
**UNR Joint Economics Working Paper Series**
**Working Paper No. 07-011**

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

**Klaus Moeltner and Randall S. Rosenberger**

**Department of Resource Economics / MS 204**
**University of Nevada, Reno**
**Reno, NV 89557**
**(775) 784-6701 | Fax (775) 784-1342**
**email:** moeltner@cabnr.unr.edu

**December, 2007**

## 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.

**JEL Classification**: C11, C15, Q51

**Keywords**: Bayesian Model Averaging; Stochastic Search Variable Selection; Meta-Analysis; Benefit Transfer; Resource Valuation

# 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

*Paper presented at the Triangle Resource & Environmental Economics Seminar Series, Research Triangle Institute, Raleigh, NC April 27, 2007, and at the Seminar on Bayesian Inference in Econometrics and Statistics, Washington University, St, Louis, MO, May 4-5, 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, $\varepsilon_{js}$ is an i.i.d. distributed normal error term with zero mean, and the $\beta$-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

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

As indicated in (5) the baseline model also includes a normally distributed study-specific random effect term $\alpha_s$ with a mean of zero and variance $V_\alpha$, and an observation-specific stochastic error term $\varepsilon_{js}$. 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 $\varepsilon_{js}$ to have observation-specific variance $\sigma^2\omega_{js}$, with $\omega_{js}$ 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 $\varepsilon_{js}$ is exactly equivalent to drawing $\varepsilon_{js}$ from a t-distribution with mean zero, scale $\sigma^2$ 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 $\sigma^2$ 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 $(0, \sigma^2)$ 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

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

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 $\alpha_s$ 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

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

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 $\delta_k$ can then be expressed as

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

where $\gamma_k$ 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 $\tau_k^2$ 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 $\tau_k^2$ could be chosen such that $\delta_k$ 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 $\tau$ in the empirical Section below.

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

$$pr\left(\mathbf{y} \mid \mathbf{X}, \mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\delta}, V_\alpha, \boldsymbol{\omega}\right) = \prod_{s=1}^{S}\left\{\left(2\pi\right)^{-n_s/2}\left|\left(\mathbf{i}_{\mathbf{n_s}} V_\alpha \mathbf{i}'_{\mathbf{n_s}} + \sigma^2\boldsymbol{\Omega}_{\mathbf{s}}\right)\right|^{-1/2}\exp\left(-\tfrac{1}{2}\left(\mathbf{y_s} - \mathbf{X_s}\boldsymbol{\theta} - \mathbf{z_s}\boldsymbol{\delta}\right)'\left(\mathbf{i}_{\mathbf{n_s}} V_\alpha \mathbf{i}'_{\mathbf{n_s}} + \sigma^2\boldsymbol{\Omega}_{\mathbf{s}}\right)^{-1}\left(\mathbf{y_s} - \mathbf{X_s}\boldsymbol{\theta} - \mathbf{z_s}\boldsymbol{\delta}\right)\right)\cdot\right\} \quad (9)$$

with $\mathbf{X_s} = \left[\mathbf{x_s} \quad \mathbf{m_s}\right]$, $\boldsymbol{\theta} = \left[\boldsymbol{\beta}'_{\mathbf{x}} \quad \boldsymbol{\beta}'_{\mathbf{m}}\right]'$, $\mathbf{X} = \left[\mathbf{X}'_1 \quad \mathbf{X}'_2 \quad \cdots \quad \mathbf{X}'_{\mathbf{S}}\right]'$, $\mathbf{Z} = \left[\mathbf{z}'_{\mathbf{1}} \quad \mathbf{z}'_{\mathbf{2}} \quad \cdots \quad \mathbf{z}'_{\mathbf{S}}\right]'$ and $\boldsymbol{\Omega} = diag\left[\omega_{1s} \quad \omega_{2s} \quad \cdots \quad \omega_{n_S S}\right]$

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

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

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 **0** and variance-covariance matrix $\mathbf{V_0}$. Equation (b) re-states the hierarchical distribution of random effect $\alpha_s$ shown above, with the common variance $V_\alpha$ following an inverse gamma distribution with shape $\varphi_0$ and scale $\gamma_0$. The same prior distribution, albeit with potentially different shape and scale parameters, holds for $\sigma^2$, the common variance component of $\varepsilon_{js}$, as shown in equation (c). As discussed above, the heteroskedastic variance component of $\varepsilon_{js}$ 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 $\gamma_k$ 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 $\gamma_k$ 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 $\delta_k$ 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 $\gamma_k$, 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 $\gamma_1$, $8/20 = 0.4$ for $\gamma_2$, and 5/20 = 0.25 for $\gamma_3$. 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: $\tau = 0.03, c = 100, \mathbf{V_0} = c^2\tau^2, \mathbf{I_{k_1}} = 9\mathbf{I_{k_1}}, \varphi_0 = \gamma_0 = \eta_0 = \kappa_0 = \frac{1}{2}, \nu_0 = 10,$ and $p = \frac{1}{2}$. As discussed in George and McCulloch ([23], [24]), a larger value of *c* and a lower value of $\tau$ 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 $\tau$ is located "too close to zero". Our choice of $\tau$ and *c* reflects these conflicting concerns. The variance terms for the prior distribution of the baseline coefficients $\boldsymbol{\beta_x}$, 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 $\sigma^2$, and $V_\alpha$. 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 $\boldsymbol{\beta}_x$ 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 20$^{th}$ 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 $\tau$, 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 $pr\left(\boldsymbol{\theta}, \boldsymbol{\delta}, \sigma^2, V_{\alpha}, \nu, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\omega} \mid \mathbf{y}, \mathbf{X}, \mathbf{Z}\right)$, 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:

