What Can We Learn from Benefit Transfer Errors?

Evidence from 20 Years of Research on Convergent Validity

Sapna Kaul, Dept. of Agricultural and Applied Economics, Virginia Tech, sapnak@vt.edu[♦]

Kevin J. Boyle, Dept. of Agricultural and Applied Economics, Virginia Tech, kjboyle@vt.edu

Nicolai V. Kuminoff, Dept. of Economics, Arizona State University, kuminoff@asu.edu

Christopher F. Parmeter, Dept. of Economics, University of Miami, cparmeter@bus.miami.edu

Jaren C. Pope, Dept. of Economics, Brigham Young University, jaren_pope@byu.edu

ABSTRACT: We develop a nonparametric approach to meta-analysis and use it to identify modeling decisions that affect benefit transfer errors. The meta-data describe the results from 31 empirical studies testing the convergent validity of benefit transfers. They evaluated numerous methodological procedures, collectively reporting 1071 transfer errors. Our meta-regressions identify several important findings, including: (1) the median absolute error is 39%; (2) function transfers outperform value transfers; (3) transfers describing environmental *quantity* generate lower transfer errors than transfers describing *quality* changes; (4) geographic site similarity is important for value transfers; (5) contingent valuation generates lower transfer errors than other valuation methods; and (6) combining data from multiple studies tends to reduce transfer errors.

KEY WORDS: benefit transfer, function transfer, convergent validity, meta-analysis

---

[♦] Corresponding author: 208 Hutcheson Hall (0401), Blacksburg, VA 24061. Phone: (540) 231-6301.

1. Introduction

Benefit transfer is one of the most common methods for conducting benefit-cost analysis at the U.S. Environmental Protection Agency [39]. When it is too time-consuming or expensive to directly estimate the monetary benefits of a policy, a surrogate measure is produced from preexisting estimates. A benefit transfer takes preexisting values from a study case (or cases) to develop a customized benefit estimate for a new policy case.[1] A "value transfer" simply substitutes a point estimate from a previous study or, in some cases, the mean or median of several point estimates. A "function transfer" predicts benefits using a previously calibrated function describing how values vary with the characteristics of people and places [21].[2]

Since a 1992 issue of *Water Resources Research* (Vol. 28, 3) raised academic interest in benefit transfers among environmental economists, at least 40 studies have investigated the empirical accuracy of this method using tests of convergent validity, which considers the difference between two different estimates of the same theoretical construct. These tests are designed to measure benefit transfer error, defined as the difference between a benefit measure estimated using original data (i.e. the policy case) and a surrogate for that benefit measure based on preexisting estimates (i.e. the study cases). Two stylized facts have emerged. Function transfers generate lower transfer errors than value transfers, and greater similarity between the study and policy cases reduces transfer errors [4][17][18][31]. The apparent lack of consensus on other methodological features of the transfer process has made it difficult to define specific protocols for the

---

[1] We adopt the phrases *study case* and *policy case* from the U.S. EPA's 2010 *Guidelines for Preparing Economic Analyses* [39]. This is a change from the conventional terminology, *study site* and *policy site*. Benefits are not always transferred to new geographic sites. Some transfers occur at the same physical location, using past values to assess current situations or predict future outcomes. Thus, benefits may be transferred to a new policy case at the same study site or to a new policy case at a different site.

[2] Readers seeking background on concepts, terminology, and methods in the benefit transfer literature are directed to the surveys prepared by Bergstrom and Taylor [2] , Boyle et al. [4], and Johnston and Rosenberger [17].

conduct of benefit transfers and to develop a clear agenda for research. Johnston and Rosenberger [17] observe, the "*complexity and relative disorganization of the (academic) literature may represent an obstacle to the use of updated (benefit-transfer) methods by practitioners.*"

The purpose of this paper is to investigate if specific modeling decisions can be identified that affect benefit transfer errors, and this is done using a new statistical approach to meta-analysis. Our quantitative research design complements and extends previous qualitative assessments of the literature [31][32]. We begin by systematically reviewing empirical studies on the convergent validity of benefit transfers conducted over the past 20 years. These studies tested a tremendous variety of methodological procedures, collectively reporting *more than one thousand* benefit transfer errors. Most of these observations come from studies conducted in the United States and Western Europe. The applications cover a wide variety of amenities. Examples include access to forest, park, and lake recreation; hunting; changes in the quantity and quality of water in lakes, rivers, and coastal areas; air quality; exposure to ultraviolet radiation; freshwater fishing in streams and rivers; proximity to various types of open space; farmland amenities; and measures of the overall ecological health of watersheds, wetlands, and rivers.

It is standard practice in meta-analyses to use linear meta-regressions with robust standard errors to distill the collective findings on important questions in the field of environmental economics [27]. However, we have some concerns about the credibility of a *linear* meta-regression in the context of our analysis. The modeling decisions that comprise a benefit transfer are represented by binary variables that can interact in complex ways to influence benefit transfer errors [3]. Any specific variable can have a unique effect on the outcome being investigated (meta-equation regressand) when combined with sets of other regressors. For example, similarity between study and policy cases may be crucial to the accuracy of a value transfer, but not as im-

portant for a function transfer that can be calibrated to policy case conditions. A statistical approach that allows for these interaction effects has the potential to provide richer insights about the literature being investigated. The typical linear meta-regression imposes separability of each of the regressors. Capturing all potential interaction effects would require adding an intractable number of regressors to the meta-equation. Therefore, we propose a new nonparametric approach to meta-analysis that does not impose the linearity and separability assumptions. We contrast the insights from our new approach with a conventional parametric approach.

Nonparametric meta-analysis is particularly useful for our application to benefit transfer because it avoids the need to impose a-priori restrictions on the functional relationship between benefit-transfer errors and the various modeling decisions that comprise the transfer process. The nonparametric approach generates response effects for the combination of regressors at every data point, which yields ranges of effects instead of the parametric point estimates of average impacts. This allows us to recognize that there are multiple possible effects for each transfer characteristic, with specific impacts depending on the empirical context defined by other methodological choices made by analysts in implementing a benefit transfer.

We follow Charles Manski's [22] "bottom-up" approach to data analysis where we start with the less restrictive nonparametric analysis and then move to a more restrictive parametric analysis. We use the nonparametric model to estimate the ranges of impacts for the benefit transfer characteristics that can be identified from variation in the meta-data. Then we add the linearity and separability restrictions that are commonly used in meta-regressions and repeat the estimation using weighted least squares. Following Manski's logic, we recognize that the credibility of inference based on our results is decreasing in the strength of the parametric restrictions that we impose on the meta-regression as findings may not be robust to relaxation of the imposed as-

sumptions in the parametric analysis.

Our parametric and nonparametric models allow us to distill several important findings from the literature. First, our analysis confirms the stylized fact that function transfers outperform value transfers. Second, benefit transfers valuing changes in environmental *quality* (e.g. an increase in river clarity) almost always have larger errors than transfers describing *quantity* changes (e.g. an increase in the quantity of domestic water availability). Third, the geographic proximity of the study and policy locations tends to reduce transfer errors, especially for value transfers. Fourth, drawing on information from multiple preexisting studies (as opposed to a single study) also tends to reduce transfer errors, especially for function transfers. Lastly, for the stated-preference applications, contingent valuation generates lower transfer errors than choice modeling. For revealed preference applications, site demand models perform better than site choice models. Meta-analysis generates transfer errors that lie in between the stated and revealed preference models.

2. Conceptual Framework

The process of benefit transfer begins by defining the relevant measure of benefits. Consider a public policy that is expected to change the quality of an environmental amenity from q^0 to q^1 at policy case j. A partial equilibrium Hicksian measure of willingness to pay (wtp) for this change is defined as:

$$V_{ij}\left(p_j, x_j, q_j^1, y_i - wtp_{ij};\ \alpha_i, d_i\right) = V_{ij}\left(p_j, x_j, q_j^0, y_i;\ \alpha_i, d_i\right), \tag{1}$$

where V_{ij} is the indirect utility for individual i at case j expressed as a function of market prices (p_j), other non-priced attributes (x_j), individual income (y_i), other demographic characteristics (d_i), and latent preferences (α_i).

Ideally, the analyst would estimate willingness to pay using the joint distributions of data describing the observable characteristics of individuals and their choices at policy case j, $G_j(y, d)$ and $F_j(p, x, q)$, respectively. If these data are unavailable, however, or if constraints on time and resources prohibit original estimation, then a benefit transfer can be used to estimate wtp_{ij} using preexisting information from a different study case, k. We use a T superscript to distinguish a benefit transfer estimate for willingness to pay $\left(\widehat{wtp}_{ij}^T\right)$ from what would be obtained from information on policy case j if there were no constraints on time, resources, or data $\left(\widehat{wtp}_{ij}\right)$. The difference between these two measures defines the benefit-transfer error (BTE):

$$
\begin{aligned}
BTE \ = \ & \widehat{wtp}_{ij}^T\big[p_j, x_j, q_j^0, q_j^1, y_i, d_i, F_k(p, x, q), G_k(y, d)|\hat{\beta}_k, v\big] \\
& - \ \widehat{wtp}_{ij}\big[p_j, x_j, q_j^0, q_j^1, y_i, d_i, F_j(p, x, q), G_j(y, d)|\hat{\beta}_j, v\big],
\end{aligned}
\tag{2}
$$

where $\hat{\beta}_k$ is a vector of parameters estimated from the study case data and v denotes the valuation methodology (e.g. travel cost, contingent valuation).

