Cross-Context Benefit Transfer: A Bayesian Search for Information Pools

Citation	Moeltner, K., & Rosenberger, R. S. (2014). Cross-context benefit transfer: a Bayesian search for information pools. American Journal of Agricultural Economics, 96(2), 469-488. doi:10.1093/ajae/aat115
DOI	10.1093/ajae/aat115
Publisher	Oxford University Press
Version	Accepted Manuscript
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsofuse

Cross-Context Benefit Transfer: A Bayesian Search for Information Pools

Klaus Moeltner and Randall S. Rosenberger

Commodity equivalence and population similarity are two widely accepted paradigms for the valid transfer of welfare estimates across resource valuation contexts. We argue that strict adherence to these rules may leave relevant information untapped. We propose a Bayesian model search algorithm that examines the probabilities with which two or more sub-sets of meta-data, each corresponding to a different combination of commodity and population, share common value distributions. Using as an example a large meta-data set of willingness-to-pay for diverse outdoor activities across different regions of the U.S., we find strong potential for contexts that would not traditionally be considered as transfer candidates to form information pools. Exploiting these commonalities leads to substantial efficiency gains for benefit estimates.

Key words: meta-analysis, Bayesian model search, benefit transfer, outdoor recreation .

JEL codes: C11, C52, Q26, Q51.

This paper questions the universal necessity of two widely accepted Benefit Transfer (BT) requirements: (i) The basic commodities under consideration must be essentially equivalent between study site and policy site, and (ii) The affected population should be similar (e.g. Boyle and Bergstrom 1992; Brouwer 2000; U.S. Environmental Protection Agency 2000; Loomis and Rosenberger 2006). These criteria, along with the requirement of similar baseline and change of environmental quality for the two sites, are often referred to as the "EPA guidelines", after they were codified in that agency's *Guidelines for Preparing Economic Analyses* (U.S. Environmental Protection Agency 2000).

Boyle et al. (2009) re-examine and formalize these requirements within a utility-theoretic structural framework. They derive what they label as "sufficient conditions" for valid preference function transfer.[1] They point out that the EPA guidelines can be relaxed to some extent as long as unobservable site characteristics and corresponding preferences are separable from observed characteristics, and the BT method controls for observable, choice-relevant differences in the underlying populations. However, their framework rests on a set of other important assumptions, most notably that the statistical distribution of error components and preferences are correctly specified and identical for study and policy contexts.

On the empirical side, the EPA conditions have found support through studies that explicitly focus on BT errors, usually in a convergent validity setting. Summaries of the numerous "BT validity" contributions are given in Johnston and Rosenberger (2010) and Kaul et al. (2013). As synthesized by Johnston and Rosenberger, p. 482: *"... there is now a fair degree of consensus that site similarity - including similarity over populations, resources, markets and other site attributes - is an important determinant of transfer validity and reliability."*

In summary, both from a theoretical and an empirical perspective the requirements for meaningful BT seem daunting. Close commodity similarity rarely holds, underlying stakeholder populations are bound to differ in many observable and unobservable aspects, and the odds of two populations sharing identical distributions for all relevant stochastic model components are likely microscopic for most BT scenarios.

Klaus Moeltner is an Associate Professor at the Department of Agricultural and Resource Economics, Virginia Tech. Randall S. Rosenberger is an Associate Professor at the Department of Forest Ecosystems and Society, Oregon State University. We thank seminar participants at the 2012 W2133 Regional Meetings in Park City, UT, for helpful comments.

[1] Their title reads "Necessary Conditions for Valid Benefit Transfer", while throughout the text they refer to these conditions as "sufficient". We believe that the latter designation is more in line with the gist of their argument.

Amer. J. Agr. Econ. 1–22; doi: 10.1093/ajae/
November 20, 2013

In this study we argue that none of these requirements are truly *necessary* to conduct policy-informative BT. This is based on the recognition that the relevant input to a policy maker's decision problem is ultimately a *value distribution* for access or change in environmental quality for a given site and stakeholder population. Given the typical valuation models used in the field, this value distribution is generally a complex function of observables, preferences, and parameters corresponding to the stochastic model components. While close similarity of *all of these elements* is naturally a *sufficient* condition for two value distributions to take similar shape and support, similar distributions can also arise via a myriad of combinations of their arguments. In other words, even if population and commodity characteristics appear to be quite different between study and policy sites, the two corresponding value distribution can still be similar if preference and error distributions counter-balance the difference in observables.[2] Moeltner et al. (2009) provide some evidence of this by showing that the Willingness-to-Pay (WTP) distributions for conserving a specific plot of farmland largely overlap for several mid-Atlantic communities, despite pronounced heterogeneity in underlying parameter estimates and population statistics.

In the assessment of recreation benefits, one of the most prolific arenas for BT, it is therefore also conceivable for two value distributions to be similar *across different activities*. This possibility has to date not been explored in a formal fashion in the BT literature. In virtually all existing applications, the underlying recreational activity has been assumed as given and constant across source studies. A notable exception is Moeltner and Rosenberger (2008), who examine the effect of generalizing the "scope" of the valued activity on BT estimates within a Meta-Regression framework. They find that for some combinations of individual activities a more robust and efficient valuation transfer can be achieved compared to considering each activity in isolation.

In this study we combine elements of both Moeltner and Rosenberger (2008) and Moeltner et al. (2009) by examining if valuation distribution can share common elements across both different populations and commodities. We split the largest currently available meta-dataset of outdoor recreation into four geographic regions and 14 outdoor activities. We then apply a Bayesian model search algorithm to explore which region/activity pairs, which we label as "contexts", share a common value distribution. We find several such clusters, with strong pooling probabilities both within and across regions and / or activities. Exploiting these pooling patterns leads to substantial efficiency gains in the estimation of expected benefits.

In a recent contribution León-Gonzalez and Scarpa (2008) address a similar challenge - preference similarity across recreation sites in the United Kingdom and Ireland- via a Bayesian model search. In their case, primary valuation data were collected for each site via an identical discrete-choice type survey instrument. Their algorithm provides guidance as to which cluster of sites are associated with similar preference parameters. We follow their general estimation strategy, broadening their distinction by "sites" to a distinction by "contexts". Importantly, our framework encompasses both multiple geographic regions (comparable to "sites" in the León-Gonzalez and Scarpa (2008) study) and multiple recreation *activities* that would traditionally be analyzed in isolation. This allows us to examine the transferability of benefits not only across regions, but also across activities. The latter would not be possible in León-Gonzalez and Scarpa's application, which focuses on a single (aggregate) activity - forest recreation.[3] However, knowledge on the feasibility of both cross-region and cross-activity transfer is of important practical relevance to policy makers, given the paucity of primary observations currently available for many region-activity combinations (that is, contexts) of interest.

We also extend the León-Gonzalez and Scarpa (2008) framework along econometric dimensions. Most importantly, while León-Gonzalez and Scarpa only consider the existence of a single pooled WTP distribution, we allow for the existence of *multiple* data pools, each associated with a different value distribution. We find that under the constraint of a

[2] In essence, this argument is the reverse of that made by Bateman et al. (2011), who note that erroneously assuming that parameter homogeneity holds across seemingly similar populations can produce vastly misleading results for BT.

[3] The authors do not elaborate on the exact mix of outdoor activities that feeds into this aggregate category. Presumably, their notion of *forest recreation* is most closely related to our activity category of *hiking*.

single pool, other, more subtle pools might be missed, leading to the erroneous inference that certain contexts must be dealt with in isolation, when in fact they share a common value distribution with other region-activity pairs. This, in turn, leads to inefficient transfer estimates.

In addition, we modify León-Gonzalez and Scarpa's Bayesian algorithm, which is geared towards a discrete choice contingent valuation context, to be suitable for a linear regression model as it is customarily employed in BT applications based on meta-regressions (e.g. Moeltner, Boyle, and Paterson 2007; Moeltner and Rosenberger 2008). This requires the specification of a different likelihood function, different prior distributions for several parameters, and different conditional posterior densities that feed into the Bayesian simulation algorithm (Gibbs Sampler). All of these technical details and modifications, along with programming code, are provided in a supplementary online appendix to facilitate replication by other researchers.

The following section outlines our econometric framework. This is followed by an empirical section that discusses the data, detailed model specifications, and estimation results. The penultimate section provides several robustness checks to examine if our results are sensitive to the level of aggregation and elicitation methods in the underlying source studies. The final section concludes.

Econometric Framework

We consider the common situation where a policy maker seeks an aggregate benefit estimate for a specific recreational activity and region - i.e. a specific context in our jargon. She has access to a broader data set that comprises $j = 1 \dots J$ individual contexts. Each of these, in turn, includes n_j observations on outcome variable y_j, and, possibly, explanatory data matrix $\mathbf{X}_j$. The sample distribution for each context is stipulated as $f\left(y_j|\boldsymbol{\theta}_j, \mathbf{X}_j\right)$, where $\boldsymbol{\theta}_j$ comprises the parameters of this density.[4]

The policy maker now has the choice of picking a single data set $\mathbf{y}_p$ that corresponds to her context of interest, or any pooled combination of the available context-specific sub-sets. The latter would be preferable, in terms of sample size and thus estimation efficiency, if all contexts in the pool share the same value distribution with each other and the policy context, i.e if $f\left(y_j|\boldsymbol{\theta}_j, \mathbf{X}_j\right) \approx f\left(y_k|\boldsymbol{\theta}_k, \mathbf{X}_k\right) \approx f\left(y_p|\boldsymbol{\theta}_p, \mathbf{X}_p\right), \quad \forall j, k.$

If the total number of available sub-sets J is relatively large, which will likely be the case if contexts are defined at a (desirable) refined level, the detection of "value pools" poses a logistic dilemma, as the number of pooling combinations becomes quickly intractable.

Let $\boldsymbol{\psi}$ be a J-dimensional vector that, for each individual context, includes a binary indicator set to 1 if context j is in the pooled portion of the data, and to 0 otherwise. Each unique combination of zeros and ones in $\boldsymbol{\psi}$ is considered a "model". The resulting total number of possible models is thus given as $M = 2^J$. This includes the no-pooling or fully independent model.[5] For example, at $J = 10$, we have 1014 possible pooling combinations. At $J = 20$, this number grows to over one million, and at $J = 31$, as is the case for our application, the total number of possible models amounts to over two billion. Our algorithm is designed to rapidly move through this large model space, with a built-in mechanism that assures that more promising models are visited more frequently.

