1-1-2006

Using Meta Analysis for Benefits Transfer: Theory and Practice

John Bergstrom
The University of Georgia

Laura Taylor
Georgia State University

REVIEW DRAFT

Using Meta Analysis for Benefits Transfer: Theory and Practice

John C. Bergstrom
Laura O. Taylor*

Abstract

Meta-analysis, or the "study of studies", attempts to statistically measure systematic relationships between reported valuation estimates for an environmental good or service and attributes of the study that generated the estimates including valuation methods, human population and sample characteristics, and characteristics of the good or service itself. In this paper, we discuss the general theory behind and practice of the emerging use of meta-analysis for benefits transfer. If carefully conducted following systematic protocols for model development, data collection, and data analysis and interpretation, we believe that meta-analysis may prove to be a useful tool for benefits transfer in particular applications. However, before widespread application of this method, more convergent validity tests are needed. One of the greatest strengths of using meta-analysis for benefits transfer is the ability to combine and summarize large amounts of information from previous studies. This strength can also lead to one of the greatest weaknesses of this method which is the loss of important valuation details across time and space in the aggregation process. Thus, application of this method to policy questions and issues should always proceed with caution.

***Senior Authorship is shared. John Bergstrom is Richard B. Russell, Jr. Professor of Public Policy and Professor, Department of Agricultural and Applied Economics, College of Agricultural and Environmental Sciences, The University of Georgia, Athens. Laura Taylor is Associate Director, Environmental Policy Program and Associated Professor, Department of Economics, Andrew Young School of Public Policy, Georgia State University, Atlanta.**

Using Meta Analysis for Benefits Transfer: Theory and Practice

1. Introduction

Meta-analysis (MA) is generally defined as a methodology for summarizing results of existing studies by estimating statistical relationships between values reported in studies to explanatory variables capturing heterogeneity within and across studies (e.g. differences in value construct measured, populations and methodology within and across studies). MA has two primary purposes; 1) test hypotheses with respect to the effects of explanatory variables on the value construct of interest; and 2) use the estimated MA model to predict estimates of the value construct across time and space. Both of these purposes can facilitate benefits transfer (BT); for example, #1 can be used to guide the "professional judgment" approach to BT and #2 provides a value estimator model or equation that can be used to statistically estimate transfer values from "study sites" to the "policy site" (Bergstrom and DeCitiva, 1999; Rosenberger and Loomis, 2000a).

The overall purpose of this paper is to review the theory behind and practice of using MA for BT. We begin in the next section by presenting a general theoretical model that provides the framework for using MA for BT. Next, we discuss more specific and empirical data collection and analysis considerations. A summary and conclusions are presented in the final section.

2. General Theoretical Model

Several previous studies have correctly emphasized the need for meta-analysis benefits transfer (MA-BT) models to be strongly based on an underlying utility theoretic model (Boyle, Poe and Bergstrom; 1994; Smith and Pattanayak, 2002; Smith, VanHoutven, and Pattanayak, 2002). Our general theoretical model will focus on estimation of willingness to pay (WTP) for some nonmarket commodity (Q) or some set of services (S) provided by Q since this is the most common and likely future application of MA-BT models. The general form of the underlying conditional indirect utility function is:

$$V = V_j(P_{j,}, M_j; Q_j, QUAL_j, SUB_j, H_j, I_j) \qquad (1)$$

where, $P_{j,}$ = prices of market goods faced by individual j, M_j = household income of individual j, Q_j = quantity on nonmarket commodity available to individual j, $QUAL_j$ = measure of quality of Q available to individual j, SUB_j = measure of substitutes for Q available to individual j; H_j = non-income characteristics of individual j's household, I_j = measure of information available to individual j.

In the absence of supply and demand uncertainty, WTP for a change in Q from some reference state-of-the-world (*R*) to some target state-of-the-world (*T*) is defined by:

$$\Delta V = V_j(P_{j,}^{T}, M_j - WTP; Q_j^{T}, QUAL_j^{T}, SUB_j^{T}, H_j, I_j) - V_j(P_{j,}^{R}, M_j; Q_j^{R}, QUAL_j^{R}, SUB_j^{R}, H_j, I_j). \qquad (2)$$

Solving (2) for WTP results in the general bid function:

$$WTP = f(P_j^{T} - P_{j,}^{R}, Q_j^{T} - Q_j^{R}, QUAL_j^{T} - QUAL_j^{R}, SUB_j^{T} - SUB_j^{R}, H_j, I_j) \qquad (3)$$

If supply and/or demand uncertainty is present, terms capturing uncertainty effects would need to be included in (2) and (3) and WTP in (3) would measure option price (Boyle, Poe and Bergstrom, 1994).

