Ecological Economics 96 (2013) 181–189

Contents lists available at ScienceDirect

Ecological Economics

journal homepage: www.elsevier.com/locate/ecolecon

Addressing dependency in the sportfishing valuation literature: Implications for meta-regression analysis and benefit transfer

Arvin B. Vista [a,*], Randall S. Rosenberger [b,1]

[a] Department of Agricultural Economics, College of Economics and Management, University of the Philippines Los Baños College, Laguna 4031, Philippines
[b] Department of Forest Ecosystems and Society, College of Forestry, Oregon State University, Corvallis, OR 97331 USA

ARTICLE INFO

Article history:
Received 26 May 2011
Received in revised form 21 October 2013
Accepted 24 October 2013
Available online 15 November 2013

Keywords:
Benefit transfer
Data dependency
Meta-regression analysis
Sportfishing valuation
Transfer error
Within-study correlation

ABSTRACT

Meta-regression analysis is a statistical summary or synthesis of a body of evidence. However, when primary studies provide more than one estimate, the presence of dependence in the metadata has implications for the statistical efficiency of estimated moderator variables. Previous meta-analyses have adjusted for within study dependence through ad hoc procedures (e.g., selecting one estimate per study and study average) or regression-based methods (e.g., weighted and panel data models). This paper defines dependency based on the underlying primary data (i.e., from the same sample) and examines the effect of different models and treatments on meta-regression estimation and implications for benefit transfer performance. The models are applied to the sportfishing literature that contains 140 papers providing 833 estimates of access values for fishing in the United States and Canada. The different methods of adjusting for dependency within the sportfishing metadata result in differences in the estimated model coefficients; hence, different transferred values and transfer errors.

1. Introduction

Meta-regression analysis is a statistical summary of a body of evidence for the purposes of synthesizing multiple empirical estimates, testing hypotheses, or benefit transfer. Benefit transfer is a method of predicting values of policy relevance for a site with limited or no data by transferring an estimated value or function from a primary research site, or by predicting an estimate from a meta-regression function[2] (Rosenberger and Loomis, 2001, 2003). Benefit transfer is a cost effective method when available time and resources preclude conducting a primary study. The body of evidence, or metadata, is comprised of empirical measures like values, elasticities, or correlation coefficients for a common good such as recreation, typically reported in published research literature. The meta-regression model estimates partial effects of the moderator variables that measure characteristics of each primary study on study outcomes. However, when a primary study provides more than one estimate, the errors may be correlated, violating the assumption of independent observations leading to biased standard errors unless this correlation is accounted for in the model (Glass et al., 1981; Nelson and Kennedy, 2009; Strube, 1987; Wood, 2008). Biased standard errors may lead to incorrect inferences about partial effects.[3] This is one of the methodological pitfalls that may affect the validity and reliability of meta-analysis (Florax, 2002). Contingent upon how data dependency is treated within a meta-analysis, the treatment may bias the estimated partial effects as well, potentially affecting the predictive performance of the meta-analysis when used for benefit transfers.

Nelson and Kennedy (2009) emphasize the need to adjust for correlated estimates, both *within* and *between* groups of studies. Between study correlation may arise from primary studies drawing from the same data source (e.g., time-series databases) or using the same study location, functional form, or explanatory variables. Between study correlation may be corrected through inclusion of moderator variables that identify the data source, location or model specification. Within study correlation may arise when primary studies report more than one estimate (Nelson and Kennedy, 2009). Corrections for within study correlation may require manipulation of the meta-data or the use of regression-based methods.

* Corresponding author. Tel.: +63 49 536 3292.
E-mail addresses: arvinvista@gmail.com (A.B. Vista), R.Rosenberger@oregonstate.edu (R.S. Rosenberger).
[1] Tel.: +1 541 737 4425.

[2] Meta-regression function is a type of benefit transfer based on the estimation of a meta-regression benefit function across multiple study sites. This function transfers then adapt the function to fit the specifics of the policy site, such as socio-economic characteristics, extent of market and environmental impact, and other measurable characteristics that systematically differ between the study site and the policy site.

[3] When multiple estimates from the same study are recorded in the metadata, standard errors of meta-regression results may be biased due to the lack of independence of observations (Glass et al., 1981; Strube, 1987; Wood, 2008). Failure to adjust for dependency will inflate the likelihood of *Type I errors* when comparisons are made *across* studies. Likewise, dependency will inflate the likelihood of *Type II errors* when comparisons are made among or between different outcomes *within* studies (Strube, 1987).

0921-8009/$ – see front matter
http://dx.doi.org/10.1016/j.ecolecon.2013.10.016

Given that dependency is likely in valuation metadata,[4] what methods are available to minimize its effects? One method is to avoid the dependency by reducing the metadata to a single estimate per primary study. For example, Lipsey and Wilson (2000) suggest drawing a single estimate per primary study based on the best evidence when multiple estimates are provided by a single primary study, or estimate the average value of multiple estimates per primary study. Another data reduction technique is to randomly draw a single estimate from among the multiple estimates per primary study (Bijmolt and Pieters, 2001). Another method is to use regression-based approaches to adjust the metadata for multiple estimates per study. Regression-based methods include panel data estimators (Jeppesen et al., 2002; Rosenberger and Loomis, 2000b), hierarchical/multilevel models (Bateman and Jones, 2003; Johnston et al., 2005), and weighted least squares (where each study is weighted by the inverse of the number of estimates it provides). A primary difference between these groups of methods is that the avoidance methods reduce the metadata to one observation per study whereas the regression-based methods retain all reported empirical estimates in primary studies.