Let a specific model (i.e. a specific combination of zeros and ones in $\boldsymbol{\psi}$) be indexed as m. We label the set of contexts that are in the pooled group for a given model mp, and the complementary set of contexts that are treated as independent as mn. Terms associated with individual contexts in mp will be subscripted with $mp, 1$,$mp, 2$ etc.. The total number of contexts in mp is denoted as J_{mp}. Analogously, J_{mn} reflects the total number of contexts in mn.

Choosing a normal density for the sample distribution of $y_j, j = 1 \dots J$, leads to the following likelihood function for the full data set, conditional on a given model m:[6] where n_j and n_{mp} denote, respectively, the sample size for individual context j and the pooled portion of contexts. Thus, a set of J_{mn} individual coefficient vectors $\boldsymbol{\beta}_j$ and error variances σ_j^2 are estimated for the independent contexts of model

[4] In practice, the free-standing contexts can be as defined at any desired level of refinement, as long as n_j remains sufficiently large to estimate parameters $\boldsymbol{\theta}_j$. In addition to econometric considerations, context definition will largely be guided by policy relevance.

[5] This corresponds to the NT model in León-Gonzalez and Scarpa (2008).

[6] In our application, the dependent variable is in log form. The resulting log-normal density for WTP in dollars is by far the most common specification in existing meta-regressions geared towards benefit transfer.

$$(1)\quad p\left(\mathbf{y}|\left\{\boldsymbol{\beta}_j\right\}_{j\in mn},\left\{\sigma_j^2\right\}_{j\in mn},\boldsymbol{\beta}_{mp},\sigma_{mp}^2,\mathbf{X},m\right)=$$
$$\prod_{j\in mn}\left(2\pi\sigma_j^2\right)^{-n_j/2}exp\left(-\frac{1}{2\sigma_j^2}\left(\mathbf{y}_j-\mathbf{X}_j\boldsymbol{\beta}_j\right)'\left(\mathbf{y}_j-\mathbf{X}_j\boldsymbol{\beta}_j\right)\right)*$$
$$\left(2\pi\sigma_{mp}^2\right)^{-n_{mp}/2}exp\left(-\frac{1}{2\sigma_{mp}^2}\left(\mathbf{y}_{mp}-\mathbf{X}_{mp}\boldsymbol{\beta}_{mp}\right)'\left(\mathbf{y}_{mp}-\mathbf{X}_{mp}\boldsymbol{\beta}_{mp}\right)\right),\quad \text{where}$$
$$\mathbf{y}_{mp}=\begin{bmatrix}\mathbf{y}'_{mp,1} & \mathbf{y}'_{mp,2} & \cdots & \mathbf{y}'_{mp,J_{mp}}\end{bmatrix}',\quad and$$
$$\mathbf{X}_{mp}=\begin{bmatrix}\mathbf{X}'_{mp,1} & \mathbf{X}'_{mp,2} & \ldots\mathbf{X}'_{mp,J_{mp}}\end{bmatrix}',$$

m. Conversely, a single coefficient vector $\boldsymbol{\beta}_{mp}$ and variance σ_{mp}^2 are estimated for the pooled portion of the model.

Since our model space also includes the fully independent case, we need prior densities for all J coefficient vectors and variances. We follow León-Gonzalez and Scarpa (2008) and choose a conjugate g-prior for the coefficient vectors. As discussed in Fernández, Ley, and Steel (2001), such a g-prior greatly reduces the number of parameters that need to be determined by the researcher a priori, facilitates the interpretation of posterior results, and assures speedy model evaluation as part of the search process.[7] Thus, we have

$$(2)\quad p\left(\boldsymbol{\beta}_j|\sigma_j^2\right)=N\left(\boldsymbol{\mu}_{0j},g_j\sigma_j^2\left(\mathbf{X}_j'\mathbf{X}_j\right)^{-1}\right)$$

where N denotes the k-variate normal density, with k indicating the length of $\boldsymbol{\beta}_j$. In our application, which proceeds without explanatory variables, $\mathbf{X}_j$ reduces to a vector of ones, $\boldsymbol{\beta}_j$ reduces to a scalar β_j, and the prior density simplifies to

$$(3)\quad p\left(\beta_j|\sigma_j^2\right)=N\left(\mu_{0j},g_j*\sigma_j^2/n_j\right).$$

As in León-Gonzalez and Scarpa (2008) we stipulate the conventional inverse-gamma (ig) prior for σ_j^2, with shape parameter ν_0 and context-specific scale parameter τ_j.[8] The exact choices for the prior parameters μ_{0j},ν_0, and τ_j, as well as the tuning parameter g_j are discussed below in our empirical section.[9]

For the parameters of the pooled portion within a given model priors are derived "on the fly" as part of the posterior simulator. Specifically, we choose the same prior density families (i.e. normal for $\boldsymbol{\beta}_{mp}$ and inverse-gamma for σ_{mp}^2), and compute prior parameters as context-weighted averages. That is

$$(4)\quad p\left(\boldsymbol{\beta}_{mp}|\sigma_{mp}^2\right)=N\left(\mu_{0,mp},g_{mp}\sigma_{mp}^2/n_{mp}\right),$$
$$p\left(\sigma_{mp}^2\right)=ig\left(\nu_0,\tau_{0,mp}\right),\quad \text{with}$$
$$\mu_{0,mp}=\sum_{j\in mp}\tfrac{n_j}{n_{mp}}\mu_{0j},\ \tau_{0,mp}=$$
$$\sum_{j\in mp}\tfrac{n_j}{n_{mp}}\tau_{0j},$$
$$g_{mp}=\sum_{j\in mp}\tfrac{n_j}{n_{mp}}g_j,\ n_{mp}=\sum_{j\in mp}n_j$$

Our framework also requires a prior model probability for each possible m. As discussed in León-Gonzalez and Scarpa (2008), a more tractable approach is to specify instead a prior pooling probability for each context, i.e. $\alpha_j=$

[7] Specifically, our conjugate normal/inverse-gamma prior for $\boldsymbol{\beta}_j$ and σ_j^2 given below assures an analytical expression for the marginal likelihood. This, in turn, is an integral component of the model selection step of the posterior simulator, as discussed below in more detail.

[8] We parameterize the *ig* density as given in Gelman et al. (2004), p.574, with expectation $\frac{\tau_j}{\nu_0-1}$ and variance $\frac{\tau_j^2}{(\nu_0-1)^2(\nu_0-2)}$.

[9] Fernández, Ley, and Steel (2001) suggest an improper (i.e not integrating to one) prior density for σ^2 to avoid prior variance parameters exerting a strong influence on posterior results. However, in their context the model search is over the inclusion or exclusion of explanatory variables, i.e. the composition of $\boldsymbol{\beta}$. Thus, there is only a single variance term in their framework and it figures in every possible model. Under these circumstances, proper posterior analysis and model comparison are possible even under an improper variance prior. In our case, a given model can include up to j different variance terms. Furthermore, the number of variance terms changes across models. This would lead to erroneous interpretation of posterior model probabilities and Bayes Factors under improper priors, a situation often referred to as the *Bartlett's paradox* (see e.g. Koop, Poirier, and Tobias 2007, ch.11).

$pr\left(\psi_j = 1\right)$. This yields a prior model probability of $p\left(m\right) = \prod_{j=1}^{J} \alpha_j^{\psi_j} \left(1-\alpha_j\right)^{1-\psi_j}$.

Combining the likelihood kernel and priors yields the model-conditioned joint posterior density. Its kernel can be written as

$$
\begin{aligned}
(5) \qquad & p\left(\left\{\boldsymbol{\beta}_j\right\}_{j\in mn}, \left\{\sigma_j^2\right\}_{j\in mn}, \right. \\
& \left. \boldsymbol{\beta}_{mp}, \sigma_{mp}^2 | \mathbf{y}, \mathbf{X}, m\right) \propto \\
& \prod_{j\in J_{mn}} p\left(\boldsymbol{\beta}_j | \sigma_j^2\right) p\left(\sigma_j^2\right) * \\
& p\left(\boldsymbol{\beta}_{mp} | \sigma_{mp}^2\right) * p\left(\sigma_{mp}^2\right) * \\
& p\left(\mathbf{y} | \left\{\boldsymbol{\beta}_j\right\}_{j\in mn}, \left\{\sigma_j^2\right\}_{j\in mn}, \right. \\
& \left. \boldsymbol{\beta}_{mp}, \sigma_{mp}^2, \mathbf{X}, m\right)
\end{aligned}
$$

As described in León-Gonzalez and Scarpa (2008) this joint posterior can be evaluated via a Gibbs Sampler (GS) with a built-in reverse-jump Metropolis-Hastings (MH) step for model selection (Green 1995). The GS proceeds by iterating sequentially through the following steps:

1) Select an initial model m^0, represented by $\boldsymbol{\psi}_{m^0}$, conditional on the data $\mathbf{y}, \mathbf{X}$, and the prior parameters as shown in (2) through (4).
2) Draw, independently, $\boldsymbol{\beta}_j$ from $p\left(\boldsymbol{\beta}_j | \sigma_j^2, \mathbf{y}_j, \mathbf{X}_j\right), \forall j \in mn$, and $\boldsymbol{\beta}_{mp}$ from $p\left(\boldsymbol{\beta}_{mp} | \sigma_{mp}^2, \mathbf{y}_{mp}, \mathbf{X}_{mp}\right)$
3) Draw, independently, σ_j^2 from $p\left(\sigma_j^2 | \mathbf{y}_j, \mathbf{X}_j\right), \forall j \in mn$, and σ_{mp}^2 from $p\left(\sigma_{mp}^2 | \mathbf{y}_{mp}, \mathbf{X}_{mp}\right)$
4) Using a reverse-jump Metropolis-Hastings (MH) step choose between the current model (in the beginning this will be m^0) and a new model m^*, and repeat steps (1)-(3).
5) Repeat steps (1)-(4) until the desired number of parameter draws is reached.

