_E} \left(\frac{Y^{\delta_E} - \alpha_E}{-\beta_E}\right)\right] - \ln(P) - 1 \right\} - \frac{\beta_E P}{\left(1 + Q_i^{\gamma_E}\right) \cdot \left(\delta_E Y^{\delta_E - 1}\right)}$ $\Delta MCS = MCS_1 - MCS_0$

[a] The corresponding inverse demand functions can be specified by solving for P:

$$P = X^{-1}(X, Q, Y; \theta)$$

Table 2. Algebraic Expressions for Hicksian WTP

Preference Specification	**Hicksian WTP** $WTP = W(P, Q_0, Q_1, Y; \theta)$
Modified CES:	$\mathrm{WTP} = \mathrm{Y} - \dfrac{\left[\left(\varphi Q_0^{\psi}\right) - \left(\varphi Q_1^{\psi}\right) + \left((P - Q_0^{\gamma})^{-\alpha} y^{\delta}\right)^{\beta}\right]^{\frac{1}{\beta \cdot \delta}}}{\left(P - Q_1^{\gamma}\right)^{\frac{-\alpha}{\delta}}}$
Linear Demand:	$WTP = \dfrac{1}{\delta}\left[\left(X_1 - \dfrac{\beta}{\delta}\right) - \left(X_0 - \dfrac{\beta}{\delta}\right) \cdot \exp\left(\dfrac{\delta\gamma}{\beta}(Q_0 - Q_1)\right) + \varphi(Q_1^{\psi} - Q_0^{\psi}) \cdot \exp\left(\dfrac{\delta}{\beta}(\beta P - \gamma Q_1)\right)\right]$
Semi-log Demand:	$\mathrm{WTP} = \mathrm{Y} - \left\{ \mathrm{Y}^{-\delta} \cdot \left[\mathrm{Y} - \dfrac{(X_1 - X_0)}{\beta/(1-\delta)}\right] - \varphi\left(Q_1^{\psi} - Q_0^{\psi}\right) \cdot (1-\delta) \right\}^{\frac{1}{(1-\delta)}}$
Log-linear Demand:	$\mathrm{WTP} = \mathrm{Y} - \left\{ \mathrm{Y}^{-\delta} \cdot \left[\mathrm{Y} - \dfrac{(X_1 - X_0) \cdot P}{(\beta - 1)/(1-\delta)}\right] - \varphi\left(Q_1^{\psi} - Q_0^{\psi}\right) \cdot (1-\delta) \right\}^{\frac{1}{(1-\delta)}}$
Stone-Geary:	$\mathrm{WTP} = \mathrm{Y} - \alpha - \beta P - \left(1 + Q^{\gamma}\right) \bullet$ $exp\left(\dfrac{\varphi\left(Q_0^{\psi} - Q_1^{\psi}\right)}{1 + Q_1^{\gamma}} + \left(\dfrac{Q_0^{\gamma}}{1 + Q_1^{\gamma}}\right) \cdot \ln\left(\dfrac{Q_0^{\gamma}}{-\beta P}\right) - \left(\dfrac{Q_1^{\gamma}}{1 + Q_1^{\gamma}}\right) \cdot \ln\left(\dfrac{Q_1^{\gamma}}{-\beta P}\right) + \left(\dfrac{1 + Q_0^{\gamma}}{1 + Q_1^{\gamma}}\right) \cdot \ln\left(\dfrac{Y - \alpha - \beta P}{1 + Q_0^{\gamma}}\right)\right)$

Table 3. Summary Estimates and Data from Water Quality Valuation Studies[a]

	Travel Cost Study (Smith et al., 1983)		Contingent Valuation Study (Desvousges et al., 1987)		
	(1)	**(2)**	**(3)**	**(4)**	**(5)**
Mean Household Income[b] (Y)	$46,400	$46,400	$41,977	$41,977	$41,977
Mean Travel Cost[b] (P)	$18	$18	n.r	n.r	n.r
Mean Number of Trips (X_0)	***7.22***	7.22	2.41	n.r	n.r
Initial WQ (Q_0)	2.5	2.5	2.5	5.1	0.5
Improved WQ (Q_1)	5.1	7.0	5.1	7.0	2.5
Mean Change in MCS[b] (ΔMCS)	***$14.57***	***$30.58***			
Mean Willingness to Pay[b] (WTP)			***$37.81***	***$26.64***	***$52.64***

n.r. = not reported

[a] The values in italics define the six conditions used to calibrate the six preference parameters

[b] In 2005 dollars

Table 4. Calibrated Parameters and Predicted Values for Six Preference Specifications

Calibrated Preference Parameters		Predicted Values	Travel Cost Study (1)	Travel Cost Study (2)	CV Study (3)	CV Study (4)	CV Study (5)
MODIFIED CES		**P**			\$44.35	\$44.35	\$44.35
α_A= 0.001640	δ_A=0.680081	**X_0**	***7.12***	7.12	2.41	2.53	2.31
β_A=0.105856	φ_A=0.000197	**X_1**	8.20	9.16	2.53	2.62	2.41
γ_A=0.901721	ψ_A=0.501745	**ΔMCS**	***\$15.73***	***\$28.16***	\$5.09	\$3.70	\$4.13
SSE = 0.029351		**WTP**	\$55.58	\$90.80	***\$41.41***	***\$24.51***	***\$50.92***
LINEAR DEMAND		**P**			\$18.22	\$18.22	\$18.22
α_B= 50.06902	δ_B= 0.000842	**X_0**	***7.22***	7.22	2.41	9.12	-
β_B= 4.910293	φ_B= 134.65826	**X_1**	13.93	18.84	9.12	14.03	2.41
γ_B= 2.582500	ψ_B= 0.229865	**ΔMCS**	***\$14.46***	***\$30.84***	\$7.88	\$11.57	\$0.59
SSE = 0.000143		**WTP**	\$44.44	\$75.74	***\$37.87***	***\$26.53***	***\$52.73***
SEMI-LOG DEMAND		**P**			\$20.76	\$20.76	\$20.76
α_C= 0.759377	δ_C= 0.686323	**X_0**	***7.19***	7.19	2.41	4.19	1.57
β_C= 0.371970	φ_C= 0.044946	**X_1**	12.52	18.79	4.19	6.29	2.41
γ_C= 0.213500	ψ_C= 0.428440	**ΔMCS**	***\$14.34***	***\$31.18***	\$4.80	\$5.64	\$2.25
SSE = 0.008004		**WTP**	\$52.24	\$89.98	***\$40.19***	***\$25.18***	***\$51.60***
LOG-LINEAR DEMAND		**P**			\$51.54	\$51.54	\$51.54
α_D= 0.000103	δ_D= 0.453201	**X_0**	***7.219***	7.219	2.407	2.407	2.406
β_D= 1.000979	φ_D= 0.0465354	**X_1**	7.220	7.220	2.407	2.407	2.407
γ_D= 0.000177	ψ_D= 1.1985167	**ΔMCS**	***\$16.78***	***\$24.23***	\$16.02	\$7.11	\$36.15
SSE = 0.069741		**WTP**	\$41.31	\$68.47	***\$39.46***	***\$25.96***	***\$50.98***
STONE-GEARY		**P**			\$16.43	\$16.43	\$16.43
α_E= 0.000414	δ_D= 0.94819	**X_0**	***7.18***	7.18	2.41	43.73	-
β_E= -1,527.578	φ_E= 0.003731	**X_1**	48.79	67.25	43.73	62.07	2.41
γ_E= 0.042245	ψ_E= 0.19086	**ΔMCS**	***\$15.26***	***\$29.37***	\$11.47	\$11.73	\$0.03
SSE = 0.008398		**WTP**	\$46.29	\$75.20	***\$39.61***	***\$25.41***	***\$52.00***

Figure 1. Predicted WTP for a Unit Change in Water Quality: Sensitivity to Baseline Water Quality

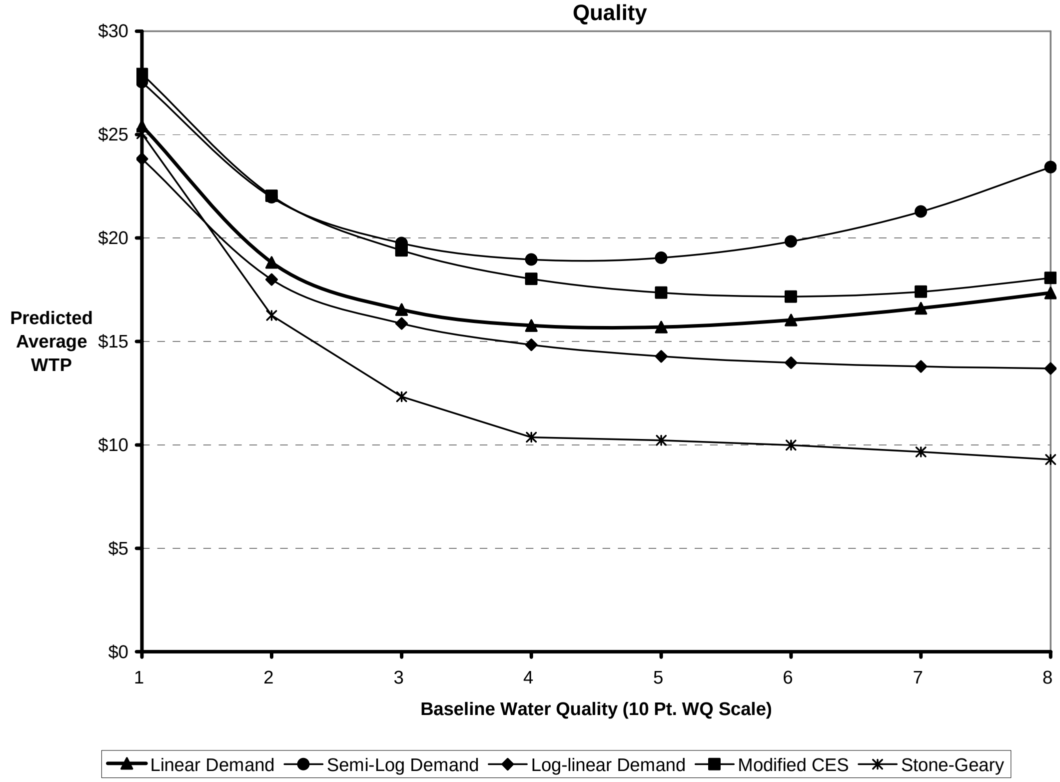


Figure 2. Predicted WTP for a Boatable-to-Fishable Water Quality Change: Sensitivity to Income

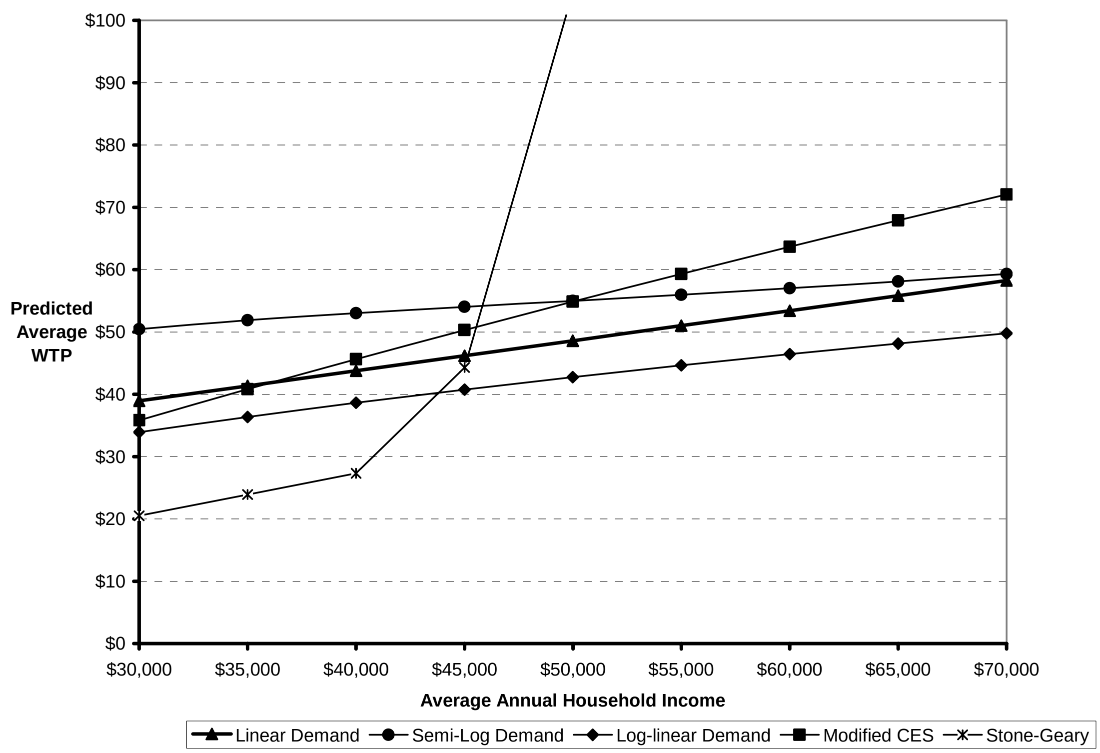
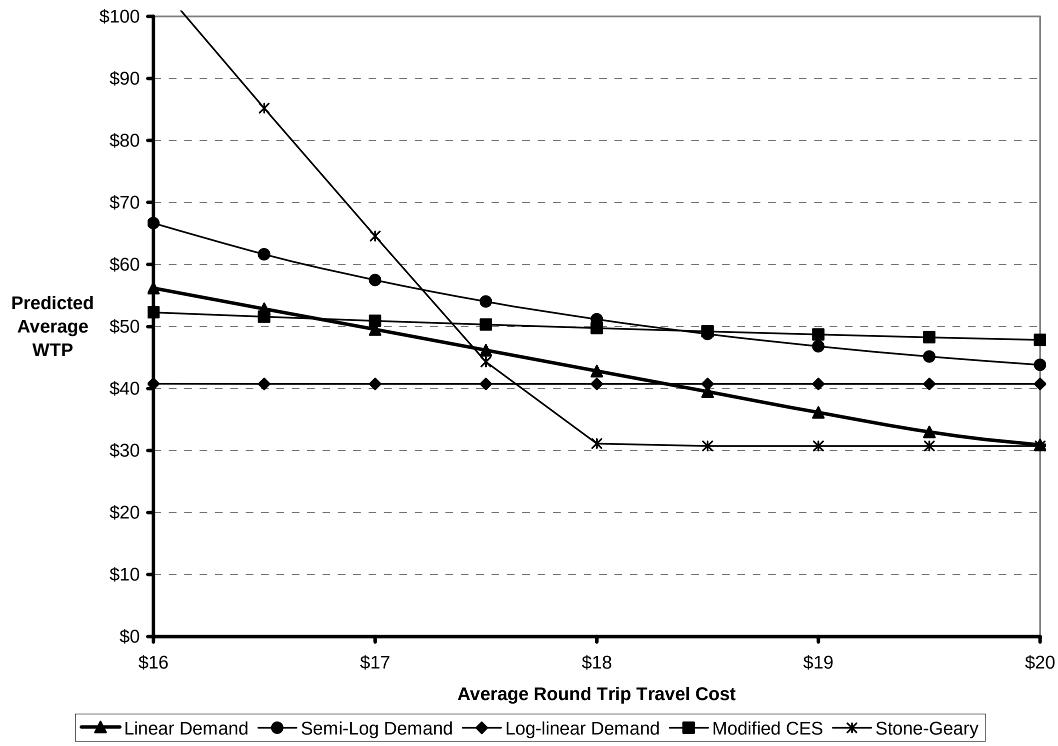


Figure 3. Predicted WTP for a Boatable-to-Fishable Water Quality Change: Sensitivity to Travel Cost



SPLIT-SAMPLE TESTS OF "NO OPINION" RESPONSES IN AN ATTRIBUTE BASED CHOICE MODEL

By

Eli P. Fenichel,[1] Frank Lupi,[2] John P. Hoehn,[3] and Michael D. Kaplowitz[4]

1 Graduate Student, Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824; fenichel@msu.edu

2 Associate Professor, Department of Agricultural Economics, and Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824-1039; Lupi@msu.edu.

3 Professor, Department of Agricultural Economics, Michigan State University, East Lansing, MI 48824; Hoehn@msu.edu

4 Associate Professor, Department of Community, Agriculture, Recreation and Resource Studies, Michigan State University, East Lansing, MI 48824-1222. Kaplowit@msu.edu

Keywords: Stated preference, contingent valuation, choice experiment, attribute based method, no opinion, internet survey.

The research reported here was supported in part by the Science to Achieve Results (STAR) Program, U. S. Environmental Protection Agency. The authors are solely responsible for any errors.

ABSTRACT

Researchers conducting surveys that elicit preferences must decide whether to formally include response options that allow respondents to express “no opinion.” Using a split-sample design, we explore the implications of alternative formats for including, or not including, “no opinion” response options in an attribute based choice experiment. We provide evidence that using multiple “no opinion” responses may help researchers differentiate between respondents who choice a “no opinion” option due to satisficing and respondents that are indifferent between alternatives. Although there is literature suggesting that “no opinion” responses can be recoded as “no” responses in the case of referendum-based contingent valuation, in our case recoding “no opinion” responses as if they were “no” responses yielded substantially disparate results.