The BTE is a measure of convergent validity that compares two estimates of the same theoretical value defined in equation (1), which is not an absolute criterion because the true value of wtp_{ij} is unknown. In an ideal setting where the policy case estimator is unbiased, such that $E\left[\widehat{wtp}_{ij}\right] = wtp_{ij}$, the BTE will provide an unbiased estimate of $\left(\widehat{wtp}_{ij}^T - wtp_{ij}\right)$. It is important to acknowledge that measurement error and other problems may cause the policy case estimator to be biased, limiting the policy relevance of the BTE statistic [32]. While this issue is important, it is beyond the scope of our study to resolve. Our research design follows the majority of the literature in separating the two problems. That is, we take $\widehat{wtp}_{ij}$ as given and focus on explaining the composite error in measuring $\widehat{wtp}_{ij}$ that is introduced by methodological features of the benefit transfer process and by differences between the study and policy cases.

The general expression for the benefit transfer error in (2) illustrates four distinct ways in which an analyst's modeling choices can influence the magnitude of the *BTE*.[3] First, the size of the *BTE* may depend on how the change in the environmental amenity, $\Delta q = q^1 - q^0$, is defined. Second, errors may stem from systematic differences between the observable characteristics of the study and policy case populations, $G_j \neq G_k$. Third, the *BTE* may vary with the valuation methodology, v. Finally, the error may depend on the transfer procedures embedded in $\widehat{wtp}_{ij}^T$. The individual modeling decisions reflected in the definitions for $\Delta q, v, G,$ and T may interact in ways that increase or decrease the *BTE*. Geographic similarity between the study and policy cases may reduce transfer errors more in value transfers than in function transfers, for example. That is, a function transfer can be calibrated to policy-case conditions by assigning levels to covariates in the transfer equation; whereas no similar calibration is possible for a value transfer so selection of the specific value to transfer that is "similar" becomes critically important.

Transfer errors are typically reported in percentage terms,

$$\text{\% Benefit Transfer Error} = \%BTE = \left[\left(\widehat{wtp}_{ij}^T / \widehat{wtp}_{ij} \right) - 1 \right] \times 100. \qquad (3)$$

Studies that do not report $\%BTE$ almost always report sufficient information for readers to make this calculation on their own.

Each study in the convergent validity literature reports one or more transfer errors conditional on a specific set of modeling decisions. We have systematically reviewed these studies and assembled a database documenting transfer errors and transfer characteristics. To extract the signals from the noise we use a meta-regression,

$$|\%BTE| = m(X) + \varepsilon, \qquad (4)$$

---

[3] A fifth potential source of error that is arguably out of the practitioner's control is any systematic difference between the latent preferences of the study and policy case populations.

where *X* includes a vector of variables that we use to describe analysts' modeling decisions (variables representing $\Delta q, v, G,$ and T).

3. Data Description

3.1. *Reviewing the Literature on Convergent Validity of Benefit Transfers*

Through an exhaustive search, we identified 40 convergent-validity studies that were published or posted online between 1990 and 2009. Thirty eight of these studies were published in peer-reviewed journals or book chapters. We ultimately excluded the two studies that were not peer reviewed because they had insufficient documentation for some of the key variables ($\Delta q, v, G,$ and T). Three peer reviewed studies were excluded for the same reason. A study by Morrison et al. [24] was excluded in order to avoid duplication of the findings reported in Morrison et al. [25]. Studies by Engel [9] and Chattopadhyay [8] were excluded because benefit transfer errors were reported as ranges rather than point estimates. In addition, a study by Eshet et al. [10] was excluded because it would have been the only one to use the hedonic methodology, and a single study would not have allowed us to identify the effect of this valuation method on the *%BTE*. These exclusions left us with a total of 31 studies with *%BTE* observations to analyze. Summary statistics describing the transfer applications and results for each included study are reported in Appendix Table I. Citations to all 40 studies are provided in the journal's online repository of supplemental material.

A uniform coding protocol was implemented to ensure that the modeling choices made by the authors of each study, and the corresponding %*BTEs*, were recorded correctly and consistently. The data were double coded by two research assistants, who then met with us to resolve discrepancies. Then we cross checked the coding a second time.

Some studies report what we refer to as "flip" error calculations. The investigators would

compute a transfer error with *j* as the policy case and *k* as the study case *a la* equation (3). Then they would flip the two cases, computing a second transfer error with *k* as the policy case and *j* as the study case. Flipping the study and policy cases changes the sign and absolute magnitude of the percentage error in (3). Unfortunately, it was not possible to infer the flip errors for most studies that did not make this calculation directly. Since we were unable to include flip errors for every study, we decided to use a single set of errors from studies that reported flips. Specifically, we used the first set of errors reported by the investigators.[4]

Each of the 31 validity studies reported multiple transfer errors, from as few as 2 to as many as 178. The error magnitudes vary with the transfer characteristics, selection of study cases and study case valuation methods, and amenity of interest. For example, Loomis et al. [20] tested the convergent validity of travel-cost estimates for the average consumer surplus associated with a single day of reservoir-based recreation in Sacramento, CA, Little Rock, AR, and Nashville, TN. After estimating travel-cost models for each of the three regions, they measured benefit transfer errors by making all possible pair-wise comparisons of estimated and transferred consumer surplus for 28 reservoir sites. This yielded 112 distinct transfer errors, half of which were flips. Excluding the flip errors left us with 56 observations on |%*BTE*| with a mean of 115%.

The final data set contains 1071 transfer errors. Of these observations, 55% describe applications in Europe, 37% are drawn from United States, and the remaining 8% are from Australia and the rest of the world.[5] The European data include observations from 12 Western Europe-

---

[4]Chattopadhyay [7] proposes an alternative formula for the %*BTE* that would avoid the flip error problem:
$|\% \ transfer \ error| = \left| \left[\left(\frac{study \ case \ value - policy \ case \ value}{study \ case \ vlaue + policy \ case \ value} \right) / 2 \right] \times 100 \right|.$
This formula produces the same measure of error, regardless of which value is defined as the study case. It would be convenient if future convergent-validity studies were to adopt this metric.

[5] We did not find evidence of systematic differences across these four regions in terms of the impact of modeling decisions on benefit transfer errors. Adding fixed effects for regions to the linear meta-regressions did not produce

an countries and the U.S. data include observations from all of the lower 48 states. The set of applications is also diverse. Eight studies considered the benefits of access to recreation sites (including forest recreation, reservoir based recreation, park recreation, and hunting); five evaluated prospective changes in the quality or quantity of water (including coastal areas, lakes, rivers, and groundwater); and three studies evaluated the benefits of reductions in sources of human health risk (air quality, water quality and ultra violet radiation). Other studies focused on opportunities for fresh water fishing (salmon, trout, big game, small game, flatfish, salmon, steelhead, walleye, pike, bass, and panfish), amenities associated with land preservation (farmland, forested land and coastal land), and the overall ecological health of watersheds, wetlands, and rivers.

3.2. The Distribution of Transfer Errors

Figure 1 illustrates the distribution of absolute benefit transfer errors. In percentage terms, the errors range from 0% to 7,496%, with a mean of 172%. However, the mean reflects a few observations with extremely large errors. The median is 39%, less than a quarter of the mean. The large difference between the mean and the median suggests a need to investigate outlying observations that could influence econometric inferences from the data.

We inspected the data for the presence of outliers using the inter-quartile range (IQR) criterion [34]. According to this criterion, an observation is classified as an outlier if $|\%BTE| < Q_1 - 1.5 \times IQR$ and/or $|\%BTE| > Q_3 + 1.5 \times IQR$, where $IQR = Q_3 - Q_1$, and Q_1 and Q_3 are the first and third quartiles of the $|\%BTE|$ distribution. This procedure detected 146 outliers.[6] Seventy-two percent of these were reported by just two studies, which also accounted for the

---

any statistically significant differences in coefficients on benefit transfer characteristics. Unfortunately, there was insufficient variation in the data to identify separate nonparametric models for each region.

[6] We also examined our models for leverage points and influential observations. There were some leverage points, but only a few were found to exert influence on OLS and WLS estimators. However, in general, it is difficult to identify influential and leverage data points when all of the explanatory variables are dichotomous. This observation, combined with the striking spread of the percentage transfer error, is what led us to adopt the IQR criterion.

largest percentage errors.[7]

Figure 2 graphs the distribution of transfer errors with outliers deleted. The mean and median |%*BTE*| are now reduced to 42% and 33%, respectively, and the maximum is reduced to 172% (equal to the mean when outliers were included). One explanation for these large reductions is that some of the validity studies were conducted for situations where the policy case and study case values were conveniently available. These comparisons may be comparisons of conveniently available value data and may not always be considered good candidates for benefit transfers, making it tempting to ignore them. On the other hand, including them in our analysis may help to identify combinations of procedures that tend to increase benefit transfer errors. To fully evaluate the sensitivity of our results to these large error observations we estimate the model with and without outliers.

3.3. Explanatory Variables

Table 1 defines the variables we use to explain the variation in benefit transfer errors, along with means and standard deviations for each variable. Of the 13 variables in the table, only |%*BTE*| is continuous. All 12 explanatory variables are binary indicators describing features of the benefit-transfer applications. They are grouped into the four categories that we defined earlier ($\Delta q, G, v,$ and T). Comparing the means and standard deviations in the last two columns of the table reveals that dropping outliers does not produce any substantial changes in the distribution of explanatory variables.

The first variable in the Δq category, QUALITYΔ, equals one if the transfer describes a change in *quality*, as opposed to a change in *quantity*. For example, changes in human health, river bank erosion, farming practices, air pollution, and water pollution are all defined as *quality*

---

[7] We considered dropping these two studies, but ultimately decided to keep them in the analysis because most of their observations are not defined as outliers by the IQR criterion (37% are outliers).

changes, whereas changes in fish catch rates, water supply, and access to recreation sites are all defined as *quantity* changes. The variable USEVALUE indicates whether Δq affects the use value of a resource as opposed to a non-use or total value.