The details for these steps are given in the supplementary online appendix. Importantly, each draw of $\boldsymbol{\beta}_{mp}$ and σ_{mp}^2 is assigned to *every* context that falls into the pooled group for model m. Thus, at the end of the sampling process, say after R iterations, each context will have allocated either R draws of $\boldsymbol{\beta}_j$ and σ_j^2, if it never ends up in a pooled group, or a mix of rn draws of $\boldsymbol{\beta}_j$ and σ_j^2, and rp draws of $\boldsymbol{\beta}_{mp}, \sigma_{mp}^2$, if it was assigned to the pooled group in rp out of R iterations.

That is, upon completion the sampling process delivers draws of coefficients and variances from their respective marginal posterior distribution, *unconditional* of any specific model. Collecting coefficients and variance for a given context in a single vector $\boldsymbol{\phi}_j$, this marginal posterior can be written as

$$
\begin{aligned}
(6) \qquad & p\left(\boldsymbol{\phi}_j | \mathbf{y}, \mathbf{X}\right) = \\
& \sum_{m=1}^{M} p\left(m | \mathbf{y}, \mathbf{X}\right) * \\
& \left[p\left(\boldsymbol{\phi}_j | \mathbf{y}_j, \mathbf{X}_j\right) I\left(j \in mn\right) + \right. \\
& \left. p\left(\boldsymbol{\phi}_{mp} | \mathbf{y}_{mp}, \mathbf{X}_{mp}\right) I\left(j \in mp\right)\right]
\end{aligned}
$$

where $p\left(m|\mathbf{y}, \mathbf{X}\right)$ is the posterior model probability, and $I\left(.\right)$ is an indicator function. Analogously, the moments of these chains also reflect this *model-averaging* process. An important advantage of our approach is that it allows for the computation of both empirical and analytical posterior model probabilities. This can be exploited to verify that the Bayesian sampler has converged, i.e. has visited a sufficiently large sub-space of total model space M for empirical and analytical model probabilities to be virtually indistinguishable (e.g. Fernández, Ley, and Steel 2001). A detailed derivation of both versions of posterior model probabilities is given in the supplementary online appendix.

What Drives Pooling?

Section B in the supplementary online appendix discusses the detailed econometric underpinnings of how prior settings and data characteristics affect pooling probabilities. In a nutshell, a given context is more likely to be assigned to the pooled category if (i) its sample size is close to the combined sample size for the existing pool (thus both individual context and pool rest on comparable empirical evidence), (ii) the context-specific sample mean lies close to the sample mean of the pooled category, (iii) the context sample mean is not too distant from the prior mean of the pooled group ($\mu_{0,mp}$ in equation (4)), and (iv) the within-context variability (as measured in squared deviations from the sample mean) is small compared to the data variability in the pooled category.

Condition (iii) gains in importance over condition (ii) as the g_j-terms, i.e. the tuning parameters in equation (3) decrease in magnitude. This increases the prior variance of β_j and thus places more weight on the actual data for a given context relative to its priors. Thus, the

more informative the context-specific data is relative to its prior, the more stringent is the requirements of mean-closeness in data to be assigned to the common pool. This is intuitively attractive, as it lends a priori more independence to contexts with relatively stronger informational content in the sample data. As we describe below in more detail, in our application we let g_j be directly related to the *original* sample size of all underlying studies that feed into a given context. This places more weight on independence over pooling for contexts that rest on original studies with large sample sizes, i.e. substantial amounts of empirical evidence.

In addition, both conditions (ii) and (iii) become more important with increasing context-specific sample size. In other words, the larger the individual sample size, the closer the context mean must be located to the pooled prior and sample means to be absorbed into the pooled category. Again, this carries the notion of leaning more towards granting independence to contexts that rest on strong empirical evidence, ceteris paribus.

Our algorithm thus strikes an intuitive balance between recognizing context-specific empirical substance (sample size, informative content of priors), penalizing for within-context noise, and observing closeness of central tendencies in both prior and sampling distributions. Section B in the supplementary online appendix provides the mathematical underpinnings for these different pooling criteria.

Allowing for Multiple Pools

One noteworthy limitation of the León-Gonzalez and Scarpa (2008) algorithm is that it only allows for a single pool. With a sufficiently large number of contexts, it is possible that multiple pools exist, each centered around a different population mean. Our preliminary runs with simulated data show that when there are multiple "true" pools, but the model allows only for a single pooled group, the algorithm tends to only recognize the pool with the smallest pooling penalties (see Section B in the supplementary online appendix), i.e the "most obvious" cluster. Naturally, this is less of an issue if the main concern of the analyst is to avoid wrongful pooling, as in León-Gonzalez and Scarpa (2008). However, in our case, we would like to identify all existing pooling patterns between contexts.

As will become evident from our empirical application, missing secondary pools can lead to serious efficiency losses for BT predictions. This is due to the fact that a naïve single-pool model erroneously classifies several contexts as "un-poolable", leaving the analyst to treat these cases as independent, and basing inference on often very small sample sizes.

In theory, a multiple-pool version of our model could be specified along the lines of a latent class, or finite mixture of normals (FMN) model, with each context being assigned a prior and posterior probability of belonging to a specific *class*, that is *information pool* (see e.g. Koop, Poirier, and Tobias 2007; Frühwirth-Schnatter 2001; Frühwirth-Schnatter, Tüchler, and Otter 2004). However, in our case this has several conceptual and computational drawbacks. First, it requires the ex ante specification of the number of expected pools, for which there is little empirical guidance. Second, such a framework would allow for a context to be assigned to multiple pools with nonzero probability. In turn, none of these pools will have a straightforward "label" as the underlying likelihood function is invariable to permutations of the latent class designations.[10] While this would not affect our ability to derive valid posterior predictive distributions of benefits, it would preempt any meaningful interpretation of the emerging pooling patterns. Perhaps the largest hurdle to the implementation of a direct multi-pool model is that the number of model transitions and corresponding transition probabilities in the Metropolis-Hastings step of the Gibbs Sampler become quickly intractable, even with as few as two or three pools.

Instead, we propose a multi-step estimation approach: We first run the Bayesian Model Averaging (BMA) algorithm using the full set of 31 contexts. We then discard all contexts that exhibit a posterior pooling probability of 90% or higher, and repeat the algorithm for the remaining sites. This is based on the observation that the contexts with such large pooling probabilities in the original run pool primarily with each other. Thus, removing these highly pooled contexts in the second round should not heavily

[10] This is the notorious "label-switching" problem in finite mixture models, as discussed e.g. in Frühwirth-Schnatter (2001), Frühwirth-Schnatter, Tüchler, and Otter (2004), and Geweke (1997). As discussed in the latter contribution, this issue is irrelevant if the labeling of the classes has no practical meaning, and for posterior constructs that are implicitly averaged over classes, such as class-unconditioned predicted outcomes.

affect the information content of the remaining data with respect to identifying further pools. The second run produces an additional cluster of contexts with pooling probabilities close to one. We set those aside and run the posterior simulator for a third time to examine if additional residual pools exist. The third run does not produce any additional pronounced pooling patterns, and thus becomes the endpoint of our analytical sequence. The discovery of a second pool leads to substantial efficiency gains in BT predictions for contexts that did not pool in the first stage, but exhibit high pooling probabilities in stage two. Details for these results are given in the next section.

Empirical Application

Data

Our starting data set includes 2,594 observations from 325 individual studies that report WTP / day for access to one of 27 outdoor recreation activities. Each estimate represents an aggregate welfare measure for a specific activity. The level of aggregation differs across observations. About a third of observations represent aggregates over individuals for a single-site context. The remaining welfare estimates are aggregated over individuals and sites, with spatial aggregation ranging from a local site-systems to the national level. The largest subset of the data (40%) represents aggregation at the State level.

We first eliminate duplicate cases, i.e. WTP estimates that are based on the exact same data[11] (17% of the starting set), observations that are not associated with a specific activity (23%), and observations with unknown underlying sample size (14%). Given the focus of this study on inter-activity and inter-regional benefit transfer within the U.S., we also drop cases with aggregation at the national level (3%), and all Canadian entries (3%). In a final cleaning step we eliminate a few isolated activities that are represented by fewer than 10 data points (0.01%).

Summary statistics for the final sample of 1,135 observations are shown in Table 1. There are 164 contributing studies, with underlying primary data collected between 1961 and 2004, and comprising a total of 14 individual outdoor activities. The data also represent four broader census regions: Northeast, Midwest, South, and West. The entries in the table are organized by these regions, and - within region - by activity. The first three regions each encompass six activities, while the West comprises 13 recreation types. Importantly, five activities are included in every region. These are wildlife viewing, running water fishing, stillwater fishing, water fowl hunting, and deer hunting. This allows for an examination if welfare estimates are transferable across regions for the same activity type. The second and third columns depict, respectively, the number of observations and the number of independent underlying studies associated with each region / activity group. Observation counts range from 11 (Northeast, saltwater fishing) to 77 (West, running water fishing). The minimum study count is two (Midwest, motor boating; West, beach), and the maximum is 25 (West, running water fishing). Column four shows the total sample size underlying the *original studies*. These figures range from close to 1,000 (Northeast, water fowl hunting) to over 200,000 (West, hiking). We will utilize this information to assign estimation weights to each category in our preferred econometric specification below.

The columns labeled "%(sp)" and "%(site)" capture the percentage of obervations within a given group that stem from Stated Preference elicitation and, respectively, the share of observations that are associated with a single-site welfare measure. As is evident from the table, these percentages vary widely across categories. We will use this information to examine if patterns of information pooling are sensitive to these study design features.

The remaining columns of Table 1 give sample statistics for WTP (in 2006 dollars) for each region / activity combination. Most group means lie in the $30 to $60 range. There are interesting region-specific patterns as to which activity generates the highest per-day welfare. For example, in the Midwest, running water fishing produces by far the highest average per-day value compared to the other recreation types. In a less pronounced fashion, this also holds for the Northeast. In the West, saltwater fishing and whitewater rafting are the premier outdoor activities based on per-day welfare. The South presents a more homogeneous pic-

[11] The most common source of duplication is within-study, when authors report multiple estimates for the same welfare measure, based on different econometric specifications. In most cases the original authors indicated their preferred specification, which we then retained for the final data set.