I. INTRODUCTION

In surveys eliciting stated preferences, some respondents do not state a preference, opting instead to answer a choice question with a response such as “don’t know”, “not sure”, or “would not vote.” These responses are variants of the “no opinion” responses discussed in more general survey research (Krosnick 2002). Treatment of “no opinion” responses in stated preference studies has largely focused on studies that use the contingent valuation method (CVM). The attribute-based method (ABM), also called choice experiments or stated choice, is a comparatively new technique that is related to, and has grown out of, CVM (Holmes and Adamowicz 2003; Foster and Mourato 2003, Louviere et al, 2000). The ABM presents respondents with a set of attributes of a good, where typically one attribute is price. The attributes and prices are varied across respondents. This differs from CVM where typically only price is varied across respondents. Thus ABM allows the researcher to value the implicit price for each attribute, much like a hedonic price study (Holmes and Adamowicz 2003). Both CVM and ABM often involve discrete choice responses, and as a result random utility models can be used in the estimation of both methods. Indeed, CVM is often considered a special case of ABM (Boxall et al. 1996).

In many ABM-based studies, respondents have been asked to choose between two or more attribute-price sets. This is similar to the referendum style questions commonly used in CVM, especially in the case where one attribute-price set is treated as a *status quo*. The National Oceanic and Atmospheric Administration (NOAA) panel recommended including a “no vote” option for binary choice CVM studies (Arrow et al. 1993). While, this recommendation has spawned a growing body of research on how to treat “would not vote” and other types of “no opinion” responses in the CVM literature, the issue has received less attention in ABM studies.

The literature on ABM does contain a related, but logically distinct, strain of research. In some ABM studies, respondents are presented with a choice set that includes several alternatives composed of varied attributes and a “none” alternative (Louviere et al, 2000) or an “opt-out” alternative (Boxall et al. 1996). In the setting of a product choice, the “none” option might be treated as a “don’t buy” decision. In a recreational site choice context, the “none” option might represent a no-trip decision or it might represent a trip to a site not included in the choice set (Banzhaf et al. 2001). In other settings, the “none” option may be considered a choice to maintain the status quo. Typically, researchers explicitly model this type of alternative as one of the elements in a multinomial choice model. In contrast, here we consider a distinct issue in the ABM, in which a failure of respondents to choose an alternative is not a choice for the status quo. Instead, we examine the instance in which respondents’ failure to choose one of the ABM alternatives is akin to a “no opinion” response.

There is growing evidence in the CVM binary choice literature that “no opinion” responses should not be treated as “for” votes (Groothuis and Whitehead 2002; Caudill and Groothuis 2005; Carson et al. 1998). However, there is not yet agreement as to whether “no opinon” responses should be treated conservatively as “against” votes (Carson et al. 1998; Kronsick 2002), or whether no opinion responses may represent cognitive difficulties, potentially resulting from an indifference in utility, and therefore should be treated as a truly unique response (Krosnick et al. 2002; Evans et al. 2003; Alberini et al. 2003; Caudill and Groothuis 2005; Champ et al. 2005). Furthermore, even those who believe that no opinion responses should be treated as unique responses largely base their argument on improving econometric efficiency with few arguing that the conservative approach yields inconsistent estimates. Groothuis and Whitehead (2003) observe that the appropriateness of treating no opinion responses as unique or "against" votes may depend on whether the study is attempting to measure willingness-to-pay (WTP) or willingness-to-accept (WTA).

Arguments for treating no opinion responses as unique are typically based on Wang's (1997) hypotheses on why a respondent may choose a no opinion response. Wang (1997) posits that there are four general categories of respondents who choose no opinion responses: 1) those who reject the CVM scenario, 2) those who know their preference and decline to answer, 3) those who make an effort and are truly unsure, and 4) those who do not make an effort and are therefore unsure.

Kronsnick et al. (2002) also present an analysis of why a respondent may choose a no opinion response. They present evidence that often no opinion responses are the result of satisficing, or simply that the "work" involved with answering the question is too great and a no opinion response involves the least work or the lowest risk.[1] Kronsnick et al. (2002) also discuss an alternative hypothesis regarding no opinion responses; the respondent's optimizing process may result in true indifference making the respondent truly unsure when the choices are "close" in terms of the associated net benefits or welfare yields. Therefore, a respondent may reply with a no opinion response because they are indifferent in a utility sense. However, it is unlikely that there is a clear line between a no opinion response resulting from optimizing and from satisficing since a respondent may begin optimizing, but may "give-up" before reaching true indifference.

More recent investigations by Alberini et al. (2003), Caudill and Groothuis (2005) and Evans et al. (2003) have aimed to improve estimation efficiency through "sorting" no opinion responses, especially focusing on identifying and making use of responses that would fall into Wang's (1997) latter two categories or that may be considered to be cases of optimizing as asserted by Krosnick et al. (2002). However, there has been little effort to sort no opinion responses that result from other phenomena; for example, no opinion responses that result from respondents being unsure due to utility indifference, and no opinion responses that result from respondents that are satisficing. Moreover, all the work to date has been based on ordinal polychotomous-choice and multi-bounded questions, which introduce other types of difficulties (Vossler and Poe 2005).

There also remains some question about the comparability of ABM studies to CVM studies (Stevens et al. 2000; Foster and Mourato 2003). ABM studies may be cognitively more difficult than CVM studies and ask respondents to explore their preferences in more detail (Stevens et al. 2000). This may result from the explicit substitutes in the ABM format. Furthermore, the multidimensional trade-offs implicit in ABM may result in a larger number of respondents who honestly “don’t know” or are closer to indifference relative to CVM. To date, there have been no studies examining whether reclassifying no opinion responses in ABM as “against” responses, considered a conservative classification in CVM, yields estimates that are consistent with similar surveys where a “no opinion” option is not offered.

This paper presents an examination of two research questions on no opinion responses in ABM studies. First, does recoding no opinion responses as “against” provide estimates consistent with those derived from surveys where there is no option of expressing no opinion? Secondly, does offering respondents with two qualitatively different no opinion responses allow expressions of welfare indifference to be sorted from those who express no opinion for other reasons? This latter issue may be generalizable to CVM because it attempts to distinguish Wang's (1997) third type of response (indifferent or too close to call) from Kronsnick et al.’s (2002) satisficing or other variants of “no opinion.”

II. SURVEY INFORMATION

A binary choice ABM survey was implemented using a web-based method with a split-sample design. In addition to the usual experimental design of the attributes, there were four unique versions of the ABM survey that differed in the response options respondents faced for their choice questions. The four sets of response formats were:

(i) "yes", "no", "too close to call" (TCC), and "not sure" (NS) (all options treatment),

(ii) "yes" and "no" (yes/no treatment),

(iii) "yes", "no", and NS (NS treatment), and

(iv) "yes", "no", and TCC (TCC treatment),

where the last expression in parentheses is what the four treatments will be called.

The TCC response is intended to reflect situations close to indifference. Collectively the NS and TCC responses are referred to as "no-opinion" responses as a shorthand to refer to respondents that did not explicitly choose yes or no in the choice scenario. The surveys that were distributed across the four groups of response categories all utilized the same experimental design for the ABM attributes.

The web-based ABM survey elicited preferences for inland, freshwater wetland mitigation. The questionnaire was developed using a series of focus groups and pretest interviews (Kaplowitz et al. 2004), and the policy setting and choice questions follow that of the paper instrument discussed in Lupi et al. (2002). Each respondent was presented with the characteristics of a common wetland that had already been approved for drainage ("drained wetland") and the characteristics of a wetland being proposed as compensation ("restored wetland") for the wetland to be drained. The attributes of the wetlands presented to respondents were wetland *type* (wooded, marsh, mixed), *size* (acres), public *access* attributes, and *habitat* attributes (see Appendix for sample choice question). The respondents were then asked, “In your opinion, is the restored wetland good enough to offset the loss of the drained wetland?” Each respondent was asked to make up to five comparisons, but each respondent was only exposed to one response option format. Details of web survey design, administration, and general results are reported in Hoehn et al. (2004).

III. RESPONSE FREQUENCY ANALYSIS

As mentioned above, the survey design incorporated four different sets of response options. Response category statistics for the completed choice questions are presented in Table 1.[2] As expected, the response treatment including all options (“all options”) resulted in the highest proportion of “no opinion” responses (25%). Chi-square tests where used to compare the probability of a “no opinion” response across the four different survey response treatments and results are presented in Table 2.[3] Table 2, section A, shows that the proportion of “no opinion” responses is significantly different when all four response options are presented to respondents as compared to instances in which one type of “no opinion” response is available to respondents. This is true at all common significance levels. It seems clear from these results that respondents are more likely to choose a “no opinion” response option when both the TCC and NS options are available to them as part of their response choice set. A chi-square test comparing the TCC survey treatment and the NS survey treatment yielded a low p-value (< 0.016). This result suggests that the TCC and NS response options are not viewed as equivalent response options by respondents, and indicates that the wording of the “no opinion” options may matter.

Carson et al. (1997) used chi-square tests to determine the effect of no opinion responses on the proportion of "yes" and "no" responses in a CVM study. A similar analysis was conducted for the ABM data, and the results are displayed in Table 2 section B. The proportion of "yes" to "no" responses was significantly different, at the 95% confidence level, between surveys that did not allow respondents to express "no opinion" and surveys that offered either TCC or NS as response options. The chi-square analysis of the proportion of responses when both "no opinion" responses were offered (the all options version) against the instances when only "yes" and "no" responses were offered yielded a p-value of 0.07. This p-value implies that the null hypothesis of no significant difference between these two proportions should not be rejected at the 95% confidence level, but may be rejected at the 90% confidence-level. This difference may not be significant at the traditional 95% confidence level but may yield different economic results. That is, the yes's and no's from these two groups may produce different estimates of WTA.

Further examining the response categories, "no opinion" responses were pooled with "no" responses, and retested against the yes-no ratio from the survey treatment that only allowed "yes" or "no" responses (Table 2 section C). All chi-square tests for all of these comparisons yielded p-values < 0.05. This result implies that pooling "no opinion" responses with "no" responses, as suggested by Carson et al. (1998), results in significantly different yes-no ratios, in contrast to the findings of Carson et al. (1998) for CVM. It remains unclear in the "all options" case, where both TCC and NS were presented as response options, whether both TCC and NS pulled equally from "yes" and "no" responses.

The distribution of yes-no ratios across response formats that allowed for a "no opinion" response was also tested (Table 2 section D). The ratio of "yes" to "no" responses did not change significantly when TCC or NS was offered as the “no opinion” response option. The distribution of yes and no responses when both NS and TCC response options where available as response choices was compared to the distribution of yes and no responses when only one “no opinion” response option was presented and were found to be significantly different at the 95% confidence level. That is, when more than one “no opinion” option was presented to respondents, the proportion of yes and no responses differed significantly.

These results indicate that survey participants may respond to the phrasing, language, or number of “no opinion” response items lending evidence to the hypothesis that various no opinion responses may represent unique types of responses. Further, these results suggest that “no opinion” responses do not pull evenly from “yes” and “no” responses and that, unlike Carson et al.’s (1998) CVM study, these responses do not consistently pull from “no” responses. It appears in this instance that no opinion responses pull more heavily from “no” responses– see Table 1. Moreover, “no opinion” responses seem to pull more evenly from “yes” and “no” responses when both TCC and NS are presented as options as opposed to when only one type of no opinion response is available (Tables 1 and 2). It appears that the marginal impact of adding a second “no opinion” response option is to pull more from “yes” than “no”, even when the first “no opinion” response option pulled more from “no” than “yes”.

There are three potential explanations for the apparent divergence in results between this ABM study and previous CVM studies. First, the underlying ABM study focuses on respondents’ WTA compensation (Groothuis and Whitehead 2003) as measured by in-kind trade-offs. Second, there may be something unique to the ABM response format that is different from CVM studies. Thirdly, it is possible that the additional “no opinion” response option causes responses to pull more evenly from both “yes” and “no.” Based on response ratios, TCC and NS responses seem to be good substitute responses when only one of the response options is available to respondents. When both TCC and NS are present, it may be presumed that a TCC response may involve, perhaps, an attempt by respondents to optimize, especially if it is assumed that this response is indeed qualitatively different from a more general NS response.[4] Next, we explore possible response category effects of welfare estimates.

IV. EFFECTS ON WELFARE

The wetlands mitigation survey used in this study asked respondents to make an in-kind tradeoff between acres of drained and restored wetlands. In essence, respondents were asked if restoration of a larger wetland would compensate for the loss of an existing wetland. This makes acres of wetlands the unit of currency for this study. Various quality attributes for the wetlands were also included in choice sets, and these act to shift demand for wetland acres. Responses were coded into 11 response variables. These variables included change in wetland acreage (effectively price), dummy variables for capturing changes in wetlands' general vegetative structure, public access, and habitat conditions for amphibians, songbirds, wading birds, and wildlife flowers (changes could be poor to good or good to excellent). Changes in wetland acres where recorded as the change in the total number of acres. Dummy variables where coded as one for a positive change, zero for no change, and -1 for a negative change. Changes from poor to excellent where indicated by both the poor to good and good to excellent dummy variables being coded as one (other coding followed this pattern). A change from no access to access was coded as one (-1 for the other direction), while changes in wetland type where coded as one if there was a change.

In-kind welfare measures can be estimated using random utility theory (Holmes and Adamowicz 2003). A random effects logit model that addresses the panel data was estimated for each of the four survey response format versions, and parameter ratios were used to calculate the minimum WTA in acres of restored wetland per acre of drained wetland (Table 3, first row). Specifically, WTA *ceteris paribus* was found by dividing the constant parameter by the negative of the marginal utility of acres. All models fit the data, with log-likelihood ratio tests against a model with a single choice dummy being significant at all common significance levels.

Each model included all variables, though not all coefficients estimated where significant at the 90% or 95% confidence level. In all models, estimates for the parameter associated with improving wild flower habitat from poor to good where not significant at that 95% confidence level (Table 3). The parameters associated with other variables that were not significant are indicated in Table 3. The parameter associated with wetland acres was significant at the 95% confidence level for all models.

Estimation results can be interpreted as the marginal implicit prices, in kind, associated with the change defined by the variable. For the constant term, the marginal implicit price is the change in acres required to maintain the same level of utility. That is, if the WTA estimate were zero then one acre restored wetland would be adequate compensation for one acre of drained wetland. In cases in which only "yes" and "no" options were presented to respondents, a restored wetland could have up to eight fewer acres for each acre of the drained wetland, *ceteris paribus*, before respondents would prefer the drained wetland (Table 3, first row). This may reflect a preference toward getting something out of a restoration project as opposed to not getting any restoration. In cases in which there were "no opinion" responses, dropping the "no opinion" responses from the analysis yielded WTA estimates that were closest to those derived from the yes/no format. The WTA estimates, *ceteris paribus,* varied greatly across response treatments. The WTA estimates showed that more than three times less compensation was demanded by respondents when “no opinion” responses were dropped as opposed to pooled with no’s. Recoding the “no opinion” responses as “no” responses in the all response options format makes the ratio of “yes” to “no” less than one (Table 1) and causes the WTA to be positive, i.e., one acre drained required more than one acre to offset the loss.[5]

An important aspect of the ABM is that allows the relative importance of the attributes to be ranked. From above we saw that recoding the data changed the yes to no ratio, and therefore, we affected the constants, as expected. However, recoding the data to address the alternative approaches for treating the no opinion responses should not affect the ranking of attributes if “no opinion” responses represent satisficing. To investigate if recoding affected the relative importance of attributes, the marginal implicit prices associated with each attribute variable (Table 3) where ranked from the largest marginal impact to lowest marginal impact (Table 4). Changing wildflower habitat from poor to good had the lowest maximum difference in rank (excluding a change in wetland type which has a negative value), though these marginal implicit prices where calculated based on parameters that were not significant at that 95% confidence level. Improving wading bird habitat from poor to good consistently ranked as having a high marginal implicit price ranging (median rank of 2 a maximum difference in ranks of 3). Changes in song bird habitat from poor to good also had a high median rank, 2, but had a maximum difference in ranks of 5. This difference is driven by the ranks associated with the TCC and all options format when the “no opinion” response are pooled with “no.” This provides some evidence that TCC response may not represent satisficing. The attribute ranks for the NS format are identical for all attributes regardless of whether the NS was pooled with “no” or dropped.

This provides some evidence that NS by itself acts more like a “no” response due to satificing than an expression of indifference.

Rank correlations between treatments provide further evidence that TCC and NS responses are not used interchangeably (Table 5). The ordering of marginal implicit prices between TCC treatments (TCC responses pooled with no and dropped) showed a correlation with the ranking of the marginal implicit prices across all treatments. However, there is a decline in the strength of the correlation across formats in the order of NS, all options, and TCC. Moreover, all formats and treatments, except *TCC pooled with no* demonstrated a rank correlation with the yes/no format. The attribute ranks from the TCC format with TCC pooled with “no” did not correlate well with either NS treatment (this approach had the three lowest correlations in the table). That said, some of the approaches gave marginal implicit prices that ranked the attributes in a manner that was highly correlated across the modeling strategies, which would be reassuring for benefits transfer of the attribute valuations.[6] The strong correlations among the yes/no format and the NS format (either treatment) indicate that NS response may represent satisficing. However, the impact of recoding TCC as “no” on the ranks of the attributes indicates that TCC responses may not simply be satisficing.

V. UNDERSTANDING NO OPINION RESPONSES

The evidence presented in the preceding sections of this paper indicates that whether or not to treat “no opinion” responses as “no” responses is not straight forward. Treating “no opinions” as “no” lead to non-positive WTA estimates because of the effect of the recoding on the constant term, and in the case of TCC greatly affect the ranks of the wetland attributes. Therefore, we do not advocate simply treating “no opinion” response as “no” in the attribute-based choice models. It is also unlikely that “no opinion” responses should be treated as “yes” responses. However, “no opinion” responses can make up a substantial portion of survey responses when a no opinion response category is present. In this studies’ survey treatment where all response options were available, 25% of the responses were either TCC or NS, and this leads to two important questions. First, is there evidence that some preference information may be recovered from “no opinion” responses? Second, is there a discernable difference between the responses with a change in wording of “no opinion” responses (i.e., “too close to call” versus “not sure”), aside from the previously discussed effect on attribute ranks?