The utility and WTP functions given in (1) and (3) above should form the basis of any MA-BT exercise. The most strict utility theoretical approach, which we will call the "strong structural utility theoretic" (SSUT) approach, would require that a specific structural form of the underlying indirect utility function in (1) be specified, and the MA-BT model is then based on the particular mathematical form of (3) that is derived from (1) and (2). In this approach, the theoretical restrictions associated with (1) and (3) are thus imposed on the MA-BT model (see Smith et al, in this issue, and Smith, VanHoutven, and Pattanayak, 2002, for a discussion of this approach). For instance, a choice of a linear functional form for the indirect utility function in (1) would indicate that income and household attributes drop out of the model and would not be included in (3). An example of this approach is described in more detail in Section 3.4.

At the other end of the spectrum on utility theoretic approaches is what we will call the "non-structural utility theoretic" (NSUT) approach. In this approach the MA-BT model generally contains explanatory variables related to economic theory, but the connections between these variables and an underlying utility function are not explicitly specified. In the NSUT approach, information from (1) – (3) do not directly inform the empirical analysis. An example of this approach in a MA-BT modeling context is given in Woodward and Wui, 2001. Although the authors "maintain an assumption that there exists an unobserved valuation function that determines a wetland's value given its physical, economic and geographic characteristics" (p258), this function is never explicitly written so that it can inform the empirical approach. In fact, the empirical model they employ links wetland values to characteristics of the wetlands and a few study characteristics, but "core economic variables" (see below) suggested by utility theory were not included.

In between the SSUT and NSUT approaches are MA-BT models where the connection between explanatory variables and an underlying utility function are explicitly specified, but only as "approximations" – we will call this the "weak structural utility theoretic" (WSUT) approach. In the WSUT approach, variables in an empirical WTP function would correspond to (3) which is assumed to be derivable from some unknown underlying utility function (1) and utility difference model (2). Although empirical specifications are informed by (3), the researcher also maintains flexibility to introduce explanatory variables into the WTP model that do not necessarily follow from the theoretical modeling structure in (1) and (2). For example, variables describing the study's methodological approach to estimating WTP may be included in the empirical model, which would not be appropriate in the SSUT approach. An example of this approach in a MA-BT modeling context is given in Boyle, Poe and Bergstrom (1994).

The NSUT approach is not suggested for MA-BT, and therefore we do not discuss it directly in the rest of this paper (although any issue relevant for the WSUT approach would apply to the NSUT approach). Both the SSUT and WSUT approaches indicate that an empirical MA-BT model should include "core economic variables". As indicted in (1)-(3), economic theory suggests that core economic variables include measures of price, income, quality, substitutes, and household characteristics. As with the commodity Q, measures of these variables need to be consistently defined across studies. For example, income (M) should be measured consistently within and across studies in terms of individual or household income.

A MA-BT model satisfies *core economic variable consistency* if it includes the basic or core demand determinants suggested by (1)-(3). Because a specific utility function is specified mathematically under the SSUT approach, this approach strictly defines the core economic variables to be included in the MA-BT model (e.g., see

Smith,VanHouten and Pattanayak, 2002). The WSUT approach provides more flexibility in the selection and specification of core economic variables.

3. Data Collection and Analysis

Just like any empirical investigation, the data used in MA may be either primary or secondary data. The typical MA in environmental valuation relies on summary statistics and analytical outcomes from individual studies as data, and can thus be considered secondary data. A MA on primary data involves pooling the original data gathered by independent research projects for analysis. Unfortunately, the availability of primary data in environmental valuation can be sparse, at best. While many journals require authors to make their data available for replication, in practice, many authors are slow to release data; or in the case of older studies and unpublished work, the data may not be available at all. In addition, even within the same valuation methodology, say contingent valuation, pooling of primary data may not be feasible because of divergent data structures resulting from different estimation methods. Thus, it will often be the case that one cannot pool data from enough studies to be sufficiently representative of the whole literature on a topic. As far as we are aware, there has not been an application of MA-BT that relies on primary data and so for the remainder of this paper, we will only focus on the issues that must be considered when conducting a MA-BT based on secondary data.

The first step is to specify a protocol for developing the MA database. This protocol should ideally be peer reviewed before data collection begins. At a minimum, a research protocol should include a statement of the MA-BT objective, the search strategy to be used to identify potentially relevant studies, the selection criteria for studies to be included, and a description of the data coding protocol to be used. In addition to peer reviewing the protocol, a list of potential studies for inclusion may be supplied to reviewers who might be able to suggest additional studies. The protocol used for data collection and coding and its implications for the final dataset should be reported in published results. For instance, in addition to stating the selection criteria used to select studies, the final report should detail the studies that met the selection criteria and those that did not. Similarly, the difficulties associated with following the coding protocol should be carefully reported so as to make transparent any potential weaknesses or unintended biases in the final MA-BT data base and model.

3.1 Objective and Search Criteria

The research objective for a MA-BT would be driven by the policy-site(s) in question. For example, X agency may wish to know the recreation values associated with Y site. Thus, a research objective would be to identify all previous studies which estimate the value of Z type recreational activities at locations with characteristics similar to Y site and to develop a MA-BT function that can be used to estimate the value of recreational activities at Y.