The second category of explanatory variables, G, assesses the similarity of the study and policy cases. POPULATION equals 1 if the study and policy case populations are essentially the same. STUDYAREA equals 1 if both cases occur in the same geographic region.

The third set of explanatory variables describes the valuation methodology. Most of the transfer errors are drawn from studies that used choice modeling (CM-31%) or contingent valuation (CV-29%). The remaining observations are based on reduced-form meta-analyses (META-17%), travel cost models of site demand (TC-12%), and random utility models of site choice (RUM-11%).[8]

Finally, we use three variables to describe the nature of the transfer procedures. VALUETRANSFER indicates whether the procedure consisted of a value transfer or a function transfer. MULTIPLESTUDY equals 1 if the study case benefit measure is a composite of results from more than one study case; and MEAN equals 1 if the transfer error is reported as an average over two or more individual transfer errors.[9]

4. Nonparametric Meta-Analysis

4.1. A Nonparametric Model for Binary Regressors

---

[8] All of the meta-analysis studies use reduced-form regressions, as opposed to the "preference calibration" approach proposed by Smith, Van Houtven, and Pattanayak [35].

[9] A referee suggested that we consider controlling for variables such as sample size and the number of explanatory variables. The problem with these variables is that they are noisy measures. For example, sample size varies with valuation method (CV observations versus CM choice occasions) and sample frames may be different (e.g., population versus users). Similarly, the range of *%BTE* logically depend on the *relevancy* of the explanatory variables, which may or may not be correlated with the number of explanatory variables. Adding these variables to the model may introduce endogeneity problems that harm, rather than help, the performance of our estimators. Nevertheless, the concept of developing objective measures of the "data quality" of original benefit transfer studies is an interesting topic for future research.

Our nonparametric approach to meta-analysis avoids the need to restrict the functional relationship between benefit-transfer errors and variables describing features of the transfer process. [10] Specifically, we adapt the kernel estimator for models with unordered discrete regressors described in Ouyang, Li and Racine [28]. Our estimation procedure recognizes that the true form of $m(X)$ in equation (4) is unknown. Equation (4) is estimated using a variant of the Aitchison-Aitken kernel function, which has a simple form and can be easily generalized. To see the mechanics of this process, let λ_r denote a smoothing parameter (or bandwidth) associated with the rth component of X. The kernel function for r is defined as

$$l(X_{ir}, X_{hr}, \lambda_r) = \begin{cases} 1 & if \ X_{ir} = X_{hr} \\ \lambda_r & if \ X_{ir} \neq X_{hr} \end{cases}, \tag{5}$$

using $i = 1, 2, \ldots, N$ to index individual observations and X_{hr} to denote the value of the rth component of X for the data point that neighbors X_i. The corresponding product kernel can be expressed as

$$L(X_{ir}, X_h, \lambda) = \prod_{r=1}^{R} l(X_{ir}, X_{hr}, \lambda_r) = \prod_{r=1}^{R} \lambda_r^{I(X_{ir} \neq X_{hr})}, \tag{6}$$

where $I(\cdot)$ is an indicator function that equals 1 iff the condition in parentheses is true. Finally, the estimator for the unknown function $m(X)$ is

$$\widehat{m}(X) = \frac{\sum_{i=1}^{N} |\%BTE|_i \ L(X_{ir}, X_h \ \lambda)}{\sum_{i=1}^{N} L(X_{ir}, X_h, \lambda)}, \tag{7}$$

The key difference between the nonparametric categorical regression estimator of $m(X)$ and a standard parametric analysis is that the introduction of local weighting allows greater flexibility in uncovering variation in *%BTE*.

---

[10] Readers are directed to Boyle, Kaul and Parmeter [5] for a detailed discussion of the nonparametric methodology used here.

To define the effect of a binary variable, let $X_{\sim r}$ denote the subset of regressors in X excluding the rth variable such that $X = [X_{\sim r}, X_r]$. The response of $\hat{m}(X)$ to changing X_r from 0 to 1 can be written as:

$$\Delta_r = \hat{m}(0, X_{\sim r}) - \hat{m}(1, X_{\sim r}), \tag{8}$$

which can be estimated at every data point. Instead of having a single point estimate for each regressor as in a linear meta-regression, we have a vector of responses that are interpreted as the change in $|\%BTE|$ for a change in X, given the values of the remaining covariates. We refer to these as "response effects".

The rate of convergence for the estimator depends on the relevance of the explanatory variables and on the selection of bandwidths. In the language of nonparametric analysis, an explanatory variable is said to be "irrelevant" if $m(X)$ is constant with respect to that particular variable; i.e. if the response effects are zero everywhere. Hall, Li and Racine [11] formalize the distinction between relevant and irrelevant variables. To see their distinction, first partition the set of explanatory variables into two components, $X = [\ddot{X}, \dot{X}]$. The variables in $\dot{X}$ are irrelevant, if the dependent variable (Y) and $\ddot{X}$ are independent of $\dot{X}$. The bandwidth for an irrelevant binary variable approaches the upper bound of "1" [28]. Therefore, if we find that $\lambda_r \approx 1$, we conclude that r is "irrelevant".

We estimate bandwidths using the data-driven least-squares, cross-validation method (LSCV). LSCV employs the leave-one-out technique to find the optimal value of the smoothing parameters. Ouyang, Li and Racine [28] demonstrate that the rate of convergence to the optimal value of λ is of the order $O_p(n^{-1})$ when all regressors are relevant and $O_p(n^{-0.5})$ when some re-

gressors are irrelevant. [11] Importantly, these convergence rates are faster than nonparametric models with continuous regressors. Thus, the nonparametric model with binary regressors is less vulnerable to the curse of dimensionality. Moreover, if all regressors are discrete and relevant, the nonparametric model converges at the same rate as the correctly specified parametric model.

In the presence of irrelevant variables, the nonparametric model for binary regressors is still consistent, but it converges slower than the correctly specified parametric model. Of course, an econometrician never knows the correct specification for a parametric model. A misspecified parametric model may not converge at all [28].

4.2. The Performance of Nonparametric Models in Information Deficient Settings

It is important to consider the space of potential modeling decisions for benefit transfers. With 5 valuation methods and 7 binary regressors describing the transfer process, there are $2^7 \times 5 = 640$ "cells" in the data "grid" describing potential approaches to conducting a benefit transfer.[12] Our data contain 1,071 observations on transfer errors, but these observations are not uniformly distributed across the data grid. Some of the cells are empty. While nonparametric models with binary regressors are not subject to the same "curse of dimensionality" as models with continuous regressors, it may still seem counterintuitive that a nonparametric model would be able to estimate bandwidths with a reasonable degree of accuracy in this environment. Fortunately, there is a wealth of recent evidence that our estimator is, in fact, well suited to "infor-

---

[11] Ouyang, Li and Racine [27] make an *iid* assumption that is not guaranteed to hold in our data. However, Li, Ouyang and Racine [19] study the nonparametric kernel regression estimator and optimal bandwidth selection in the presence of weakly dependent data. Their results are similar to the iid results from Ouyang, Li, and Racine. While Li, Ouyang and Racine do not explicitly consider the case where irrelevant variables are included in the regression model, they note, "… we fully expect that the results in Hall et al.'s [9] analysis carry through to weakly dependent data settings. We present simulation results that support this contention …" (page 699).

[12] The valuation methodologies can be modeled as separate binary variables or as a categorical variable with META as the base category. Both these techniques generate qualitatively similar results. In this paper, we present the results with methodological variables as a category recognizing that we lose the ability to identify separate bandwidths for methodological variables.

mation deficient" environments [14][28][29][33]. We summarize technical details in the appendix.

Of course, the nonparametric model does not allow us to calculate response effects for empty cells. Greater numbers of empty cells in the support of a particular variable make it harder to determine the full range of impacts of that variable on the benefit transfer error. Adding parametric restrictions to the meta-regression reduces the number of cells that have to be filled in order to identify a variable's impact, while simultaneously increasing the scope for functional form misspecification to bias the estimator. With this tradeoff in mind, we proceed in two stages. First we estimate the meta-regression nonparametrically. The results allow us to characterize ranges of response effects for most variables. Then we estimate a conventional linear meta-regression using weighted least squares, recognizing that the credibility of inferences based on our results is decreasing in the strength of the functional form assumptions we maintain [22].

5. **Results**

5.1. Nonparametric Model

LSCV bandwidths were estimated separately for the full data set (N=1071) and for the data set excluding outliers (N=925). The two sets of bandwidths differed substantially. Therefore, we adapted Hartarska et al.'s [12] procedure for dealing with outliers in nonparametric models. Following their logic, we mitigate the influence of outliers by first estimating the optimal bandwidths without outliers and then using the estimated bandwidths to estimate $m(X)$ on the full data set (including outliers). Table 2 reports the resulting LSCV bandwidths, along with summary statistics for the distribution of response effects (mean, median, 25th and 75th quar-

tile).[13]   Using the wild bootstrap procedure we estimate cluster robust standard errors to correct for potential heteroskedasticity and correlation of the errors within each original study.[14]

The estimated quartiles of the response effect distributions reveal five clear results. The first result is one of the stylized facts in the literature; value transfers tend to generate larger transfer errors than function transfers. The second result follows logically from what is known about the complexity of research design; transfer errors tend to be larger for studies that consider changes in environmental quality, rather than the quantity, of a particular amenity. Third, contingent-valuation applications generate lower transfer errors than choice modeling. Fourth, travel cost models of site demand are associated with lower transfer errors than random utility models of site choice. Lastly, contingent valuation generates lower transfer errors than meta-analysis, which in turn performs better than choice modeling, travel cost and random utility models. All of these results are large in magnitude, statistically significant, and remarkably robust; the interquartile ranges and means all suggest consistent directions for the response effects. (We will interpret these results in Section 5.3, after summarizing results from the WLS model.)