To address these questions, we used parameter estimates derived from the simple yes/no model to predict “yes” responses for the data that was held aside or reserved for model assessments (see footnote 2). The 1,865 unused (reserved) responses served as a set of “true” responses for testing purposes and were all from the treatment containing all four response options (all options treatment). The model parameters were used to predict the probability of a yes response for the reserved data. If the model has the ability to discern yes from no votes, then for respondents that actually answered yes, we would expect the mean predicted probability of a yes to be larger than the mean predicted probability of a yes for those respondents that actually chose no. Further, if respondents chose either TCC or NS as a result of an attempt to optimize but found the welfare yield to be “close” to their level of indifference, then we would expect the mean predicted value associated with TCC and NS responses to be between the mean predicted value associated with “yes” and “no” responses.[7] This is indeed the case as shown in Table 6.

To test if these means are significantly different from one another, a single factor ANOVA was used. The group mean square is 7.94 and the error mean square is 0.03 yielding an F-statistic = 90.36 with 3 and 1,823 degrees of freedom, which yields a p-value that is essentially zero. This implies that the mean associated with at least one response type is significantly different from the mean associated with at least one other response type. If the model has predictive power, then it should be expected, that at least “no” and “yes” responses were significantly different.

The Tukey test, also known as “the honestly significant difference test” and “wholly significant difference test,” was used to identify the response options that had significantly different means in a set of *post hoc*, pair-wise comparisons (Zar 1996). Tukey tests allow one to determine if there are pairs of means such that the null hypothesis of no difference would not be rejected if just those two means where tested alone. Results are presented in Table 7. The critical value for the Tukey test with error degrees of freedom of 1,823, and four categories at the 95% confidence level is 3.633. All comparisons yielded a Tukey q-statistic greater than the critical value except the NS-TCC comparison (q = 2.954). This result supports the hypothesized expectation that the predicted mean associated with “yes” and “no” responses are indeed different. It is also interesting to note, that these results indicate that both “no opinion” responses are significantly different from both “yes” and “no” responses – implying the model has predictive power. This indicates that “no opinion” responses may indeed reflect that “no-opinion” respondents are near their utility indifference.

An alternative explanation for the means associated with “no opinion” responses lying between the means of “yes” and “no” responses is that the predicted variance associated with “no opinion” responses is significantly large. However, the ANOVA results show that the means are indeed significantly different. In light of these results, in future analyses it may be possible to gleam extra information by treating the “no opinion” responses as a unique answer. It is also possible that by including multiple “no opinion” responses, respondents that would otherwise satisfice are forced to examine their preferences, at least enough to choose between TCC and NS.

VI. CONCLUSION

To our knowledge, this is the first paper to explore the treatment of “no opinion” responses in an ABM setting that tries to differentiate between alternative types of no opinion responses. The differences and similarities between ABM and CVM are well documented (Boxall et al. 1996; Holmes and Adamowicz 2003). Research on how to treat no opinion responses in the CVM literature has been advancing since the NOAA commission made its recommendation to include a "no-vote" option. The work presented in this paper provides contrary evidence regarding conventional wisdom that “no opinion” responses should be treated as “no” responses as in the CVM literature (Carson et al. 1998).

There are two alternative hypotheses that may be used to explain the results presented here. First, the ABM response format may be different enough so that no opinion responses represent optimizing and not satisficing. This may be because the tabular form lessens the cognitive work asked of the respondent (Viscusi and Magat 1987) and facilitates making tradeoffs (Hoehn et al, 2004). However, it may be that the results presented here have more to do with the WTA framing of our choice question, supporting Groothuis' and Whitehead's (2002) findings.

Dropping “no opinion” responses appears to yield results most consistent with surveys that do not offer no opinion response options. As the number of no opinion options increased so too did respondents’ use of those responses. It does seem likely that the inclusion of two no opinion responses eliminate many respondents that may be leaning in a given direction, and potentially would have answered "yes" or "no." It is also likely that by adding a second no opinion response option a disproportionate number of would-be "yes" voters switch to one of the no opinion responses (this may be true even if a disproportionate number of would-be "no" voters would choose “no opinion” when only one no opinion option is available). This result seems to present a tradeoff for researchers. If there is a way to recover information from some no opinion responses, then adding an additional response option may be beneficial. However, if no such tool exists then sample size may be greatly reduced.

In this paper, we present evidence that when two no opinion response options are offered respondents may have used these options differently, to express indifference that may have resulted from optimizing (“too close to call”) as opposed to uncertainty that may have resulted from satisficing (“not sure”). The effect that TCC had on attribute ranks indicates that this option is affecting the decision making process in ways consistent with indifference.

Understanding how to treat response options that allow respondents to express “no opinion” is important to the future development and refinement of attribute-based stated preference techniques. These techniques are increasingly contributing to our ability to measure preferences for goods and services that have non-use values or potential attributes that extend beyond current conditions. This paper provides a first step in understanding how to treat “no opinion” responses in the ABM format, but more work in this area is still needed. Specific areas of future study include investigating if estimating the probability of a “too close to call” response can be used to estimate indifference and improve the ability to predict choices. However more than anything else, more case studies need to be examined, especially cases involving WTP.

VII. APPENDIX I. SAMPLE SURVEY.

Wetlands Scorecard #3:
How do the Drained and Restored Wetlands Compare?

Wetland Features	Drained Wetland	Restored Wetland
Is it marsh, wooded, or a *mix* of marsh and woods?	Wooded	Mixed
How large is it?	12 acres	24 acres
Is it open to public?	Yes	Yes
Are there trails and nature signs?	No	Yes
How good is the habitat for different species?		
Amphibians and reptiles like frogs and turtles	good	excellent
Small animals like raccoon, opossum, and fox	good	good
Songbirds like warblers, waxwing, and vireo	good	--
Wading birds like sandpiper, heron, or crane	--	excellent
Wild flowers?	good	good

What do the habitat ratings mean?

excellent:	The wetland habitat supports these species in better than average numbers and variety; a casual observer is very *likely to see a variety* of these species.
good:	The wetland habitat supports these species in average numbers and variety; a casual observer is *likely to see a few* of these species.
--	The wetland habitat supports these species in very small numbers or not at all; a trained observer is *unlikely to find any* of these species.

Wetland Case #3

The scorecard on the left page compares the natural features of the drained and restored wetlands.

In your opinion, is the restored wetland good enough to offset the loss of the drained wetland in Case #3? (Circle the letter next to your decision)

a. Yes, the restored wetland offsets the loss of the drained wetland

b. No, the restored wetland does not offset the loss of the drained wetland

c. Too close to call

d. Not sure

The Fine Print:

The drained and filled wetlands...

...are *common* wetland types.

...do *not* contain any rare species or rare habitat.

...are the *same* in terms of features not mentioned in the scorecard.

VIII. REFERENCES

Alberini, A., K. Boyle, and M. Welsh. 2003 "Analysis of contingent valuation data with multiple bids and response options allowing respondent to express uncertainty." *Journal of Environmental Economics and Management* **45**: 40-62

Arrow, K., R. Solow, P.R. Portney, E.E. Learner, R. Radner, and H. Schuman. "Report of the NOAA Panel on Contingent Valuation." *Federal Register* **58:10** (Jan. 15, 1993): 4601-4614

Bahnzaf, M. R., Johnson F R, and Mathews K E. 2001. "Opt-out alternatives and anglers' stated preferences," in Bennett J. and Blamey R (eds.) The Choice Modelling Approach to Environmental Valuation. Cheltenham: Edward Elgar Publishing Company.

Boxall, P.C., W.L. Adamowicz, J. Swait, M. Williams, and J. Louviere. 1996 "A comparison of stated preference methods for environmental valuation." *Ecological Economics* **18**: 243-253

Caudill, S.B. and P.A. Groothuis. 2005. "Modeling hidden alternatives in random utility models: an application to "don't know" responses in contingent valuation." *Land Economics* **81**: 445-454

Carson, R.T., W.M. Hanemann, R.J, Kopp, J.A. Krosnick, R.C. Mitchell, S. Presser, R.A. Ruud, V.K. Smith, M. Conaway, and K. Martin. 1998. "Referendum design and contingent valuation: the NOAA panel's no-vote recommendation." *Review of Economics and Statistics* **80**: 484-487

Champ, P.A., A. Alberini, and I. Correas. 2005. "Using contingent valuation to value a noxious weed control program: the effects of including an unsure response category." *Ecological Economics* **55**: 47-60

Evens, M.F., N.E. Flores, and K.J. Boyle. 2003 "Multiple-bounded uncertainty choice data as probabilistic intentions." *Land Economics* **79**: 549-560

Foster, V. and S. Mourato. 2003. “Elicitation Format and Sensitivity to Scope.” *Environmental and Resource Economics* **24**: 141-160

Groothuis, P.A. and J.C. Whitehead. 2002. “Does don’t know mean no? Analysis of ‘don’t know’ response in dichotomous choice contingent valuation questions.” *Applied Economics* **34**: 1935-1940

Hoehn, J.P., F. Lupi, and M.D. Kaplowitz. “Web-based methods for valuing wetland services.” U.S. Environmental Protection Agency report **R827922**. April 19, 2004

Hoehn, J.P., F. Lupi, and M. D. Kaplowitz, 2004. "The effects of questionnaire formats on elicited preferences and values in stated preference experiments," in *Benefits and Costs of Resource Policies Affecting Public and Private Land*, (D. McLeod), Western Regional Research Publication.

Holmes, T.P. and W.L. Adamozicz 2003. “Attribute-based methods” in A Primer on Nonmarket Valuation. eds. Champ. P.A., K.J. Boyle, and T.C. Brown. Kluwer Academic Publishers, Boston, MA. p. 171-219

Kaplowitz, M. D., F. Lupi, and J.P. Hoehn 2004. “Multiple-methods for developing and evaluating a stated choice survey for valuing wetland ecosystems.” In Questionnaire Development, Evaluation, and Testing Methods. Pgs. 503-24. A.S. Presser, ed., John Wiley & Sons, Inc., New York.

Krosnick, J. A. 2002. “The causes of no-opinion responses to attitude measures in surveys: They are rarely what they appear to be.” In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), Survey Nonresponse. New York: Wiley. p. 87-100

Krosnick, J.A., A.L. Holbrook, M.K. Berent, R.T. Carson, W.M. Hanemann, R.J. Kopp, R.C. Mitchell, S. Presser, P.A. Ruud, V. K. Smith, W.R. Moody, M.C. Green, and M. Conaway.

2002. "The impact of "on opinion" response options on data quality, non-attitude reduction or an invitation to satisfice?" *Public Opinion Quarterly* **66**: 371-403

Louviere, J.J., D.A. Hensher, J.D. Swait. 2000. Stated Choice Methods: Analysis and Applications. Cambridge University Press. Cambridge, U.K.; New York.

Lupi, Frank, Michael D. Kaplowitz, and John P. Hoehn. 2002. "The Economic Equivalency of Drained and Restored Wetlands in Michigan," *American Journal of Agricultural Economics* **84**:1355-1361.

Stevens, T.H., R. Belkner, D. Dennis, D. Kittredge, and C. Willis. 2000. "Comparison of contingent valuation and conjoint analysis in ecosystem management." *Ecological Economics* **32**: 63-74

Viscusi, W. K., and W.A. Magat. 1987. Learning about Risk. Harvard University Press, Cambridge, MA.

Vossler, C.A. and G.L. Poe. 2005. "Analysis of contingent valuation data with multiple bids and response options allowing respondents to express uncertainty: a comment." *Journal of Environmental Economics and Management* **49**: 197-200

Wang, H. 1997. "Treatment of "don't know" responses in contingent valuation surveys: a random valuation model." *Journal of Environmental Economics and Management* **32**: 219-232

Zar, J.H. Biostatistical Analysis 3rd edition. Prentice Hall, Upper Saddle River, NJ. 1996, pp. 139, 212-215,492-494

Endnotes

[1] The work requirements may range from physically reading the survey to understanding the question to actually evaluating preferences.

[2] A total of 4,865 responses where received for the “all options” treatment, (i). However 1,865 of these observations where randomly selected and reserved for later use in assessing the model predictions.

[3] All chi-square tests use the Yates correction for variables coming from a binomial distribution (Zar 1996).

[4] It could be argued that one effect of including no opinion options would be to lower item non-response rates. The item non-response rates were as follows: all options (0.8%), Yes/No (0.7%), NS (0.4%), and TCC (1.5%). These differences are statistically significant (p-value < 0.0001). Interestingly, the TCC rate stands out as being higher than the NS rate which provides some further evidence that TCC is being used by respondents differently than NS. Overall though, none of the treatments had substantial levels of item non-response.

[5] Recall under MLE in the logit model, the estimated constant parameter will ensure that the average of the predicted probabilities of yes answers will match the sample share of yes answers. To the extent that the “no opinion” answers are not being explained by the other parameters of the models, then under the recoding of responses the new estimated constant must adjust to match the new sample shares. Thus, recoding of “no opinion” responses as “no” responses has the clear effect of lowering the constant and hence lowering the marginal implicit prices.

[6] Clearly the latter did not apply to the valuations for the constant term where use in benefits transfer would be warrant caution.

[7] Indeed, in a model that had perfect ability to discriminate the choices in accord with the theory, we would expect the predicted yes probabilities to be > 0.5 (<0.5) for the yes answers (for the no answers). For those that selected TCC, we’d expect predicted yes probabilities to be about equal to 0.5 – they point of indifference implied by RUM theory. Hypotheses for the NS option expectations are less clear. Suppose the NS represent saticficing, then really easy choices would get answered and the easiest to answer are the choices where they are clearly yes or clearly no. In this case, we’d expect that NS tends to act like TCC but does so with larger variance.

Table 1. Responses data, TCC = too close to call, NS = not sure .

Survey version/ Response treatment		**Total responses**	**Proportion of**						
			of Yes	**of No**	**of TCC**	**of NS**	**Proportion “no opinion” (NS + TCC)**	**Ratio of "yes" to "no"**	**Ratio of "yes" to "no pooled with no opinion”**
All options	i	3000	0.467	0.287	0.164	0.082	0.25	1.63	0.88
Yes/No	ii	1586	0.590	0.410	-	-	0.00	1.44	-
NS	iii	1619	0.553	0.288	-	0.159	0.16	1.92	1.24
TCC	iv	1683	0.537	0.272	0.191	-	0.19	1.97	1.16

Table 2. Chi-square test results.

A. Probability of an "no opinion" response

comparison	NS v. TCC	All options v. NS	All options v. TCC
χ^2 statistic	5.8360	47.1749	18.3050
p-value	0.0157	0.0000	0.0000

B. The ratio of Yes to No for "no opinion" formats compared to Yes/No treatement

comparison	YES/NO and All options	YES/NO and NS	YES/NO and TCC
χ^2 statistic	3.2734	13.6764	16.4712
p-value	0.0704	0.0002	0.0000

C. The ratio of Yes to No when "no opinion" are pooled with "no"

comparison	YES/NO and All options	YES/NO and NS	YES/NO and TCC
χ^2 statistic	62.4845	4.4130	9.3238
p-value	0.0000	0.0357	0.0023

D. The ratio of Yes to No with "no opinion" responses compared among "no opinion" formats

comparison	TCC and NS	NS and All options	TCC and All options
χ^2 statistic	0.0961	4.9850	6.8678
p-value	0.7566	0.0256	0.0088

Table 3. Marginal implicit prices of attributes associated with the WTA in-kind acres compensation for drained wetlands.*

Survey type	Yes/ No	TCC pooled with No	TCC discarded	NS pooled with No	NS discarded	All options pooled with No	All options discarded
WTA ,ceteris paribus	-8.42	-2.54[b]	-11.96	-3.55[b]	-11.56	4.81	-5.16
change of wetland type	5.86	5.79	7.76	4.80	4.45	2.75	2.66[a]
access	-6.65	-4.89	-5.53	-7.48	-7.03	-5.64	-7.37
amphibian $p \rightarrow g$	-7.40	-3.25[a]	-4.02	-8.78	-9.44	-7.32	-6.74
song bird $p \rightarrow g$	-10.70	-3.36	-4.84	-10.24	-11.02	-4.97	-6.96
wading bird $p \rightarrow g$	-7.69	-7.68	-7.56	-9.50	-10.20	-5.99	-6.72
wild flower $p \rightarrow g$	-3.79[b]	-3.16[a]	-2.44[b]	-3.66[a]	-3.44[b]	-2.23[a]	-1.55[b]
amphibian $g \rightarrow e$	-5.49	-4.39	-3.95	-6.92	-6.79	-5.24	-5.37
song bird $g \rightarrow e$	-5.47	-3.41	-2.19[a]	-1.89[b]	-2.90	-5.24	-5.30
wading bird $g \rightarrow e$	-5.49	-2.75	-3.57	-4.53	-4.98	-3.37	-3.50
wild flower $g \rightarrow e$	-2.27[b]	-4.12	-4.52	-4.26	-3.81	-1.53[a]	-1.59[a]

* Values in this table represent the parameter estimate associated with the listed variable divided by the parameter on percent changes in areas. $p \rightarrow g$ = poor to good, and $g \rightarrow e$ = good to excellent. [a] and [b] indicate ratios using parameter estimates that were NOT significant at the 95 and 90%confidence levels respectively.