Table 1. Meta-data sample statistics

region, activity	obs.	studies	s.size	% (sp)	% (site)	wtp mean	wtp sd	wtp min	wtp max
Northeast, wildlife viewing	41	7	7,459	98%	10%	51.86	42.78	2.56	171.04
Northeast, running water fishing	18	7	15,116	6%	44%	69.48	38.12	14.44	149.57
Northeast, stillwater fishing	31	11	31,522	68%	23%	34.56	21.03	5.11	86.87
Northeast, saltwater fishing	11	5	49,514	64%	18%	36.42	44.94	2.41	132.99
Northeast, water fowl hunting	17	5	962	100%	6%	36.30	22.91	17.04	111.42
Northeast, deer hunting	47	9	6,946	100%	2%	56.85	33.22	11.51	161.98
Midwest, wildlife viewing	39	5	6,934	100%	3%	35.89	17.33	12.46	104.57
Midwest, running water fishing	22	3	2,035	14%	18%	106.05	94.41	17.25	390.45
Midwest, stillwater fishing	72	13	13,673	36%	63%	23.08	19.39	0.71	90.81
Midwest, water fowl hunting	24	3	2,382	96%	0%	31.00	14.77	4.23	65.25
Midwest, deer hunting	58	7	10,073	97%	3%	55.79	19.58	12.59	122.25
Midwest, motor boating	14	2	1,979	0%	100%	16.91	27.92	2.31	90.16
South, wildlife viewing	66	9	9,703	89%	14%	53.76	57.41	2.80	364.73
South, running water fishing	21	6	3,509	29%	81%	58.33	39.25	5.13	176.76
South, stillwater fishing	41	11	14,648	80%	17%	45.92	37.90	3.84	242.27
South, water fowl hunting	30	4	4,969	87%	13%	56.07	44.48	21.91	179.66
South, deer hunting	73	6	9,147	99%	0%	57.53	22.29	8.95	138.19
South, motor boating	14	5	7,129	50%	93%	22.61	16.96	3.05	62.35
West, hiking	45	21	222,684	16%	67%	48.94	57.57	2.61	273.21
West, camping	46	13	6,161	15%	70%	22.31	22.60	1.81	139.53
West, wildlife viewing	76	15	20,496	83%	13%	54.10	45.49	4.24	328.48
West, running water fishing	77	25	25,112	21%	64%	68.59	59.36	8.66	312.71
West, stillwater fishing	51	12	17,076	76%	33%	51.58	39.85	2.56	208.24
West, saltwater fishing	16	3	48,444	0%	88%	141.87	119.35	5.80	372.60
West, water fowl hunting	26	7	3,735	77%	23%	39.67	28.28	3.58	133.26
West, small game hunting	26	5	6,402	19%	69%	44.97	69.18	1.38	305.66
West, deer hunting	71	15	36,426	70%	24%	66.39	41.73	6.39	265.42
West, elk hunting	20	8	9,506	95%	0%	80.01	27.69	37.94	126.86
West, beach	11	2	11,384	0%	100%	14.55	24.39	1.68	84.03
West, whitewater rafting	19	9	2,917	53%	100%	108.01	122.22	5.83	420.06
West, pleasure driving	12	4	3,752	25%	17%	40.74	37.79	16.18	156.12
All	1,135	164	611,795	64%	32%	52.28	49.79	0.71	420.06

obs. = number of obervations in the meta-data; s.size = total underlying sample size in the original studies
%(sp) = percentage of meta-observations associated with Stated Preference elicitation
%(site) = percentage of meta-observations associated with single-site values

ture, with similar mean WTP values for most activities.

Several activities exhibit similar values for mean and standard deviation across regions (e.g. wildlife viewing for the Northeast, South, and West; deer hunting for the Midwest and South). The same holds for different activities within the same region (e.g. stillwater fishing and waterfowl hunting in the Northeast; wildlife viewing and stillwater fishing in the West), and even for different activities across regions (e.g. stillwater fishing in the Midwest and motor boating in the South; water fowl hunting in the South and wildlife viewing in the West). Thus, these sample statistics hint at ample opportunities for borrowing and transferring information across regions, activities, or both. We will explore this possibility more formally using our econometric framework.

Model Specification

Following the bulk of recent meta-regression contributions, we specify our dependent variable as log(WTP) in all subsequent models. This forces predicted WTP to remain in the positive domain, which makes intuitive sense given the "site-access" interpretation of the value estimates captured in our meta-data. Given the lack of meaningful explanatory variables that are observed for all 31 contexts, we opt to specify our empirical models without any regressors. Our framework thus relates to an Analysis-of-Variance (ANOVA), as it considers both the within-context and across-context variability in the dependent variable in its evaluation of pooling potential.[12]

In spirit, our approach also shares common ground with the non-parametric meta-analysis recently suggested by Kaul et al. (2013). As do Kaul et al. (2013), we avoid the risk of mis-specifying the meta-relationship between the outcome of interest and explanatory variables. This also circumvents the "N vs. K" dilemma discussed in Moeltner, Boyle, and Paterson (2007), i.e. the need to truncate the meta-sample due to the lack of regressors for some observations.

This simplification implies that the prior mean of β_j, μ_{0j}, signifies the prior expectation of the outcome variable y_j, i.e. $log(WTP_j)$. This makes it difficult to assign an arbitrary value to this parameter without using the data. Certainly, the popular value of zero for the prior mean of the regression intercept would be a rather extreme choice, given that the observed WTP estimates in our data generally take larger values than \$1. We compromise on this issue by using the sample statistics reported in Table 1 of Walsh, Johnson, and McKean (1992) (p.708) to formulate priors. This table depicts aggregate WTP per day for a variety of outdoor activities that largely overlap with ours. These values are based on 287 individual estimates flowing from 120 studies conducted between 1968 and 1988. Thus, the Walsh, Johnson, and McKean (1992) data set can be interpreted as a sub-set of the earlier portion of our meta-data. We use their reported mean WTP estimates as prior means for β_j, after converting to 2006 dollars and taking logs. Since Walsh, Johnson, and McKean (1992) do not provide dis-aggregated results by census region, we assign the same activity-specific value to each region.[13]

We proceed in similar fashion with the specification of the prior shape and scale parameters for the inverse-gamma density of the error variance σ_j^2. First, we set the shape ν_0 to a value of 2 for all contexts. Given our parameterization, this implies that the prior expectation of σ_j^2 is equal to the prior scale, τ_{0j}. We then use the (implicitly) reported standard deviations in Table 1 of Walsh, Johnson, and McKean (1992) to assign prior values to these shape parameters, after converting to 2006 currency and adjusting for the logged form of our outcome variable.[14].

The final parameter that needs to be chosen a priori is g_j. It enters directly into the expression for the prior variance of β_j (see (3)), and also regulates the relative influence of the prior mean μ_{0j} on model selection, as described in Section B of the supplementary online appendix. Specifically, a large setting for g_j will

[12] However, in stark contrast to classical ANOVA we do not aim for a binary decision rule on the hypothesis that two or more contexts share a common population mean, but rather derive a posterior *probability* for each possible pooling pattern.

[13] Walsh, Johnson, and McKean (1992) do not provide an explicit list of their underlying original studies. However, 61 (37%) of our studies that are associated with original data collected *after* the end of the Walsh, Johnson, and McKean (1992) time frame of 1968-1988. This assures that there is at best partial overlap between the two data sets.

[14] Specifically, in our case τ_{0j} denotes the expectation of the variance of log(WTP). The Walsh, Johnson, and McKean (1992) standard deviations, however, refer to WTP in absolute (un-logged) terms. A conversion is given via $var\left(logWTP_j\right) = \tau_{0j} = log\left(1 + \frac{var(WTP_j)}{E(WTP_j)^2}\right)$

place relatively less weight on μ_{0j} during model estimation and comparison.

León-Gonzalez and Scarpa (2008) follow a recommendation by Fernández, Ley, and Steel (2001) and other empirical BMA contributions and set $g_j = n_j$. For our application this would imply that $var(\beta_j) = \sigma_j^2$, i.e. the prior variance for β_j corresponds to the sampling variance for a single data point within context j (see (3)). The expectation of this variance is given by τ_{0j}, as discussed above. These τ_{0j} terms fall into a range of 0.2 - 0.8 for most contexts. This centers the prior of β_j relatively tightly around μ_{0j}. To allow for more prior variability we prefer setting g_j equal to n_j times the log of the total number of original observations underlying all studies comprised in context j (third column of Table 1 above). This places relatively more weight on the meta-data and less weight on the prior settings for contexts that are associated with a large amount of primary data, regardless of the number of meta-observations n_j. We choose this specification primarily to lend more empirical importance to contexts with large underlying sample sizes. At the same time this also illustrates the ability of a Bayesian meta-analytical framework to incorporate prior information that would be difficult to capture in a classical estimation context.[15] We will henceforth refer to our preferred specification as "M1". Detailed prior settings for all contexts are given in Section D of the supplementary online appendix.

We set the prior pooling probability, α_j, to 0.5 for all contexts. This yields an equal prior probability of 0.5^{31} for all possible models. This uniform prior model probability facilitates model evaluation in the MH step of our algorithm (see Section A of the supplementary online appendix).

We assess convergence of the posterior algorithm by computing the correlation between empirical and analytical posterior model probabilities, as suggested in Fernández, Ley, and Steel (2001). This correlation coefficient exceeds 0.99, and thus suggests that the posterior algorithm provided adequate model coverage.

For comparison purpose, we also estimate a fully independent model, labeled "M2", that estimates population moments separately for each context. The prior settings for β_j and σ_j and the number of burn-ins and retained draws are as for model M1. To assess convergence for the J individual contexts we use Geweke's (1992) convergence diagnostics (CD). These scores clearly indicate convergence for all parameters. To gauge the degree of (undesirable) serial correlation in our Markov chains we also compute Inefficiency Factors (IEF) for all parameters as suggested in Chib (2001). All IEF scores are close to one, which indicates that our posterior simulator has efficient mixing properties.