The consensus among MA practitioners, regardless of field of study, is to be as comprehensive as possible in terms of the studies included in a meta analysis. Rather than using researcher judgment about the quality of individual studies to limit those included in a MA, the preferred approach is to include all existing information and incorporate questions of “quality” into the analysis (Rosenberger and Loomis, 2003; Stanley 2001). To exclude studies from a MA is equivalent to applying a zero weight to the information in the study, which is likely inappropriate.

To avoid "publication bias," studies from the "grey literature" should also be included in the MA. The editorial screening process may result in research that has insignificant results, or is duplicative of earlier work in either its study area or methods. Searching for working papers, Masters and PhD dissertations, and technical reports are greatly facilitated by current on-line research databases and Internet search engines. In addition, personal contact with researchers active in the area may provide additional studies and should be pursued.

3.2 Selection Criteria

Identifying studies relevant to the research objective (i.e., relevant for the policy site) can be tricky when complex environmental commodities are the commodity of interest. The assembled studies and data pooled from these studies should, at a minimum, satisfy both commodity and welfare consistency.

Commodity consistency requires that the commodity (Q) being valued, must be fundamentally the same within and across studies. How do we know or judge whether or not Q is consistent within or across studies? In some instances, it is relatively straightforward to assemble studies that are consistent with respect to the commodity they are valuing. For instance, Smith and Osborne (1996) conducted a MA of contingent valuations studies valuing changes in visibility across the US. For this MA, the commodity is a relatively homogenous environmental good, and thus the task of deciding which studies could be pooled together was straightforward. However, the commodity of interest is considerably more complex in many situations. For example, consider the case of wetland valuation. Wetlands perform complex ecological processes such as chemical cycling that function together to provide a rich environment for fish and shellfish birth and growth. End products of these ecological processes and functioning include commercial and recreational fish harvests. Some wetlands studies may value end products such a commercial or recreational fish harvest while others may value the natural characteristics and capacity of wetlands to remain healthy ecosystems (as indicated by ecological processes and functioning). Pooling estimation results across these studies is more problematic since the object of valuation may be difficult to consistently define and reconcile across studies.[1]

How we know or judge when the object of valuation across studies is sufficiently similar to allow pooling? One approach is to employ Lancasterian utility theory to guide this process. If we assume that people derive utility or satisfaction from characteristics or services provided by Q rather than Q itself, then we may rewrite equation (1) as:

$$V = V_j(P_{j,}, M_j; S(Q_j, QUAL_j), SUB_j, H_j, I_j) \qquad (4)$$

where $S(Q_j, QUAL_j)$ represents the services j receives from Q and QUAL, and all other variables are as defined in (1). If we consider two different observations of Q, Q^1 and Q^2, then if $S(Q^1{}_j, QUAL_j) \approx S(Q^2{}_j, QUAL_j)$ and the utilities associated with these services are approximately the same:

$$V^1{}_j(P_{j,}, M_j; S(Q^1{}_j, QUAL_j), SUB^1{}_j, H_j, I_j) \approx$$
$$V^2{}_j(P_{j,}, M_j; S(Q^2{}_j, QUAL_j), SUB^2{}_j, H_j, I_j), \qquad (5)$$

we can then reasonably pool WTP observations for Q^1 and Q^2. In other words, if Q^1 and Q^2 provide the same fundamental set of services to a person we can surmise that the person's preferences over Q^1 and Q^2 would be consistent and it therefore would be valid

---

[1] For an example of a protocol for pooling estimation results for ecological processes and functioning and human benefits across wetlands valuation studies for the purpose of MA, see Brower et al, 1999.

to pool WTP observations for Q^1 or Q^2 (or the services provided by Q^1 and Q^2) into the same MA function.

For example, suppose we are interested in pooling WTP observations for different types of recreational activities (Q^1, Q^2, . . . Q^n) into the same MA equation. For this pooling to meet the criteria of *commodity consistency*, we would have to determine or assume that all of these activities provide a typical person with the same fundamental set of services (e.g., exercise, excitement, entertainment). If recreation activities provide fundamentally different sets of services to people (e.g., downhill skiing vs. picnicking), pooling WTP observations for these activities within and across studies would violate the commodity consistency criterion.

Of course, determining what is "approximately the same" is not always easy, and researcher knowledge of the resource, its services, and the relevant user populations will be key to developing defensible aggregations. Two particularly important aspects of environmental resources that must be considered in this identification process relate to the *spatial and temporal scale* of the resource being valued. The spatial scale of a resource refers to the geographic extent of Q and changes in Q and S(Q) considered in a study. For example, a change in resource quantity or quality may cover only a relatively small local area (e.g., single county or smaller) or extend over a larger regional (e.g., state, multi-state) or national area. The temporal scale of a study refers to the time frame of the resource changes being considered in a study. Because the spatial and temporal scales determine when and where services provided by change in Q will occur, these scales would be expected to influence WTP for a change in Q and the services it provides

(Brouwer et al, 1999; Poe, Boyle and Bergstrom, 2001), and thus should be considered when developing and following a selection protocol.