Avoidance of dependency through the use of a single best estimate or average estimate has not yet been fully addressed through empirical research in environmental economics meta-analyses. However, in their analysis using marketing metadata, Bijmolt and Pieters (2001) reject avoidance approaches (i.e., selecting single estimates per primary study) because they are inefficient, lose information and degrees of freedom, and provide biased estimates of partial effects of moderator variables. In environmental economics, Ghermandi et al. (2010) evaluated data dependency by estimating four meta-regression models on wetland valuation metadata, including fully-specified and parsimonious models on all observations ($n = 416$), a model weighted by study ($n = 416$), and a randomly selected single estimate per study model ($n = 169$). Their conclusion, largely based on differences in adjusted-R^2, was that the weighted and single-estimate models did not result in improvements over the parsimonious model with all observations treated as being independent.

Previous meta-analyses that applied a regression-based method to address data dependency (e.g., weighted independent estimates or panel model dependent estimates) that provide efficient standard errors, found modest changes on estimated partial effects (Bateman and Jones, 2003; Bijmolt and Pieters, 2001; Ghermandi et al., 2010; Jeppesen et al., 2002; Johnston et al., 2005). Rosenberger and Loomis (2000b) evaluated panel data models on recreation use values metadata using three different sources of data dependency, including by study, by researcher and by data structure. In all three panel stratifications, they reject the panel structure in favor of independent observations for the full metadata (Rosenberger and Loomis, 2000a, 2000b).

This paper evaluates regression-based and avoidance methods to adjust for potential data dependency when one or more of the studies in the metadata contain multiple estimates; i.e., within-study correlation. It departs from previous dependency corrections that identify the study as the source of dependency by classifying the underlying data as the source of dependency within studies with multiple estimates. Beginning with the full metadata, or *all-set* metadata, all of the reported benefit estimates are used in the regressions. Two regression-based approaches for accounting for data dependency in the *all-set* metadata include weighting of the metadata by the number of observations per study and using panel data estimators. Two treatments of the metadata for avoiding dependency include a *best-set* metadata (comprised of the best available benefit measures reported in a study as identified by the author of the primary study or based on methodological and statistical criteria) and an *average-set* metadata (comprised of the average of the estimates reported in the primary studies). These alternative approaches to dealing with sample level dependence will result in different values of standard errors of the meta-regression functions, which could affect the predictions in benefit transfer applications. These approaches are tested by comparing the associated percentage transfer errors in an out of sample benefit transfer exercise. The next section outlines the data dependency conceptual framework, followed by a description of the metadata and model results. This paper concludes with recommendations for study selection and methodological protocol when conducting a meta-analysis and its use in benefit transfers.

2. Approaches to Multiple Estimates

Let the meta-regression (MR) model be defined as:

$$y_i = \alpha + \beta_k \mathbf{X}_i + \varepsilon_i \tag{1}$$

where y is the dependent variable, which is the vector of estimates reported across the individual primary studies, i that indexes each estimate; $\mathbf{X}$ is a matrix of moderator variables (i.e., the identifiable characteristics among the different studies) that account for systematic components explaining the variation in y; and ε is a random error component with mean zero and variance σ^2_ε. The parameters α (constant term common across all observations) and β (slopes) are estimated by:

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \left(\mathbf{X}'T^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'T^{-1}Y \tag{2}$$

where $T = \mathbf{W}\sum\mathbf{W}$, $\mathbf{W}$ is a diagonal matrix with weights and $\sum$ is a block-diagonal matrix with error variances.

There are at least three approaches to deal with multiple estimates within studies: 1) *all-set* approach; 2) *best-set* approach; and 3) *average-set* approach.[5] The *all-set* approach codes all of the estimates reported in primary studies, treating each estimate as an independent observation. On the other hand, the *best-set* approach codes only the best available benefit measures reported in a study as identified by methodological and sample criteria, while an *average-set* approach codes the average of benefit measures reported in the primary studies.

When estimates are independent, then the studies are as well; i.e., all the covariance terms are zero and the estimate of the standard deviation of the linear combinations of studies is based solely on the main diagonal of the variance–covariance matrix (Strube, 1987). In this *all-set* approach, all weights equal 1, $\mathbf{W} = \mathbf{I}_M$, where $\mathbf{I}_M$ is the identity matrix of order $M \times M$, and the error at the study level is assumed zero. The parameters, α and β, are estimated via ordinary least squares (OLS). Assuming multiple estimates from a single study are independent is likely an unwarranted assumption that can bias model results (Nelson and Kennedy, 2009).

When primary studies report more than one estimate, the use of weights is suggested—studies with many estimates have a larger effect on the results of the meta-analysis than studies with fewer estimates (Rosenthal, 1991). In this case, multiple estimates are treated as independent weighted estimates, where weights are defined by: $\mathbf{W}_{mj} = M / M_j J$, for all $j = 1, 2, \ldots, J$. When the number of estimates from each primary study is used as the weight, the within-study weights sum to one (see also Johnston et al., 2006; Mrozek and Taylor, 2002). Hence, each primary study, rather than each observation, has an equal weight in determining the regression coefficients. Weights are applied by placing the square root of the weights onto the diagonal of an $M \times M$ matrix, thereby forming the diagonal matrix $\mathbf{W}$, with zero's

[4] In Nelson and Kennedy's (2009, p.351) survey of meta-analyses in environmental and resource economics, they note that "110 of 140 studies use more than one estimate from each primary study."

[5] Another method is to randomly select an estimate for each study (Bijmolt and Pieters, 2001; Ghermandi et al., 2010). This method is not evaluated in this paper since it will lead to the selection of estimates that are less favored for their measurement of underlying values and variability in the metadata and estimated partial effects.