Other nonparametric response effects are less robust. Convergent validity studies where the study and policy cases describe the same geographic area are associated with smaller transfer errors sometimes, but not always. The 25th and 50th quartiles of the response effect distribution

---

[13] We perform a kernel based nonparametric test of significance. This test is analogous to a t-test in linear regression [30]. Most of the variables are highly significant except for MEAN that is significant at 10% and USEVALUE, which is insignificant.

[14] Cameron, Gelbach, and Miller [7] demonstrate that these standard errors perform well in making robust inferences in the presence of clustered observations. After estimating equation (7), a wild bootstrap sample is generated as $\%|BTE|_s^* = \hat{m}(X)_s + \hat{u}_s \epsilon^*, s = 1, ... ,31$, where s is the study id and ϵ^* is a white noise term and $\hat{m}(X)$ and $\hat{u}$ are the nonparametric fitted value and residual terms, respectively. The white noise term is defined such that $E(\epsilon^*) = 0$ and $E(\epsilon^{*2}) = 1$. Values are randomly selected using a two point distribution given by: $\epsilon^* = \frac{1-\sqrt{5}}{2}$ with probability $p = \frac{1+\sqrt{5}}{2\sqrt{5}}$ and $\epsilon^* = \frac{(1+\sqrt{5})}{2}$ with probability $(1-p)$. The bootstrap sample is then used to estimate new response effects. This procedure is repeated 999 times. Thus, for each estimated response effect of the relevant variables, we will have 999 values. The standard error for an estimated response effect is calculated by taking the standard deviation of the corresponding 999 bootstrapped response effects.

are negative, but the 75th quartile is positive. Likewise, in situations where the study case value was defined using data from multiple studies, the negative mean response effect is clearly driven by the observations in the left tail of the distribution of response effects. Finally, the data do not allow us to recover precise estimates for the response effect distributions associated with POPU-LATION, MEAN, and USEVALUE. The estimated bandwidth for the first variable is 1, and the cluster robust standard errors for the other two variables suggest that their response effect distributions cannot be estimated precisely (MEAN that is significant at 10% and USEVALUE is insignificant). Therefore, in the hope of recovering reasonable approximations to the means of their response effect distributions, we repeat the estimation after adding the conventional linearity and separability restrictions to the meta-regression.

5.2. Parametric Model: Weighted Least Squares

Recall that the 31 convergent validity studies vary considerably in the number of values they report for the %*BTE* (from 2 to 178). Since ordinary least squares estimation assigns equal weight to each observation, studies that provide more observations have greater influence on the results from linear estimation. To mitigate this influence, we estimate equation (4) using weighted least squares (WLS) with cluster robust standard errors. Each observation is weighted by the total number of observations contributed by the corresponding study. Thus, individual observations from studies that provide more observations receive less weight in the estimation. The last two columns of Table 2 report the WLS results, with and without outliers. Not surprisingly, dropping outliers decreases the absolute magnitudes of point estimates for the regression coefficients and improves model fit.[15] The modeling decisions we observe explain three quarters of the variation in the percentage transfer error when outliers are removed.

---

[15] We also estimated the model using OLS. The results were much more sensitive to outliers. This is partly due to the fact that the study with the largest error (7496%) also had the largest number of observations.

17

Imposing the linearity and separability restrictions on the model produces substantial changes in the estimated mean response effects. Nevertheless, with or without outliers, the WLS point estimates are still consistent with most of the qualitative findings from the nonparametric regression. The most striking changes are in the coefficients for the variables describing study case valuation methods. The WLS model suggests that RUM, travel cost, contingent valuation, and choice modeling all tend to produce smaller transfer errors than meta-analysis (the excluded category). While our nonparametric model is consistent with this conclusion for contingent valuation, it provides conflicting evidence that travel cost and meta-analysis produce smaller transfer errors than RUM and choice modeling. A consistent kernel test [15] rejects the null hypothesis that the WLS model is correctly specified against the nonparametric alternative. The differences in sign and magnitude for the mean response effects for RUM, TC, and CM reveal the empirical importance of the misspecification bias in the functional form of the linear model. Moreover, the width of the interquartile range of nonparametric response effects suggests that the linear model overlooks important features of the data.

5.3. Interpretation of the Results

We use a novel approach to graphical analysis—45 degree plots of whisker figures—to help us interpret our findings. The idea for the plots is simple. Both axes of a 2–dimensional diagram are used to represent the same range of response effects for an independent variable [13]. This makes it possible to visualize the entire distribution of response effects (and their confidence intervals) along the 45-degree line. To see this, consider Figure 3. The solid square represents our WLS estimate for the impact of value transfers relative to function transfers, using the full data set. Its level (31%) can be seen from the horizontal axis. Its whiskers define a 95% confidence interval around our point estimate, measured on the vertical axis. The solid circle

18

and its whiskers summarize our WLS estimate when we exclude outliers. Each remaining verti-

cal line represents a cluster of nonparametric response effects. The number at the center of each

line reports the share (percentage) of effects clustered at that point (e.g. 49% of the effects are

clustered at a point near a 200% error). To understand why clustering occurs, recall that all re-

gressors are binary, making $\hat{m}(X)$ discontinuous. As a result, the response effects tend to be

clustered at values that correspond to specific combinations of explanatory variables. When

multiple clusters are too dense for their individual shares to be legible, brackets are used to indi-

cate the cumulative share for the group. For example, 26% of response effects are located in a

series of clusters just above zero. Finally, the text box indicates that 88% of the response effects

are positive.

Visual inspection of the response effect distribution can reveal important features of the

data that cannot be seen from the mean response effects and their quartiles in table 2. For exam-

ple, seeing the right tail of the distribution may be especially important if a benefit transfer prac-

titioner wants to avoid benefit-transfer practices that create the possibility of extreme errors.

Seeing the entire distribution also enables the analyst to assess the robustness of WLS mean ef-

fects.

5.3.1. Function Transfer versus Value Transfer

Figure 3 confirms the stylized fact that function transfers outperform value transfers [6]

[32] [31]. Eighty-eight percent of the response effects are in the positive quadrant and so are the

point estimates from both WLS models. While moving from WLS to the nonparametric model

does not change our conclusion about the sign of the mean response effect, it does suggest two

additional findings. First, the superior performance of function transfers is extremely robust.

Second, the distribution of response effects has a fat right tail. Moving from function transfer to

value transfer often increases the benefit transfer error by as much as 200%. This effect is several times as large as the mean response effects suggested by the WLS models. The superior performance of function transfers is likely due, in part, to the presence of spatial and temporal variation in amenities and variation in household income and preferences (Boyle et al. [3]). Function transfers adjust for features of this heterogeneity, whereas value transfers do not.

5.3.2. Quantity Changes versus Quality Changes

An equally robust result is that benefit transfers are almost always generate lower transfer errors when they describe quantity changes, rather than quality changes. Figure 4 illustrates that 96% of the nonparametric response effects are positive, along with both estimates from the WLS model. Nearly two thirds of the response effects exceed 100%. These findings make sense. Quantity changes are usually easier to describe than quality changes. Anglers can easily understand a change in catch rates or a permanent closure of a fishing site, for example. In contrast, it may be difficult for them to assess a change in water quality, especially when the change is not visible and the effect on fishing is not explicit. Furthermore, consumers may perceive quality differently at the study and policy cases, increasing the transfer error. If the benefit-transfer practitioner is unable to control whether their assessment is framed as a quantity or quality change, the results in Figure 4 suggest that extra caution and additional sensitivity analyses are warranted if the transfer involves a change in environmental quality.

5.3.3. Geographic Similarity

The nonparametric model reveals important features of the response effect distribution for STUDYAREA that are obscured by the linearity and separability assumptions of a conventional meta-analysis. To see this, first notice that our WLS point estimates are both negative, but close to zero (-7% to -14%) and statistically insignificant. In contrast, the mean response effect

20

in the nonparametric model (-52%) is larger and statistically different from zero. It indicates that geographic similarity between the study and policy cases substantially reduces transfer errors, consistent with the stylized facts in the literature [4][17][31].

However, the mean nonparametric response effect summarizes a wide distribution. While more than half of the response effects are negative, 36% are clustered at a large positive value (nearly 200%) with a very small confidence interval. This counterintuitive cluster of extreme positive values reflects two attributes of the nonparametric analysis. First, the response effects in Figure 5 are only calculated for the observations with STUDYAREA=1.[16] This is a small share of the data (26%). Second, of this 26%, a large cluster of observations come from a single study that happens to have an extreme response effect, which may suggest a poor comparison application was chosen for the validity investigation.

The response effects for STUDYAREA can be divided into two groups – applications with value transfers or function transfers. Within each group, the average response effect is negative but the direction of the effect is far more robust for value transfers. Thus, our results reinforce the general importance of geographic similarity between study and policy cases when value transfers are conducted. These results also suggest that function transfers facilitate calibration, based on explanatory variables that might overcome differences in the policy and study cases.

5.3.4   Combining Data from Multiple Study Cases

Intuition would suggest that combining data from multiple study cases would reduce transfer errors as long as the selected study cases are appropriate for the transfer. Moreover,

---

[16] Similar to dummy variables in a linear model, calculation of response effects requires there to be a base group with a value of 0 for each variable. We defined the base group to be the subset of observations that did not make the modeling decision represented by a '1' for the corresponding variable. If we were to switch the base group, we could calculate response effects for the remaining observations, allowing us to "fill in" the response effect distribution. However, the qualitative results would be unchanged.