Table 4. Ranking of the absolute value of the marginal implicit prices and the maximum difference in ranking across models. Rankings of one had the largest marginal effect and 10 had the lowest marginal effect.

Answer coding	YES/ NO	TCC pooled with No	TCC dropped	NS pooled with No	NS dropped	No opinions pooled with No	No opinions dropped	Median rank	Maximum difference
change of wetland type	10	10	10	10	10	10	10	10	0
access	4	2	2	4	4	3	1	3	3
amphibian $p \rightarrow g$	3	7	5	3	3	1	3	3	6
song bird $p \rightarrow g$	1	6	3	1	1	6	2	2	5
wading bird $p \rightarrow g$	2	1	1	2	2	2	4	2	3
wild flower $p \rightarrow g$	8	8	8	8	8	8	9	8	1
Amphibian $g \rightarrow e$	6	3	6	5	5	4	5	5	3
song bird $g \rightarrow e$	7	5	9	9	9	5	6	7	4
wading bird $g \rightarrow e$	5	9	7	6	6	7	7	7	4
wild flower $g \rightarrow e$	9	4	4	7	7	9	8	7	5

Table 5. Rank correlation results between treatments (TCC = too close to call, NS = not sure).

	YES/NO	TCC pooled with No	TCC discarded	NS pooled with No	NS discarded	No opinions pooled with No	No opinions discarded
YES/NO	**1.00**	0.39	0.72	0.94	0.94	0.75	0.87
TCC pooled with No	0.39	**1.00**	0.75	0.49	0.49	0.59	0.59
TCC discarded	0.72	0.75	**1.00**	0.85	0.85	0.56	0.75
NS pooled with No	0.94	0.49	0.85	**1.00**	1.00	0.68	0.84
NS discarded	0.94	0.49	0.85	1.00	**1.00**	0.68	0.84
no opinions pooled with No	0.75	0.59	0.56	0.68	0.68	**1.00**	0.81
no opinions discarded	0.87	0.59	0.75	0.84	0.84	0.81	**1.00**

Table 6. Summary statistics for predicted probability of yes by actual response in the reserved data.

	Actual Response			
	YES	NO	TCC	NS
Mean	0.669	0.514	0.585	0.602
Standard deviation	0.162	0.187	0.165	0.185
Total responses	916	490	303	118

Table 7. Tukey test results. The critical value at the 95% confidence level is 3.633.

Comparison	NO - YES	NO - NS	NO - TCC	YES - TCC	YES - NS	NS - TCC
Difference of means	0.155	0.088	0.070	0.085	0.068	0.017
Standard Error	0.003	0.005	0.004	0.003	0.004	0.006
Tukey q-statistic	48.107	17.872	16.356	24.515	17.948	2.954

Meta-Regression and Benefit Transfer: Data Space, Model Space, and the Quest for 'Optimal Scope'

Klaus Moeltner
Department of Resource Economics,
University of Nevada, Reno

Randall S. Rosenberger
Department of Forest Resources,
Oregon State University

Draft March 13, 2007

* Contact Information:

Klaus Moeltner
Department of Resource Economics / MS 204
University of Nevada, Reno
Reno, NV 89557-0105
phone: (775) 784-4803
fax: (775) 784-1342
e-mail: moeltner@cabnr.unr.edu

Meta-Regression and Benefit Transfer: Data Space, Model Space, and the Quest for 'Optimal Scope'

Abstract:

Meta-functional Benefit Transfer, while conceptually attractive, is often plagued by the paucity of available source studies and related small sample problems. A broadening of scope of the Meta-Regression Model by adding data from "related, yet different" contexts or activities may circumvent these issues, but may not necessarily enhance the efficiency of transfer functions if the different contexts do not share policy-relevant parameters. We illustrate how different combinations of contexts can be interpreted as 'data spaces' which can then be explored for the most promising transfer function using Bayesian Model Search techniques. Our results indicate that for some scope-augmented data spaces model-averaged benefit predictions can be more efficient than those flowing from the baseline context and data.

Key words: Bayesian Model Averaging, Stochastic Search Variable Selection, Meta-Analysis, Benefit Transfer, Resource Valuation

JEL codes: C11, C15, Q51

Acknowledgement: This research was supported in part by funds from the U.S. EPA STAR grant #RD-832-421-01 to Oregon State University, and by the Nevada Agricultural Experiment Station.

I) Introduction

Benefit Transfer, i.e. the synthesis of existing resource valuation results and the transfer of these findings to a new policy site or context continues to grow in popularity with policy makers and resource managing agencies. For example, in a recent insiders' assessment of the role of Benefit Transfer (BT) at the U.S. Environmental Protection Agency (EPA), Iovanna and Griffiths [1] illustrate how BT has been employed in recent years on numerous occasions in the agency's enforcement of the Clean Water Act. The authors further predict that due to the triple constraints of expediency, financial strains, and administrative hurdles "original assessment studies will remain a rare exception" in future EPA valuation efforts.

It is not surprising, therefore, that the concept of BT has stirred increasing interest amongst resource economists in the U.S. and abroad, and spurred research efforts on both the theoretical underpinnings of BT (e.g. [2, 3]) and its econometric and computational implementation (e.g. [4-6]). This study focuses on the latter aspect of BT. Specifically, we examine the issue of 'optimal scope', i.e. the optimal size and composition of a meta-dataset when BT estimates are to be produced via a meta-regression approach.

In most situations that call for BT some information on the policy context, such as basic site characteristics or aggregate demographics for the underlying population of interest, will be available to the analyst. In that case, empirical findings generally support the use of functional BT over point ("value") transfers or simple aggregation of existing estimates (e.g. [7, 8]). If there exists a study for which the physical and temporal context and the composition of underlying stakeholders are very similar to those for the targeted policy application, parameter estimates from this single source can in theory be combined with policy site attributes to form the transfer function. In practice, however, a close correspondence across multiple dimensions for a study site and policy site is unlikely. Therefore, researchers have increasingly resorted to meta-analytical approaches to derive parameter estimates for function transfer.

The primary rationale for combining information from multiple existing sources in a meta-dataset and using a Meta-Regression Model (MRM) to derive parameter estimates for BT is that each source context will likely overlap with the policy scenario in one or several dimensions of site or population characteristics. In essence, the MRM produces parameters that apply to the “prototypical” context or site, and this prototypical context can be expected to more closely correspond to the policy setting than any single context alone. In addition, MRMs allow disentangling the effects of site attributes, user characteristics, and study-methodological factors on welfare estimates from underlying source studies.

As can be expected, this approach is not without flaws or pitfalls. Common shortcomings of the MRM-BT approach range from weak links with underlying economic theory ([2, 9]), difficulties in identifying appropriate source studies and collecting sufficient and adequate data ([10]), and econometric challenges related to data gaps and small sample issues ([6]).

Perhaps one of the most important, yet least analyzed challenges in meta-analytical BT is the question of ‘optimal scope’ of the MRM, given a specific target policy application. For example, if welfare measures associated with the reduction in sulfur dioxide are sought, could or should the MRM also include values corresponding to a reduction in, say, nitrogen oxides or carbon monoxide? If the value of a day of trout fishing is of primary interest, should the meta-model also include data on bass or salmon fishing? In econometric terms, the question of optimal scope can be interpreted as the exact definition of the dependent variable in the MRM, which, in turn, defines the set of source studies to be considered for inclusion in the meta-dataset. This issue has been briefly raised at various points in time in the literature (e.g. [9, 11, 12])[1], but has not yet been examined in depth in existing contributions.

This study aims to fill this gap. We discuss the exact nature of the optimal scope problem and illustrate the associated econometric dilemmas (next Section), develop an econometric framework that can aid in the determination of optimal scope (Section III), and apply this framework to simulated and actual meta-data (Section IV). Section V summarizes our findings and offers concluding remarks.

II) Optimal Scope: Conceptual and Econometric Considerations

Optimal Scope, Data Space, and Model Space

The question of optimal scope is best illustrated with a brief example: Imagine a resource planner that is considering improving habitat and access along a specific river segment and managing it as a recreational coldwater fishery[2]. The costs of the project are relatively clear, but, as usual, expected economic benefits to potential users are more difficult to assess. Time and funding considerations call for a BT approach. For simplicity, assume that the only relevant and well-predictable characteristic of the new fishery, other than the basic identifiers "coldwater fishery" and "running water", is the expected daily catch rate, x_p ("*p*" stands for "policy site"). A thorough literature search reveals a set of S_0 studies comprising n_0 observations that report welfare results for coldwater fishing at running water[3]. This suggests the following simple MRM:

$$y_{js} = \beta_0 + \beta_1 x_{js} + \varepsilon_{js} \tag{1}$$

where y_{js} *is a* welfare measure for a day of coldwater / river fishing at site *j* reported in study *s*, x_{js} is the catch rate for that site, ε_{js} is an i.i.d. distributed normal error term with zero mean, and the β -terms are meta-regression coefficients to be estimated by the MRM. For simplicity, we will abstract for the moment from econometric considerations such as study-specific unobservables and heteroskedasticity, as well as from the potential effect of study-methodological characteristics on reported welfare estimates. A Benefit Transfer measure for the policy site could then be computed as

$$\hat{y}_p = \hat{\beta}_0 + \hat{\beta}_1 x_p \,, \tag{2}$$

i.e. by combining MRM parameter estimates with known attributes of the policy site, in this case simply the expected catch rate x_p.

However, the analyst may have reservations taking this approach due to the following possible (and, in practice, commonly observed) reasons: (i) The sample size n_0 may be too small to estimate the parameters in (1) with any degree of precision, and / or (ii) the studies included in set S_0 have a narrow geographic distribution, a narrow definition of underlying visitor populations, or are in other ways too context specific to allow for the construction of a robust BT function. To attenuate these problems the analyst may want to combine the original set of studies with another available set S_1, comprising n_1 observations, that report welfare results and catch rates associated with *warmwater* / running water fisheries[4]. A natural rationale for combining the two data sets would be the hopeful anticipation that the regression intercept and the marginal effect of catch rates may be similar for both fishery types (reflecting similarity in underlying angler preferences), and that in that case a pooled MRM of the form (1), but with sample size $n = n_0 + n_1$ can be expected to generate more efficient parameter estimates, and thus a more efficient BT function.

The added studies deviate in one identifying dimension ("type of fishery") from the policy context. In other words, the *scope* of the MRM has been broadened to include both coldwater and warmwater fisheries, and the definition of the dependent variable has changed from, say, "WTP for a day of coldwater fishing at a river" to "WTP for a day of fishing (cold- or warmwater) at a river". In our terminology, this constitutes a re-definition (and augmentation) of the *data space* underlying the MRM. For notational convenience we will label the original ("baseline") dataset as d_0 , the added dataset as d_1 , the original *data space* as D_0 , and the augmented data space as D_1. Thus, we have $D_0 = \{d_0\}$ and $D_1 = \{d_0, d_1\}$.

Naturally, imposing any pooling constraints on the augmented MRM a priori would be risky. If the two activity types do not pool on the intercept, catch rate, or both, using (1) would amount to a model mis-specification, producing biased parameter estimates and misleading BT predictions for the policy context. A more prudent approach would be to start with the most general specification, i.e.

$$y_{js} = \beta_{0,c} + \beta_{1,c} x_{js} + \beta_{0,w} W_{js} + \beta_{1,w} W_{js} x_{js} + \varepsilon_{js} \tag{3}$$

where W_{js} is a 0/1 indicator for observations associated with the warmwater sub-set, $\beta_{0,w}$ captures the deviation in intercept for warmwater observations, and $\beta_{1,w}$ measures the differentiated marginal effect of catch rate on WTP for warmwater cases compared to the baseline effect for coldwater observations (now indexed by subscript *c*).

In the terminology of this study, equation (3) implicitly defines the *model space* for data space D_1. Specifically, the augmentation of scope of the MRM has ex ante added two additional regressors to the MRM – W_{js} and $W_{js}x_{js}$. This implies $2^2 = 4$ possible models, since each new regressor can either emerge as significant (and should thus be included in the augmented model) or not (and could thus be dropped from the augmented model). Indexing inclusion by "1", and exclusion by "0", the model space corresponding to data space D_1 can then be described as

$$\mathrm{M}_1 = \begin{bmatrix} M_{1,1} \\ M_{1,2} \\ M_{1,3} \\ M_{1,4} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{bmatrix} \tag{4}$$

In a classical framework, statistical *in*significance of estimates for both $\beta_{0,w}$ and $\beta_{1,w}$ would lend support for the $M_{1,1} = \begin{bmatrix} 0 & 0 \end{bmatrix}$ case, leading to the pooled model (1), with augmented sample size *n*. This model would then be a logical candidate to generate the BT function. Decision rules for cases $\mathrm{M}_{1,2} - \mathrm{M}_{1,4}$ are less clear-cut. Since the BT function will *always be solely based on estimates of the baseline parameters* (here $\beta_{0,c}$ and $\beta_{1,c}$), the added model coefficients constitute de facto nuisance terms which will add noise to the estimation of the parameters that are actually needed to construct the Benefit Transfer. In this case, broadening the scope of the MRM will only improve the efficiency of the BT function if the gain in sample size offsets the loss in degrees of freedom and estimation noise associated with the introduction of the nuisance terms.

Econometric theory provides only limited guidance as to these countervailing effects. In most cases the analyst will have to take an empirical approach to identify the optimal scope of the MRM. For example, a reasonable strategy would be to estimate both model (1) with original data space D_0 and the applicable version of model (3) for data space D_1, and to compute BT predictions and confidence intervals for both cases. The prediction with the tighter confidence interval could then be chosen to guide policy decisions for the new context.

Finally, assume another dataset d_2 exists for a second related activity that deviates in a different single dimension from the baseline context, say "coldwater fishing at stillwaters" (lakes, ponds, etc). This enables the analyst to define two additional data spaces, $D_2 = \{d_0, d_2\}$ and $D_3 = \{d_0, d_1, d_2\}$. The model selection procedure outlined above has to be repeated for each new space, as the trade-off between increase in sample size and efficiency loss due to nuisance parameters will be different in each case. Note that D_3 yields the MRM with the broadest scope, i.e. "WTP for a day of coldwater or warmwater fishing at rivers or stillwaters".

Classical Challenges and Bayesian Approaches

As conveyed in the above example, the classical strategy to determine the optimal scope of the MRM is conceptually straightforward: (i) Compile a baseline meta-dataset D_0 that corresponds exactly to the policy context, (ii) specify a baseline MRM that includes explanatory variables with known values for the policy site, (iii) identify "related, but different" activities or resource amenities and collect corresponding meta-data, (iv) specify the most general MRM for the resulting augmented data space, (v) through a series of specification tests, determine which activities share common parameters with the baseline context and impose the corresponding equality constraints on the augmented model, (vi) compute BT predictions for the baseline model and the augmented model, (vii) repeat steps (iii) – (vi) for other related activities and resulting data spaces, and (viii) choose the data space and MRM that produces the most efficient BT predictions.

However, there are several problematic issues with this approach. As can be imagined, the number of additional indicators and interaction terms (which become nuisance parameters if found significant in the augmented model) proliferates rapidly with both the number of initial regressors for which policy site information is available, and the related activities or amenities considered. To illustrate, given the availability of $d_a, a = 1 \cdots A,$ additional data sets corresponding to "related" activities or amenities, the number of possible additional data spaces $D_t, t = 1 \cdots T$, amounts to $T = A + \binom{A}{2} + \binom{A}{3} + \cdots + \binom{A}{A-1} + 1 = \sum_{j=1}^{A-1} \binom{A}{j}$, i.e. the number of all single data sets that can be combined with the baseline data, all possible combinations of pairs of data sets that can be combined with d_0, all possible combinations of triplets, etc., until the final space that combines all available data. The last column in Table 1 shows the number of data spaces resulting from adding up to five activities to the baseline model.

Each data space requires the specification of a separate MRM and a corresponding series of specification tests to identify pooling restrictions. For each added activity, the MRM must ex ante include a deviation term for the intercept and interaction terms with all other baseline regressors, as shown in (3). For example, for k_1 original regressors, and *a* added activities, the resulting augmented MRM will include $k_1 \cdot a$ additional covariates[5]. The upper half of Table 1 depicts this product for up to five added activities and baseline regressors. While these figures appear manageable, the associated *model spaces* will comprise $2^{(k_1 \cdot a)}$ elements, i.e. all possible combination of included and excluded terms. Thus, model spaces and therefore the number of possible pooling restrictions can quickly take on formidable dimensions, even for a modest number of baseline regressors and added activities, as shown in the lower half of Table 1.

In a classical estimation framework, this poses the dilemma of either (i) embarking on a time-consuming battery of specification tests with the usual risks of propagating decision errors and other problems related to 'pretest estimators' (e.g. [13]), (ii) ex ante imposing pooling constraints, thus risking model mis-specification, or (iii) facing small sample problems by falling back on the baseline model or an MRM with reduced data space. Furthermore, with increasing data fragmentation, some cell counts for specific interaction terms may become too small for specification test to provide any meaningful guidance.

A related problem in a classical estimation framework arises through its reliance on asymptotic theory. Regardless of scope, a realistically specified MRM will at the very least have to control for intra-study error correlation and heteroskedastic error variances (e.g. [4], [6], [14]) This departure from the basic linear regression model and thus from well understood small sample properties requires invoking asymptotic theory in the interpretation of model and test results. However, for augmented MRMs with lower dimensional scope sample sizes may still be too small to have much confidence in asymptotic test results. This further complicates the model selection process within a given data space and thus the search for optimal scope.