Estimation Results, Round 1

We implement our search algorithm via a Markov-Chain Monte Carlo (MCMC) program with 100,000 discarded "burn-in" draws and 200,000 retained draws for all model parameters. We start the chain with the fully independent model, i.e. with a zero-vector for model identifier ψ. The algorithm visits 2,669 distinct models in the retained iterations. Posterior probabilities for the top ten models lie in the 0.02 to 0.04 range. While this leaves a considerable degree of model uncertainty for this application, these posterior model probabilities are of orders of magnitude larger than the prior probability of 0.5^{31}. More importantly, the model search produces clear signals for cross-context pooling.

Estimation results for models M1 and M2 are given in table 2. For ease of interpretation the table also repeats key sample statistics (first triplet of columns). The remainder of the table captures posterior pooling probabilities for M1 (column four), and the posterior mean and lower and upper bound of the 95% credible interval for expected WTP, in dollars, for both models.[16] The most important result captured in the table is that there are several contexts that almost always fall into the pooled category, i.e that have posterior pooling probability of close to one. These include *wildlife viewing* in the *Northeast*, *South*, and *West*, *still-water fishing* in the *South* and *West*, two additional non-motorized water activities in the *South* (*running water fishing, water fowl hunt-*

[15] We also estimate a more traditional model with $g_j = n_j$. The results are very similar to our preferred specification, and are thus not reported in the interest of brevity.

[16] The posterior results for expected WTP in dollars were obtained by computing $E(y_{j,r}) = exp\left(\beta_{j,r} + 0.5 * \sigma_{j,r}^2\right)$ for all $r = 1 \ldots R$ draws of parameters flowing from the Gibbs Sampler, and all $j = 1 \ldots J$ contexts. The 95% credible interval comprises the area between the 2.5^{th} and the 97.5^{th} percentile of the posterior distribution for expected WTP.

Table 2. Estimation Results, Round 1

		sample		pr.	Expected WTP M1 (BMA)			M2 (Independent)		
region, activity	obs	mean	std	pool	mean	lower	upper	mean	lower	upper
Northeast, wildlife viewing	41	51.86	42.78	1.00	55.27	48.98	61.32	54.69	41.84	73.28
Northeast, running water fishing	18	69.48	38.12	0.79	59.87	50.00	89.98	75.45	55.87	105.11
Northeast, stillwater fishing	31	34.56	21.03	0.53	45.48	29.69	58.77	35.80	28.64	45.63
Northeast, saltwater fishing	11	36.42	44.94	0.00	53.42	16.48	145.23	50.70	16.44	144.60
Northeast, water fowl hunting	17	36.30	22.91	0.42	45.39	31.31	59.16	38.75	30.44	50.36
Northeast, deer hunting	47	56.85	33.22	0.56	56.96	49.81	67.18	58.51	49.55	69.78
Midwest, wildlife viewing	39	35.89	17.33	0.00	36.63	31.90	42.46	36.58	31.94	42.24
Midwest, running water fishing	22	106.05	94.41	0.13	106.35	55.67	171.82	113.84	77.46	174.86
Midwest, stillwater fishing	72	23.08	19.39	0.00	26.63	19.95	36.38	26.64	19.96	36.41
Midwest, water fowl hunting	24	31.00	14.77	0.30	39.77	26.74	57.61	33.94	26.29	44.85
Midwest, deer hunting	58	55.79	19.58	0.00	57.18	51.85	63.36	57.18	51.84	63.34
Midwest, motor boating	14	16.91	27.92	0.00	15.14	8.15	30.34	15.11	8.12	30.10
South, wildlife viewing	66	53.76	57.41	0.99	55.25	48.91	61.34	54.77	43.48	70.18
South, running water fishing	21	58.33	39.25	0.98	55.44	49.03	61.84	64.23	46.16	92.62
South, stillwater fishing	41	45.92	37.90	0.98	55.08	48.21	61.21	47.10	38.13	59.26
South, water fowl hunting	30	56.07	44.48	0.94	55.37	48.78	61.96	56.50	44.95	72.45
South, deer hunting	73	57.53	22.29	0.00	58.80	53.47	64.89	58.82	53.51	64.97
South, motor boating	14	22.61	16.96	0.00	25.39	16.31	42.56	25.26	16.28	41.29
West, hiking	45	48.94	57.57	0.01	53.07	36.29	80.45	53.09	36.29	80.52
West, camping	46	22.31	22.60	0.00	22.98	17.52	30.81	22.99	17.52	30.86
West, wildlife viewing	76	54.10	45.49	0.99	55.27	49.01	61.30	54.33	46.35	64.23
West, running water fishing	77	68.59	59.36	0.88	57.67	50.72	77.29	71.03	57.64	88.76
West, stillwater fishing	51	51.58	39.85	1.00	55.27	48.98	61.30	53.22	43.43	66.26
West, saltwater fishing	16	141.87	119.35	0.00	170.82	97.56	317.31	171.10	98.63	317.07
West, water fowl hunting	26	39.67	28.28	0.84	53.36	36.53	61.15	44.47	31.46	65.20
West, small game hunting	26	44.97	69.18	0.00	49.53	26.07	100.19	49.52	26.07	99.78
West, deer hunting	71	66.39	41.73	0.42	64.12	53.38	80.35	69.22	58.84	82.30
West, elk hunting	20	80.01	27.69	0.00	84.94	72.35	100.71	84.93	72.31	100.82
West, beach	11	14.55	24.39	0.00	13.58	6.48	30.52	13.56	6.45	30.56
West, whitewater rafting	19	108.01	122.22	0.02	131.64	59.46	293.79	133.02	63.76	296.52
West, pleasure driving	12	40.74	37.79	0.92	54.15	37.87	61.18	41.80	30.30	59.59

ing), and one additional activity in the *West* (*pleasure driving*).

For most of these eight contexts, sample means ($50 - 55) and standard deviations ($38-45) are of comparable magnitude, which favors posterior pooling. Two of the contexts, *West, pleasure driving* and *South, stillwater fishing* have somewhat lower sample means, in the $40-45 range. For *West, pleasure driving*, the meta-sample size (12) is small enough to dampen the pooling penalty that derives from the deviation of the context-specific sample mean ($40.74) from the pooled sample mean (about $55) to be sorted into the pooled group by the posterior algorithm (see (**??**) in Section B of the supplementary online appendix). For *South, stillwater fishing* the context-specific sample mean is close enough to the pooled sample mean such that the efficiency penalty that comes with independent estimation exceeds the penalty of mean-deviation, as illustrated in Section B of the supplementary online appendix.

In general, the algorithm categorically rejects pooling for contexts that have extreme sample means (e.g. *Midwest, running water fishing*,*West, saltwater fishing*, *West, beach*), or a large enough meta-sample size such that even a minor deviation from the pooled sample mean translates into prohibitive pooling penalties (e.g. *Midwest, South, deer hunting*). The latter tendency is consistent with the finding in León-Gonzalez and Scarpa's simulation exercise, where pooling probabilities decrease with increasing sample size for sites known to be independent.

The posterior pooling probabilities captured in table 2 indicate how often a given context was assigned to the pooled category, but do not provide information on pooling associations, i.e. which other contexts were included in the pool at a given iteration. To visualize pairwise pooling probabilities, we create a diagram akin to "heat maps" used in the physical sciences. It is a 2 by 2 symmetric grid where each cell corresponds to a specific pair of contexts. The intensity of the shading of the cell indicates the magnitude of pairwise pooling probabilities, i.e. how often the specific two contexts were included in the pooled group *together*. Figure 1 displays these patterns. The legend at the bottom of the figure show abbreviations for regions and activities, as well as a key for pooling intensity. The diagonal line of cells captures the own-pooling probabilities, i.e. how often a context was assigned to the pooled category. The exact numerical values for these probabilities are given in the "pr.pool" column of table 2. The figure confirms that the contexts with extremely high pooling probabilities listed above are naturally also highly pooled in each pairwise combination. A second pooling pattern that emerges is that for three of the four regions, *Northeast, South* and *West*, *wildlife viewing* pools at least moderately high with water-based fishing and hunting activities, which, in turn, pool with one another. The figure also visualizes the weak pooling patterns associated with *any* activity in the Midwest, and the complete absence of any pooling for a variety of contexts. However, as illustrated below, some of these seemingly independent contexts will form their own pool in our second estimation round.

The last six columns of table 2 compare posterior results for expected WTP (in dollars) for the BMA (M1) and the Independent model (M2). For contexts that never pool the results for the two models are close to identical. Remaining differences are solely related to random simulation noise. The payoff for recognizing pooling patterns becomes visible when comparing credible intervals for contexts with high pooling probabilities. The efficiency gains in terms of tighter credible intervals are staggering, ranging from 15-70% for contexts with pooling probability of 0.8 or higher. This pronounced information gain is visualized in figure 2, which depicts posterior densities for expected WTP for four contexts with high pooling probabilities along with their independent counterparts.[17]

Estimation Results, Round 2

We discard all contexts from the first round with posterior pooling probabilities in excess of 90%. This facilitates the detection of a second pool, if it exists. We then re-run the posterior simulator for model M1. The results are shown in table 3. There are now five new contexts with posterior pooling probability of close to one: *Northeast, saltwater fishing; Midwest, stillwater fishing; South, motor boating; West, camping, small game hunting.* All of them had zero pooling probability in the first round. This illustrates the benefit of our multi-step

[17] Similar efficiency gains hold for median WTP, computed as $med\,(y_{j,r}) = exp\,(\beta_{j,r})$ for all $r = 1 \ldots R$ draws of parameters flowing from the Gibbs Sampler. These results and corresponding figures are provided in the supplementary online appendix.