Smith, Van Houtven, and Pattanayak's [35] logic for "preference calibration" suggests there are gains from using multiple study cases to span the relevant portion of the policy case benefit function. Indeed, the WLS models suggest a modest reduction in *%BTE* and the mean nonparametric response effect for MULTIPLESTUDY in Table 2 appears to confirm this finding.

Figure 6 reveals some interesting heterogeneity underlying the mean response effect. Just over three quarters of the effects are positive, but they are all less than 1% with tight confidence intervals. The negative response effects are at least an order of magnitude larger; the mean is -28% and the maximum is -9%. Focusing on the negative response effects, 86% correspond to function transfers and 73% correspond to studies that evaluated quantity changes. In contrast, almost all of the value transfers have positive (but near zero) response effects. These findings indicate that combining values from several studies has greater scope to improve the accuracy of function transfers that can adjust for differences between the sites.

5.3.5 *Valuation Methodology*

The nonparametric response effects for study case value methods are the most difficult to interpret. Designs with more choice alternatives (e.g., choices in CM and sites in RUM) appear to be associated with larger errors. In the context of stated preference research, contingent valuation is associated with smaller errors than choice modeling. On the revealed preference side, travel cost models of site demand tend to have smaller errors than random utility models of site choice. The mean response effects range from -70% for CV to 349% for CM. Meta-analysis lies in between these extremes, which is not surprising since the meta-analyses in our data are mostly based on study case value estimates derived from two or more of these four valuation methods.

One interpretation of these results is that increasing the number of choice alternatives leads to more complex econometric models, increasing the scope for modeling decisions to in-

troduce bias and increase the transfer error.[17] However, we must be extremely cautious with this interpretation. It is important to keep in mind that, on average, CV studies generate the fewest transfer errors per study.[18] In comparison, CM and META studies often test convergent validity by comparing each transfer estimate with each of the remaining 'N-1' alternatives. The "extra" transfer errors generated by CM, RUM, META and TC based studies may be econometric targets of opportunity that would not be considered best practices in a real world policy evaluation.

More importantly, our measure for the benefit transfer error does not say anything about the relative performance of these methods in estimating the actual wtp at the policy and study cases using case-specific data. At most our results only illustrate that CV has performed relatively better in benefit transfer applications than other methods. Following this result, an important area for future research would be to compare the performance of study case valuation methodologies in the context of benefit transfer. This can be achieved by conducting a benefit transfer exercise for the same set of study and the policy cases using more than one valuation method.

6   Conclusions

What are the practical implications from the past 20 years of research on the convergent validity of benefit transfers? Out of 1071 transfer errors reported by 31 studies, the median absolute error is 39% for all data and 33% with the outliers removed. Individual transfer errors range from close to 0% to larger than 200%. Both the median transfer error and its range are similar in magnitude to certain types of errors associated with original studies. For example, Murphy et al.

---

[17] As Boyle et al. [3] demonstrate, specification of a utility function for an econometrically consistent RUM or CM transfer requires that four "S" assumptions are satisfied. The study site model must be correctly *specified*; unobserved attributes of the study and policy sites must be *separable* from observed attributes in the utility function; consumers must not be *sorted* across the study and policy sites according to unobserved attributes of their preferences; and the data on demographic attributes of policy site consumers must not present any form of *selection* bias.

[18] A total of twelve contingent valuation studies generate 29% of the benefit transfer errors, therefore, the average transfer errors per study is about 2.4%. The average transfer errors per study exceed 3% for the rest of the study case valuation methods.

23

[26] assess the importance of "hypothetical bias" in stated preference research by analyzing results from 28 laboratory and field experiments on the difference between hypothetical statements of value and actual statements of value based on binding commitments. The median hypothetical value overstated actual value by 35%, with the majority of estimates for hypothetical bias falling between 0% and 200%.

Our results demonstrate characteristics of benefit transfers that are associated with greater or lesser errors. These insights can help to guide future analyses on where caution is important in the conduct of a benefit transfer and robustness checks are warranted. The evidence overwhelmingly supports the stylized fact that function transfers outperform value transfers. The literature also suggests that benefit transfers are better able to predict the willingness to pay for quantity changes than for changes in environmental quality. The former is a choice made by the benefit-transfer practitioner, while the latter is defined by the policy question being addressed. That said, if the practitioner must perform a value transfer then it becomes especially important to ensure that the study-case value estimate matches the policy-case value definition. We also find that using information from multiple and geographically similar study cases can reduce benefit transfer errors. Site similarity tends to be more important for value transfers, but function transfers appear better equipped to exploit information from multiple studies.

It is important to distinguish between the way that meta-analysis is used for benefit transfers and the way that we have used the methodology in this study. We have developed a new nonparametric approach to meta-analysis and demonstrated that it can extract important signals from the data that remain hidden in conventional linear models. This advancement continues a long tradition of refining the econometrics of meta-analysis in order to distill key findings from research on important questions in environmental economics [1][37][38][41]. In contrast, when

meta-analysis is used to conduct a benefit transfer the methodology must provide policy-relevant transfer estimates. Parametric estimation is desirable for providing point estimates of meta-equations parameters, but the robustness of these parameter estimates should be investigated using nonparametric estimation. Our results on the performance of methodological variables need to be carefully interpreted. For benefit transfer, we find that contingent valuation and travel cost models perform better than choice modeling and random utility models, respectively. However, these results do not suggest anything about the relative performance of these methods in estimating the actual wtp at the policy and study cases using case-specific data. An important area for future investigation would be to test the performance of benefit transfer applications for specific valuation methods (e.g., CV, RUM, META, etc.) holding the study and policy case applications constant. Our results on META indicate that the performance of meta-analysis based on specific valuation methods (e.g. CV and TC), in the context of benefit transfer, should be investigated in the future.

Finally, we have some recommendations for future studies evaluating the validity of benefit transfers:

- In our review of the literature, we observed several cases where transfer characteristics were not clearly documented. Future benefit-transfer validity analyses must be more thorough in their documentation of data and analysis procedures.

- Wide ranges of transfer errors were sometimes presented, but not explained. A documentation protocol is needed in order for future convergent-validity studies to enhance the credibility of benefit transfers. Each study should define the criteria they use to identify study and policy cases that are good candidates for benefit transfers, and then justify each comparison in the context of those criteria.

- A protocol for computing errors should be used, such as presented by Chattopadhyay [8], so that benefit transfer errors are invariant to the specification of study and policy cases.

- Investigators need to go beyond simply reporting transfer errors to explore why some comparisons have small errors and others have large errors.

Taking these steps would provide insights to refine the criteria for when appropriate data are available to conduct benefit transfers and enhance the credibility of benefit transfers.

References

[1] H.S. Banzaf, V.K. Smith, Meta-analysis in model implementation: choice sets and the valuation of air quality improvements, Journal of Applied Econometrics 22 (6) (2007) 1013-1031.

[2] J.C. Bergstrom, L.O. Taylor, Using meta-analysis for benefits transfer: theory and practice, Ecological Economics 60(2) (2006) 351-360.

[3] K.J. Boyle, N.V. Kuminoff, C.F. Parmeter, J.C. Pope, Necessary Conditions for Valid Benefit Transfers, American Journal of Agricultural Economics 91 (2009) 1328-1334.

[4] K.J. Boyle, N.V. Kuminoff, C.F. Parmeter, J.C. Pope, The benefit-transfer challenges, Annual Review of Resource Economics 2 (1) (2010) 161-182.

[5] K. J. Boyle, S. Kaul, C. F. Parmeter, Meta-analysis − advances and new perspectives toward data synthesis and estimation robustness; in R. Johnston, J. Rolfe, R. Rosenberger, R. Brouwer (eds.), Benefit Transfer of Environmental and Resource Values: A Handbook for Researchers and Practitioners, forthcoming, 2013.

[6] R. Brouwer, F. A. Spaninks, The validity of environmental benefits transfer: further empirical testing, Environmental and Resource Economics 14 (1) (1999) (95-11).

[7] A. C. Cameron, J. B. Gelbach, D. L. Miller, Bootstrap-based improvements for inference with clustered errors, The Review of Economics and Statistics, 90(3) (2008) 414–427.

[8] S. Chattopadhyay, A repeated sampling technique in assessing the validity of benefit transfer in valuing non-market goods, Land Economics 79 (4) (2003) 576-596.

[9] S. Engel, Benefit function transfers versus meta-analysis as policy making tools: a comparison, in: R. J. G. M. Florax, Peter Nijkamp, Kenneth George Willis (eds.), Comparative Environmental Economic Assessment, Edward Elgar, Cheltenham, UK, 2002.

[10] T. Eshet, M.G. Baron, M. Shechter, Exploring benefit transfer: disamenities of waste transfer stations, Environmental and Resources Economics 37(3) (2007) 521-547.

[11] P. Hall, Q. Li, J. Racine, Nonparametric estimation of regression functions in the presence of irrelevant regressors, The Review of Economics and Statistics 89(4) (2007) 784–789.

[12] V. Hartarska, C. F. Parmeter, D. Nadolnyak, B. Zhu, Economies of scope for microfinance: differences across output measures, Pacific Economic Review 15 (4) (2010) 464–481.

[13] D. J. Henderson, S. C. Kumbhakar, C. F. Parmeter, A simple method to visualize results in nonlinear regression models, Economic Letters 117 (3) (2012) 578 – 581.

[14] D. J. Henderson, C. Papageorgiou, C. F. Parmeter, Growth empirics without parameters, The Economic Journal 122 (559) (2012) 125-154.