We therefore propose a Bayesian approach to model search in this study. The general rationale of Bayesian Model Search (BMS) techniques is to assign a posterior model probability to each possible specification as part of the overall estimation process. Rather than assessing the superiority of one model over another through pair-wise hypothesis tests, the Bayesian approach either selects the model with the highest posterior probability, or, more frequently, creates a weighted average of model results for inference purposes. The latter strategy is labeled Bayesian Model Averaging (BMA). Hoeting et al [15], Chipman et al. [16], and Koop [17], Ch. 11, provide a good overview of these concepts and techniques.

The BMA approach controls for *model uncertainty*, i.e. the notion that even with extensive theoretical guidance the researcher can never be completely certain which of a set of competing model specifications best describes the underlying data. Rather than selecting a potentially inferior model, the researcher may then prefer to base any econometric inference on a weighted average over all models. This will naturally give more weight to "more likely" models, and low weight to models with low posterior probabilities. Not surprisingly, a common application of BMS and BMA is within the context of identifying the best set of explanatory variables in large regression models (e.g. ,[18, 19], [20, 21]) which, in essence, is also the problem at hand for this study.

Based on the exact computational strategy to generate posterior model probabilities BMS techniques can be grouped into two broad categories: (i) Strategies that require the computation of the marginal likelihood for each model to generate model weights (e.g. [22],[19]), and (ii) Strategies that assign mixture priors to each coefficients, and base model selection and weights on the posterior probabilities that a given coefficient should be included in the model (e.g. [23, 24]).

Since the derivation of the marginal likelihood is computationally burdensome for specifications other than the basic linear regression model[6], we will follow the second strategy to examine the model space for each MRM within a given data space, and, ultimately, to identify the MRM that generates the most efficient BT predictions. Specifically, we will employ George and McCulloch's [23] Stochastic Search Variable Selection (SSVS) algorithm to examine the plausibility of pooling restrictions in a given augmented MRM. We use the search results to assign posterior weights to each model in the MRM's model space, and illustrate how these results can be used to either select a single specification to generate the BT function, or to produce model-averaged BT predictions in cases where no single model receives overwhelming posterior support. The details of this approach are described in the next Section.

III) Econometric Framework

The baseline MRM

As point of departure we specify a baseline MRM that relates welfare measures for the activity or amenity of primary interest reported in study *s* for site *j*, y_{is}, to site and population characteristics for which information is also available for the policy context, $\mathbf{x_{js}}$, and study-methodological indicators $\mathbf{m_s}$. The importance of including these methodological indicators to avoid omitted variables problems has been acknowledged numerous times in meta-analytical research related to resource valuation. For a recent discussion see Johnston et al. [26] and Moeltner et al. [6].[7] The baseline model is thus given as

$$
\begin{aligned}
&y_{js} = \mathbf{x}'_{\mathbf{js}}\boldsymbol{\beta}_{\mathbf{x}} + \mathbf{m}'_{\mathbf{s}}\boldsymbol{\beta}_{\mathbf{m}} + \alpha_s + \varepsilon_{js} \qquad \text{with} \\
&\alpha_s \sim n\left(0, V_\alpha\right) \qquad \varepsilon_{js} \sim n\left(0, \sigma^2\omega_{js}\right), \quad \text{where} \quad \omega_{js} \sim ig\left(\tfrac{v}{2}, \tfrac{v}{2}\right).
\end{aligned}
\tag{5}
$$

As indicated in (5) the baseline model also includes a normally distributed study-specific random effect term α_s with a mean of zero and variance V_α, and an observation-specific stochastic error term ε_{js}. Since most source studies report multiple welfare measures reflecting several sites or applications, the random effect term will capture study-specific unobservables and intra-study correlation. To control for heteroskedasticity, we specify ε_{js} to have observation-specific variance $\sigma^2\omega_{js}$, with ω_{js} drawn from an inverse-gamma distribution with shape and scale equal to $v/2$.[8] In essence, this stochastic structure corresponds to Geweke's [27] Student-t linear model with the added feature of a random effects term. As shown in that study the hierarchical specification of the variance of ε_{js} is exactly equivalent to drawing ε_{js} from a t-distribution with mean zero, scale σ^2 and v degrees of freedom. This allows for higher probabilities of large error variances than would be expected for a basic normal model, a likely occurrence in a meta-regression context. To be specific, for any given σ^2 a small value of v (say 5 to 10) implies a heavy-tailed distribution, while, as is well known, the t-distribution approaches normality for larger values of v. As discussed in Koop [17], Ch. 6, for v>100 the t-distribution becomes virtually indistinguishable from the normal (0, σ^2) density.

Allowing for heteroskedasticity and the possibility of large differences in error variances across observations and studies is of integral importance for our application. Specifically, it may well be possible that a given activity shares common marginal effects of regressors with the baseline context, yet differs substantially in the mix and magnitude of unobservables that enter the reported welfare measures. This may further improve the efficiency of data-augmented BT functions if variance terms for the added activity are generally smaller than those for the baseline model, but could also introduce additional noise into the MRM and thus the transfer function if error variances are larger than those for the baseline case. These effects and trade-offs become clearly visible in our empirical application. At the same time, our specification of heteroskedasticity follows the paradigm of parameter sparseness – it only requires the estimation of a single additional parameter, v. This is important given our objective of searching model space rapidly and efficiently, and the corresponding requirement to keep run-times for individual models as short as possible.

At the panel (= study) level, the baseline model can be written as

$$\mathbf{y_s} = \mathbf{x_s}\boldsymbol{\beta}_\mathbf{x} + \mathbf{m_s}\boldsymbol{\beta}_\mathbf{m} + \mathbf{i}_{\mathbf{n_s}}\alpha_s + \boldsymbol{\varepsilon}_\mathbf{s} \quad \text{with}$$
$$\boldsymbol{\varepsilon}_\mathbf{s} \sim mvn\left(\mathbf{0}, \sigma^2\boldsymbol{\Omega}_\mathbf{s}\right) \quad \text{and} \quad \boldsymbol{\Omega}_\mathbf{s} = diag\left[\omega_{1s} \quad \omega_{2s} \quad \cdots \quad \omega_{n_s,s}\right], \tag{6}$$

where $\mathbf{i}_{\mathbf{n_s}}$ is a vector of ones with length n_s, i.e. the total number of observations furnished by study *s*. It should be noted that conditional on α_s and $\boldsymbol{\Omega}_\mathbf{s}$, $\mathbf{y_s}$ remains multivariate-normally distributed with expectation $\left(\mathbf{x_s}\boldsymbol{\beta}_\mathbf{x} + \mathbf{m_s}\boldsymbol{\beta}_\mathbf{m} + \mathbf{i}_{\mathbf{n_s}}\alpha_s\right)$ and variance-covariance matrix $\left(\sigma^2\boldsymbol{\Omega}_\mathbf{s}\right)$.

Scope augmentation and the SSVS algorithm

Let us now combine the baseline data d_0 with meta-data for a related activity, d_1, as discussed in the previous Section. This adds a deviation indicator and a set of interaction terms to the original model, yielding

$$y_{js} = \mathbf{x}'_{\mathbf{js}}\boldsymbol{\beta}_\mathbf{x} + \mathbf{m}'_\mathbf{s}\boldsymbol{\beta}_\mathbf{m} + \mathbf{z}'_{\mathbf{js}}\boldsymbol{\delta} + \alpha_s + \varepsilon_{js} \quad \textit{with}$$
$$\mathbf{z_{js}} = \left[I\left(js \in d_1\right) \quad I\left(js \in d_1\right)x_{1,js} \quad I\left(js \in d_1\right)x_{2,js} \quad \cdots \quad I\left(js \in d_1\right)x_{k_1,js}\right]', \tag{7}$$

where *I(.)* is an indicator function taking a value of one if observation *js* belongs to the added data set.[9] The objective at hand is now to examine which of the elements in $\mathbf{z_{js}}$ are "close enough" to zero to call for a pooling restriction.

This is precisely the intuition behind the SSVS algorithm ([23, 24]). The basic idea of this approach is to assign a mixture prior to model parameters with uncertain explanatory importance, i.e. the elements of vector $\boldsymbol{\delta}$ in our case. Specifically, we model each coefficient in $\boldsymbol{\delta}$ to have a prior probability *p* of coming from a "well behaved" normal distribution with mean zero and "large" variance, and probability *(1-p)* of following a close-to-degenerate normal distribution with mean zero and a "very small" variance. The resulting mixture prior for, say, element δ_k can then be expressed as

$$pr\left(\delta_k\right) = \gamma_k \cdot n\left(0, c_k^2\tau_k^2\right) + \left(1-\gamma_k\right)\cdot n\left(0,\tau_k^2\right) \quad \text{with}$$
$$pr\left(\gamma_k\right) = bern\left(p\right), \tag{8}$$

where γ_k is a Bernoulli-distributed indicator term taking a value of one with probability *p*, and a value of zero with probability *(1-p)*. We follow standard SSVS notation by labeling the "small" variance as τ_k^2 and the "large" variance as $c_k^2 \tau_k^2$ [10].

As indicated in (8) and discussed in [24], each element of $\boldsymbol{\delta}$ could in theory be assigned its own variance priors, perhaps based on "thresholds of practical significance". In other words, c_k^2 and τ_k^2 could be chosen such that δ_k is assigned to the degenerate distribution with high probability whenever its absolute value falls below a threshold beyond which it no longer affects the dependent variable for all practical purposes. While such coefficient-specific thresholds are meaningful in the medical field and related sciences, they are ex ante difficult to assess in our application. We thus follow a common alternative strategy by setting $c_k = c, \tau_k = \tau, \forall k,$ and standardizing *all* regressors in (7) to allow model coefficients to have the common interpretation of "marginal effect on y_{js} due to a 1-standard deviation movement away from the mean" for a given regressor (e.g. [17], Ch. 11). We will discuss the exact choice of c and τ in the empirical Section below.

The likelihood function for our full Bayesian specification for a scope-augmented MRM thus emerges as

$$
\begin{aligned}
&pr\left(\mathbf{y} \mid \mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\delta}, V_\alpha, \boldsymbol{\omega}\right) = \\
&\prod_{s=1}^{S} \left\{ (2\pi)^{-n_s/2} \left| \left(\mathbf{i}_{\mathbf{n}_s} V_\alpha \mathbf{i}'_{\mathbf{n}_s} + \sigma^2 \boldsymbol{\Omega}_{\mathbf{s}} \right) \right|^{-1/2} \exp\left(-\tfrac{1}{2} \left(\mathbf{y}_{\mathbf{s}} - \mathbf{X}_{\mathbf{s}} \boldsymbol{\theta} - \mathbf{z}_{\mathbf{s}} \boldsymbol{\delta} \right)' \left(\mathbf{i}_{\mathbf{n}_s} V_\alpha \mathbf{i}'_{\mathbf{n}_s} + \sigma^2 \boldsymbol{\Omega}_{\mathbf{s}} \right)^{-1} \left(\mathbf{y}_{\mathbf{s}} - \mathbf{X}_{\mathbf{s}} \boldsymbol{\theta} - \mathbf{z}_{\mathbf{s}} \boldsymbol{\delta} \right) \right) \cdot \right\} \\
&\text{with} \quad \mathbf{X}_{\mathbf{s}} = \begin{bmatrix} \mathbf{x}_{\mathbf{s}} & \mathbf{m}_{\mathbf{s}} \end{bmatrix}, \quad \boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\beta}'_{\mathbf{x}} & \boldsymbol{\beta}'_{\mathbf{m}} \end{bmatrix}', \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}'_1 & \mathbf{X}'_2 & \cdots & \mathbf{X}'_{\mathbf{S}} \end{bmatrix}', \quad \mathbf{Z} = \begin{bmatrix} \mathbf{z}'_{\mathbf{1}} & \mathbf{z}'_{\mathbf{2}} & \cdots & \mathbf{z}'_{\mathbf{S}} \end{bmatrix}' \quad \text{and} \\
&\boldsymbol{\omega} = diag\begin{bmatrix} \omega_{1s} & \omega_{2s} & \cdots & \omega_{n_S S} \end{bmatrix}
\end{aligned}
\tag{9}
$$

where S is the total number of studies included in the MRM. For notational convenience we have collected original regressors $\mathbf{x_s}$ and study-methodological indicators $\mathbf{m_s}$ into a common panel matrix $\mathbf{X_s}$, with corresponding combined coefficient vector $\boldsymbol{\theta}$. It should be noted that SSVS vector $\boldsymbol{\gamma}$ does not enter the likelihood function. This will facilitate the posterior updating for this vector as shown in Appendix A. The full set of priors for the augmented Bayesian MRM is given as

$$
\begin{array}{ll}
(a) & pr(\boldsymbol{\theta}) = mvn(\mathbf{0}, \mathbf{V_0}) \\
(b) & pr(\alpha_s \mid V_\alpha) = mvn(0, V_\alpha), \forall s \qquad pr(V_\alpha) = ig(\varphi_0, \gamma_0) \\
(c) & pr(\sigma^2) = ig(\eta_0, \kappa_0) \\
(d) & pr(\omega_{js} \mid v) = ig\left(\tfrac{v}{2}, \tfrac{v}{2}\right), \forall js \qquad p(v) = g\left(1, \tfrac{1}{v_0}\right) \\
(f) & pr(\delta_k \mid \gamma_k) = \gamma_k \cdot n\left(0, c^2\tau^2\right) + (1-\gamma_k) \cdot n\left(0, \tau^2\right), k = 1 \cdots k_z \\
 & pr(\gamma_k) = bern(p), \forall k,
\end{array}
\tag{10}
$$

where k_z indicates the total number of regressors in $\mathbf{z_s}$. Equation (a) indicates that the prior for all coefficients not subjected to SSVS scrutiny is multivariate normal with mean vector $\mathbf{0}$ and variance-covariance matrix $\mathbf{V_0}$. Equation (b) re-states the hierarchical distribution of random effect α_s shown above, with the common variance V_α following an inverse gamma distribution with shape φ_0 and scale γ_0. The same prior distribution, albeit with potentially different shape and scale parameters, holds for σ^2, the common variance component of ε_{js}, as shown in equation (c). As discussed above, the heteroskedastic variance component of ε_{js} follows an inverse-gamma distribution with identical shape and scale parameter $v/2$, with the hyper-prior distribution of v given as gamma with shape 1 and inverse scale $1/v_0$. In our parameterization, this corresponds directly to the exponential distribution with inverse scale $1/v_0$. As discussed in Koop [17], Ch. 6, this choice of hyper-prior distribution for v is computationally convenient and assures the required condition of $v > 0$. Finally, equation (f) reiterates the hierarchical prior distribution for γ_k as discussed above. The likelihood in (9) and the priors in (10) also apply to variants of our model that do not call for the SSVS algorithm (see below). Naturally, a standardization of regressors and use of prior (f) are no longer needed in that case.

The Bayesian framework then combines likelihood function and priors to derive marginal posterior distributions for all parameters. We use a Gibbs Sampler (GS) along the lines suggested in Koop [17], Ch. 6, to simulate these distributions. The details of this algorithm are given in Appendix A.

Model weights and BT predictions

For each element of $\boldsymbol{\delta}$ and for each draw $r = 1 \ldots R$ of the GS, the posterior simulator produces a binary draw of γ_k based on its posterior probability, $pr(\gamma_k \mid \mathbf{y}, \mathbf{X}, \mathbf{Z})$, as outlined in detail in Appendix A. This draw will take the value of one if there is posterior support that δ_k belongs to the large-variance distribution and should thus be included in the augmented model, and a value of zero otherwise. For example, if $\boldsymbol{\delta}$ and thus $\boldsymbol{\gamma}$ have three elements, a GS sequence of 20 consecutive posterior draws of γ_k, k=1…3, could look like this:

$$\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \end{bmatrix} \tag{11}$$

In the first round of this hypothetical GS sequence none of the coefficients in $\boldsymbol{\delta}$, and thus none of the variables in $\mathbf{z_{js}}$ were chosen for inclusion in the model, in the second and third round only the first element of $\mathbf{z_{js}}$ was selected, in the fourth round the first two elements were selected, and so forth.

This information can then be used to examine how often, out of *R* repetitions, *a given element of* $\boldsymbol{\gamma}$ is set to "1", i.e. how often the underlying explanatory variable is selected for inclusion in the model. In our simple example above, these empirical shares are $11/20 = 0.55$ for γ_1, $8/20 = 0.4$ for γ_2, and 5/20 = 0.25 for γ_3. This provides a quick first look at the *relative importance* of each ex ante questionable regressor. However, as shown e.g. in George and McCulloch [23] and Chipman [18] a more thorough examination of this sequence is needed to draw conclusions on model weights and model selection. As illustrated in the previous Section (equ. (4)), the number of elements in $\boldsymbol{\gamma}$ de facto determine the model space M_t for the added regressors in data space D_t. Thus, sequence (11) also contains information on the empirical probabilities for each possible model in M_t. In our simple example above there are $2^3 = 8$ possible models. For example, model [0,0,0] was selected 4/20 times and would thus receive model weight 0.2. Model [0 0 1] was selected only once, yielding a model weight of 1/20 = 0.05, and so on.

The researcher can then select a single model as the "most promising specification" if model weights are distributed such that a specific model receives overwhelming support. Alternatively, if these posterior weights are more uniformly distributed and thus less discriminating, the analyst may want to use these weights to form model-averaged posterior inferences. Since the latter scenario is more likely in the context of MRM and BT, and since the selection of a single model is a trivial special case of forming model-averaged predictions, we will focus on the model-averaging approach in this analysis.