$pr\left(\boldsymbol{\xi} \mid \mathbf{y}, \mathbf{X}, \mathbf{Z}, \sigma^2, V_{\alpha}, \boldsymbol{\omega}\right) = mvn\left(\boldsymbol{\mu_1}, \mathbf{V_1}\right)$ where

$$\mathbf{V_1} = \left( \mathbf{V}_{\boldsymbol{\xi}}^{-1} + \sum_{s=1}^{S} \mathbf{Xz}_s' \left( \mathbf{i}_{n_s} V_{\alpha} \mathbf{i}_{n_s}' + \sigma^2 \boldsymbol{\Omega}_s \right)^{-1} \mathbf{Xz}_s \right)^{-1} \text{ and } \quad \boldsymbol{\mu_1} = \mathbf{V_1} \left( \sum_{s=1}^{S} \mathbf{Xz}_s' \left( \mathbf{i}_{n_s} V_{\alpha} \mathbf{i}_{n_s}' + \sigma^2 \boldsymbol{\Omega}_s \right)^{-1} \mathbf{y}_s \right).$$

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

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

$$pr\left(\alpha_s \mid \mathbf{y}_s, \mathbf{Xz}_s, \boldsymbol{\xi}, \sigma^2, \boldsymbol{\omega}\right) = mvn\left(\boldsymbol{\mu_1}, \mathbf{V_1}\right) \text{ where } \quad \mathbf{V_1} = \left( \mathbf{V}_{\boldsymbol{\alpha}}^{-1} + \mathbf{i}_{n_s}' \left( \sigma^2 \boldsymbol{\Omega}_s \right)^{-1} \mathbf{i}_{n_s} \right)^{-1} \text{ and } \quad \boldsymbol{\mu_1} = \mathbf{V_1} \left( \mathbf{i}_{n_s}' \left( \sigma^2 \boldsymbol{\Omega}_s \right)^{-1} \tilde{\mathbf{y}}_s \right).$$

Step 3: Draw $V_\alpha$

Given the vector of random effects, the conditional posterior distribution for $V_\alpha$ 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 $\sigma^2$

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 density of $\boldsymbol{\omega}$, 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[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

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

Step 7: Draw $\boldsymbol{\gamma}$

As shown in Koop et al [37], Ch. 16, conditional on $\delta_k$, the conditional posterior of $\gamma_k$ remains Bernoulli with an updated success probability (i.e. $pr\left(\gamma_k = 1 \mid \delta_k\right)$) of $\frac{p\phi\left(\delta_k; 0, c^2\tau^2\right)}{p\phi\left(\delta_k; 0, c^2\tau^2\right) + (1-p)\phi\left(\delta_k; 0, \tau^2\right)}$, where $\phi$ 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 $pr\left(\gamma_k = 1 \mid \delta_k\right) > u$, $\gamma_k$ 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 $y_{p,r,h} = \mathbf{x_p'\beta_{x,r}} + \mathbf{m_h'\beta_{m,r}} + \alpha_{p,r} + \varepsilon_{p,r}, h = 1 \ldots H$, where $\mathbf{m_h}$ represents a specific mix of methodological indicators. We then compute the weighted average over methodologies to obtain $y_{p,r} = \sum_{h=1}^{H}\left(\mathbf{x_p'\beta_{x,r}} + \mathbf{m_l'\beta_{m,r}} + \alpha_{p,r} + \varepsilon_{p,r}\right)\pi = \mathbf{x_p'\beta_{x,r}} + \pi\sum_{h=1}^{H}\mathbf{m_l'\beta_{m,r}} + \alpha_{p,r} + \varepsilon_{p,r}$.

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 $\nu/2\left(\left(\nu/2\right)-1\right)^{-1}=\frac{\nu}{\nu-2}$, and $2\left(\frac{\nu}{2}\right)+1=\nu+1$ 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 $\gamma_k$-term and the resulting hierarchical setup for the mixture distribution of $\delta_k$ corresponds well to the Bayesian notion of "hierarchical priors", i.e. the prior of $\delta_k$ depends on another *model* parameter $\gamma_k$, 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 99.9$^{th}$ 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 $std/\sqrt{\left(R_p\right)}$ 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

| Data Space | Model | weight | relevant coeff's for prediction const. | catch | inc | tc | exponentiated distribution of predictions 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 |
| $D_1$ (n=300) | 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 |
| | 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 $\geq 200$

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)