[15] C. Hsiao, Q. Li, J. Racine, A consistent model specification test with mixed discrete and continuous data, Journal of Econometrics 140(2) (2007) 802–826.

[16] R.J. Johnston, Choice experiments, site similarity and benefits transfer, Environmental and Resources Economics 38(3) (2007) 331-351.

[17] R.J. Johnston, R.S. Rosenberger, Methods, trends and controversies in contemporary benefit transfer, Journal of Economic Surveys 24 (3) (2010) 479-510.

[18] S. Kirchhoff, B.G. Colby, J.T. LaFrance, Evaluating the performance of benefit transfer: an empirical inquiry, Journal of Environmental Economics and Management 33 (1) (1997) 75-93.

[19] Q. Li, D. Ouyang, J. S. Racine, Categorical semi parametric varying-coefficient models, Journal of Applied Econometrics, *Forthcoming*.

[20] J.B. Loomis, B. Roach, F. Ward, R. Ready, Testing transferability of recreation demand models across regions: a study of corps of engineer reservoirs, Water Resources Research

31 (3) (1995) 721–730.

[21] J.B. Loomis, R. S. Rosenberger, Reducing barriers in future benefit transfers: needed improvements in primary study design and reporting, Ecological Economics 60 372-378.

[22] C.F. Manski, Identification for Prediction and Decision, Harvard University Press, Cambridge, 2007.

[23] K. Moeltner, R. S. Rosenberger, Meta-regression and benefit transfer: data space, model space, and the quest for 'Optimal Scope', B. E. Journal of Economic Analysis and Policy, 8(1) (2008).

[24] M. Morrison, J. Bennett, R. Blamey, J. Louivere, Choice modeling, non-use values and benefit transfer, Economic Analysis and Policy 30 (1) (2000) 13-32.

[25] M. Morrison, J. Bennett, R. Blamey, J. Louviere, Choice modeling and tests of benefit transfer, American Journal of Agricultural Economics 84 (1) (2002) 161-170.

[26] J. J. Murphy, P. G. Allen, T. H. Stevens, D. Weatherhead, A meta-analysis of hypothetical bias in stated preference valuation, Environmental and Resource Economics, 30(3) (2005) 313–325.

[27] J.P. Nelson, P.E. Kennedy, The use (and abuse) of meta-analysis in environmental and natural resource economics: an assessment, Environmental and Resource Economics 42(3) (2009) 345-377.

[28] D. Ouyang, Qi Li, J.S. Racine, Nonparametric estimation of regression functions with discrete regressors, Econometric Theory 25(1) (2009) 1-42.

[29] C. F. Parmeter, Z. Zheng, P. McCann, Cross-validated bandwidths and significance testing; in Q. Li and J. S. Racine (eds.), Advances in Econometrics: Nonparametric Methods, Elsevier Science, 2009.

[30] J. S. Racine, J. Hart, Q. Li, Testing the significance of categorical predictor variables in nonparametric regression models, Econometric Reviews 25 (2006) 523-544.

[31] R. S. Rosenberger, T. Phipps, Correspondence and convergence in benefit transfer accuracy: meta-analytic review of the literature; in S. Navrud and R. Ready (eds.), Environmental Value Transfer: Issues and Methods, 2007.

[32] R. S. Rosenberger, T. D. Stanley, Measurement, generalization, and publication: Sources of error in benefit transfers and their management, Ecological Economics 60 (2) (2006) 372–378.

[33] O. Y. Savchuk, J. D. Hart, S. J. Sheather, Indirect cross-validation for density estimation, Journal of the American Statistical Association, 105 (489) (2010) 415-424.

[34] N. C. Schwertman, M. A. Owens, R. Adnan, A simple more general boxplot method for identifying outliers, Computational Statistics & Data Analysis 47 (1) (2004) 165–174.

[35] V.K. Smith, G. Van Houtven, S. K. Pattanayak, Benefit transfer via preference calibration: 'Prudential algebra' for policy, Land Economics 78 (1) (2002)132-152.

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

[37] V.K. Smith, Y. Kaoru, Signals or noise? Explaining the variation in recreation benefit estimates, American Journal of Agricultural Economics 72(2) (1990) 419-433.

[38] V.K. Smith and L. Osborne, Do Contingent Valuation Estimates Pass a Scope Test? A Meta Analysis. Journal of Environmental Economics and Management 31 (1996): 287-301.

[39] U.S. Environmental Protection Agency, Guidelines for Preparing Economic Analyses, EPA 240-R_10-001 (prepublication edition), 2010. http://yosemite.epa.gov/ee/epa/eed.nsf/pages/Guidelines.html/$file/Guidelines.pdf, ac-

cessed December 31, 2010.

[40] U.S. Executive Office of the President, Executive Order 12866: Regulatory Planning and Review, Federal Register 58(190): 51735- 51744, 1993.

[41] R. G. Walsh, D. M. Johnson, J. R. McKean, Issues in nonmarket valuation and policy application: a retrospective glance, Western Journal of Agricultural Economics 14(1) (1989) 178-188.

<div align="center">

Table 1. Variables, Definitions and Summary Statistics

</div>

| Variables	Definitions	All Data	Without Outliers
		Mean (Standard Deviation)	
|%Benefit Transfer Error|	$\left\|\left[\left(\dfrac{\widehat{wtp}_{ij}^T}{\widehat{wtp}_{ij}}\right) - 1\right] \times 100\right\|$	171.97 (555.19)	42.39 (36.39)
Amenity (Δq)			
QUALITYΔ	1 if a change in quality; 0 if change in quantity	0.47 (0.50)	0.45 (0.50)
USEVALUE	1 if use value; 0 if non-use value	0.66 (0.47)	0.69 (0.46)
Study and Policy Cases (G)			
POPULATION	1 if study and policy case populations are the same; 0 otherwise	0.09 (0.29)	0.11 (0.31)
STUDYAREA	1 if study and policy cases geographic area are the same; 0 otherwise	0.26 (0.44)	0.24 (0.43)
Valuation Methodology (v) *(META is the base/omitted category in estimation)*			
META	1 if the valuation method is a meta-analysis	0.17 (0.38)	0.19 (0.39)
RUM	1 if the valuation method is a random-utility model	0.11 (0.32)	0.09 (0.29)
TC	1 if the valuation method is a travel-cost model	0.12 (0.32)	0.12 (0.32)
CV	1 if the valuation method is contingent valuation	0.29 (0.45)	0.33 (0.47)
CM	1 if the valuation method is choice modeling	0.31 (0.46)	0.27 (0.44)
Transfer Procedures (T)			
VALUETRANSFER	1 if value transfer; 0 if function transfer	0.38 (0.48)	0.36 (0.48)
MULTIPLESTUDY	1 if two or more study cases are used to estimate study-case value; 0 otherwise	0.27 (0.45)	0.31 (0.46)
MEAN	1 if transfer error is reported as a mean of two or more transfer errors; 0 otherwise	0.15 (0.36)	0.17 (0.37)
N	Number of observations	1071	925

Table 2. Nonparametric and Parametric Meta-Regression Results

	BW	**Nonparametric Regression[a]**				**WLS[b]**	
		Response Effect Quartiles				**All Data**	**Without Outliers**
		Mean	**25%**	**50%**	**75%**		
QUALITYΔ	0.00	263.36 (135.50)	27.04 (0.54)	172.72 (57.73)	590.85 (69.32)	49.84[*] (26.27)	24.85[*] (13.09)
USEVALUE	0.15	-3.43 (1.48)	-8.44 (6.47)	-8.44 (6.47)	-2.13 (7.68)	26.33 (18.28)	4.25 (9.76)
POPULATION	1	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	37.79 (38.56)	19.82 (18.20)
STUDYAREA	0.00	-51.86 (17.09)	-247.29 (115.43)	-70.33 (38.72)	194.22 (0.19)	-14.17 (15.22)	-7.10 (10.85)
RUM	0.00	172.36 (0.20)	194.23 (0.20)	194.23 (0.20)	194.23 (0.20)	-132.97[***] (48.75)	-60.34[***] (23.28)
TC	0.00	18.82 (0.19)	6.13 (0.46)	27.00 (43.07)	27.00 (43.07)	-130.62[**] (51.21)	-68.48[***] (24.60)
CV	0.00	-70.17 (17.08)	-156.34 (2 6.26)	-37.70 (18.10)	-37.70 (18.10)	-150.10[***] (36.95)	-65.94[***] (16.56)
CM	0.00	349.71 (135.50)	172.72 (57.73)	329.62 (135.50)	503.10 (69.97)	-91.60[**] (36.30)	-29.89[*] (16.28)
VALUETRANS-FER	0.41	121.79 (91.37)	1.17 (1.96)	156.91 (91.37)	202.61 (87.16)	31.22[*] (17.44)	4.79 (4.96)
MULTIPLES-TUDY	0.02	-8.80 (0.64)	0.19 (2.81)	0.19 (2.81)	0.19 (2.81)	-23.00[**] (9.76)	-9.66 (6.62)
MEAN	0.00	-24.61 (17.09)	-15.14 (18.10)	-15.14 (18.10)	-15.14 (18.10)	-21.44 (15.23)	-9.24 (10.67)
Intercept						126.42[***] (38.88)	77.02[***] (18.16)
$\bar{R}^2$						0.33	0.76
F - value						47.95	269.72
N	925	1071	1071	1071	1071	1071	925