Thus, our generation of BT predictions associated with a given data space D_t proceeds in two steps: First, we standardize all regressors and implement the SSVS algorithm to derive individual model weights as described above. Second, after recording these weights, we re-run all models in model space M_t with non-standardized regressors, using the modified Geweke [27] model *without* the SSVS component. For each model, we then derive a posterior distribution of BT predictions, and then average these predictions over models using the model weights collected from Step 1. Analytically, this posterior distribution of BT prediction y_p given policy site descriptors $\mathbf{x_p}$ is given as

$$pr\left(y_p \mid \mathbf{x_p}\right) = \sum_{m=1}^{M} \left\{ \int_{\mathbf{\Gamma}} \left[\sum_{h=1}^{H} pr\left(y_p \mid \mathbf{x_p}, M_m, \mathbf{\Gamma}, m_h\right) pr\left(m_h\right) \right] pr\left(\mathbf{\Gamma} \mid \mathbf{y}, \mathbf{X}, \mathbf{Z}, M_m\right) d\left(\mathbf{\Gamma}\right) \right\} pr\left(M_m \mid \mathbf{y}\right) \qquad (12)$$

where subscript m indexes a specific model in M_t, M denotes the total number of models in M_t, m_h labels a specific combination of methodological indicators, H is the total number of such combinations, and $\mathbf{\Gamma}$ comprises all model parameters as introduced in (10), with the exception of $\boldsymbol{\gamma}$, which is no longer needed for Step 2.

Equation (12) indicates that the posterior predictive distribution of y_p, conditional only on policy descriptors $\mathbf{x_p}$ is derived by marginalizing conditional draws of y_p over (i) methodological indicators, (ii) model parameters, and (iii) all models in M_t. The practical implementation of (12) is described in Appendix B. The statistical properties of the model-averaged posterior distribution of $\mathbf{y_p} \mid \mathbf{x_p}$ can then be examined for each available data space and compared to analogous predictions for the baseline model. We will illustrate this final step in selecting a transfer function within the context of our empirical applications in the next Section.

IV) Empirical Implementation

Simulated application

We first illustrate our approach using simulated data. To examine the performance of the SSVS algorithm under different sample sizes and error distributions we generate eight simulated data sets with degrees-of-freedom parameter v set to either 40 or 10 for each of four sample sizes, 2000, 1000, 600, and 300. These scenarios are captured in the first column of Table 2. We ex ante hypothesize that the ability of the SSVS algorithm to discern "true" models will diminish with smaller sample size and heavier tails of the error distribution (i.e. a smaller value of v). The n = 300, v = 10 scenario is designed to mimic some key aspects of typical meta-data traditionally employed for BT purposes, i.e. small to moderate sample size and considerable error noise.

For each simulation scenario we first create a baseline data set d_0 composed of S_0 "studies" with n_{s0} observations on "WTP" and three explanatory variables, yielding an initial baseline sample size n_0. For ease of communication and close correspondence with the empirical application below we label these variables "catch rate", "income" and "travel cost". Catch rate is computed as the log of a uniform (0.8, 20) variate, income is generated as $\log\left(\frac{1}{1000}\text{uniform}\left(30000, 200000\right)\right)$, and travel cost is derived as $\log\left(\frac{1}{10}\text{uniform}\left(10, 200\right)\right)$. We then add a constant term and combine these regressors with the coefficients given in the first row of Table 2. We further add a random effects term drawn from the standard normal distribution, and an error term drawn from a t-distribution with mean 0, scale 1, and v degrees of freedom. A dependent variable $\mathbf{y_0}$ is then computed following equation (5) (without methodological indicators).

Next, we create a second data set d_1 of same panel structure and sample size as the baseline, with regressors, random effects, and error terms drawn from the exact same distributions as hold for the baseline data. However, we specify regression coefficients that deviate from those stipulated for the baseline model in the slope coefficients for "catch rate" and "travel cost", as shown in the second row of

Table 2. This yields dependent variable $\mathbf{y_1}$. We then combine the two data sets in an augmented model with sample size *n* by stacking vectors $\mathbf{y_a}$, a = 1,2, and the two sets of explanatory variables, adding an indicator for the d_1- set, and its interactions with the three explanatory variables. This yields the specification given in equation (7) (without methodological indicators).

For each *n* / *v* scenario, we standardize these regressors and apply the SSVS algorithm to derive model weights for the $2^4 = 16$ individual models contained in the augmented model space M_1. We use the following prior values: $\tau = 0.03, c = 100, \mathbf{V_0} = c^2\tau^2, \mathbf{I_{k_1}} = 9\mathbf{I_{k_1}}, \varphi_0 = \gamma_0 = \eta_0 = \kappa_0 = \frac{1}{2}, v_0 = 10,$ and $p = \frac{1}{2}$. As discussed in George and McCulloch ([23], [24]), a larger value of *c* and a lower value of τ implies a sharper distinction between the two normal densities in the mixture prior for $\boldsymbol{\delta}$. However, the authors recommend keeping the ratio of the two variances, i.e. c^2, at or below 10,000 to avoid convergence problems. Such problems will also arise if τ is located "too close to zero". Our choice of τ and *c* reflects these conflicting concerns. The variance terms for the prior distribution of the baseline coefficients $\boldsymbol{\beta}_\mathbf{x}$, i.e. the diagonal elements of $\mathbf{V_0}$, are chosen to correspond to the variance of the non-degenerate distribution of $\boldsymbol{\delta}$. The shape and scale parameters for the inverse-gamma priors imply diffuse distributions for σ^2, and V_α. Given our parameterization of the gamma prior for *v* in (10), the inverse scale v_0 also constitutes the expectation for this distribution, and v_0^2 denotes the variance. A value of 10 for v_0 implies that *v* is a priori expected to take this value, leading to a moderately heavy-tailed *t*-prior for the regression errors. At the same time, a variance of v_0^2 =100 keeps the prior distribution for *v* sufficiently diffuse to assign adequate weight to the data in posterior updating. Finally, the choice of 0.5 for the Bernoulli parameter *p* implies an equal prior weight of $\left(\frac{1}{2}\right)^{k_z}$ for each possible model contained in a given data space. For each scenario, the standard deviation of the proposal density for *v* in the Metropolis Hastings algorithm contained in the GS (denoted as s_v in Appendix A) is set to achieve an optimal acceptance rate of 44-50% (see e.g. [28] Ch. 11). All models are estimated using 15000 burn-in draws and 10000 retained draws in the Gibbs Sampler. The decision on the appropriate amount of burn-ins was guided by Geweke's [29] convergence diagnostic (CD).

The lower half of Table 2 shows the SSVS acceptance shares for each coefficient associated with the added regressors. A perfectly discriminating GS run would *always* select the interacted coefficients for "catch rate" and "travel cost", and never select the deviation from the constant term and the interacted coefficient for "income". As can be seen from Table 2, our simulated models with large sample sizes come close to this ideal notion of "perfect discrimination". For both the n = 2000 and n = 1000 cases acceptance shares are at 100% for "catch rate", and close to 90% for "travel cost", while the coefficients of deviation for the constant term and "income" are only selected in 5-8% of draws. The lower share of "hits" for "travel cost" compared to "catch rate" may be a result of the somewhat more subtle absolute difference between baseline and added data with respect to the travel cost coefficient, or it may simply be a manifestation of the relative lower information content for this variable in the generated data. It is also clear from the Table that a lower value for v, i.e. a more diffuse distribution of the regression error, results in a subtle but systematic further reduction in acceptance shares for "travel cost" for the two large-sample scenarios.

As is evident from the last four rows of the Table, the SSVS routine essentially loses its ability to identify the difference in coefficients between baseline and added data for "travel cost", while acceptance shares for "catch rate" remain fairly high even for the n = 300, v =10 scenario. Overall, this first examination of simulation results suggests that the ability of the SSVS algorithm to correctly identify regressors that should be included in a given model (i) generally diminishes with sample size, (ii) slightly diminishes with lower values of v, and (iii) can be variable-specific, depending on how informative the underlying data are for each individual regressor.

Data space, model combinations, and empirical model weights flowing from the SSVS analysis for the n = 300, v = 10 case are given in Table 3. The first row simply lists the baseline model, which, by definition, does not include any added regression terms. The last column depicts the empirical model weights assigned by the SSVS routine to each of the 16 possible models in data space D_1. Clearly, no single model receives overwhelming posterior support. The highest weight (0.48) is assigned to the partially correct model M_5, which stipulates a difference in coefficients for “catch rate”, but a shared coefficient for “travel cost”. The second largest share (0.267) is allocated to the null model M_1 while the correct model M_{11} only receives a very small posterior weight of 0.007. In our simulated context high weights for the null model and low weights for the correct augmented model simply imply that the underlying data lack sufficient information to identify structural parameter differences.

Overall, given our empirical context these results convey two important messages regarding the interpretation of model weights flowing from the SSVS algorithm: (i) A high weight for the null model, which a hopeful analyst may interpret as “perfect poolability” of two activities or contexts, may simply be indicative of noise in the underlying data, and (ii) the most appropriate model may not receive considerable posterior weight. This suggests a model averaging approach to generate BT predictions.

The results for the second step of our analysis are provided in Table 4. For ease of interpretation the first three columns reiterate data space, model labels, and model weights, respectively. The next four columns show the posterior means for the BT-relevant coefficients, i.e. the elements of $\boldsymbol{\beta}_{\mathbf{x}}$ in equation (7). The last six columns depict key statistical features of the posterior predictive distribution of BT prediction y_p. We follow the steps outlined in Appendix B to generate these predictions. For each of the R = 10,000 parameter draws from the original GS, we draw a set of r_p = 100 predicted values for policy outcome y_p. We then keep every 20th of these draws to reduce autocorrelation in our sequence. Thus, we retain 50,000 posterior predictive draws for our analysis[11]. To mimic our sport fishing application below and derive “realistic” WTP figures the statistics in Table 4 refer to the exponentiated version of this predictive distribution.

The first row in Table 4 gives the results for the baseline model. For our purposes the key features of these results are a mean predicted benefit of 32.5, with a numerical standard error (nse) of 0.5.[12] The last three columns show the lower bound, upper bound, and width of the corresponding 95% numerical confidence interval. As can be seen from the Table, the posterior means for BT-relevant coefficients generated by models in the D_1 space differ from those for the baseline model primarily in the estimated intercept. Given our random effects specification, this intercept is somewhat more difficult to estimate under small sample sizes. The baseline model grossly under-predicts the true value of -2.5 (see Table 2). The D_1 models, while still considerably off-target, are closer to the true values. Also, the added data reduces posterior noise in the BT predictions, as evidenced by the substantially smaller posterior standard deviation for all D_1 models compared to the baseline specification. Given the *known* shortcomings of the baseline model and the noticeably reduced posterior variability in the scope augmented models, the model averaged predictive distribution, given in the bottom row of the Table, would clearly be a more robust choice to form BT predictions than the baseline model. It also generates more efficient predictions than the baseline specification, as evident from the smaller *nse* and corresponding interval width.

Sport fishing application

To illustrate our methodology with actual meta-data, we selected a baseline set of studies that report aggregate estimates of consumer surplus for a day of *coldwater* fishing in a *running water* environment. All welfare observations are associated with all-or-nothing site values to allow for a clear association of WTP estimates with status quo site characteristics. The studies are drawn from two sources: an updated outdoor recreation meta-data set described in Rosenberger and Loomis [30], and the sport fishing meta-data collected by Boyle et al. [31]. These two sources combined constitute arguably the largest collection of recreational meta-information currently available. Yet, as shown in Table 5, we could only identify 15 studies comprising a total of 73 observations that satisfy our "policy context" criteria. This creates a realistic setting for the desire to augment the data with related activities.

We consider a scope augmentation along the dimensions used in our introductory example: *warmwater* fisheries, and *stillwater* environments. This yields four possible data spaces, as summarized in Table 5. As can be seen from the table, augmenting the scope of the data produces a marked increase in sample size, especially for the saturated data space D_4, which comprises 37 studies and 229 observations.

Our methodological indicators are “journal” (1 = journal article), “report” (1 = government report), “dc” (1 = dichotomous choice framework), “oe” (1 = open ended, iterative bidding, or payment card framework), “substitute” (1 = study addressed or incorporates substitute sites), and “sample 200” (1 = underlying sample size $\geq$ 200). The implicit baseline categories for publication source and elicitation format are “technical report, thesis, or dissertation”, and “travel cost method”, respectively. All data spaces have reasonable cell counts for these methodological categories, as shown in the second half of Table 5. To assure a positive value for WTP we model the dependent variable in log form.

For an illustrative implementation of our approach we require continuous baseline variables that – ideally - are reported for all observations. Given the data gaps traditionally encountered in meta-sets (see [6]) this proved to be a major challenge. We ultimately chose daily catch rate and annual household income (both in log form) to represent site attributes and population characteristics, respectively. We replaced missing observations for income (approximately 70% of cases) with State-level census information, and missing observations on catch rates (approximately 50% of cases) with predicted values flowing from an auxiliary regression model relating available catch rates to regional indicators, water types, and fish species. The derivation of daily catch rates was further complicated by the fact that many studies reported this attribute in units other than “per-day”, which required additional conversion steps reliant on aggregate information. Despite these shortcomings our meta-dataset is still suitable to illustrate our conceptual and estimation framework.

The priors and number of GS draws for the standardized model with the SSVS components are the same as for the simulated case, except for the value of τ, which is increased to 0.3 to improve the convergence properties of the Gibbs Sampler . The standard deviation for the proposal density in the MH component varies from $10\sqrt{\frac{1}{n}}$ to $45\sqrt{\frac{1}{n}}$ to yield a uniform acceptance rate of 45-50% for all data spaces. Table 6 shows the composition of individual models for each data space. The one-dimensionally augmented data spaces D_1 and D_2 each include eight models, while this number increases sharply to 64 for the saturated space D_3. For the latter, only models with empirical weights $\geq$ 1% are listed in the Table

6 for ease of exposition. For each augmented data space, the first model (M_1) denotes the "null" model, i.e. the fully pooled specification.

The last column of Table 6 shows the posterior weights for each model produced by the first-stage SSVS analysis. For each data space, the null model carries by far the largest weight, with all other specifications receiving relatively minor weight shares. At this stage it might be tempting to embrace the null model and ignore all other specifications for BT purposes. However, this would be risky for two reasons: (i) The weight shares for the fully pooled version, while substantial, are far from overwhelming, and (ii) as seen from the simulated example, a large weight for the null model may simply indicate a lack of explanatory power in the underlying data. Overall, thus, there still exists a considerable degree of model uncertainty for all augmented data spaces, which again suggests a model-averaging approach.

Therefore, we subject all data spaces and models to the second step of our analysis. For D_3, we only estimate the models with probability weight of 1% or higher to conserve on computing time[13]. As for the simulated data we set $\mathbf{V_0} = 100 \cdot \mathbf{I_{k_1}}$ for this step. The results from this second stage analysis are captured in Table 7. The layout for Table 7 is the same as for Table 4. As can be seen from the first row the baseline model generates a posterior distribution of WTP with a mean \$67.13, a standard deviation of 94.14, and numerical standard error of 0.42. Augmenting the baseline scope of the MRM with observations on warmwater fishing reduces posterior noise as evidenced by a significantly smaller posterior standard deviation for all models in D_1. In contrast, posterior noise increases compared to the baseline model for models in D_2 and D_3.

Clearly, thus, WTP estimates associated with stillwater environments carry more error noise than estimates corresponding to warmwater fishing, ceteris paribus. Also, the point estimates for the posterior mean of y_p are systematically higher than the baseline result for all models in D_2 and most models in D_3. Therefore, the overall picture that emerges is that the context of *warmwater fishing in a running water environment* is more compatible with the baseline scenario than the context of *coldwater fishing in a stillwater environment*. Even the substantial gain in sample size for the fully saturated space D_3 cannot compensate for this lack of affinity with the baseline context and the added noise through larger regression errors. This is also evidenced by the larger standard deviation and *nse* for the model-averaged distribution for D_2 and D_3 compared to the baseline result.

In contrast, and this is perhaps the most important finding flowing from this analysis, the model-averaged predictive distribution for data space D_1 has slightly more efficient properties than the baseline posterior, as indicated by a smaller posterior standard deviation (79.9 vs. 94.1) and nse (0.36 vs. 0.42). We can thus conclude that a more efficient BT function is derived if the scope of the baseline data is augmented along the dimension "*warmwater fishing*", but not along the dimension of "*stillwater*".

V) Conclusion

We illustrate in this study how Bayesian Model Search and Model Averaging techniques can be used to better utilize existing information on resource values for BT predictions. Specifically, we employ George and McCulloch's [23] SSVS algorithm to assign posterior probability weights to different model versions in a scope-augmented Meta-Regression. We show how these weights can then be used to derive model-averaged BT predictions for the augmented data space. Our approach circumvents typical classical challenges that arise when combining different data sets, such as the reliance on asymptotic theory for the interpretation of test results in a small-sample environment, the risk of compounding Type I or Type II decision errors in series of specification tests, and small cell counts for different context combinations. Our empirical findings indicate that for some augmented MRMs resulting model-averaged BT functions can be more efficient than those flowing from a baseline model with a narrower scope and smaller sample size.

While our meta-data are based on aggregate estimates of welfare and aggregate values for site and user characteristics, it should be noted that our methodology is also applicable to individual-level source data. In that case small sample problems may be less pressing. However, the general question of 'optimal scope' remains, and with it the classical challenges associated with rapidly proliferating model spaces in augmented data. The application of our approach to such refined and richer meta-data will be subject to future research.