Another consideration with respect to pooling Q's for MA is the *range* of the change in Q or Q services valued across studies. If the range of changes is very different across studies (for example, one study measured WTP for a 1,000 acre increase in wetlands protection and another measured WTP for a 100,000 increase in wetlands protection) either adjustments would need to be made in the dependent variable (e.g., converting to WTP per acre) or the different range of changes would need to be controlled for on the right-hand side of the MA-BT equation. Adjusting the dependent variable prior to estimation may be problematic if it fundamentally changes the nature of what people valued (or perceived they valued) in the original studies.

Pooling of Q's across studies should also consider *framing effects*. Framing effects refer to how the specified reference and target levels of Q influence WTP for a particular change in Q or Q services. Framing effects are important to consider because even if the range of a change in Q or Q services is the same within and across studies, WTP for the change in Q and services provided may be sensitive to the starting point (e.g., reference level) and end point (e.g., target level) specified within and across studies (Brouwer et al, 1999; Tversky and Kahneman, 1991). At a minimum, framing effects may need to be controlled for on the right-hand side of the MA-BT equation.

In addition to satisfying commodity consistency, assembled studies should satisfy *welfare change measure consistency*. In all MA-BT modeling approaches, the value construct of interest – the dependent variable in MA-BT empirical models – is willingness-to-pay (WTP) for changes in the quality, quantity or services of an environmental or natural resource. Economic theory indicates that WTP for a change in Q should represent a Hicksian exact welfare change measure of consumer surplus defined by (2) above (Freeman 1999; Just, Hueth and Smith, 2004). To satisfy *welfare change measure consistency*, all measures of WTP within and across studies included in a MA-BT model should represent the same Hicksian exact welfare change measure, or *ex-post* calibrations should be made to account for theoretical differences between welfare change measures estimated within and across studies (Boyle, Poe and Bergstrom, 1994; Smith and Pattanayak, 2002; Smith, Van Houtven and Pattanayak, 2002).

For example, conceptually, the contingent valuation method (CVM) measures Hicksian consumer surplus whereas the travel cost method (TCM) measures Marshallian consumer surplus. If measures of WTP, say for outdoor recreation, are pooled into the same MA-BT model, *welfare change measure consistency* would require that calibrations be made to account for theoretical differences between Hicksian and Marshallian measures of consumer surplus for a change in Q unless one can argue that for theoretical reasons the differences would be expected to be negligible for the change in Q of interest (Randall and Stoll, 1980; Willig, 1976).

The difference between Marshallian and Hicksian surplus associated with a change in the quantity of an environmental good, ΔQ, can be written as (Randall and Stoll, 1980):

$$(M - H)/M \approx (\zeta M)/2Y \qquad (6)$$

where M is the Marshallian surplus changes associated with ΔQ and is equal to the area under the implicit Marshallian demand, or $M = \int_{\Delta Q} P(Q, Y)dQ$; H is the Hicksian surplus changes associated with ΔQ; ζ is the price flexibility of income[2] and is assumed to be constant; and Y is the numeraire. Because $\zeta = [\partial P(Q,Y)/\partial Y]\cdot [Y/P(Q,Y)]$, it is easy to see that the divergence in Marshallian and Hicksian surplus measures is a function of both income and ΔQ.

Calibration of Marshallian surplus measures would require sufficient reporting and appropriate functional form choices in the original studies to allow either the use of duality theory to recover Hicksian welfare measures or adjustments of Marshallian surplus following (6) above. In most cases, the information requirements will not be met to calibrate Marshallian measures. Thus, one alternative is to only include theoretically consistent welfare measures in the MA. The obvious disadvantage of this approach is the loss of sample size.

A second approach to addressing welfare consistency is to recognize the potential difference and try to control for these differences in the MA regression. Shrestha and Loomis (2003) argue in their MA of recreation demand values that differences in Hicksian and Marshallian estimates of recreation demand values from contingent valuation (CV) and travel-cost (TC) models, respectively, are likely to be small because income effects are either not present or very small in travel cost models (i.e., they assume that $\zeta \approx 0$ in (6)). The authors thus pool data from these two types of studies and include a dummy variable indicating the type of study from which the WTP estimate was drawn (see also Loomis and White, 1996, Rosenberger and Loomis, 2000a, and Walsh, Johnson and McKean, 1992). The assumption is that the dummy variable will capture differences in WTP measures that arise from the idiosyncrasies of each study method, but these

[2] The price flexibility of income indicates how changes in income affect the amount the consumer is willing to spend to consume a specific, parametric, amount of Q and is an analogous concept to the income

differences are not because one type of measure is Hicksian and the other Marshallian. However, (6) suggests that in absence of direct evidence for the size of income effects, a more consistent approach would be to allow for variation in WTP estimates based on both changes in income and Q, rather than through a simple scalar effect.