[a] The variable META is excluded from this table as it is the base category for valuation methods. Bandwidths are estimated with outliers excluded. Robust cluster wild bootstrapped standard errors are given in parentheses. Bandwidths for RUM, CV, TC, and CM are identical as they included as one categorical variable with META as the base category. The fact that the 25th, 50th and 75th quartiles in the distribution of response effects are identical for some explanatory variables reflects the discreteness in variable space. The bandwidths for variables QUALITYΔ, STUDYAREA, RUM, CV, TC, CM, and MEAN suggest that their response effects are generated from a frequency type estimation. The intercept for nonparametric model can be easily computed, however, they are not useful for our analysis and are not reported them in this table.
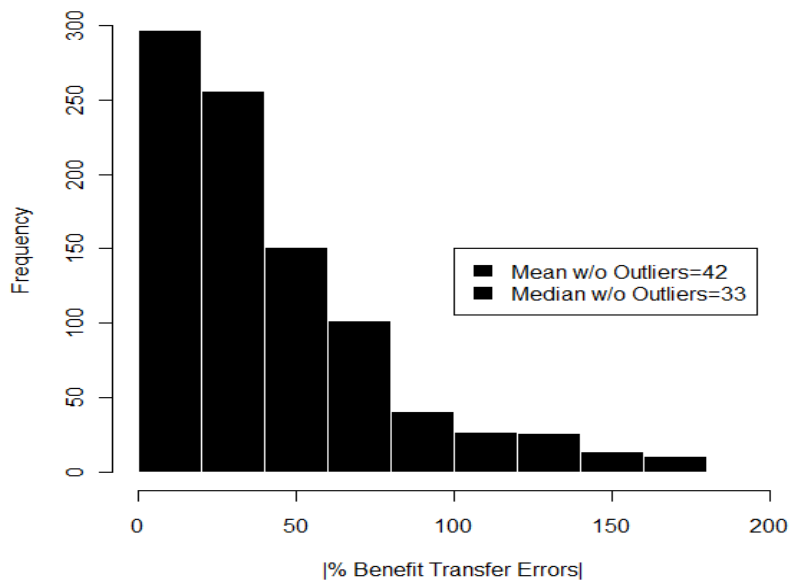
[b] Significance codes for WLS models: '***' 0.01, '**' 0.05, '*' 0.1.

# Figure 1. Distribution of Percentage Transfer Errors (N=1071)[a]
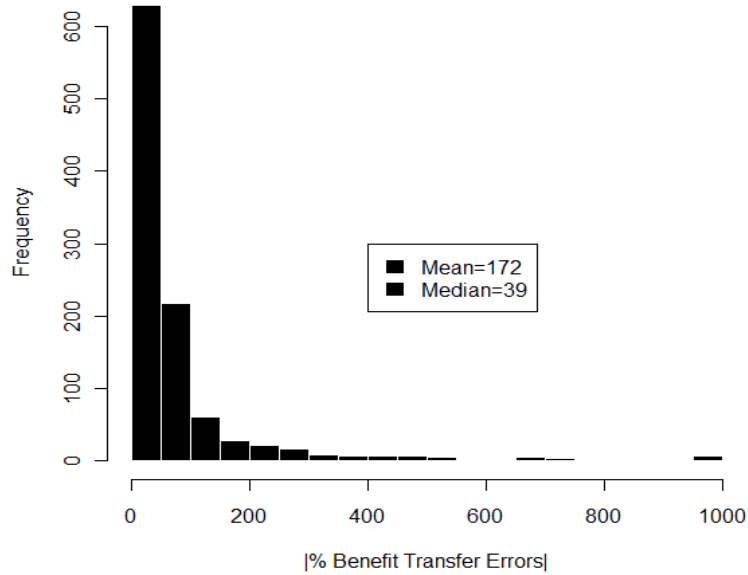
[a] The scaling of the horizontal axis excludes 3.83% of observations with errors exceeding 1000%.

# Figure 2. Truncated Distribution of Percentage Transfer Errors, Excluding Outliers (N= 925)[a]

[a] This histogram is drawn for |%BTE| without outliers. Of 1071 observations 13.63% are outliers according to the interquartile range criterion.

Figure 3. Benefit Transfer Error Response Effects for VALUETRANSFER[a]
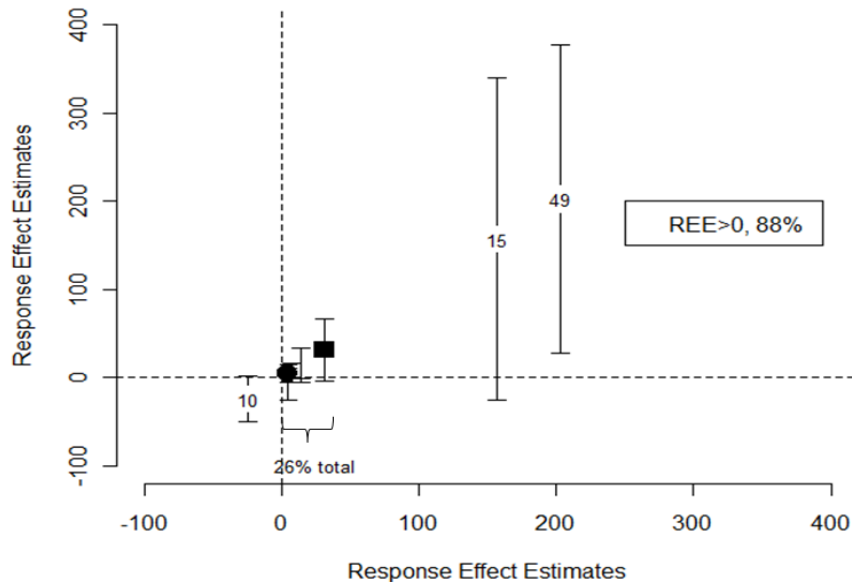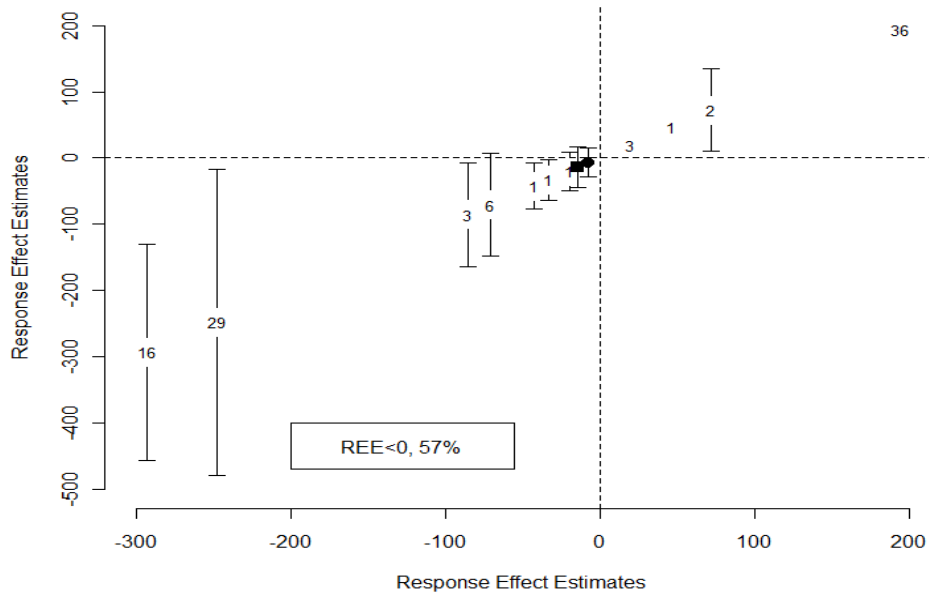

[a] The figure plots WLS point estimates for the VALUETRANSFER variable, nonparametric response effect esti-mates (REE), and 95% confidence intervals for each. The square and circle represent the WLS point estimates based on the data with and without outliers, respectively. The horizontal bars above and below denote 95% confidence intervals. The numbers indicate the share of the response effects at the point where the number is located. Cluster-ing of REEs occurs because all of the independent variables are binary.

Figure 4. Benefit Transfer Error Response Effects for QUALITYΔ [a]
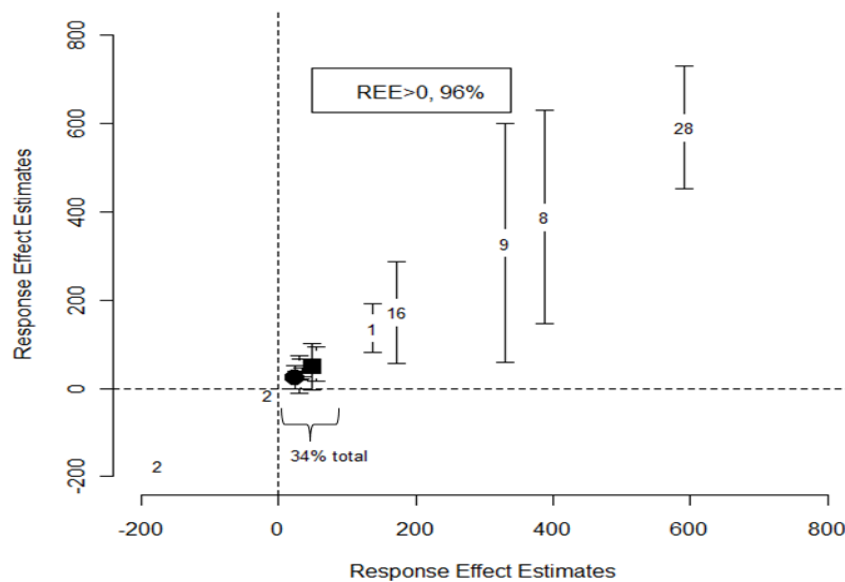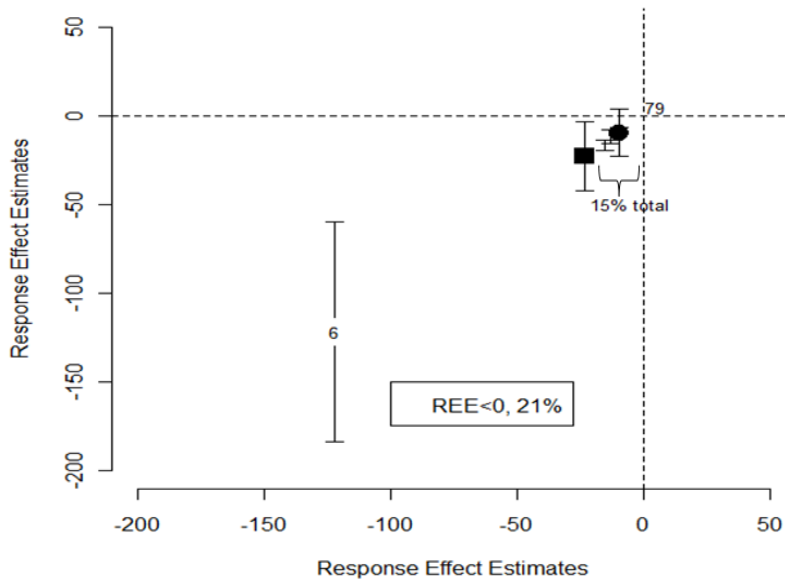