APPENDIX A

This Appendix outlines the detailed steps of the Gibbs Sampler (GS) for the random effects regression model with t-distributed errors and an embedded SSVS routine for a subset of coefficients. It is convenient to apply Tanner and Wong's [32] concept of data augmentation and treat draws of $\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_S \end{bmatrix}$ and $\boldsymbol{\omega} = \begin{bmatrix} \omega_{11} & \omega_{21} & \cdots & \omega_{n_S S} \end{bmatrix}$ as additional data. As in the main text, we label the regression coefficients subjected to SSVS scrutiny as $\boldsymbol{\delta}$ and the remaining coefficients as $\boldsymbol{\theta}$. This yields the augmented joint posterior $pr\left(\boldsymbol{\theta}, \boldsymbol{\delta}, \sigma^2, V_{\boldsymbol{\alpha}}, v, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\omega} \mid \mathbf{y}, \mathbf{X}, \mathbf{Z}\right)$, which the GS breaks down into consecutive draws of conditional components.

Step 1: Draw $\boldsymbol{\theta}$, $\boldsymbol{\delta}$

It is convenient to stack $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$ into a single coefficient vector $\boldsymbol{\xi}$ and to conformably combine data **X** and **Z** into common matrix **XZ**, with panel (= study) specific component $\mathbf{Xz_s}$. The prior variance of $\boldsymbol{\xi}$ can then be compactly written as $\mathbf{V}_{\boldsymbol{\xi}} = diag\left[\mathbf{V_0}, \mathbf{V}_{\boldsymbol{\delta}}\right]$, where $\mathbf{V}_{\boldsymbol{\delta}} = diag\left[\gamma_k \cdot n\left(0, c^2\tau^2\right) + \left(1-\gamma_k\right) \cdot n\left(0, \tau^2\right), k = 1 \cdots k_z\right]$. To avoid highly correlated draws and to expedite convergence we will draw $\boldsymbol{\xi}$ unconditional on the random effects $\boldsymbol{\alpha}$, along the lines suggested in Chib and Carlin [33]. This leads to the following conditional posterior:

$$pr\left(\boldsymbol{\xi} \mid \mathbf{y}, \mathbf{X}, \mathbf{Z}, \sigma^2, V_\alpha, \boldsymbol{\omega}\right) = mvn\left(\boldsymbol{\mu}_1, \mathbf{V}_1\right) \quad \text{where}$$

$$\mathbf{V_1} = \left(\mathbf{V}_{\boldsymbol{\xi}}^{-1} + \sum_{s=1}^{S} \mathbf{Xz}_s' \left(\mathbf{i}_{\mathbf{n_s}} V_\alpha \mathbf{i}_{\mathbf{n_s}}' + \sigma^2 \boldsymbol{\Omega}_s\right)^{-1} \mathbf{Xz_s}\right)^{-1} \text{ and } \quad \boldsymbol{\mu}_1 = \mathbf{V_1}\left(\sum_{s=1}^{S} \mathbf{Xz}_s' \left(\mathbf{i}_{\mathbf{n_s}} V_\alpha \mathbf{i}_{\mathbf{n_s}}' + \sigma^2 \boldsymbol{\Omega}_s\right)^{-1} \mathbf{y_s}\right).$$

Step 2: Draw $\boldsymbol{\alpha}$

Defining the conceptual regression model $\tilde{\mathbf{y}}_\mathbf{s} = \mathbf{y_s} - \mathbf{Xz_s}\boldsymbol{\xi} = \mathbf{i}_{\mathbf{n_s}} \alpha_s + \boldsymbol{\varepsilon}_\mathbf{s}$ and applying standard results for posterior moments for Gaussian regressions (e.g. [34]), we obtain

$$pr\left(\alpha_s \mid \mathbf{y_s}, \mathbf{Xz_s}, \boldsymbol{\xi}, \sigma^2, \boldsymbol{\omega}\right) = mvn\left(\boldsymbol{\mu}_1, \mathbf{V}_1\right) \quad \text{where} \quad \mathbf{V_1} = \left(\mathbf{V}_{\boldsymbol{\alpha}}^{-1} + \mathbf{i}_{\mathbf{n_s}}' \left(\sigma^2 \boldsymbol{\Omega}_s\right)^{-1} \mathbf{i}_{\mathbf{n_s}}\right)^{-1} \text{ and } \quad \boldsymbol{\mu}_1 = \mathbf{V_1}\left(\mathbf{i}_{\mathbf{n_s}}' \left(\sigma^2 \boldsymbol{\Omega}_s\right)^{-1} \tilde{\mathbf{y}}_\mathbf{s}\right).$$

Step 3: Draw V_α

Given the vector of random effects, the conditional posterior distribution for V_α can be derived in straightforward fashion as $pr\left(V_\alpha \mid \boldsymbol{\alpha}\right) = ig\left(\varphi_1, \gamma_1\right)$ with $\varphi_1 = \left(S + 2\varphi_0\right)/2$ and $\gamma_1 = \left(\boldsymbol{\alpha}'\boldsymbol{\alpha} + 2\gamma_0\right)/2$.

Step 4: Draw σ^2

Expressing the vector of random effects for the full sample as $\tilde{\boldsymbol{\alpha}}$ and applying standard results for generalized regression models, we obtain

$pr\left(\sigma^2 \mid \mathbf{y}, \mathbf{X}, \mathbf{Z}, \boldsymbol{\xi}, \boldsymbol{\omega}\right) = ig\left(\eta_1, \kappa_1\right)$ with $\eta_1 = \left(n + 2\eta_0\right)/2$ and

$\kappa_1 = \frac{1}{2}\left(\left(\mathbf{y} - \mathbf{XZ}\boldsymbol{\xi} - \tilde{\boldsymbol{\alpha}}\right)' \boldsymbol{\Omega}^{-1}\left(\mathbf{y} - \mathbf{XZ}\boldsymbol{\xi} - \tilde{\boldsymbol{\alpha}}\right) + 2\kappa_0\right)$.

Step 5: Draw v

The relevant kernel for draws of v is its prior times the segment of the likelihood in (9) that includes this parameter, i.e. $pr\left(v \mid \boldsymbol{\omega}\right) = \frac{1}{v_0}\exp\left(-\frac{v}{v_0}\right) \cdot \prod_{s=1}^{S}\prod_{js=1}^{n_s} \frac{\left(\frac{v}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)} \omega_{js}^{-\left(\frac{v}{2}+1\right)} \exp\left(-\frac{v}{2\omega_{js}}\right)$. This is a non-standard density, and we use a random walk Metropolis-Hastings algorithm (MH, [35], [36]) to take draws from this kernel. Specifically, we draw a candidate value of v_c in the r$^{\text{th}}$ round of the GS from a truncated-at-zero normal proposal density with mean v_{r-1}, i.e. the current value of *v*, and standard deviation s_v, and accept the draw as the new current value with probability $\alpha_v = \min\left(\frac{pr\left(v_c \mid \boldsymbol{\omega}\right)}{pr\left(v_{r-1} \mid \boldsymbol{\omega}\right)}, 1\right)$. The standard deviation of s_v is chosen (after some trial and error in preliminary runs) to yield an acceptance probability in the 45-50% range, as suggested by Gelman et al. [28], Ch. 11.

Step 6: Draw $\boldsymbol{\omega}$

For this step we note that $\frac{\varepsilon_{js}}{\sigma} \sim n\left(0, \omega_{js}\right)$. We can then use again standard results for the Gaussian regression model to obtain $pr\left(\omega_{js} \mid y_{js}, \mathbf{xz_{js}}, \boldsymbol{\xi}, \sigma^2, v, \alpha_s\right) = ig\left(\psi, \zeta\right)$ with $\psi = \left(v + 1\right)/2$ and

$$\zeta = \frac{1}{2}\left(\left(y_{js} - \mathbf{x}\mathbf{z}'_{\mathbf{js}}\boldsymbol{\xi} - \alpha_s\right)^2 / \sigma^2 + v\right).$$

<u>Step 7:</u> Draw $\boldsymbol{\gamma}$

As shown in Koop et al [37], Ch. 16, conditional on δ_k, the conditional posterior of γ_k remains Bernoulli with an updated success probability (i.e. $pr\left(\gamma_k = 1 \mid \delta_k\right)$) of $\frac{p\phi\left(\delta_j;0,c^2\tau^2\right)}{p\phi\left(\delta_j;0,c^2\tau^2\right)+\left(1-p\right)\phi\left(\delta_j;0,\tau^2\right)}$, where ϕ denotes the normal density. In practice, draws from this updated Bernoulli are obtained by comparing this expression to a random draw u form the uniform [0,1] distribution. If $pr\left(\gamma_k = 1 \mid \delta_k\right) > u$, γ_k is set to one, and it is set to zero otherwise.

APENDIX B:

To derive the posterior predictive distribution of $y_p \mid \mathbf{x_p}$ we proceed as follows:

Step 1: The methodological indicators comprised in $\mathbf{m_s}$ delineate a set of H possible methodological combinations. We follow [6] and assign equal probabilities $\pi_h = \pi = 1/H$ to each combination.

Step 2: For a given draw of parameters within model M_m in the r^{th} round of the original GS we first draw a random effect $\alpha_{p,r}$ from $n\left(0, V_{\alpha,r}\right)$, then an error term $\varepsilon_{p,r}$ from $t\left(0, \sigma_r^2, v_r\right)$, and compute $y_{p,r,h} = \mathbf{x}'_\mathbf{p}\boldsymbol{\beta}_\mathbf{x,r} + \mathbf{m}'_\mathbf{h}\boldsymbol{\beta}_\mathbf{m,r} + \alpha_{p,r} + \varepsilon_{p,r}, h = 1 \ldots H$, where $\mathbf{m_h}$ represents a specific mix of methodological indicators. We then compute the weighted average over methodologies to obtain $y_{p,r} = \sum_{h=1}^{H}\left(\mathbf{x}'_\mathbf{p}\boldsymbol{\beta}_\mathbf{x,r} + \mathbf{m}'_\mathbf{l}\boldsymbol{\beta}_\mathbf{m,r} + \alpha_{p,r} + \varepsilon_{p,r}\right)\pi = \mathbf{x}'_\mathbf{p}\boldsymbol{\beta}_\mathbf{x,r} + \pi\sum_{h=1}^{H}\mathbf{m}'_\mathbf{l}\boldsymbol{\beta}_\mathbf{m,r} + \alpha_{p,r} + \varepsilon_{p,r}$.

Step 3: We repeat Step 2 r_p times to obtain multiple draws of $y_{p,r}$ for each set of parameters. While this is optional, it is computationally inexpensive and improves the efficiency of the predictive distribution.

Step 4: Repeat Steps 2 and 3 for each set of original parameter draws, i.e. for each $\Gamma_r, r = 1 \ldots R$. The resulting sequence of $r_p \cdot R$ draws of $y_{p,r}$ can then be examined to assess the properties of BT predictions associated with model M_m.

Step 5: To generate a model-averaged posterior predictive distribution of $y_p \mid \mathbf{x_p}$, we repeat Steps 2- 4 for each model M_m in the model space M_t of data space D_t, multiply each model-specific sequence by the model-specific weight flowing from the SSVS analysis as shown in Section III, and sum over sequences.

Notes:

[1] Bergstrom and Taylor [9] deem this issue alternatively "commodity consistency" across source studies.

[2] Coldwater fisheries traditionally include species such as trout, steelhead, salmon, mountain whitefish, and grayling.

[3] For simplicity and ease of exposition we will abstract in this example and in the remainder of this study from data gap issues and resulting "N vs. K" dilemmas as discussed in Moeltner et al. [6]. In other words, we assume that all source studies include information on all policy-relevant explanatory variables. It would be straightforward to incorporate "N vs. K" corrections into the econometric framework outlined in this analysis.

[4] In the U.S., common warmwater fish are crappies, small and largemouth bass, sunfish, yellow perch, and catfish.

[5] For simplicity and without loss in generality, we abstract from any higher order interactions in this study. Naturally, the proliferation of regressors and required specification tests would further accelerate with the consideration of such terms.

[6] As described in Raftery [25] there exist a variety of mathematical approximations for the marginal likelihood that can be used to ease computational requirements in posterior simulators. However, these approximations all rely on asymptotic theory for consistency. As mentioned in Chipman et al. [16], such approximations can become unreliable in small sample-cases. Since small-sample issues are important in this study, we refrain from using BMS methods based on approximated marginal likelihoods.

[7] Naturally, the baseline model could also include other regressors than methodological indicators for which no information is available for the policy context, but which may be important for model stability. Just like the elements of $\mathbf{x}_{js}$ these additional covariates would then have to be interacted with activity indicators as new data sets are added to avoid mis-specification errors. Furthermore, since there are no known values for the policy site to insert for these covariates when generating BT predictions, BT

predictions would have to be marginalized over these regressors, in analogy of our treatment of methodological indicators (see also [6]). To avoid these straightforward but tedious computational additions we will abstract from such variables in this analysis.

[8] In our parameterization, this implies an expectation of $\frac{v}{2}\left(\left(\frac{v}{2}\right)-1\right)^{-1}=\frac{v}{v-2}$, and $2\left(\frac{v}{2}\right)+1=v+1$ degrees of freedom.

[9] To avoid a proliferation of interaction terms and added computational complexity in generating BT predictions we assume that the effect of methodological covariates does not change significantly across activities. For most "related activities" that one would traditionally consider in a data-augmented model this is likely a relatively robust assumption.

[10] While seemingly adding notational clutter, the introduction of the γ_k -term and the resulting hierarchical setup for the mixture distribution of δ_k corresponds well to the Bayesian notion of "hierarchical priors", i.e. the prior of δ_k depends on another *model* parameter γ_k, which, in turn has a hyper-prior distribution with parameter *p*. It is also a natural and logical setup to allow for the derivation of a *posterior* probability for the event $\gamma_k = 1$, which is of crucial importance in our case.

[11] To guard against dramatic outliers, we further truncate this distribution at the 99.9th percentile, i.e. we discard the 50 largest observations. This final adjustment is implemented in identical fashion for all models. Intuitively, this correction could be interpreted as "imposing income constraints" on the predicted WTP values.

[12] The nse is computed as $std/\sqrt{(R_p)}$ where *std* is the standard deviation of the predicted distribution and R_p is the length of the series. A numerical 95% confidence interval is obtained as (posterior mean $\pm 1.96 \cdot nse$).

[13] The 13 models in D_3 listed in Tables 6 and 7 have a combined model weight of 0.85. For model-averaging purposes we calibrate each individual model weight by this total to preserve the adding-up condition for the posterior probability mass function of these weights.

References

[1] R. Iovanna and C. Griffiths, Clean water, ecological benefits, and benefits transfer: A work in progress at the U.S. EPA, *Ecological Economics* (2006), **60**, 473-482.

[2] V. K. Smith and S. K. Pattanayak, Is Meta-Analysis a Noah's Ark for Non-Market Valuation?, *Environmental and Resource Economics* (2002), **22**, 271-296.

[3] V. K. Smith, S. K. Pattanayak and G. V. Houtven, Structural Benefits Transfer: An example using VSL estimation, *Ecological Economics* (2006), **60**, 361-371.

[4] R. S. Rosenberger and J. B. Loomis, Panel stratification in meta-analysis of economic studies: An investigation of its effects in the recreation valuation literature, *Journal of Agricultural and Applied Economics* (2000), **32**, 459-470.

[5] C. J. Léon, F. J. Vazquez-Polo, N. Guerra and P. Riera, A Bayesian Model for Benefits Transfer: Application to National Parks in Spain, *Applied Economics* (2002), **34**, 749-757.

[6] K. Moeltner, K. Boyle and R. Paterson, Meta-analysis and benefit transfer for resource valuation - Addressing classical challenges with Bayesian modeling, *Journal of Environmental Economics and Management* (2007), **53**, 250 - 269.

[7] R. Brouwer and F. Spaninks, The validity of environmental benefits transfer: Further empirical testing, *Environmental and Resource Economics* (1999), **14**, 95-117.

[8] S. Kirchhoff, B. G. Colby and J. T. LaFrance, Evaluating the performance of benefit transfer: An empirical inquiry, *Journal of Environmental Economics and Management* (1997), **33**, 75-93.

[9] J. C. Bergstrom and L. O. Taylor, Using meta-analysis for benefits transfer: Theory and practice, *Ecological Economics* (2006), **60**, 351-360.

[10] J. B. Loomis and R. S. Rosenberger, Reducing barriers in future benefit transfers: Needed improvements in primary study design and reporting, *Ecological Economics* (2006), **60**, 343-350.

[11] K. Segerson, The benefits of groundwater protection: discussion, *American Journal of Agricultural Economics* (1994), **76**, 1076-1078.

[12] S. Engel, Benefit function transfer versus meta-analysis as policy tool: A comparison, *in* "Comparative environmental economic assessment" (R. J. G. M. Florax, P. Nijkamp and K. G. Willis, Eds.), Edward Elgar, Cheltenham, U.K. and Northampton, MA (2002).

[13] E. E. Leamer, Let's take the Con out of Econometrics, *The American Economic Review* (1983), **73**, 31-43.

[14] I. J. Bateman and A. P. Jones, Contrasting conventional with multi-level modeling approaches to meta-analysis: Expectation consistency in U.K. woodland recreation values, *Land Economics* (2003), **79**, 235-258.

[15] J. A. Hoeting, D. Madigan, A. E. Raftery and C. T. Volinsky, Bayesian Model Averaging: A Tutorial, *Statistical Science* (1999), **14**, 382-417.

[16] H. Chipman, E. I. George and R. E. McCulloch, The practical implementation of Bayesian Model Selection, *in* "IMS Lecture Notes - Monograph Series" (2001), **38**, 67-116.

[17] G. Koop, Bayesian Econometrics John Wiley & Sons, Ltd (2004).

[18] H. Chipman, Bayesian Variable Selection with Related Predictors, *Canadian Journal of Statistics* (1996), **24**, 17-36.