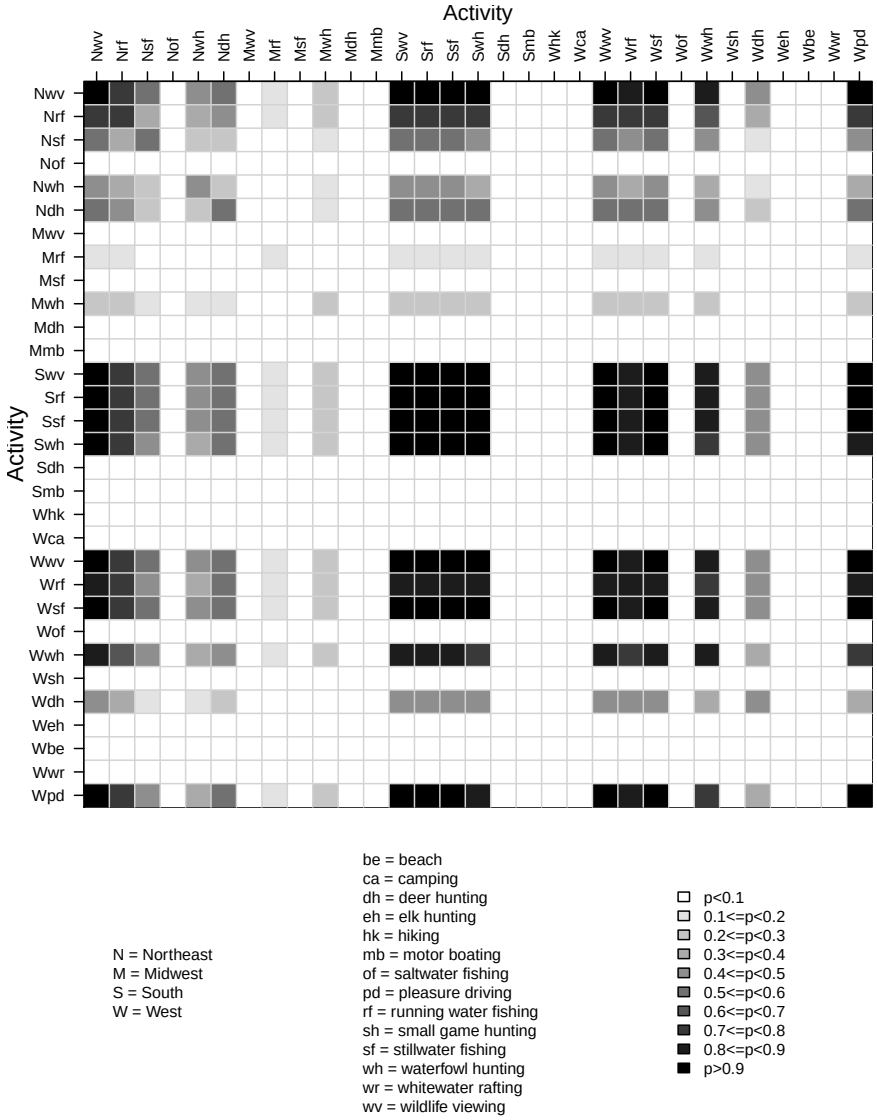

Figure 1. Heat Map for Pairwise Pooling Probabilities, Round 1

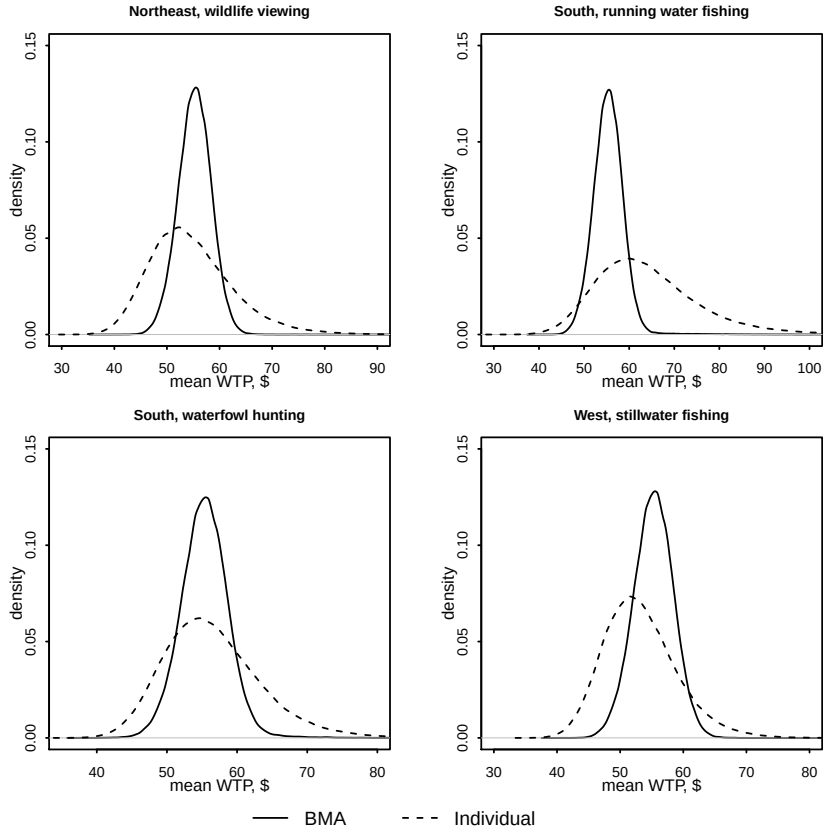

Figure 2. Posterior distributions for expected WTP, Round 1

Table 3. Estimation Results, Round 2

					Expected WTP					
		sample		prob.	M1 (BMA)			M2 (Independent)		
region, activity	obs.	mean	std	pool	mean	lower	upper	mean	lower	upper
Northeast, running water fishing	18	69.48	38.12	0.00	75.48	55.94	104.86	75.45	55.87	105.11
Northeast, stillwater fishing	31	34.56	21.03	0.01	35.81	28.63	45.62	35.80	28.64	45.63
Northeast, saltwater fishing	11	36.42	44.94	0.99	31.46	23.51	40.27	50.70	16.44	144.60
Northeast, water fowl hunting	17	36.30	22.91	0.00	38.77	30.48	50.32	38.75	30.44	50.36
Northeast, deer hunting	47	56.85	33.22	0.00	58.52	49.56	69.89	58.51	49.55	69.78
Midwest, wildlife viewing	39	35.89	17.33	0.00	36.59	31.92	42.23	36.58	31.94	42.24
Midwest, running water fishing	22	106.05	94.41	0.00	113.94	77.53	174.84	113.84	77.46	174.86
Midwest, stillwater fishing	72	23.08	19.39	1.00	31.25	23.53	39.68	26.64	19.96	36.41
Midwest, water fowl hunting	24	31.00	14.77	0.03	33.92	26.30	44.76	33.94	26.29	44.85
Midwest, deer hunting	58	55.79	19.58	0.00	57.20	51.86	63.35	57.18	51.84	63.34
Midwest, motor boating	14	16.91	27.92	0.82	27.97	10.12	38.73	15.11	8.12	30.10
South, deer hunting	73	57.53	22.29	0.00	58.81	53.49	64.92	58.82	53.51	64.97
South, motor boating	14	22.61	16.96	0.94	30.88	21.96	39.74	25.26	16.28	41.29
West, hiking	45	48.94	57.57	0.73	38.19	27.28	68.45	53.09	36.29	80.52
West, camping	46	22.31	22.60	0.96	30.77	22.12	39.09	22.99	17.52	30.86
West, running water fishing	77	68.59	59.36	0.00	71.02	57.65	88.78	71.03	57.64	88.76
West, saltwater fishing	16	141.87	119.35	0.00	171.16	98.63	318.54	171.10	98.63	317.07
West, water fowl hunting	26	39.67	28.28	0.37	40.33	28.24	61.81	44.47	31.46	65.20
West, small game hunting	26	44.97	69.18	1.00	31.34	23.61	39.96	49.52	26.07	99.78
West, deer hunting	71	66.39	41.73	0.00	69.24	58.85	82.29	69.22	58.84	82.30
West, elk hunting	20	80.01	27.69	0.00	84.92	72.27	100.81	84.93	72.31	100.82
West, beach	11	14.55	24.39	0.83	28.02	8.47	39.06	13.56	6.45	30.56
West, whitewater rafting	19	108.01	122.22	0.03	130.40	47.46	292.73	133.02	63.76	296.52