3.3 Data Coding

The data coding protocol involves both the selection of the appropriate information that must be extracted from each study as well as specifying how this information should be coded so that it provides a consistent numerical metric by which heterogeneous studies may be compared with each other systematically. In a WSUT approach, two types of information would be extracted from studies; those implied by equation (3) and those which are used to account for heterogeneity across studies (and welfare estimates) which are not due to differences in the core variables presented in (3).

For core variables (those described in the theoretical WTP function), identification of variables to be extracted is guided directly by (3). In addition to consistent measures of the commodity being valued, Q, and the WTP for changes in Q, a general utility theoretic model is likely to indicate that measures of income, demographic characteristics, and substitutes/complements related to the resource being valued should be included in the coding. Issues relating to the development of consistent welfare measures and resource measurement have already been discussed. Ensuring consistency in coding of income and demographic characteristics across and within studies is likely to be easier than coding substitutes/complements, but attempts should be made to do so. For example, if one is conducting a MA-BT study of CV studies for water quality, it

---

elasticity of demand for goods which are purchased facing a parametric price.

would be appropriate to code an indicator variable equal to one if the original CV survey mentioned substitutes/complements to the survey respondent.

Variables which capture heterogeneity within and across studies for reasons other than variations in the core economic variables, referred to as study design variables, must also be included in the analysis. A MA-BT model satisfies *study design variable consistency* if it controls for heterogeneity within and across studies attributable to how the study authors designed and conducted a study. Important study design variables include the WTP valuation method, WTP elicitation method, and WTP calculation method.

Major WTP valuation methods include stated preference methods (CVM, contingent choice or conjoint analysis) and revealed preference methods (TCM, hedonic price method, averting behavior). The WTP elicitation method refers to the manner in which monetary preferences (e.g., bids) are elicited from participants in a stated preference study. Examples of elicitation methods include dichotomous-choice questions, open-ended questions, payment cards and iterative bidding. Previous meta-analysis studies indicate the sensitivity of WTP estimates within and across studies to the valuation and bid elicitation methods used and offer explanations for this sensitivity (Carson et al, 1996; Walsh, Johnson and McKean, 1992; Rosenberger and Looms, 2000; Shrestha and Loomis, 2003).

The WTP calculation method refers to the statistical procedures used in a stated or revealed preference study to estimate mean or median WTP from collected valuation data. For example, a previous CVM study may have used a Hanemann (1984) truncated integration method that allows only nonnegative measures of individual WTP. Another previous study may have used the Cameron (1988) WTP equation method that allows for nonnegative measures of WTP. Previous meta-analysis studies indicate that estimates of mean WTP for the same Q and Q change are sensitive to bid calculation techniques that include or exclude the possibility of negative bids (Poe, Boyle and Bergstrom, 2001).

As indicated above, much information is needed to completely characterize a MA-BT function. Ideally, studies reviewed would have comprehensive information available regarding the commodity (Q), services (S) provided by Q, the range or scope of Q and S changes, the reference and target levels of Q, core economic variables and study design variables. Practically, however, this type of comprehensive data availability is likely to be very rare. For instance, Rosenberger and Loomis, 2000a report that 20% of the studies they reviewed did not report enough information to be included in their analysis, resulting in a 10% decrease in sample size. In these cases, it is important to contact original authors to attempt to fill-in the missing information. Mrozek and Taylor (2002) were successful in this endeavor increasing their sample size by over 10% after authors of original studies provided additional information relating to their empirical models.

In addition to attempting to complete missing data, the researcher may need to make assumptions and adjustments with respect to variable definitions and coding of apparent data in each study when it is not clear exactly how the information provides maps into the pre-determined variable definitions for the MA. The subjectivity related to this process adds error to the pooling of variables and data across studies. At a minimum, researchers should report any deviations from their research protocol. Ideally, two independent coding of the data would be undertaken and then compared for consistency before analyzing the data. This type of double-blind coding of data is commonly applied in MA applications in the medical field. For variables/studies in which there are discrepancies across each researcher's subjective assessment of how the information should be coded, a process of reconciling the differences should be identified and followed. The obvious disadvantage of following a double-blind coding process is the major commitment of time and resources to complete the task. However, this systematic process can help to overcome a major potential weakness of MA-BT studies; namely researcher subjectivity in the specification and coding of data for the MA-BT function.