[19] A. E. Raftery, D. Madigan and J. A. Hoeting, Bayesian model averaging for linear regression models, *Journal of the American Statistical Association* (1997), **92**, 179-191.

[20] C. Fernandez, E. Ley and M. F. J. Steel, Model uncertainty in cross-country growth regression, *Journal of Applied Econometrics* (2001), **16**, 563-576.

[21] C. Fernandéz, E. Ley and M. F. Steel, Bayesian modeling of catch in a north-west Atlantic fishery, *Applied Statistics* (2002), **51**, 257-280.

[22] D. Madigan and J. York, Bayesian graphical models for discrete data, *International Statistical Review* (1995), **63**, 215-232.

[23] E. George and R. McCulloch, Variable selection via Gibbs Sampling, *Journal of the American Statistical Association* (1993), **88**, 881-889.

[24] E. I. George and R. McCulloch, Approaches for Bayesian variable selection, *Statistica Sinica* (1997), **7**, 339-373.

[25] A. E. Raftery, Approximate Bayes Factors and accounting for model uncertainty in Generalized Linear Models, *Biometrika* (1996), **83**, 251-266.

[26] R. J. Johnston, E. Y. Besedin and M. H. Ranson, Characterizing the effects of valuation methodology in function-based benefits transfer, *Ecological Economics* (2006), **60**, 407-419.

[27] J. Geweke, Treatment of the independent student-t linear model, *Journal of Applied Econometrics* (1993), **8**, S19-S40.

[28] A. Gelman, J. B. Carlin, H. S. Stern and D. B. Rubin, Bayesian data analysis Chapman&Hall/CRC Boca Raton, London, New York, Washington, D.C. (2004).

[29] J. Geweke, Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments, *in* "Bayesian Statistics 4" (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, Eds.), Oxford University Press, Oxford, UK (1992).

[30] R. S. Rosenberger and J. B. Loomis, Benefit Transfer, *in* "A Primer on Nonmarket Valuation" (P. A. Champ, K. J. Boyle and T. C. Brown, Eds.), Kluwer Academic Press, Dordrecht / Boston / London (2003).

[31] K. J. Boyle, R. Bishop, J. Caudill, F. J. Charbonneau, D. Larson, M. A. Markowski, R. E. Unsworth and R. W. Paterson, A database of sport fishing values (1998).

[32] M. A. Tanner and W. H. Wong, The calculation of posterior distributions by data augmentation (with discussion), *Journal of the American Statistical Association* (1987), **82**, 528-550.

[33] S. Chib and B. P. Carlin, On MCMC sampling in hierarchical longitudinal models, *Statistics and Computing* (1999), **9**, 17-26.

[34] D. V. Lindley and A. F. M. Smith, Bayes Estimates for the Linear Model, *Journal of the Royal Statistical Society, Series B (Methodological)* (1972), **34**, 1-41.

[35] W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* (1970), **57**, 97-109.

[36] S. Chib and E. Greenberg, Understanding the Metropolis-Hastings algorithm, *American Statistician* (1995), **49**, 327-335.

[37 G. Koop, D. J. Poirier and J. L. Tobias, Bayesian Econometric Methods, Cambridge University Press (2007).

Table 1: Proliferation of Data Space and Model Space

		number of baseline regressors					
		1	2	3	4	5	number of data spaces
		number of additional terms in the MRM					
number of added activities ("data sets")	1	1	2	3	4	5	1
	2	2	4	6	8	10	3
	3	3	6	9	12	15	7
	4	4	8	12	16	20	15
	5	5	10	15	20	25	31
		number of possible models					
number of added activities ("data sets")	1	2	4	8	16	32	
	2	4	16	64	256	1,024	
	3	8	64	512	4,096	32,768	
	4	16	256	4,096	65,536	1,048,576	
	5	32	1,024	32,768	1,048,576	33,554,432	

Table 2: Coefficients and SSVS Acceptance Shares for Simulated Data

	constant	catch	income	travel cost
true coefficients for baseline data	-2.500	1.000	0.600	-0.400
true coefficients for added data	-2.500	1.400	0.600	-0.200
simulation scenario	acceptance shares			
n = 2000, v = 40	0.052	1.000	0.047	0.889
n = 2000, v = 10	0.065	1.000	0.047	0.850
n = 1000, v = 40	0.079	1.000	0.072	0.857
n = 1000, v = 10	0.073	1.000	0.076	0.693
n = 600, v = 40	0.143	0.998	0.087	0.058
n = 600, v = 10	0.182	0.993	0.098	0.069
n = 300, v = 40	0.105	0.620	0.092	0.074
n = 300, v = 10	0.106	0.597	0.100	0.079

Table 3: Data Space, Model Space and Empirical Model Weights for Simulated Data

		interaction terms (1 = included)					
data space	model	d1	d1*catch	d1*inc	d1*tc	n	model weight
D_0	M1	-	-	-	-	150	N/A
D_1	M1	0	0	0	0	300	0.267
	M2	0	0	0	1	300	0.036
	M3	0	0	1	0	300	0.045
	M4	0	0	1	1	300	0.005
	M5	0	1	0	0	300	0.479
	M6	0	1	0	1	300	0.028
	M7	0	1	1	0	300	0.033
	M8	0	1	1	1	300	0.002
	M9	1	0	0	0	300	0.039
	M10	1	0	0	1	300	0.004
	M11	**1**	**0**	**1**	**0**	**300**	**0.007**
	M12	1	0	1	1	300	0.000
	M13	1	1	0	0	300	0.045
	M14	1	1	0	1	300	0.004
	M15	1	1	1	0	300	0.007
	M16	1	1	1	1	300	0.001

d1 = indicator for added data
catch = catch rate
inc = income
tc = travel cost
“correct model” shown in boldface

Table 4: Estimated Coefficients and Predictions for Simulated Data

			relevant coeff's for prediction				exponentiated distribution of predictions					
Data Space	Model	weight	const.	catch	inc	tc	mean	std	nse	low	up	width
D_0 (n=150)	M1	-	-0.593	0.979	0.314	-0.395	32.500	110.803	0.496	31.528	33.472	1.944
	M1	0.267	-1.055	1.150	0.380	-0.355	29.469	67.062	0.300	28.881	30.057	1.176
	M2	0.036	-1.050	1.149	0.379	-0.432	25.096	56.402	0.252	24.602	25.591	0.989
	M3	0.045	-1.061	1.148	0.321	-0.356	21.941	50.697	0.227	21.496	22.385	0.889
	M4	0.005	-1.068	1.152	0.328	-0.374	21.916	49.394	0.221	21.483	22.349	0.866
	M5	0.479	-1.157	1.007	0.392	-0.360	19.786	42.485	0.190	19.414	20.159	0.745
	M6	0.028	-1.164	1.007	0.393	-0.361	20.327	47.280	0.212	19.912	20.742	0.830
	M7	0.033	-1.148	0.997	0.401	-0.361	21.183	48.234	0.216	20.760	21.606	0.846
D_1 (n=300)	M8	0.002	-1.184	1.003	0.410	-0.371	20.819	46.345	0.207	20.413	21.226	0.813
	M9	0.039	-1.341	1.147	0.384	-0.356	22.589	52.436	0.235	22.129	23.049	0.920
	M10	0.004	-1.298	1.147	0.383	-0.376	22.745	53.107	0.238	22.279	23.210	0.931
	M11	0.007	-0.982	1.152	0.305	-0.360	22.021	49.699	0.222	21.585	22.457	0.872
	M12	0.000	-	-	-	-	-	-	-	-	-	
	M13	0.045	-1.063	0.989	0.394	-0.360	22.572	52.329	0.234	22.114	23.031	0.917
	M14	0.004	-0.968	0.986	0.391	-0.393	22.078	49.731	0.223	21.642	22.514	0.872
	M15	0.007	-0.724	0.992	0.321	-0.362	22.267	50.384	0.225	21.825	22.709	0.884
	M16	0.001	-0.626	0.989	0.318	-0.397	22.161	51.172	0.229	21.712	22.610	0.898
D_1, weighted average	-	-	-	-	-	-	23.017	51.239	0.229	22.569	23.466	0.897

catch = catch rate
inc = income
tc = travel cost
mean = posterior mean / std = standard deviation / nse = numerical standard error / low (up) = lower (upper) bound of numerical 95% confidence interval for the mean / width = (up – low)

Table 5: Data Space Composition and Methodological Indicators for Sport Fishing Data

	data space composition					
	fishery		water type			
data space	cold	warm	river	still	studies	obs.
D_0	x		x		15	73
D_1	x	x	x		21	94
D_2	x		x	x	28	112
D_3	x	x	x	x	37	229

	cell counts for methodological indicators					
	journal	report	dc	oe	subst	samp200
D_0	13	23	38	5	22	35
D_1	16	41	40	10	23	35
D_2	37	27	39	8	47	41
D_3	51	105	52	21	49	53

dc = dichotomous choice method
oe = open ended, iterative bidding, payment cards
subst = substitute sites are addressed or included
samp200 = sample size ≥ 200

Table 6: Data Space, Model Space and Empirical Model Weights for Sport Fishing Data

		interaction terms (0 = excluded, 1 = included)							
data space	model	warm	warm*catch	warm*inc	still	still*catch	still*inc	n	model weight
D_0	M1	-	-	-	-	-	-	73	N/A
D_1	M1	0	0	0	-	-	-	94	0.589
	M2	0	0	1	-	-	-	94	0.116
	M3	0	1	0	-	-	-	94	0.066
	M4	0	1	1	-	-	-	94	0.013
	M5	1	0	0	-	-	-	94	0.109
	M6	1	0	1	-	-	-	94	0.085
	M7	1	1	0	-	-	-	94	0.013
	M8	1	1	1	-	-	-	94	0.009
D_2	M1	-	-	-	0	0	0	112	0.519
	M2	-	-	-	0	0	1	112	0.116
	M3	-	-	-	0	1	0	112	0.098
	M4	-	-	-	0	1	1	112	0.034
	M5	-	-	-	1	0	0	112	0.104
	M6	-	-	-	1	0	1	112	0.082
	M7	-	-	-	1	1	0	112	0.027
	M8	-	-	-	1	1	1	112	0.021
D_3 (all models with weight >=0.01)	M1	0	0	0	0	0	0	229	0.373
	M2	0	0	0	0	0	1	229	0.051
	M3	0	0	0	0	1	0	229	0.041
	M5	0	0	0	1	0	0	229	0.053
	M6	0	0	0	1	0	1	229	0.037
	M9	0	0	1	0	0	0	229	0.075
	M10	0	0	1	0	0	1	229	0.010
	M13	0	0	1	1	0	0	229	0.013
	M17	0	1	0	0	0	0	229	0.045
	M33	1	0	0	0	0	0	229	0.073
	M34	1	0	0	0	0	1	229	0.010
	M35	1	0	0	0	1	0	229	0.011
	M41	1	0	1	0	0	0	229	0.060

warm = indicator for warmwater fishery
still = indicator for stillwater environment
catch = catch rate
inc = income

Table 7: Estimated Coefficients and Predictions for Sport Fishing Data

Data Space	Model	n	model weight	relevant coeff's for prediction: const	ln(catch)	ln(inc)	exponentiated distribution of predictions: mean	std	nse	low	up	width
D_0	M1	73	-	2.101	-0.091	0.116	67.127	94.143	0.421	66.302	67.953	1.651
D_1	M1	94	0.589	1.278	-0.070	0.198	75.260	89.731	0.401	74.473	76.047	1.574
	M2	94	0.116	1.814	-0.036	0.133	58.446	64.415	0.288	57.881	59.011	1.130
	M3	94	0.066	0.301	-0.189	0.302	67.063	74.234	0.332	66.412	67.714	1.302
	M4	94	0.013	1.016	-0.095	0.214	58.540	65.788	0.294	57.963	59.117	1.154
	M5	94	0.109	1.503	-0.031	0.160	58.437	64.237	0.287	57.873	59.000	1.127
	M6	94	0.085	2.117	-0.034	0.104	58.431	63.823	0.286	57.872	58.991	1.119
	M7	94	0.013	0.886	-0.095	0.226	58.816	65.485	0.293	58.242	59.390	1.148
	M8	94	0.009	1.444	-0.097	0.175	57.883	64.508	0.289	57.317	58.448	1.131
D_1, weighted average	-	112	-	-	-	-	68.925	79.923	0.358	68.224	69.626	1.402
D_2	M1	112	0.519	3.711	0.066	-0.050	76.073	141.499	0.633	74.832	77.314	2.482
	M2	112	0.116	3.982	0.060	-0.070	79.788	141.514	0.633	78.547	81.030	2.483
	M3	112	0.098	3.738	0.061	-0.057	75.880	139.043	0.622	74.661	77.100	2.439
	M4	112	0.034	4.331	-0.106	-0.085	86.487	140.202	0.627	85.257	87.716	2.459
	M5	112	0.104	4.113	0.060	-0.081	81.361	148.390	0.664	80.060	82.662	2.602
	M6	112	0.082	3.552	0.056	-0.028	81.897	139.779	0.625	80.671	83.122	2.451
	M7	112	0.027	4.218	-0.103	-0.074	87.100	148.214	0.663	85.800	88.400	2.600
	M8	112	0.021	4.044	-0.099	-0.058	89.162	153.572	0.687	87.815	90.509	2.694
D_2, weighted average	-	112	-	-	-	-	78.440	142.073	0.636	77.194	79.687	2.493
D_3 (all models with weight >=0.01)	M1	229	0.373	0.827	-0.072	0.231	80.799	120.148	0.538	79.746	81.853	2.107
	M2	229	0.052	0.98	-0.057	0.219	83.936	127.849	0.572	82.815	85.057	2.242
	M3	229	0.041	1.305	-0.021	0.186	85.828	130.985	0.586	84.679	86.977	2.298
	M5	229	0.054	1.131	-0.059	0.205	83.622	124.39	0.557	82.532	84.713	2.181
	M6	229	0.037	0.757	-0.054	0.239	82.748	122.661	0.549	81.672	83.824	2.152
	M9	229	0.075	-0.154	-0.082	0.307	66.299	93.872	0.42	65.476	67.122	1.646
	M10	229	0.010	-0.022	-0.07	0.296	68.065	96.538	0.432	67.218	68.912	1.694
	M13	229	0.013	0.001	-0.069	0.294	68.135	98.28	0.44	67.273	68.997	1.724
	M17	229	0.045	1.132	-0.046	0.201	83.019	127.803	0.572	81.898	84.14	2.242
	M33	229	0.074	-0.47	-0.08	0.335	65.914	91.694	0.41	65.11	66.718	1.608
	M34	229	0.010	-0.275	-0.067	0.319	68.08	97.649	0.437	67.224	68.936	1.712
	M35	229	0.011	0.142	-0.018	0.277	70.804	104.73	0.469	69.886	71.723	1.837
	M41	229	0.060	0.903	-0.08	0.21	66.723	93.643	0.419	65.902	67.544	1.642
D_3, weighted average*	-	229	-	-	-	-	77.448	114.190	0.511	76.446	78.450	2.004

mean = posterior mean / std = standard deviation / nse = numerical standard error / low (up) = lower (upper) bound of numerical 95% confidence interval for the mean / width = (up – low)

Session I: Benefits Transfer

Comments by Matt Massey on:

Benefits Transfer of a Third Kind: An Examination of Structural Benefit Transfer, George Van Houtven, Subhrendu Pattanayak, Sumeet Patil, and Brooks Depro

Meta-Regression and Benefit Transfer: Data Space, Model Space, and the Quest for 'Optimal Scope', Klaus Moeltner and Randall Rosenberger

Split-Sample Tests of "No Opinion" Responses in an Attribute Based Choice Model, Eli Fenichel, Frank Lupi, John Hoehn, and Michael Kaplowitz

Benefit Transfer

- Both VHPPD and MR investigate ways to conduct benefit transfers in situations where there are only a small number of “appropriate” studies (and a potentially larger number of “related” studies”)

- In some ways, the strength of each study is the weakness of the other

Benefit Transfer

- Both Studies start by:

1. Choosing a specific form for the utility or welfare function
2. Then collect all appropriate studies

- Then the methods start to diverge

Structural BT

3. Starting from the utility function specified in Step 1, expressions for the results reported in the studies from Step 2 are derived (i.e. WTP, number of trips, …)

4. The reported results from the studies in Step 2 are then plugged into the expression from Step 3 and the expressions are then solved for the coefficient values that return the reported results

5. The coefficient values from step 4 are then used in the utility function specified in Step 1 and used to solve for the desired welfare effects

Structural BT

- Strengths
  - Utility theoretic
  - Can deal with small (and large) sample sizes
  - Relatively quick and easy to do
- Weaknesses
  - No specific guidance on how to select the appropriate model

Bayesian Model Search

3. Add “related activity” studies to the dataset and re-specify the model to include the necessary new variables
4. Use SSVS algorithm to assign prior probabilities to all model parameters with uncertain explanatory importance
5. The priors from Step 4 are then combined with the likelihood function to derive posterior distributions for all parameters
6. For each element in the model, the posterior distributions from Step 5 are used to predict whether or not a variable belongs in the model
7. Step 6 is repeated for multiple draws and the percentage of times a variable is predicted to be included in the model can then be used to either identify a dominant model or to create a weights for each model specification

Bayesian Model Search

8. Next all model specifications are then rerun without the SSVS component.
9. For each model then derive posterior distributions of BT predictions
10. Average the predictions from Step 10 using the model weights collected in Step 7

Bayesian Model Search

- Strengths
  - Provides specific guidance on how to select the appropriate model
  - Can help to augment small sample sizes by determining what “related” information can help improve estimation
- Weaknesses
  - Relatively complicated and hard to do