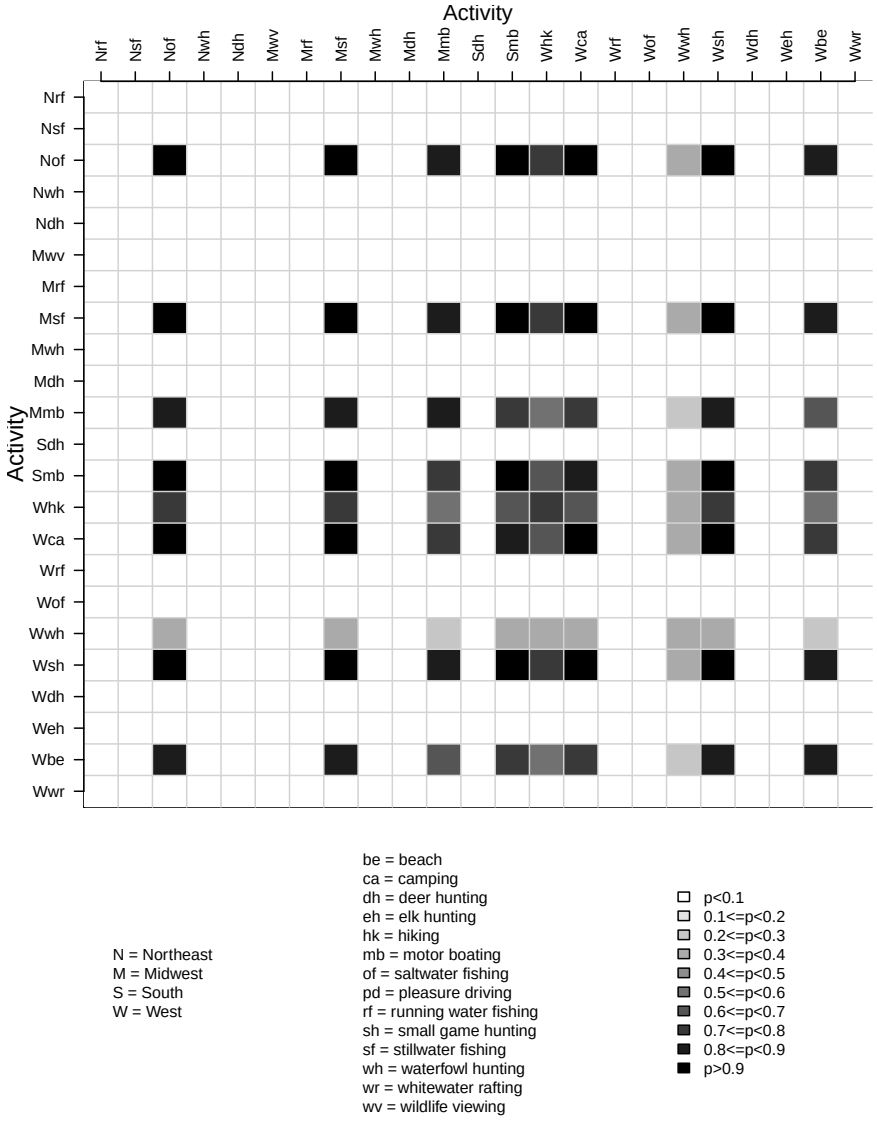

Figure 3. Heat Map for Pairwise Pooling Probabilities, Round 2

approach, as the new pool would have otherwise remained undetected. The "heat map" with pairwise pooling probabilities for round two is given in figure 3. It shows that three additional activities, *Midwest, motor boating* and *West, hiking, beach* also figure relatively frequently in the pooled category. A comparison of posterior densities for expected WTP for the BMA and the Independent model mirrors the finding from the first round that information borrowing across contexts can produce sizable efficiency gains. For example, the credible intervals for *Northeast, saltwater fishing* and *West, small game hunting* generated by the independent model are almost twice as wide than those produced by the pooled version. Figure 4 compares posterior densities for four of the five contexts that did not pool in the first round, but formed a high-probability information pool in round two. As can be seen from the figure, the context-specific (individual) models generally produce WTP predictions with much higher posterior variability compared to the pooled version. This effect is especially pronounced for *Northeast, saltwater fishing* (upper left panel) and *West, small game hunting* (lower right panel).

In summary, extending the León-Gonzalez and Scarpa (2008) model to allow for multiple information pools has important practical implications, as it can greatly reduce the plausible range for predicted benefits.

Robustness Checks

Ideally, we would like to interpret the pooling patterns that emerge from our analysis as evidence of value similarities across activities and regions. However, it is also possible that our results are largely driven by commonalities in study design or implementation, usually referred to as "methodological factors" (e.g Johnston et al. 2006; Moeltner, Boyle, and Paterson 2007) or "study design features" (Boyle et al. 2010) in meta-analytical work. Furthermore, common levels of aggregation underlying individual meta-observations could in theory affect pooling patterns.

Our data set includes two variables that allow for at least a cursory check if such undesirable methodological pooling effects might be present: an indicator variable for Stated Preference elicitation and an indicator variable for WTP values that derive from site-specific valuation, as opposed to regional aggregates over multiple sites. The proportion of observations falling into each category for a given context are listed in table 1 under the heading of *%sp* and *%site*, respectively.

At first glance, table 1 shows that the highly pooled contexts in rounds 1 and 2 of our original analysis exhibit strong variability for both methodological categories. For the eight contexts with pooling probabilities over 90% from the first round, the within-context proportion of *sp* observations ranges from 25% to 100%, and the proportion of *site* specific observations from 13% to 81%. For the five highly pooled contexts from round 2, *sp* proportions range from 15% to 68%, and *site* proportions from 17% to 93%. Thus, based on this purely descriptive inspection, there does not appear to exist an obvious pattern of methodology or aggregation-driven pooling.

To explore this possibility more formally we repeat our analysis for the two separate subsets of our meta-data associated with positive entries for *sp* and *site*, respectively. In both cases, we follow the strategy for composing the original data set and eliminate contexts with fewer than ten remaining observations after applying the respective filters. Table 4 contains sample statistics and estimation results for the *sp*-only set. On average, the filtering by *sp* led to a 20% loss in observations across clusters. The set includes eight contexts that pooled at 80% or higher in the original round 1, and only one context that pooled highly in the original round 2. Thus, in absence of any methodological effects, we would again expect our round 1 pattern to emerge for this sub-set. This is indeed the case. Despite an average observation loss of 25% for the highly-pooled cases, the original pooling pattern largely survives, with six of the eight contexts pooling again at 79% or higher. Thus, we conclude that the original primary pooling pattern is relativley robust to the elicitation format used in the underlying studies.

Sample statistics and estimation results for the *site*-only subset are given in table 5. The set contains 14 remaining clusters, with an average observation loss of 31% compared to the unfiltered data. The set includes four contexts that pooled at 80% or higher in our original round 1, and six contexts that pooled at 80% or higher in our original round 2. Thus, we expect emerging pooling patterns to be dominated by the latter group, in absence of any aggregation

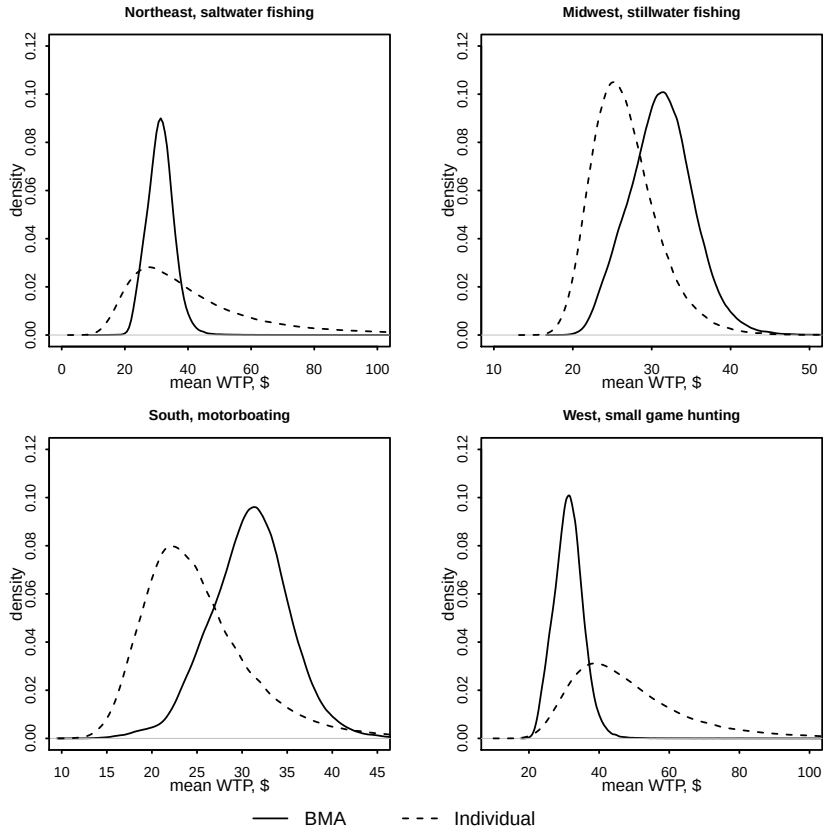

Figure 4. Posterior distributions for expected WTP, Round 2

Table 4. Results for the SP-elicitation subset

region, activity	obs	sample mean	std	pr. pool	Expected WTP mean	lower	upper
Northeast, wildlife viewing	40	50.35	42.21	0.99	50.09	43.70	57.27
Northeast, stillwater fishing	21	27.08	16.95	0.17	31.46	22.26	50.78
Northeast, water fowl hunting	17	36.30	22.91	0.65	45.54	32.21	55.32
Northeast, deer hunting	47	56.85	33.22	0.41	55.50	46.82	68.24
Midwest, wildlife viewing	39	35.89	17.33	0.05	36.97	31.93	46.17
Midwest, stillwater fishing	26	32.33	12.91	0.38	39.42	28.68	52.75
Midwest, water fowl hunting	23	31.88	14.44	0.68	44.66	29.00	54.96
Midwest, deer hunting	56	56.12	18.76	0.00	57.25	52.50	62.61
South, wildlife viewing	59	50.03	57.64	0.93	50.08	43.32	57.69
South, stillwater fishing	33	38.57	15.51	0.81	47.85	36.29	55.66
South, water fowl hunting	26	40.31	17.42	0.08	42.55	35.91	51.30
South, deer hunting	72	57.67	22.42	0.00	58.98	53.59	65.22
West, wildlife viewing	63	54.59	44.57	0.89	50.83	44.66	59.21
West, running water fishing	16	78.84	72.77	0.79	58.28	45.07	112.39
West, stillwater fishing	39	54.13	44.36	0.99	50.16	43.79	57.51
West, water fowl hunting	20	43.56	25.32	0.40	47.11	37.88	56.15
West, deer hunting	50	69.58	25.74	0.00	72.73	63.62	83.78
West, elk hunting	19	77.78	26.53	0.00	82.74	70.36	98.20
West, whitewater rafting	10	139.91	106.85	0.01	183.22	83.70	400.47

Table 5. Results for the SITE-specific subset

region, activity	obs	sample mean	std	pr. pool	Expected WTP mean	lower	upper
Midwest, stillwater fishing	45	17.09	18.53	0.86	23.28	15.34	35.64
Midwest, motor boating	14	16.91	27.92	0.84	22.63	10.41	36.09
South, running water fishing	17	49.20	27.25	0.12	56.44	39.57	80.86
South, motor boating	13	19.56	13.04	0.72	23.73	16.03	37.72
West, hiking	30	57.53	67.30	0.27	60.09	27.49	106.82
West, camping	32	24.03	25.62	0.85	24.54	17.61	38.12
West, wildlife viewing	10	66.71	52.59	0.36	67.04	22.78	145.69
West, running water fishing	49	62.77	55.67	0.12	65.55	50.45	87.52
West, stillwater fishing	17	52.34	45.05	0.12	54.86	39.37	76.77
West, saltwater fishing	14	154.19	121.43	0.00	167.25	109.74	266.06
West, small game hunting	18	44.54	83.36	0.88	26.80	17.51	58.58
West, deer hunting	17	37.55	28.41	0.12	41.95	28.30	68.03
West, beach	11	14.55	24.39	0.85	22.56	8.63	36.47
West, whitewater rafting	19	108.01	122.22	0.11	125.25	56.36	289.17

effects. This is confirmed by the results captured in the table. Five of the six highly pooled contexts in original round 2 pool again at 80% or higher, despite an average observation loss of 20% compared to the full data. Moreover, when we eliminate these cases and re-run the model, a pooling pattern emerges that closely resembles our original round 1 results. Specifically, all four originally pooled contexts pool again with probabilities of 85-99%. Thus, we do not find any evidence of pronounced aggregation effects that may drive the identified pooling patterns.