3.4 Empirical Estimation

The empirical strategy in a MA-BT based on a WSUT approach differs considerably from that for a MA-BT based on the SSUT approach. The WSUT approach, which has been the most commonly employed, treats the data gathered as a cross-sectional or panel data set. If more than one WTP estimate is provided by a study, as is often the case, the data can be considered panel data. Even if only one observation is provided per study, the data may still be considered a panel data set if observations are correlated across studies that are, for example, written by the same author. Or it may be appropriate to stratify the data by study site. Particular attention must be given to identifying the correct panel structure of the data to avoid bias and incorrect inference (see Rosenberger and Loomis, 2000b). Unlike data used in other applications, such as labor market data, which follows a cross section of individuals over time, the nature of the data stratification may not be obvious in a MA analysis dataset. WTP data for a MA may be stratified by common study, common authors, common study sites, or by common secondary-data sources.

Rosenberger and Loomis, 2000b explore panel stratifications in a recreational demand MA database. They explore stratification of the data by study, by researcher (i.e., a common first author), and by the relationship between WTP estimates across studies. Mrozek and Taylor (2002) stratified WTP estimate data by two levels: common secondary data sources and common study. Mrozek and Taylor (2002) also estimated weighted least squares models in which each observation is weighted by the number of observations included in the meta data from each study. This approach gives equal weight to each study, not each observation, in determining the coefficient estimates.

A second concern in meta data is the information contained in the original WTP estimates. Ideally, one could account for information differences across WTP estimates by, say, weighting each observation by the inverse of the variance of the WTP estimate. This method is commonly employed in models in which the dependent variable is, say, a sample outcome such as the log-odds ratio and allows observations with more precise WTP estimates to weight more heavily in determining the MA regression coefficient estimates. Depending on the method used to estimate WTP, the variance of WTP estimates may not be consistently available across studies. In these instances, one can approximate significance level (or joint significance level) by weighting by the observation by the significance level (or joint significance level) of the coefficient estimate(s) upon which the WTP estimate is based.

In a SSUT approach, the researcher uses the information from the original studies to estimate the parameter values of an assumed common underlying utility function. Smith and Pattanayak (2002) first illustrate the SSUT approach with an application to visibility valuation studies (a CV study and a hedonic price analysis study are considered). Smith and Pattanayak assume the following form for utility (p286, equation 4):

$$V = (\theta_1 v)^b + \Sigma_{j=2 \text{ to } k} (\theta_j A_j)^b + [p^{-\alpha}(m - R(v, A_1,\ldots A_k))]^b \qquad (7)$$

where v is visibility in miles of visual range, A are the housing and location attributes, m is household income, R(.) is the hedonic price function expressed as annual rent, and P is a composite price index for other goods. The structural utility parameters are $\theta_1,\ldots\theta_k$, α, and b.

From equation (7), two expressions for WTP for a change in visibility can be derived. The first is WTP as computed following the logic in equations (2) and (3) in Section 2. Given the specification of the indirect utility function chosen by Smith and Pattanayak (2002), WTP is given by (p287, equation 6a):

$$WTP = (m - R(.)) - [(\theta_1 v^0)^b - (\theta_1 v^1)^b + [(m - R(.))]^b]^{1/b} \qquad (8)$$

Data for WTP estimates, income, v^0, and v^1 are obtained by a CV survey by Brookshire et al (1981).

The second measure of WTP that can be obtained from (7) is a marginal valuation in terms of rental prices an individual is willing to pay for a marginal change in visible range. This is obtained by totally differentiating (7) and yields (Smith and Pattanayak, 2002, page 288, equation 6b):

$$\partial R/\partial v = [\theta_1(v^3)^{b-1}]/[(m - R(.))^{b-1}] \qquad (9)$$

Data on marginal rental prices, and visible range (v^3), and household income are available to Smith and Pattanayak (2002) from a hedonic study by Beron et al (2001).

To estimate the parameters in equation (8) and (9), an error component is added to the model, the two equations are stacked, and a restricted distance estimator is used where θ^1 and b are restricted to be the same in the two equations (Smith and Pattanayak, p 288).

Note how the SSUT approach dictates both the exact functional form for the estimating equations and the exact form for covariates. In equation (8), the baseline and improved level of visibility must be available, not just the change in visible range. In equation (9) the visible range need not equal the visibility levels in equation (8). In addition, the net household income must be available (annual income minus annual rent). R(.) was not directly available, so Smith and Pattanayak (2002) proxy for it by using a Consumer Expenditure survey to compute the fraction of income spent on housing services.

While not actually computed in Smith and Pattanayak (2002), the logic is then to use the estimates of θ^1 and b in equations (8) or (9) to compute value estimates for visible range changes in a policy site. Data on visibility, income, and annual housing expenditures are all that would be needed.[3] In a second example, Smith, Van Houtven and Pattanayak (2002) apply the SSUT approach to estimate benefits of water quality improvements. Data from three studies, each representing a major valuation method (CV, Travel Cost and Hedonic price analysis), are combined to impute (rather than estimate) the parameters in equations (8) and (9) above. Starting with the same assumed form for the conditional indirect utility function as given in (7), the authors derive three equations which link WTP of some form – whether it be a Marshallian consumer surplus or a statement of Hicksian WTP – to observable data and utility parameters. The data from each study are then used to compute the utility parameter estimate that is consistent with the observed data. Unit values for changes in water quality are then derived, varying only by income. Presumably, the preference function captures all utility parameters important for valuation. The examples by Smith and Pattanayak (2002) and Smith, Van Houtven and Pattanayak (2002) assume only net consumption (income less housing expenditures) are important sources of variation in values.