[a] The figure plots WLS point estimates for the QUALITYΔ variable, nonparametric response effect estimates (REE), and 95% confidence intervals for each. Response effects in the negative quadrant have very small standard errors resulting into small confidence intervals, which are not legible on this graph. See the footnote to Figure 3 for addi-tional explanation.

Figure 5. Benefit Transfer Error Response Effects for STUDYAREA [a]

[a] The figure plots nonparametric response effect estimates (REE), and 95% confidence intervals for the STUDYAREA variable. Majority of the response effects in the positive quadrant have very small standard errors resulting into small confidence intervals that are not legible on this graph. See the footnote to Figure 3 for additional explanation.

Figure 6. Benefit Transfer Error Response Effects for MULTIPLESTUDY [a]

[a] The figure plots nonparametric response effect estimates (REE), and 95% confidence intervals for the MULTIPLESTUDY variable. 79% of response effects in the positive quadrant are valued at 0.19%. These response effects have very small standard errors resulting into small confidence intervals that are not legible on this graph. See the footnote to Figure 3 for additional explanation.

Appendix I. Nonparametric Estimation in Information Deficient Settings

A well-known paradox in the theoretical study of nonparametric cross validation methods is that "the harder the estimation problem, the better cross-validation performs" (see Savchuk, Hart and Sheather [33] for a summary and citations to the literature). This paradox is exemplified by the 'information deficient' setting in which we are working. In Theorems 2.1 and 3.1 of their paper, Ouyang, Li, and Racine [28] demonstrate that the convergence rate of the bandwidth for our estimator does not depend on the number of empty cells. Thus, the presence of empty cells will not deteriorate the empirical performance of cross-validation relative to the theoretical insights.

Recent studies have also produced simulation-based evidence on the performance of both the nonparametric estimator and the cross-validated bandwidths. While none of the simulation environments provide an exact match to our data, some have considered problems that are *harder* than the one we face. The strongest evidence can be found in Henderson, Papageorgiou and Parmeter [14]. Their simulations investigate model performance in a setting with 10 covariates (5 continuous, 5 discrete) using sample sizes ranging from 409 to 719. They find that cross validation performs well (in terms of the bandwidths) and that the nonparametric estimator outperforms misspecified linear models. Importantly, the mixed discrete-continuous nature of their data would imply slower rates of convergence for the bandwidths than in our study. Parmeter, Zheng and McCann [29] report similar results in terms of the cross-validated bandwidths' ability to correctly smooth out irrelevant variables. Thus, the collective evidence from the recent literature suggests that having a large number of covariates and empty cells does not prevent the cross-validated bandwidths from being able to distinguish relevant from irrelevant covariates; nor does it prevent the estimator from identifying the response effects for explanatory variables in the cells that are filled.

Appropriate selection of the smoothing parameter also alleviates the information defi-

ciency issue by combining empty or sparsely filled cells with nearby cells containing data thereby leveraging information from cells that are deemed "close". Consider the category of study case valuation method, "*v*". We combine the variables in this category (META, RUM, TC, CV and CM) into a single categorical variable taking values in (0, 1, 2, 3, 4). In the presence of many discrete variables combining variables that define the same procedure is a viable strategy for improving the success of bandwidth estimation; however, we dispense the ability to identify separate bandwidths for each of these variables as they are combined into a single category. Although we cannot identify separate bandwidths, Ouyang, Li and Racine [28] demonstrate this will yield qualitatively similar insights into each cell's "effect" on the response of interest implying that we can consistently estimate each cell's effect.

Now suppose that no observations exist for category '2'. In this case we can still present function estimates using the smoothing approach. Our nonparametric categorical regression estimator for category '2' would be:

$$\widehat{m}(2) = \frac{\sum_{i=1}^{n} |\%BTE|_i L(X_i, 2, \lambda)}{\sum_{i=1}^{n} L(X_i, 2, \lambda)} = \frac{\sum_{|X_i-2|=1} |\%BTE|_i \lambda + \sum_{|X_i-2|=2} |\%BTE|_i \lambda^2}{\sum_{|X_i-2|=1} \lambda + \sum_{|X_i-2|=2} \lambda^2} \tag{9}$$

where λ is the bandwidth used to smooth over the five cells. This shows that the nonparametric smoothing estimator is a weighted average of nearby cells (0, 1, 3, 4). The weight depends upon the distance of a cell from the cell of interest, $|X_i - 2|$. In the parametric approach, there would only be four categories (0, 1, 3, 4), and the category '2' would have been placed in the baseline category '0', meaning that the function estimate for true '0' members would also be applied to '2' members. However, in case of nonparametric regression, the estimator leverages information from the nearby cells.

Appendix Table I. Characteristics of 31 Benefit Transfer Validity Studies Included in the Meta Analysis

Authors (year)	Valuation Methods	Resources Valued	N	% transfer error mean (min, max)	|% transfer error| mean (min, max)
Barton (2002)	CV	Beach Quality	8	-20 (-23,-10)	20 (10,23)
Bergland, Magnussen and Navrud (1995)	CV	Water Quality	2	-21 (-24,-18)	21 (18,24)
Brouwer and Bateman (2005)	CV	Human Health	85	17 (-41,123)	34 (0.4,123)
Brouwer and Spaninks (1999)	CV	Farm Land	8	3 (-59,60)	42 (22,60)
Colombo and Hanley (2008)	CM	Farm Land	178	680 (2, 7496)	680 (2,7496)
Colombo, Calatrava-Requens, and Hanley (2007)	CM	Soil Conservation	54	110 (8,1148)	110 (8,1148)
Groothuis (2005)	CV/TC	Deer Hunting	120	-10 (-75,136)	30 (0.1,136)
Hanley, Wright, and Alvarez-Farizo (2006)	CM	Ecosystem Health	2	-72 (-78,-67)	72 (67,78)
Jiang, Swallow, and McGonagle (2005)	CM	Coastal Land	5	-68 (-85,-53)	68 (53,85)
Johnston and Duke (2009)	CM	Farm Land	4	-76 (-100,-29)	76 (29,100)
Johnston (2007)	CM	Mixed Resources	24	-12 (-101,58)	37 (7,101)
Kerr and Sharp (2006)	CM	Ecosystem Health	22	79 (-63,704)	120 (2,704)
Kristofersson and Navrud (2007)	CV	Fishing/ Ecosystem Health	21	125 (7,319)	125 (7,319)
Lindhjem and Navrud (2008)	META	Multiple Use Forestry	16	73 (25,266)	73 (25,266)
Loomis et al. (1995)	TC	Reservoir	56	106 (-50,475)	115 (0.5,475)

Loomis (1992)	TC	Sport Fishing	10	0.2 (-18,9)	6 (1, 18)
Matthews, Hutchinson, and Scarpa (2009)	CV	Forest Recreation	84	12 (-42,125)	27 (0.0,125)
Morrison and Bennett (2006)	CM	Ecosystem Health	28	-25 (-171,30)	35 (1,171)
Morrison et al. (2002)	CM	Wetlands	9	-32 (-66,-4)	32 (4,66)
Parsons and Kealy (1994)	RUM	Water Recreation	11	-4 (-66,75)	21 (1,75)
Piper and Martin (2001)	CV	Water Supply	8	35 (-9,149)	39 (3,149)
Ready and Navrud (2007)	CV	Human Health	2	39.1 (37.7,41.9)	39.1 (37.7,41.9)
Ready et al. (2004)	CV	Human Health	21	37 (20,83)	37 (20,83)
Rosenberger and Loomis (2000)	META	Mixed Resources	115	17 (-79,319)	49 (0.0,319)
Rozan (2004)	CV	Air Quality	4	-27.5 (-28,-27)	27.5 (27,28)
Shrestha and Loomis (2003)	META	Mixed Resources	34	60 (-74,411)	84 (12,411)
Shrestha and Loomis (2001)	META	Outdoor Recreation	18	6 (-46,81)	28 (0.5,81)
Stapler and Johnston (2009)	META	Sport Fishing	4	228 (64,572)	228 (64,572)
Vandenberg, Poe and Powell (2001)	CV	Ground Water	8	29 (16,44)	29 (16,44)
Zanderson, Termansen and Jensen (2007a)	RUM	Forest Recreation	6	4 (-66,55)	30 (4,66)
Zanderson, Termansen and Jensen (2007b)	RUM	Forest Recreation	104	180 (-73,1111)	194 (1.3,1111)