Discussion

As mentioned above, our results are probably most meaningful when applied to a BT situation where a general, or aggregate value estimate is needed for a given activity and region. For example, a decision maker may seek the value of a "typical day of motor boating in the South", perhaps in the context of a regional economic impact analysis of new legislation on motorized boating. The available *direct* evidence consists of 14 observations from 5 underlying studies (see table 1), producing an estimated expectation of $25.28 (table 3). However, our Bayesian model suggests that the value distribution specific to this activity / region pair is practically indistinguishable from value distributions associated with four or five other contexts, as discussed above. The Bayesian pooled expectation of $30.88 is based on an implicit sample of 169 observations, involving 41 original studies. Thus, the BMA estimate is likely a more reliable indicator of boating values in the South, even though the bulk of contributing observations come from other activities and regions.

As noted in León-Gonzalez and Scarpa (2008), the benefits of exploiting the pooling patterns made transparent by the BMA algorithm are largest in small sample situations for the target site or - in our case - contexts. The ability to borrow information from other contexts is less critical when own-context sample sizes are relatively large. For example, a policy maker interested in the per-day value of wildlife viewing in the West, perhaps in context of a benefit-cost analysis of a new regional wildlife management plan, can resort to 76 observations from 15 original studies to derive this estimate ($54.35, table 2). As is evident from table 2, little is gained by substituting this figure for the Bayesian pooled mean of $55.29.

Overall, our analysis presents a promising picture for the potential of cross-context information borrowing for outdoor recreation. Of all 31 activity / region pairs in our meta-data, 26 pool at least occasionally with other combinations in one of the two estimation rounds. Only four contexts are persistently reluctant to pool with any other activity / region pairs. These are *wildlife viewing* in the Midwest, *water fowl hunting* in the south, and *saltwater fishing* and *elk hunting* in the west. In those cases the decision maker is left to make do with context-specific data to draw inference on recreation values.

Naturally, this raises the question as to which contexts ought to be included in the meta-data in the first place.[18] As described above, our algorithm makes pooling decisions purely based

[18] We thank one of our reviewers for raising this point.

on statistical criteria, such as within-sample variability and closeness of sample means. In theory, the meta-model could include a hodgepodge of contexts from any realm of socioeconomic activity or even purely scientific processes, each of which may be pooled with any other if its statistical properties allow for it. However, this is exactly akin to the notion of which variables to include in a regression model if predictive fit is the primary analysis goal. In both cases the analyst's judgment is needed to decide on a reasonable scope for the model. In regression modeling, the ultimate inclusion criterion is likely the - at least remote - plausibility of a causal relationship with the dependent variable. In our case, it would probably be prudent to only include contexts that are - at least at some level - conceptually related, say along the notion of *potentially* flowing from common structural preferences. For our application, all contexts can be lumped into the umbrella of "leisure activities", or - even more narrowly - "outdoor recreation activities". It is not unreasonable to assume that structural preferences, say marginal utilities, might be similar across the activities and/or regions included in our data.

A more practical decision rule for the analyst might be based on the (hypothetical) questions if the strength of pooling between two, potentially very disparate, contexts would likely increase or diminish with increasing context-specific sample size. Our algorithm is quite sensitive to own-sample sizes, and is more likely to reject pooling as context-specific sample sizes increase, ceteris paribus. Additional work along these lines and, more generally, addressing the issue of "optimal scope" of the meta-data when a search for information pools for BT purposes is the primary goal could be very beneficial.

Conclusion

We adapt the Bayesian Model Search algorithm by León-Gonzalez and Scarpa (2008) to explore if different outdoor recreation contexts based on different underlying populations may nonetheless share the same value distribution. Using a large meta-data set comprising 14 outdoor activities across four U.S. regions we find strong evidence of value similarity across multiple activities and regions. Exploiting these revealed pooling patterns allows for the derivation of more reliable, and in some cases vastly more efficient, WTP distributions for a specific context. This ability to borrow information across contexts is especially beneficial in small sample settings, which is often the norm if only meta-data are available.

We modify León-Gonzalez and Scarpa's econometric framework along two primary dimensions. First, we relax their constraint of a single information pool. Using a simple and intuitive multi-step approach we allow for the emergence of secondary, more subtle pools that would otherwise remain undiscovered. This leads to further efficiency gains in BT predictions for contexts that would have erroneously been classified as "un-poolable" in a single pool framework. Second, we adapt León-Gonzalez and Scarpa's Bayesian algorithm, which is geared towards a discrete choice / contingent valuation approach, to accommodate the likelihood function and priors for a standard linear meta-regression model.

Our Bayesian estimation framework exhibits numerous desirable features, such as the ability to quickly identify pooling patterns for a relatively large context space, and to capture information spillovers even under imperfect, i.e. partial pooling. However, we would like to re-iterate the main objective of this analysis, that is to highlight the feasibility and potential practical benefits of cross-context BT. If only a handful of contexts are under consideration, the analyst could in theory proceed within a classical estimation framework, for instance by estimating a general model with context-specific parameters, and then testing for potential cross-context pooling restrictions. We primarily hope that our work will encourage researches to consider a broader mix of contexts when specifying that "general model".

On a final note, we motivate our cross-context analysis by arguing that value distributions can converge across contexts despite differences in site characteristics, population features, or preferences. By the same token, we cannot infer preference similarity from observationally identical value distributions, as tempting as it may be. However, if subsequent analysis, overcoming the limitations of our meta-data, should reveal that two or more highly pooled contexts from our BMA model *also* exhibit strong similarities in population and site characteristics, similarity in value distributions may indicate similarity in preferences. This would be a logical extension of our analysis and a fruitful avenue for future research.

References

Bateman, I., R. Brouwer, S. Ferrini, M. Schaafsma, D. Barton, A. Dubgaard, B. Hasler, S. Hime, I. Liekens, S. Navrud, L. De Nocker, R. Ščeponavičiūtė, and D. Semėnienė. 2011. Making benefit transfer work: Deriving and testing principles for value transfers for similar and dissimilar sites using a case study of the non-market benefits of water quality improvements across Europe. *Environmental and Resource Economics* 50: 365–387.

Boyle, K., and J. Bergstrom. 1992. Benefit transfer studies: myths, pragmatism, and idealism. *Water Resources Research* 28: 657–663.

Boyle, K., N. Kuminoff, C. Parmeter, and J. Pope. 2009. Necessary conditions for valid Benefit Transfers. *American Journal of Agricultural Economics* 91: 1328–1334.

Boyle, K., N. Kuminoff, C. Parmeter, and J. Pope. 2010. The Benefit Transfer Challenges. *Annual Review of Resource Economics* 2: 161–182.

Brouwer, R. 2000. Environmental value transfer: State of the art and future prospects. *Ecological Economics* 32: 137–152.

Chib, S. 2001. Markov Chain Monte Carlo Methods: Computation and Inference. In *Handbook of Econometrics*, volume 5. Elsevier.

Fernández, C., E. Ley, and M. Steel. 2001. Benchmark priors for Bayesian model averaging. *Journal of Econometrics* 100: 381–427.

Frühwirth-Schnatter, S. 2001. Markov Chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association* 96: 194–209.

Frühwirth-Schnatter, S., R. Tüchler, and T. Otter. 2004. Bayesian analysis of the heterogeneity model. *Journal of Business & Economic Statistics* 22: 2–15.

Gelman, A., J. Carlin, H. Stern, and D. Rubin. 2004. *Bayesian Data Analysis, Second Edition*: Chapman & Hall/CRC.

Geweke, J. 1992. Evaluating the Accuracy of Sampling-based Approaches to the Calculation of Posterior Moments. In *Bayesian Statistics 4*. Oxford University Press, Oxford, UK.

Geweke, J. 1997. Interpretation and inference in mixture models: Simple MCMC works. *Computational Statistics and Data Analysis* 51: 3529–3550.

Green, P. 1995. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82: 711–732.

Johnston, R., M. Ranson, E. Besedin, and E. Helm. 2006. What determines willingness-to-pay per fish? A meta-analysis of recreational fishing values. *Marine Resource Economics* 21: 1–32.

Johnston, R., and R. Rosenberger. 2010. Methods, trends and controversies in contemporary benefit transfer. *Journal of Economic Surveys* 24: 479–510.

Kaul, S., K. Boyle, N. Kuminoff, C. Parmeter, and J. Pope. 2013. What can we learn from Benefit Transfer Errors? Evidence from 20 years of research on convergence validity. *Journal of Environmental Economics and Management* 66: 90Ũ–104.

Koop, G., D. Poirier, and J. Tobias. 2007. *Bayesian Econometric Methods*: Cambridge.

León-Gonzalez, R., and R. Scarpa. 2008. Improving multi-site benefit functions via Bayesian model averaging: A new approach to Benefit Transfer. *Journal of Environmental Economics and Management* 56: 50Ũ–68.

Loomis, J., and R. Rosenberger. 2006. Reducing barriers in future benefit transfers: Needed improvements in primary study design and reporting. *Ecological Economics* 60: 343–350.

Moeltner, K., K. Boyle, and R. Paterson. 2007. Meta-Analysis and Benefit-Transfer for Resource Valuation: Addressing Classical Challenges with Bayesian Modeling. *Journal of Environmental Economics and Management* 53: 250–269.

Moeltner, K., R. Johnston, R. Rosenberger, and J. Duke. 2009. Benefit transfer from multiple contingent experiments: A flexible two-step model combining individual choice data with community characteristics. *American Journal of Agricultural Economics* 91: 1335–1342.

Moeltner, K., and R. Rosenberger. 2008. Meta-Regression and Benefit Transfer: Data Space, Model Space, and the Quest for 'Optimal Scope'. *B.E. Journal of Economic Analysis & Policy* 8.

U.S. Environmental Protection Agency 2000. *Guidelines for preparing economic analyses*. United States Environmental Protection Agency, Office of Water.

Walsh, R., D. Johnson, and J. McKean. 1992. Benefit transfer of outdoor recreation demand studies, 1968–1988. *Water Resources Research* 28: 707–713.