4. Summary and Conclusions

Previous convergent validity tests for benefits transfer (BT) suggest that value estimator models are equations are a more accurate and reliable tool for BT (Bergstrom and DeCitiva, 1999; Rosenberger and Loomis, 2000a). Meta-analysis (MA) models are a particular type of value estimator model and therefore may provide a viable tool for BT. In this paper we have referred to such models as MA-BT models.

As in any research process, the major steps for estimating a MA-BT model are: 1) theoretical model development; 2) data collection and 3) data analysis and interpretation. With respect to theoretical model development, we argue in this paper that MA-BT models should satisfy *core economic variable consistency, commodity consistency* and *welfare change measure consistency*. Each of these three major types of consistency involve numerous important theoretical sub-points and considerations which are discussed in detail in the paper.

The primary form of data collection for a MA-BT model is pooling of secondary data (e.g., WTP estimation results) from previous studies rather than pooling of primary data from previous studies. We argue that the selection of previous studies as data sources and the coding and pooling of secondary data from these studies for the purpose of

[3] Note, Smith and Pattanayak, 2002 assumed that the composite price index does not vary across households in the two studies used (because they are from the same area, Los Angeles) and so this index is

estimating MA-BT models should follow strict, systematic protocols, primarily to reduce researcher subjectivity bias. In addition to satisfying the three major types of theoretical consistency mentioned above, coding and pooling of secondary data for estimation of MA-BT models should also satisfy *study design variable consistency*. This type of consistency controls for heterogeneity across studies due to major differences in study design or methodology.

With respect to data analysis, we recommend either the weak structural utility theoretic (WSUT) approach or strong structural utility theoretic (SSUT) approach. The primary advantage of the SSUT approach is assurance that the MA-BT model strictly follows utility theory. A disadvantage of this approach is that restrictions imposed by the utility function specification may preclude some explanatory variables of interest from the MA-BT model. A primary advantage of the WSUT approach is that is provides more flexibility in the choice of explanatory variables while ensuring that the MA-BT model includes core economic variables. A disadvantage of this approach is the lack of direct (e.g., mathematical) connection between the MA-BT model and its underlying utility function. Thus, to some extent, strict theoretical consistency may be compromised.

To our knowledge, the only empirical tests of MA-BT models to date have estimated models based on the WSUT approach (Rosenberger and Loomis, 2000a; Shrestha and Loomis, 2003). Both of these studies tested for convergent validity using outdoor recreation economic value data bases. The convergent validity tests showed promising results especially for national level models.

Although convergent validity tests were not conducted as part of the analysis, MA models with intended or potential application for benefits transfer have been developed

---

set equal to one in deriving (8) and (9).

and published for outdoor recreation (Rosenberger and Loomis, 2001; Walsh, Johnson and McKean, 1990), water quality (Boyle, Poe and Bergstrom, 1994), air quality (Smith and Huang, 1995), threatened and endangered species (Loomis and White, 1996) and wetlands (Brouwer et al, 1999). All of these studies followed the WSUT approach. In addition, as described in more detail in this paper, Smith et al, 2002 and Smith and Pattanayak, 2002 illustrate the use of the SSUT approach to develop MA-BT models for water quality and air quality, respectively.

Before widespread application of MA-BT models, there is a need for additional MA-BT convergent validity tests across different types of natural resources and environmental commodities. The specification and estimation of the MA-BT models used in these tests should follow the protocols and procedures outlined in this paper. As the development and testing of MA-BT models progresses, on the continuum of BT applications discussed by Bergstrom and DeCivita (1999), we expect that MA-BT models will be sufficiently accurate tools for applications requiring low to moderate accuracy (e.g., screening, minor policy decisions). For applications requiring moderate to high accuracy (e.g., litigation, major policy decisions), primary data studies will probably still be preferred over MA-BT models. Part of the reason may be that the greatest strengths of MA-BT models, that of compiling and explaining results from many studies, may also be one of its greatest weaknesses. For in aggregating results across studies, MA-BT models may lose much rich detail from individual studies which help us to understand and apply time-and-space specific values to specific policy questions and issues.

References Cited

Bergstrom, J.C. and P. DeCivita, 1999. Status of benefits transfer in the united states and canada: a review. Canadian Journal of Agricultural Economics, 47: 79-87.

Beron, K., J. Murdoch and M. Thayer, 2001. The benefits of visibility improvement: New evidence from the Los Angeles metropolitan area, Journal of Real Estate Finance and Economics, 22(March/May):319-338.

Boyle, K.J., G.L. Poe and J.C. Bergstrom. 1994. What do we know about groundwater values? Preliminary implications from a meta analysis of contingent-valuation studies, American Journal of Agricultural Economics, 76 (5):1055-61.

Brouwer, R., I.H. Langford, I.J. Batemen, T.C. R.K. Turner. 1999. A meta-analysis of wetland contingent valuation studies, Regional Environmental Change, 1 (November):47-57.

Carson, R.T., N.E. Flores, K.M. Martin, and J.L. Wright, 1996. Contingent valuation and revealed preference methodologies: comparing the estimates for quasi-public goods. Land Economics, 72 (1): 80-99.

Just, R.E., D.L. Hueth and A. Schmitz. 2004. The welfare economics of public policy: a practical approach to project and policy evaluation. Edward Elgar, Cheltenham, U.K. and Northampton, MA, USA.

Loomis, J.B. and D.S. White, 1996. Economic benefits of rare and endangered species: summary and meta-analysis. Ecological Economics, 18: 197-206.

Freeman, A. Myrick III. 2003. The measurment of environmental and resource values: theory and methods, second edition. Resources for the Future, Washington, D.C.

Mrozek, J.R. and L.O. Taylor. 2002. What determines the value of life? a meta-analysis. Journal of policy analysis and management. 21(2):253-270.

Platt, J. and E. Ekstrand, 2001. Meta analysis of water recreation. Report Number EC-2001-05, Economics Group, Bureau of Reclamation, U.S. Department of the Interior, Denver, CO.

Poe, G.L., K.J. Boyle and J.C. Bergstrom. 2001. A preliminary analysis of contingent Values for ground water quality revisited. In Bergstrom, J.C., K.J. Boyle and G.L. Poe (editors), The Economic Value of Water Quality, Edward Elgar: Cheltenham, U.K. and Northampton, MA, USA.

Randall, A. and J.R. Stoll. 1980. Consumer’s surplus in commodity space. American Economic Review. 71(3):449-57.

Rosenberger, R.S. and J.B. Loomis, 2000a. Using meta-analysis for benefit transfer: in-sample convergent validity tests of an outdoor recreation database. Water Resources Research, 36(4): 1097-1107.

Rosenberger, R.S. and J.B. Loomis, 2000b. Panel stratification in meta-analysis of Economic studies: an investigation of its effects in the recreation valuation literature, Journal of Agricultural and Applied Economics, 32(3): 459-470.

Rosenberger, R.S. and J.B. Loomis. 2001. Benefits transfer of outdoor recreation use values. General Technical Report RMRS-GTR-72, Rocky Mountain Research Station, U.S.D.A. Forest Service, Fort Collins, CO.

Rosenberger, R.S. and J.B. Loomis. 2003. Benefit Transfer. In Champ, P.A., K.J. Boyle, and T.C. Brown (editors), A Primer on Nonmarket Valuation, Kluwer: Dordrecht, The Netherlands and Norwell, MA, USA.

Shrestha, R.K. and J.B. Loomis. 2003. Meta-analytic benefit transfer of outdoor recreation economic values: testing out-of-sample convergent validity. Environmental and Resource Economics, 25: 79-100.

Smith, V.K. and Y. Karou. 1990. Signals or noise? explaining the variation in recreation benefit estimates. American Journal of Agricultural Economics, 72: 419-433.

Smith, V.K. and L.L. Osborne. 1996. Do contingent valuation estimates pass a “scope” test? a meta-analysis. Journal of Environmental Economics and Management, 31: 287-301.

Smith, V.K. and S.K. Pattanayak. 2002. Is meta-analysis a noah’s ark for non-market valuation? Environmental and Resource Economics, 22: 271-296.

Smith, V.K., G. VanHoutven and S. K. Pattanayak. 2002. Benefits transfer via Preference calibration: “prudential algebra” for policy. Land Economics, 78(1): 132-152.

Stanley, T.D. 2001. Wheat from chaff: meta-analysis as quantitative literature review. Journal of Economic Perspectives, 15(3): 131-152.

Tverksy, A. and D. Kahneman. 1991. Loss aversion in riskless choice: a reference-dependent model. Quarterly Journal of Economics, 106:1039-1061.

Vaughan, W.J. and C.S. Russell. Valuing a fishing day: an application of a systematic varying parameter model. Land Economics, 58 (4): 451-461.

Walsh, R.G., D.M. Johnson, and J.R. McKean, 1992. Benefit transfer of outdoor Recreation demand studies, 1968-1988. Water Resources Research, 28 (3): 707-713.

Willig, R.D. 1976. Consumer's surplus without apology. American Economic Review. 66(4):589-97.

Woodward, R.T. and Yong-Suhk Wui. 2001. The economic value of wetland services: a meta-analysis. Ecological Economics, 37: 257-270.